

TRAVAIL DE BACHELOR 2018

Projet VASCO - Plateforme publique de visualisation de données

Juillet 2018

Étudiante : Audrey Michel

Professeur : Florian Evéquoz

Déposé le : 30.07.2018

RÉSUMÉ

« Le Big Data », de nos jours, ces trois petits mots ressortent très souvent dans nos discussions, à la télévision et dans les journaux. Nous avons de plus en plus de données provenant de diverses ressources telles que les téléphones portables, les réseaux sociaux, les objets connectés qui nous suivent dans notre quotidien. Ces données sont incompréhensibles pour la plupart des personnes. Cependant, il existe déjà des outils tels que D3.js, HighCharts, Plotly, Vega, etc permettant de visualiser des données par le biais de graphiques. Malheureusement, ils sont parfois très complexes à utiliser. Cette problématique nous amène à réaliser ce projet afin de créer une plateforme publique de visualisation de données. Celle-ci devra être facile d'utilisation et de compréhension pour des personnes n'ayant pas de grandes connaissances dans l'analyse de données.

Le but final de ce travail est d'avoir une plateforme en ligne afin que chacun puisse y accéder rapidement. L'utilisateur peut y importer des fichiers au format CSV. Les données sont chargées et automatiquement une liste de graphiques est affichée. De plus, il est possible d'exporter les graphiques en SVG, PNG et en JSON.

Ce travail a été réalisé en plusieurs parties : Tout d'abord un état de l'art technique a fait ressortir le besoin de créer une plateforme qui automatise des visualisations de toutes pièces. L'étudiante a choisi de travailler avec la librairie Vega-Lite qui produit des graphiques avec très peu de lignes de code en JSON. Ensuite, l'implémentation terminée, la plateforme a été testée par des utilisateurs lambdas. Ils ont alors attiré notre attention sur le fait que certains éléments n'étaient pas assez visibles, si l'utilisateur n'est pas familiarisé avec le site internet. Moyennant cette phase de tests, nous avons pu confirmer que l'objectif de créer une plateforme compréhensible et facile d'utilisation a été atteint.

Cependant, ce projet peut être amélioré en diminuant le temps de chargement des visualisations, en élargissant le choix de format, le type et la taille de fichiers à importer. Ainsi qu'en ajoutant quelques fonctionnalités telles que le changement de style de graphique, de la couleur, d'ajouter des commentaires, etc.

MOTS-CLÉS

Visualisation de données, graphiques, plateforme web, Vega-Lite

AVANT-PROPOS

Ce document a été effectué dans le cadre du travail de Bachelor durant le 6^{ème} semestre en Informatique de Gestion à la Haute École Spécialisée de Suisse occidentale (HES-SO) Valais-Wallis.

Le sujet de ce travail, « Projet VASCO - Plateforme publique de visualisation de données », a été proposé par Monsieur Florian Evéquoz. Ce travail est le début du projet VASCO développé à l'institut Informatique de Gestion de la HES-SO Valais-Wallis. J'ai choisi ce sujet car il m'a intrigué et je voulais en savoir plus sur les librairies de visualisation de données. De plus, c'est un sujet d'actualité qui se développera de plus en plus dans le futur. En effet, les données sont en expansions continues et il est très intéressant et utile de pouvoir les analyser et surtout de les comprendre.

L'objectif principal de ce travail est de créer un *proof of concept* (POC) d'une plateforme publique de visualisation de données facile d'utilisation et de compréhension.

Ce rapport est basé sur le guide de présentation et de réalisation des travaux écrits pour les étudiants des filières Économie d'entreprise, Informatique de gestion et Tourisme de la Haute École de Gestion & Tourisme de la HES-SO Valais. Ce travail est basé sur les normes de mise en forme de la 6^e édition du manuel de publication de l'*American Psychological Association* (APA).

REMERCIEMENTS

Je tiens à remercier plusieurs personnes qui ont contribué à la bonne réalisation de ce projet. Pour commencer, je remercie infiniment mon professeur de travail de Bachelor, Florian Evéquoz qui a pris la peine de me suivre et de me donner de bons conseils tout au long du projet.

Je remercie aussi Christophe Pignat qui a pris le temps nécessaire pour me créer une place sur le serveur de la Hes-So afin d'héberger mon site internet ainsi que d'avoir pris le temps de m'expliquer le fonctionnement de celui-ci.

Un grand merci aux personnes qui ont pris la peine de tester la plateforme.

Je remercie également Rafael Peixoto, Pedro Gil Ferreira et Maxime Bétrisey qui ont pris la peine de m'expliquer comment résoudre certains problèmes face auxquels je me suis confrontée durant l'implémentation du projet.

Et pour finir, je tiens à remercier mes proches et ma famille pour leur soutien et leurs encouragements lors de réalisation de ce travail et tout au long de ma formation.

TABLE DES MATIÈRES

LISTE DES FIGURES	VIII
LISTE DES TABLEAUX.....	IX
LISTE DES ABRÉVIATIONS.....	X
INTRODUCTION	1
OBJECTIFS	1
SOLUTION PROPOSÉE	2
PHASES DU PROJET	2
Phase d'analyse	2
Phase de développement.....	2
MÉTHODOLOGIE	3
1. ÉTAT DE L'ART.....	5
1.1. OPEN SOURCE	5
1.2. CRITÈRES D'ANALYSE	6
1.3. OUTILS EXISTANTS	6
1.3.1 Data-Illustrator	6
1.3.2 Datavoyager.....	8
1.3.3 RAWGraphs	10
1.3.4 Datawrapper	12
1.3.5 Tableau	16
1.4. CONCLUSION	18
2. ANALYSE ET CHOIX TECHNOLOGIQUES	20
2.1 LIBRAIRIES	20
2.2 JAVASCRIPT	20
2.3 CRITÈRES DE SÉLECTION	21
2.3.1 D3.js.....	22
2.3.2 C3.js.....	23
2.3.3 Plotly	24
2.3.4 Chart.js	26
2.3.5 Chartist.js	27
2.3.6 HighCharts JS	28
2.3.7 Vega	29
2.3.8 Vega-Lite	30
2.4 CONCLUSION	31
3. TECHNOLOGIES UTILISÉES	33

3.1	LANGAGES DE PROGRAMMATION.....	33
3.1.1	HyperText Markup Language.....	33
3.1.2	Pug.....	33
3.1.3	Cascading Style Sheets.....	34
3.1.4	Node.js.....	34
3.2	LIBRAIRIES.....	34
3.2.1	JQuery.....	34
3.2.2	Convert-csv-to-json.....	35
3.2.3	CompassQL.....	35
3.2.4	Vega-tooltip.....	36
3.2.5	Vega-embed.....	36
3.2.6	Express.js.....	36
3.2.7	Multer.....	36
3.2.8	Express-session.....	36
3.3	FRAMEWORK.....	37
3.3.1	Materialize.....	37
4.	DÉFINITION DES BESOINS.....	38
4.1	FONCTIONNALITÉS SOUHAITÉES PAR LE DÉVELOPPEUR.....	38
4.2	FONCTIONNALITÉS SOUHAITÉES PAR L'ADMINISTRATEUR.....	38
4.3	FONCTIONNALITÉS SOUHAITÉES PAR LES UTILISATEURS.....	39
5.	MODÉLISATION ET CRÉATION DE LA PLATEFORME.....	40
5.1	MOCKUP.....	40
5.2	SÉLECTION DE DONNÉES.....	42
5.2.1	Format CSV.....	42
5.2.2	Type de données.....	42
5.2.3	Données quantitatives.....	43
5.2.4	Données qualitatives.....	43
6.	TESTS UTILISATEURS.....	44
6.1	RESSOURCES NÉCESSAIRES.....	44
6.2	UTILISATEURS CIBLES.....	44
6.2.1	Utilisateurs dans le domaine de l'analyse de données.....	45
6.2.2	Utilisateurs non-initiés.....	45
6.3	TYPE DE TÂCHES.....	45
6.4	DEGRÉ DE DIFFICULTÉS.....	45
6.5	SCÉNARIOS.....	45
6.6	RÉSULTATS DE LA PREMIÈRE PHASE DE TEST.....	46

6.7	MODIFICATIONS APPORTÉES À L'INTERFACE	49
6.8	RÉSULTATS DE LA DEUXIÈME PHASE DE TESTS.....	51
6.9	RÉSULTATS FINAUX	52
7.	DIFFICULTÉS RENCONTRÉES	53
7.1	MANQUE DE DOCUMENTATION.....	53
7.2	TAILLE DES FICHIERS	53
7.3	DONNÉES SÉLECTIONNÉES	54
7.4	VALEURS MANQUANTES	54
7.5	VARIABLES DE SESSION	54
	CONCLUSION	55
	AVIS TECHNIQUE	55
	AVIS PERSONNEL	56
	AMÉLIORATIONS FUTURES	56
	RÉFÉRENCES.....	57
	ANNEXES	64
	ANNEXE I : CAHIER DES CHARGES.....	64
	ANNEXE II : PRODUCT BACKLOG	67
	ANNEXE III : GUIDE DE L'UTILISATEUR.....	69
	ANNEXE IV : GUIDE TECHNIQUE.....	76
	ANNEXE V : DONNÉES SÉLECTIONNÉES - CÉRÉALES	78
	ANNEXE VI : DONNÉES SÉLECTIONNÉES - AIRBNB À GENÈVE	79
	ANNEXE VII : DONNÉES SÉLECTIONNÉES - MEILLEURES COMPAGNIES EN INFORMATIQUE ENTRE 2014 ET 2015	80
	ANNEXE VIII : QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO	81
	ANNEXE IX : ANALYSE DU QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO	85
	ANNEXE X : RÉPONSE DU QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO AVANT MODIFICATION	89
	ANNEXE XI : RÉPONSE AU QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO APRÈS MODIFICATION	97
	DÉCLARATION DE L'AUTEUR.....	104

Liste des figures

Figure 1 - Processus de la méthodologie Scrum	3
Figure 2 - Logo Data-Illustrator (Data Illustrator, 2018)	6
Figure 3 - Plateforme de Data-Illustrator avec un exemple de visualisation	7
Figure 4 - Logo datavoyager (Datavoyager, 2015)	8
Figure 5 - Plateforme de Datavoyager avec un exemple de visualisation	9
Figure 6 - Logo RAWGraphs (RAWGraphs, 2018).....	10
Figure 7 - Plateforme de RAWGraphs avec des exemples de visualisation	11
Figure 8 - Logo Datawrapper (Datawrapper, 2018)	12
Figure 9 - Prix des abonnements chez Datawrapper	13
Figure 10 - Plateforme de Datawrapper avec un exemple de visualisation	14
Figure 11 - Logo Tableau (Tableau, 2018)	16
Figure 12 - Exemple de visualisation sur Tableau Public.....	17
Figure 13 - Logo JavaScript (PluralSight, 2018)	20
Figure 14 - Logo D3.js (D3.js, 2017)	22
Figure 15 - Exemple de visualisation C3.js.....	23
Figure 16 - Logo Plotly (Plotly, 2017).....	24
Figure 17 - Prix des abonnements de Plotly	25
Figure 18 - Exemple de visualisation Chart.js	26
Figure 19 - Site internet de Chartist.js.....	27
Figure 20 - Logo de HighCharts JS (Highcharts, 2018).....	28
Figure 21 - Exemple de visualisation Vega-Lite	30
Figure 22 - Logo Node.js (Node.js, s.d.)	34
Figure 23 - Étape de Sélection version 1	41
Figure 24 - Étape de Sélection version 2	41
Figure 25 - Étape de Modification version 1	41
Figure 26 - Étape de Modification version 2	41
Figure 27 - Site vitrine VASCO	41
Figure 28 - Exemple d'un fichier CSV des meilleures entreprises dans l'informatique entre 2014 et 2015 ...	42
Figure 39 - placeholder de la page importation.....	46
Figure 40 - Page de Sélection avant la phase de tests - synthèse	49
Figure 41 - Page de Sélection après la phase de tests - synthèse	49
Figure 42 - Page de Sélection avant la phase de tests - liste des graphiques.....	49
Figure 43 - Page de Sélection après la phase de tests - liste des graphiques	49
Figure 44 - Page de Modification avant la phase de tests.....	50
Figure 45 - Page de Modification après la phase de tests.....	50

Liste des tableaux

Tableau 1 - Récapitulatif des Sprints	4
Tableau 2 - Résumé de Data Illustrator	8
Tableau 3 - Résumé datavoyager.....	10
Tableau 4 - Résumé de RAWGraphs.....	12
Tableau 5 - Résumé Datawrapper	15
Tableau 6 - Résumé Tableau.....	18
Tableau 7 - Récapitulatif de l'État de l'art.....	19
Tableau 8 - Résumé D3.js	23
Tableau 9 - Résumé C3.js	24
Tableau 10 - Résumé Plotly	25
Tableau 11 - Résumé Chart.js	26
Tableau 12 - Résumé Chartist.js	27
Tableau 13 - Résumé Highcharts	28
Tableau 14 - Résumé de Vega	29
Tableau 15 - Résumé Vega-Lite.....	31
Tableau 16 - Récapitulatif des libraires JavaScript.....	31
Tableau 17 - Synthèse des résultats de la première partie des tests	48
Tableau 18 - Synthèse des résultats de la deuxième partie des tests	51

Liste des abréviations

CSS	Cascading Style Sheets
CSV	Comma-separated values
HES-SO	Haute École Spécialisée de Suisse occidentale (HES-SO)
HTML	Hypertext Markup Language
IDC	International Data Corporation
JS	JavaScript
JSON	JavaScript Object Notation
PNG	Portable Network Graphics
POC	Proof Of Concept
SVG	Scalable Vector Graphics
VASCO	Assisted Visualisation Generation Platform for AnalySing Data COllaboratively
Zo	Zettaoctets

INTRODUCTION

De nos jours, les données sont primordiales dans notre vie quotidienne alors qu'auparavant, elles contribuaient simplement au bon développement de notre vie habituelle. En effet, les données sont en augmentation continue de jour en jour, cependant notre performance d'analyse et de compréhension de celles-ci pour acquérir des connaissances et un savoir-faire augmente à vitesse très réduite. Il existe quelques outils permettant l'analyse de données, mais ils sont assez complexes. Il faut donc être formé sur la matière afin de pouvoir tirer profit de ces données et ainsi créer de nouvelles connaissances en les exploitant à bon escient.

IDC (International Data Corporation) prévoit que d'ici 2025, la sphère de données mondiales atteindra 163 zettaoctets (ZO). C'est dix fois plus que les 16 ZO de données générées en 2016 (IDC, 2017). Jeff Fochtman, responsable marketing chez Seagate, raconte : « Ce qui est étonnant, ce n'est pas que la production de données à stocker augmente, mais le rythme effréné de cette augmentation. Nous en sommes nous-mêmes surpris. Et la vague de l'Internet des objets ne fait que commencer. » (Les Echos, 2017)

OBJECTIFS

Ce travail de Bachelor a comme but principal de créer un *proof of concept* d'une plateforme publique de visualisation de données. De manière à accomplir correctement ce but, l'étudiant va être amené à atteindre plusieurs objectifs :

- Faire un état de l'art technique
- Création d'un *Product Backlog*
- Analyser et déterminer les technologies à utiliser pour le projet
- Apprendre de nouveaux langages de programmation
- Développer un prototype
- Apporter un grand soin à l'architecture logicielle et à la documentation
- Réaliser une phase de tests
- Respecter les différentes échéances

L'objectif final est de créer une plateforme qui aide les personnes n'ayant aucune expérience dans l'analyse de données, à visualiser toutes sortes de données et à créer des connaissances en toute simplicité.

SOLUTION PROPOSÉE

Le projet final sera une plateforme déployée pour que tout le monde puisse y accéder rapidement. Sur celle-ci, un utilisateur pourra tout d'abord importer des données au format CSV. Puis, une fois les données chargées, il arrivera sur une nouvelle page avec des propositions de graphiques ainsi qu'un petit tableau récapitulatif ayant comme information les noms des colonnes, la somme, la moyenne et le nombre de lignes du fichier. Le but est qu'il puisse avoir une vue d'ensemble globale des données importées afin qu'il puisse les analyser et les comprendre facilement. Il pourra finalement exporter les graphiques qu'il souhaite en SVG, PNG et en JSON.

PHASES DU PROJET

Le travail se déroule en deux grandes phases : la phase d'analyse et la phase de développement. La première phase d'analyse se compose de l'état de l'art, des choix technologiques et de l'analyse des besoins. La deuxième phase de développement concerne toute la partie pratique qui est d'implémenter la plateforme selon l'analyse faite à la phase précédente et pour finir réaliser des tests et améliorer celle-ci.

PHASE D'ANALYSE

Nous allons faire un état de l'art des outils existants afin de ressortir les avantages et désavantages de ceux-ci. Cette analyse nous donnera des idées afin de réaliser un projet remplissant nos objectifs qui sont de créer une plateforme facile d'utilisation et de compréhension. Puis, nous allons faire une analyse des technologies afin de choisir avec lesquelles nous allons implémenter le projet, ainsi qu'une analyse des besoins. Pour finir, nous allons créer minutieusement des mockup en veillant à ce que la plateforme soit *user-friendly*. Ceci ayant pour but que les personnes ne venant pas du monde de l'informatique la comprennent facilement.

PHASE DE DÉVELOPPEMENT

Nous allons développer un prototype de la plateforme en utilisant les bibliothèques choisies durant la phase d'analyse. Nous mettrons la plateforme en ligne sur un serveur de la HES-SO afin qu'elle soit rendue publique et que tout le monde puisse y avoir accès.

Tout au long de la création du projet, nous faisons attention à réaliser une plateforme très facile de compréhension. Pour vérifier cela, nous procéderons à une phase de tests réalisés par des personnes lambda afin qu'elles nous disent leur ressenti.

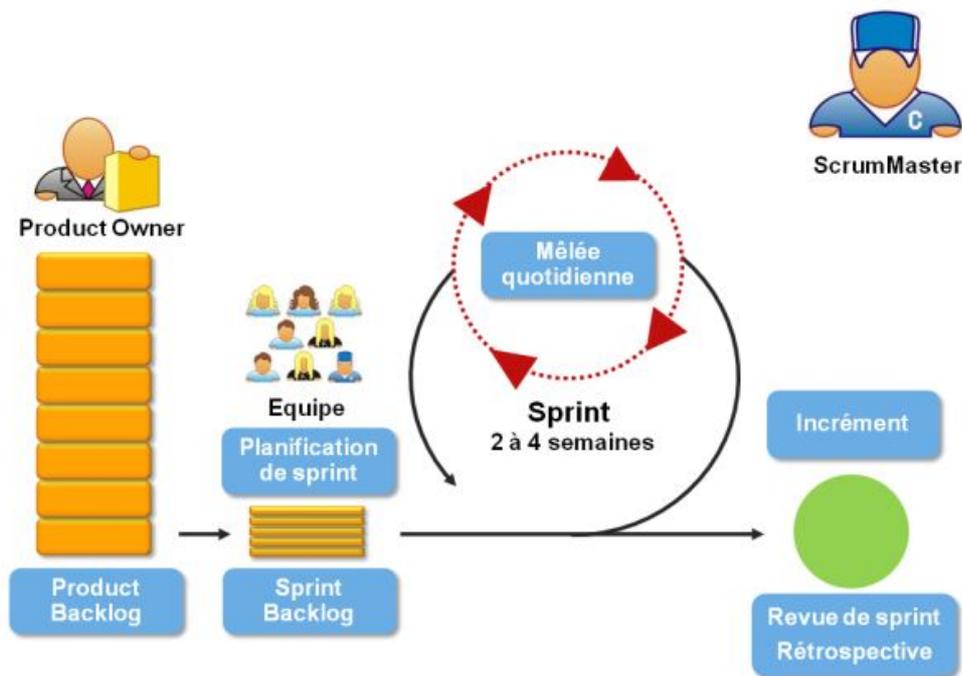
MÉTHODOLOGIE

La gestion de ce travail a été réalisée avec la méthodologie Scrum de la méthode Agile qui est une méthode de gestion de projets la plus utilisée (Florent, s.d.). Elle est basée sur une méthode itérative de livrables du produit. Au lieu d’attendre que le projet soit 100% finalisé pour le livrer au client, l’équipe délivre des tronçons “utilisables” du projet au cours du temps (Littlefield, 2017).

Par contre, nous l’avons simplifiée, car nous n’avons que deux rôles. L’étudiant était le Scrum Master et l’équipe de développement et le professeur était le *Product Owner*.

Dans le cadre de ce travail, nous avons fait des *Sprints* (période de travail d’une durée de 2 à 4 semaines) de deux semaines et à chaque fin de *Sprint* l’étudiant et le *Product Owner* se voyaient afin de présenter le travail réalisé et définir les prochaines étapes grâce aux *user stories* du *Product Backlog* (Le *Product Backlog* se trouve dans les annexes).

Figure 1 - Processus de la méthodologie Scrum



Source : (Florent, s.d.)

Cette méthodologie permet, dès le début du projet d’avoir une vue d’ensemble sur toutes les étapes du travail. De plus, cela permet de gérer le temps et de pouvoir anticiper et réajuster en cas de problèmes.

Tableau 1 - Récapitulatif des Sprints

Sprint	Début	Fin	Durée	Objectifs principaux
0	30.04.18	14.05.18	2 semaines	Création du Product Backlog et cahier des charges
1	14.05.18	04.06.18	2 semaines	État de l'art et choix technologiques
2	04.06.18	18.06.18	3 semaines, car une en Finlande pour les cours	Mock-up et choix de data set + préparation de l'environnement
3	18.06.18	02.07.18	2 semaines	Implémentation du Back-end
4	02.07.18	16.07.18	2 semaines	Continuer l'implémentation du Back-end + front-end + rapport
5	16.07.18	30.07.18	2 semaines	Déploiement + tests utilisateurs + rapport

Données de l'auteur

Comme on peut le constater dans le tableau récapitulatif, le projet s'est bien composé de 6 Sprints de 2 semaines chacun mise à part pour le Sprint 2 qui a duré 3 semaines car nous étions au séminaire en Finlande.

1. ÉTAT DE L'ART

Pour créer de manière adéquate la plateforme publique de visualisation de données selon nos buts, il est important de faire une analyse des outils existants qui ont partiellement ou les mêmes fonctionnalités que les nôtres.

Par le biais de cet état de l'art, nous comprendrons mieux les difficultés de la réalisation d'une plateforme de visualisation de données. Cette analyse a pour but, premièrement, de cibler plus précisément les objectifs des outils existants sur le marché afin de déterminer la différence de celles-ci avec notre projet, ce que l'on peut apporter de plus. Deuxièmement, cet état de l'art permet de nous donner de nouvelles idées cohérentes pour la création de notre future plateforme.

Cette analyse va se faire en plusieurs étapes. Pour commencer, un petit résumé descriptif sera effectué pour chaque outil. Puis, les étapes de création de visualisation seront listées et décrites afin de mieux connaître les propriétés de celle-ci. Pour finir, une petite conclusion sera faite en décrivant les points clés de l'outil qui seront intéressants d'exploiter lors de la création de notre plateforme. Enfin, notre avis personnel sera donné et un tableau récapitulatif de ces fonctionnalités sera réalisé.

1.1. Open Source

La désignation Open Source est une déclaration qui définit certains droits dans le monde de l'informatique. Le fait d'être certifié Open Source ne signifie pas seulement l'accès au code source, mais aussi, selon le site officiel opensource.org, les conditions de distribution des logiciels libres doivent respecter les critères suivants : (Opensource, 2007)

1. Redistribution gratuite
2. Source code
 - a. Le programme doit inclure le code source et doit permettre la distribution sous forme de code source ainsi que sous forme compilée.
3. Œuvres dérivées
 - a. La licence doit permettre les modifications et les œuvres dérivées, et permettre aussi leur distribution dans les mêmes conditions que la licence du logiciel original.
4. Intégrité du code source de l'auteur
5. Pas de discrimination à l'égard des personnes ou des groupes
6. Pas de discrimination à l'encontre des domaines d'activités de l'entreprise
7. Distribution de la licence
8. La licence ne doit pas être spécifique à un produit
9. La licence ne doit pas restreindre d'autres logiciels
10. La licence doit être neutre sur le plan technologique (Opensource, 2007)

1.2. Critères d'analyse

Il est important de se rappeler que notre objectif principal est de faire une plateforme pour une personne non-initiée à l'analyse et la visualisation de données. Donc l'état de l'art se porte principalement sur la facilité de compréhension et d'utilisation du service proposé.

Les principaux critères d'analyse de l'état de l'art qui suit sera si l'outil est open source, s'il est *User Friendly*, facile de compréhension. Quel est son prix ? De quand date sa création ? Comment se fait l'importation des données ainsi que l'exportation ? Quels types de fichiers accepte-t-il ?

Il est important que les étapes de l'importation à l'exportation soient simples et compréhensibles, car nos utilisateurs cibles ne sont pas des informaticiens.

1.3. Outils existants

Nous allons faire une analyse de 5 outils d'analyse de visualisation de données qui ressemblent sous certains aspects à notre projet.

1.3.1 Data-Illustrator

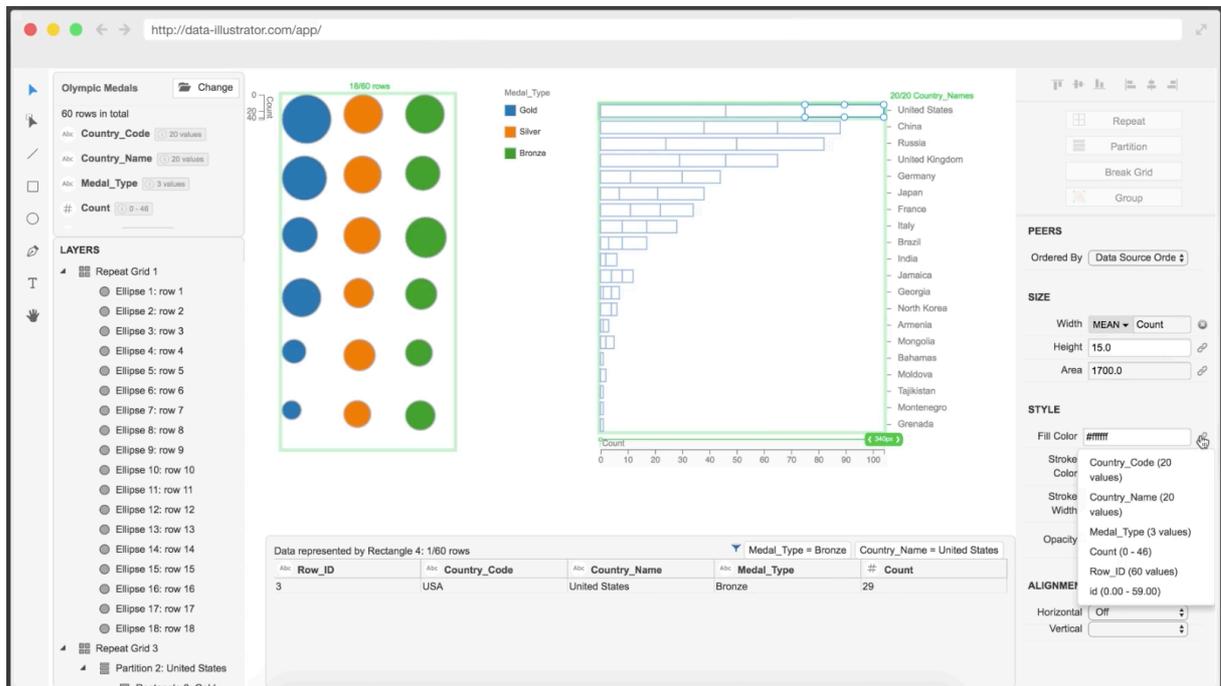


Figure 2 - Logo Data-Illustrator (Data Illustrator, 2018)

Data-Illustrator est une plateforme en ligne de visualisation de données. Il s'agit d'un projet de recherche qui est sorti en avril 2018. C'est pourquoi il est difficile de trouver de la documentation sur cet outil, mais il existe un document scientifique décrivant l'outil (Liu, et al., 2018). Selon leur site internet, Data Illustrator est une collaboration de recherche entre Adobe Systems Inc. et Georgia Institute of Technology (Data Illustrator, 2018).

La plateforme offre la possibilité de créer de toutes pièces les visualisations désirées, mais elle ne propose pas des exemples de graphiques prédéfinis. Leur objectif principal est de proposer la création de visualisation sans programmer.

Figure 3 - Plateforme de Data-Illustrator avec un exemple de visualisation



Source : (Data Illustrator, 2018)

Voici les étapes principales pour créer une visualisation :

1. Importer le fichier de données (n'accepte que les fichiers csv)
2. Créer manuellement le graphique désiré
 - Pas possible de revenir en arrière
3. Exporter la visualisation en SVG

Pour conclure, cette plateforme est, pour une personne non-initiée à la création de visualisation, très difficile de compréhension, car l'utilisateur doit construire de toutes pièces les graphiques et la plateforme n'est pas très *user-friendly*. Il serait utile d'avoir plus de documentation, comme c'est un projet tout récent, il y en a que très peu pour l'instant. L'outil a plutôt pour but de créer des graphiques de toutes pièces que de permettre directement l'analyse de visualisation. Par conséquent, nous ne nous baserons que peu sur cet outil, car notre objectif principal est l'analyse de données par le biais de visualisations automatisées directement sur la plateforme et non de créer des graphiques intégralement.

Voici un petit tableau récapitulatif :

Tableau 2 - Résumé de Data Illustrator

Data-Illustrator	
Site internet	http://data-illustrator.com/
Création	Avril 2018
Origine	USA
Prix	Gratuit
Open Source	Non
Librairies utilisées	Paper.js, RxJs, D3.js, Datalib, Backbone.js
User Friendly	Non
Importation	CSV
Exportation	SVG
Documentations	Non
Graphiques interactifs	Non

Source : Tableau de l'auteur provenant de sources multiples
 a) (Data Illustrator, 2018)
 b) (Liu, et al., 2018)

1.3.2 Datavoyager

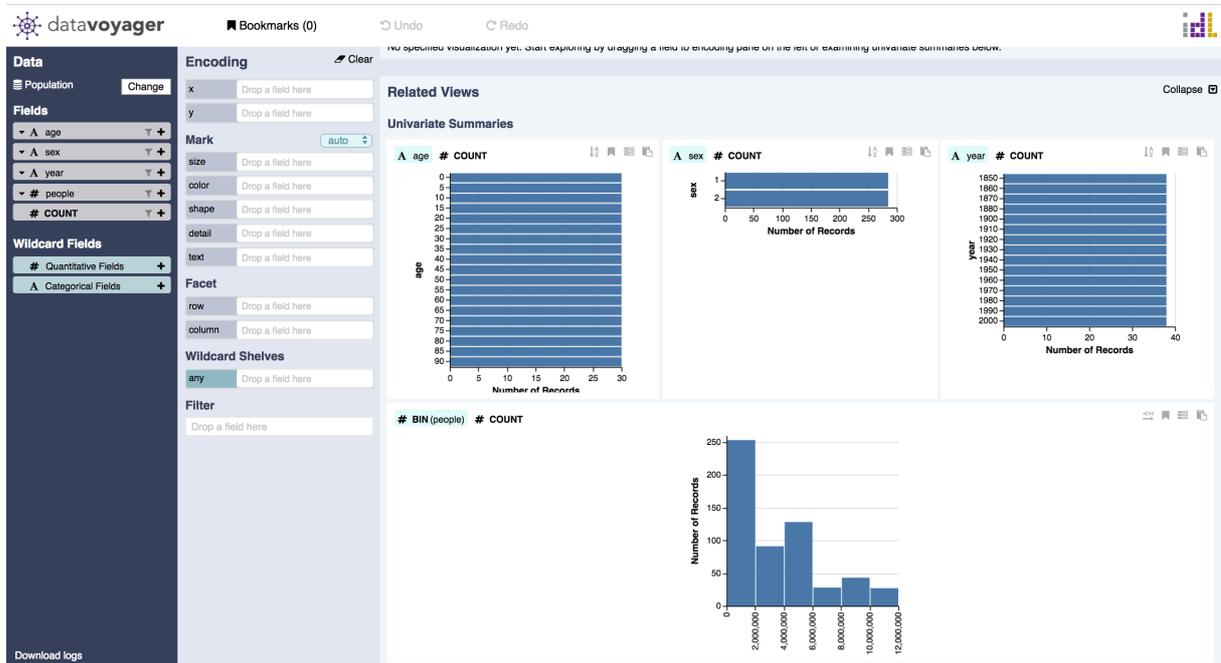


Figure 4 - Logo datavoyager (Datavoyager, 2015)

Datavoyager est une plateforme en ligne de visualisation de données. Elle a été créée de base pour un projet de recherche à l'Université de Washington (Vallandingham, 2015).

La plateforme offre la possibilité de visualiser des exemples de graphiques prédéfinis liés avec les fichiers importés. L'utilisateur peut avoir une visualisation immédiate sur l'ensemble de ses données importées. Cette option est avantageuse pour les personnes n'ayant aucune connaissance dans le domaine.

Figure 5 - Plateforme de Datavoyager avec un exemple de visualisation



Source : (Datavoyager, 2015)

Voici les étapes principales pour créer une visualisation :

1. Importer le fichier de données (accepte les fichiers CSV, les URL en JSON, CSV ou TSV ou le « copier-coller » des données dans un éditeur de texte)
2. Afficher immédiatement plusieurs exemples de graphiques avec les données
 - a. Les graphiques sont interactifs
 - b. Possibilité de revenir en arrière
 - c. Possibilité d'ajouter des commentaires
 - d. Possibilité de trier les données
3. Exporter la visualisation en appuyant sur un bouton qui copie en JSON (Vega-lite), enregistrer l'image grâce au clic droit ou télécharger tous les graphiques en une fois en CSV.

Pour conclure, cette plateforme se démarque de ses concurrents en affichant directement des exemples de graphiques avec les données importées liées à ceux-ci. Cela permet aux utilisateurs d'avoir rapidement une vue d'ensemble des données importées. En revanche, la suite, pour modifier les graphiques, devient un peu plus complexe. Il est malheureusement difficile de trouver de la documentation, car c'est un projet de recherche. Il est cependant possible de consulter un document scientifique décrivant l'outil (Wongsuphasawat, et al., 2017).

Cet outil se rapproche énormément de ce qu'on souhaite offrir sur notre plateforme : le fait d'afficher automatiquement les données chargées sous plusieurs exemples de visualisation. Cela permet à l'utilisateur d'avoir directement une vue d'ensemble des *data sets* chargés. Nous allons nous inspirer fortement de cette étape de visualisation globale sur notre projet.

Voici un petit tableau récapitulatif :

Tableau 3 - Résumé datavoyager

Datavoyager	
Site internet	https://github.com/vega/voyager
Création	2015
Origine	USA, Washington
Prix	Gratuit
Open Source	Oui
Librairies utilisées	Datalib, Vega-Lite, Vega-Lite-ui, zepto.js, angular.js
User Friendly	Non
Importation	CSV et URL en JSON, CSV ou TSV
Exportation	JSON (Vega-lite), PNG
Documentations	Non
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

- a) (Datavoyager, 2015)
- b) (Wongsuphasawat, et al., 2017)
- c) (Voyager 2 Github, 2018)

1.3.3 RAWGraphs

RAWGraphs

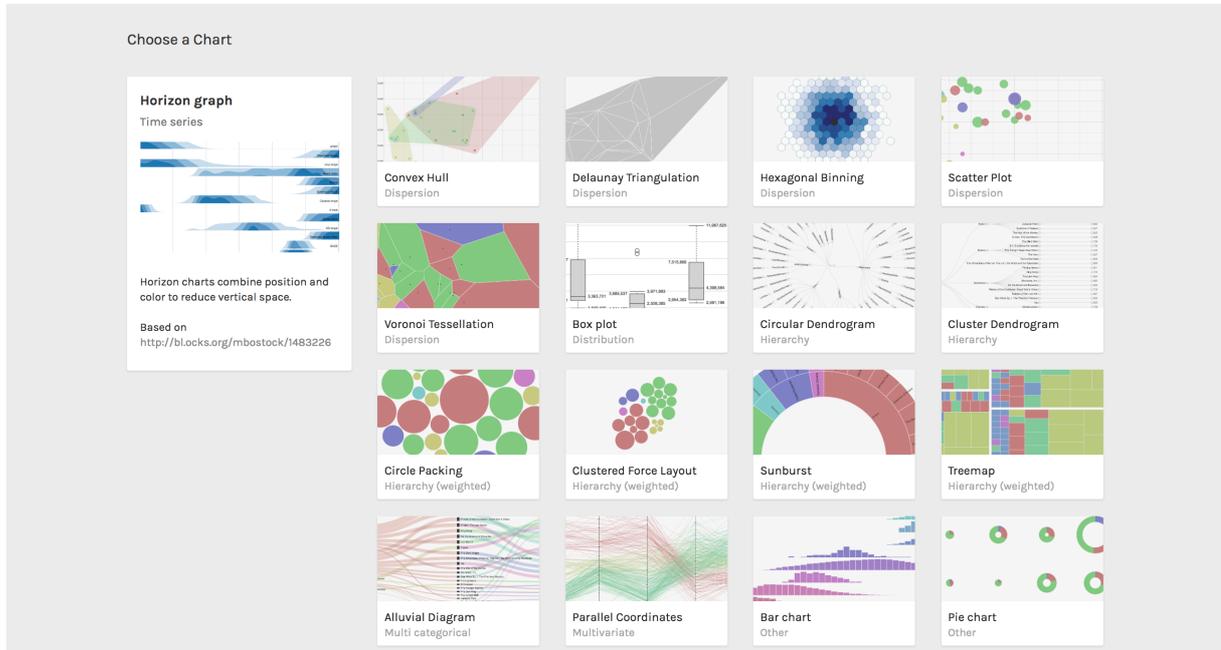
Figure 6 - Logo RAWGraphs (RAWGraphs, 2018)

RAWGraphs est une plateforme en ligne pour la visualisation de données. Leur but principal est de faciliter la représentation visuelle de données complexes pour tout le monde. Le projet est né au DensityDesign Lab en 2013, et est maintenant maintenu par le laboratoire en collaboration avec Calibro in Milan (Mauri, 2013).

La plateforme propose une gamme de style de graphiques personnalisables, c'est-à-dire que l'utilisateur choisit un style de design puis choisit l'agencement de celui-ci. De plus, il est possible de

créer de nouveaux graphiques personnalisés. RAWGraphs met à disposition le téléchargement local du produit. Pour finir, un de leurs points forts est qu'ils ne stockent pas les données importées.

Figure 7 - Plateforme de RAWGraphs avec des exemples de visualisation



Source : (Caviglia, Mauri, Uboldi, & Azzi, 2017)

Voici les étapes principales pour créer une visualisation :

1. Importer le fichier de données (accepte les fichiers CSV, TSV, DSV, JSON, fichiers Excel et les URL ou copier-coller les données dans un éditeur de texte)
2. Choisir un style de design de graphique
 - a. Possibilité de créer son propre graphique
3. Cartographier les données du graphique en choisissant les axes, couleurs, etc
 - a. Visuel immédiat juste en dessous
4. Exporter la visualisation au format vectoriel (SVG), images (PNG) et modèles de données (JSON).

Pour conclure, RAWGraphs est une très bonne plateforme, car elle est *user friendly*, elle est gratuite et s'adapte à l'utilisateur : celui-ci, ne connaissant pas très bien dans le domaine, peut quand même arriver à un résultat et le plus initié pourra assouvir ses envies en customisant les visualisations. De plus, elle ne stocke pas les données lors de l'importation de celles-ci. Par contre, les graphiques

sont très « colorés » et cela peut déplaire à certains. Nous allons nous inspirer du design très *user friendly* de la plateforme.

Voici un petit tableau récapitulatif :

Tableau 4 - Résumé de RAWGraphs

RAWGraphs	
Site internet	https://rawgraphs.io/
Création	2013
Origine	Milan
Prix	Gratuit
Open Source	Oui
Librairies utilisées	D3.js, angular.js, JQuery, is.js, sheet.js, bootstrap, ...
<i>User Friendly</i>	Oui
Importation	CSV, TSV, DSV, JSON, fichiers Excel et les URL
Exportation	SVG, PNG, JSON
Documentations	Oui
Graphiques interactifs	Non

Source : Tableau de l'auteur provenant de sources multiples

- a) (RAWGraphs, 2018)
- b) (Mauri, 2013)
- c) (Caviglia, Mauri, Uboldi, & Azzi, 2017)
- d) (Predictive Analytics Today Review Desk, 2018)

1.3.4 Datawrapper



Figure 8 - Logo Datawrapper (Datawrapper, 2018)

Datawrapper est une plateforme en ligne de visualisation de données. Elle a été lancée en mai 2012 en Allemagne par des journalistes pour des journalistes. Aujourd'hui, les salles de presse du monde entier, telles que ABC, New York Times, Daily Nation, etc, utilisent Datawrapper pour construire leurs graphiques et leurs cartes (Datawrapper, 2018).

Datawrapper propose de créer des graphiques, mais aussi des cartes interactives. La plateforme est gratuite, mais il est possible de payer un abonnement mensuel selon les besoins de chacun

(freemium). De plus, l’outil est partiellement open-source, c’est-à-dire que certains plugins le sont, mais pas tous (Datawrapper Github, 2018).

Figure 9 - Prix des abonnements chez Datawrapper

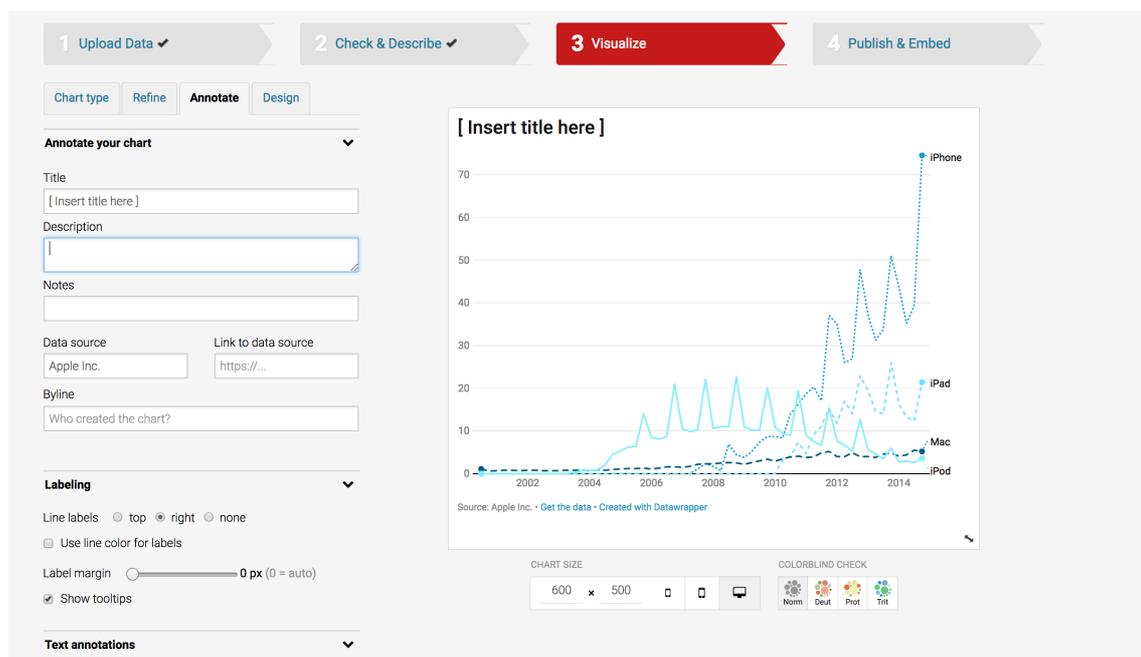
	Popular option				
	Single 10k free	Single Flat 29€/month	Team 129€/month	Custom 279€/month	Enterprise 879€+
User type	Occasional use for a small blog	Use daily	Professional use in a team	Full solution with custom branding	Fully customized solution with custom maps, print-export and CMS integration. Email us to learn more.
Monthly chart views	10,000	Unlimited	Unlimited	Unlimited	
Number of users	1	1	Unlimited	Unlimited	
Export as PNG image & PDF		✓	✓	✓	
Option: Without "Created with Datawrapper" Attribution			✓	✓	
We customize your chart according to your brand (fonts, colors, etc.)				✓	
Cancellation policy	Cancel anytime, your charts always stay online.				

Source : (Datawrapper, 2018)

Voici les étapes principales pour créer une visualisation :

1. Importer le fichier de données (accepte les fichiers CSV, les fichiers Excel, les données de Google Sheet et les URL ou copier-coller les données dans un éditeur de texte).
2. Vérifier et décrire les données :
 - a. Possibilité de les modifier.
 - b. Possibilité de choisir la langue des données (format des chiffres, des dates, etc).
3. Visualiser les graphiques :
 - a. Choisir un style de graphique.
 - b. Affichage immédiat du graphique choisi avec les données.
 - c. Choisir ce qu'on veut afficher selon les données.
 - d. Graphiques interactifs.
 - e. Possibilité de changer plusieurs paramètres (couleurs, axes, styles de trait, le titre, ajouter une description, les labels leurs placements, la grandeur, etc).
 - f. Possibilité de choisir la taille du graphique et choisir la résolution (téléphone, tablette, web)
4. Publier :
 - a. Envoi par email.
 - b. Cliquer sur un lien dans l'email.
 - c. Mettre un mot de passe ce qui crée obligatoirement un compte.
 - d. Puis, donne soit un lien web soit un code html.

Figure 10 - Plateforme de Datawrapper avec un exemple de visualisation



Source : (Datawrapper, 2018)

Pour conclure, Datawrapper est une excellente plateforme de visualisation pour les journalistes, mais, sans abonnement, elle est très rétreinte. Par exemple, pour la version gratuite, l'exportation du graphique est très compliquée et il n'est pas possible d'avoir la visualisation en PNG. De plus, comme les données importées sont stockées sur le site internet, cela peut déplaire à certaines personnes en raison de la confidentialité des données. Par contre, Datawrapper a énormément de documentation ce qui est très positif. Nous allons nous inspirer de l'importation des données et de la structure des étapes de la plateforme.

Voici un petit tableau récapitulatif :

Tableau 5 - Résumé Datawrapper

Datawrapper	
Site internet	https://www.datawrapper.de/
Création	Mai 2012
Origine	Allemagne
Prix	<i>Freemium</i>
Open Source	Certains plugins mais pas tous
Librairies utilisées	Dw.js, d3-light.js, RaphaelJS
<i>User Friendly</i>	Oui
Importation	CSV, fichiers Excel, Google Sheet et les URL
Exportation	Email et dessus un lien web soit un code html
Documentations	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

- a) (Datawrapper, 2018)
- b) (Datawrapper Github, 2018)
- c) (Predictive Analytics Today Review Desk, 2017)
- d) (Médiacademie, 2013)
- e) (Lorenz)

1.3.5 Tableau



Figure 11 - Logo Tableau (Tableau, 2018)

Tableau est un logiciel pour la visualisation de données. L'entreprise a été fondée en 2003 par trois amis de Stanford. « Au départ, il y avait un brillant informaticien, un professeur primé et un dirigeant d'entreprise ingénieux et passionné par les données. Ils ont uni leurs talents pour relever l'un des plus grands défis logiciels : permettre aux utilisateurs ordinaires de comprendre leurs données. Voilà comment a commencé Tableau Software » (Tableau, 2018).

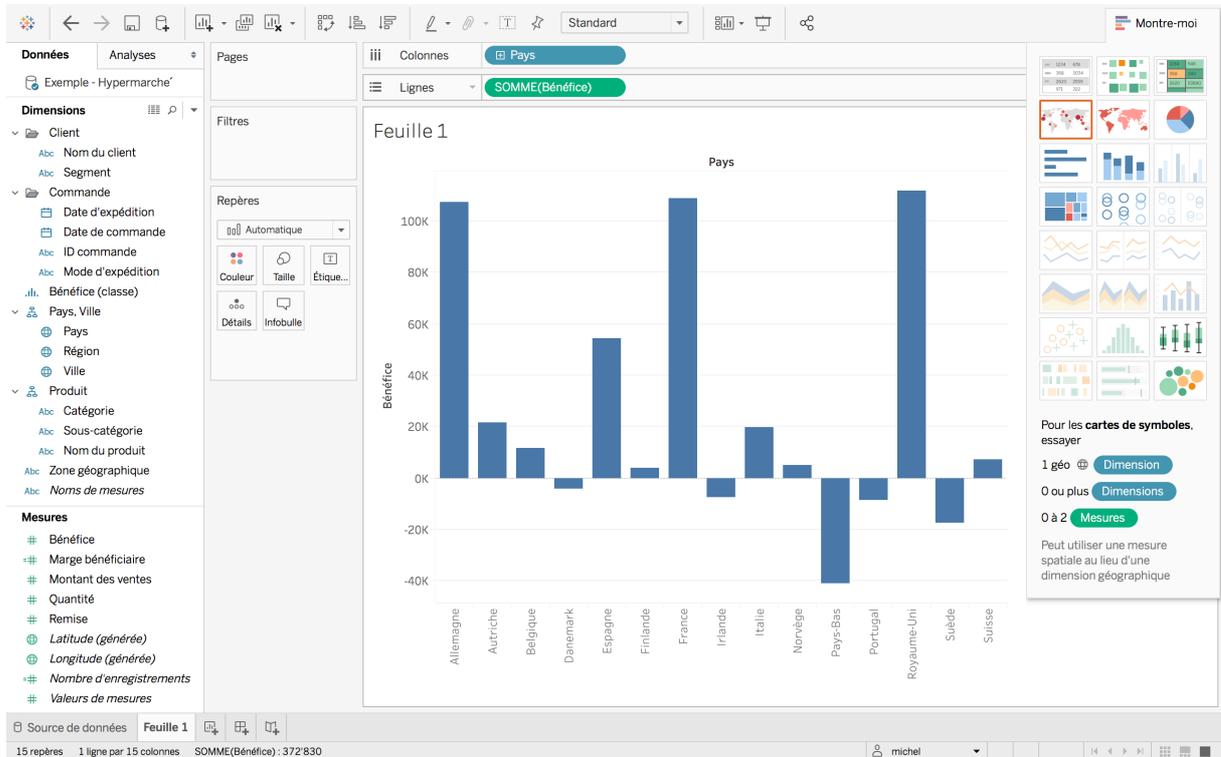
La société propose plusieurs outils :

- Tableau Desktop permet la visualisation de données pour tout public :
 - Période d'essai gratuite de 14 jours (Tableau, 2018).
 - 70 dollars par mois pour les utilisateurs individuels (la licence comprend aussi Tableau Server, Tableau Prep).
- Tableau Prep permet la préparation des données :
 - Période d'essai gratuite de 14 jours (Tableau, 2018).
 - 70 dollars par mois pour les utilisateurs individuels (la licence comprend aussi Tableau Server, Tableau Desktop).
- Tableau Server permet la visualisation de données pour les entreprises :
 - Entre 13 dollars et 70 dollars par mois selon les besoins (Tableau, 2018).
 - Essai gratuit avec une version d'évaluation (Tableau, 2018).
- Tableau Online permet la visualisation de données pour les entreprises dans le cloud :
 - Entre 15 dollars et 70 dollars par mois selon les besoins (Tableau, 2018).
 - Essai gratuit en s'inscrivant.
- Tableau Public permet la visualisation de données via une application :
 - Gratuit.

Nous avons testé Tableau Public. Voici les étapes principales pour créer une visualisation :

1. Télécharger l'application sur l'ordinateur.
2. Importer un fichier Excel, texte, PDF, JSON.
3. Visualisation des données sous forme de tableau.
4. Création manuellement de visualisation :
 - a. Choisir un style de graphiques ou de cartes interactives.
5. Exportation sous Tableau en fichier twbx.

Figure 12 - Exemple de visualisation sur Tableau Public



Source : (Tableau, 2018)

Pour conclure, Tableau est un excellent logiciel qui vise des publics cibles très variés avec ses différents types d'outils. Cependant, il faut quand même avoir quelques compétences dans le domaine de l'analyse de données. De plus, il faut télécharger un logiciel ce qui peut déplaire à certaines personnes. Et pour finir, l'exportation des graphiques n'est pas variée. Cependant, nous allons quand même nous en inspirer pour le design lors de la création des visualisations : L'affichage des dimensions du *data set* importé. C'est une fonction est intéressante car l'utilisateur pourra avoir un aperçu global sur les données.

Voici un petit tableau récapitulatif :

Tableau 6 - Résumé Tableau

Tableau	
Site internet	https://www.tableau.com/
Création	2003
Origine	USA (Seattle)
Prix	<i>Freemium</i>
Open Source	Non
Librairies utilisées	
<i>User Friendly</i>	Non
Importation	Fichiers Excel, texte, PDF, JSON, données spatiales
Exportation	Tableau (twbx)
Documentations	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

a) (Tableau, 2018)

b) (Predictive Analytics Today Review Desk, 2017)

1.4. Conclusion

Par le biais de cette analyse, nous avons remarqué qu'à présent, il existe deux sortes de plateformes. La première possibilité qui la plus utilisées est lorsque les données sont chargées et que l'utilisateur peut créer de toutes pièces les graphiques. Cette option est pratique pour les utilisateurs ayant des compétences dans le domaine de l'analyse de données, mais pour des personnes qui n'en ont pas, il est compliqué de comprendre le fonctionnement de l'outil. La deuxième possibilité est lorsque les données sont chargées, la plateforme génère automatiquement plusieurs propositions de graphiques. L'utilisateur peut avoir une visualisation immédiate sur l'ensemble de ses données importées. Cette option est avantageuse pour les personnes n'ayant aucune connaissance dans le domaine. Nos utilisateurs cibles sont des personnes ayant des lacunes dans le domaine de l'analyse de données, c'est pourquoi nous allons opter pour cette dernière façon de faire.

Nous avons pu remarquer qu'il est très important d'avoir une bonne documentation pour les utilisateurs afin qu'ils puissent utiliser et comprendre la plateforme rapidement. Nous allons donc faire attention de créer une documentation claire et précise.

De plus, les étapes de la création d'une visualisation doivent être simples, car si elles sont trop complexes, les utilisateurs seront vite perdus dans leur démarche. Nous allons faire attention à réaliser une plateforme simple d'utilisation et *user friendly*.

En ce qui concerne le projet, nous allons le faire gratuit et open source, car nous trouvons important que celui-ci soit accessible à tous.

Pour finir, nous avons fait un petit tableau récapitulatif des 5 outils analysés.

Tableau 7 - Récapitulatif de l'État de l'art

	Data-Illustrator	Datavoyager	RAWGraphs	Datawrapper	Tableau
Création	Avril 2018	2015	2013	Mai 2012	2003
Origine	USA	USA, Washington	Italie, Milan	Allemagne	USA (Seattle)
Open Source	Non	Oui	Oui	Partiellement	Non
Prix	Gratuit	Gratuit	Gratuit	Version gratuite et payante (<i>freemium</i>)	Version gratuite et payante (<i>freemium</i>)
Librairies utilisées	Paper.js, RxJs, D3.js, Datalib, Backbone.js	Datalib, Vega-Lite, Vega-Lite-ui, zepto.js, angular.js	D3.js, angular.js, JQuery, is.js, sheet.js, bootstrap	Dw.js, d3-light.js, RaphaelJS	
Documentations	Non mais articles scientifiques	Non mais articles scientifiques	Oui	Oui	Oui
Importation	CSV	CSV et URL en JSON, CSV ou TSV	CSV, TSV, DSV, JSON, fichiers Excel et les URL	CSV, fichiers Excel, Google Sheet et les URL	Fichiers Excel, texte, PDF, JSON, données spatiales
Exportation	SVG	JSON (Vega-lite), PNG	SVG, PNG, JSON	Email et dessus un lien web soit un code html	Tableau (twbx)
Graphiques interactifs	Non	Oui	Non	Oui	Oui

Source : Tableau de l'auteur provenant de sources multiples

- a) (Tableau, 2018)
- b) (Data Illustrator, 2018)
- c) (Datavoyager, 2015)
- d) (Datawrapper, 2018)
- e) (RAWGraphs, 2018)

2. ANALYSE ET CHOIX TECHNOLOGIQUES

De nos jours, il existe des milliers de librairies utilisant JavaScript. Le choix de celles-ci doit être assez rigoureux. Il est indispensable de faire une analyse minutieuse, car la tendance est souvent de se fier à l'aspect de l'interface, alors qu'il faut faire attention à ce qui se cache derrière. En effet, il est primordial que la librairie soit bien structurée et qu'elle ait de la documentation afin de bien la comprendre pour la maîtriser au mieux.

Dans cette analyse, nous allons faire un rapport de certaines librairies JavaScript de visualisation afin de relever les aspects positifs et négatifs de celles-ci. Nous allons premièrement expliquer ce que fait la librairie et, deuxièmement, faire un petit tableau récapitulatif avec des critères de sélection. Ils porteront surtout sur la documentation fournie avec la librairie. Si elle est open-source ? Quel est son prix ? Quelles sont ses performances techniques ? Quelle est sa facilité de mise en œuvre ? Quelle est la possibilité de la personnalisation des classes ?

2.1 Librairies

Une librairie ou bibliothèque logicielle regroupe un ensemble de fonctions déjà paramétrées. L'avantage est que les développeurs n'ont pas besoin de coder leur programme de toutes pièces. En effet, ils peuvent installer des librairies qui ont un « annuaire » de fonctions leur évitant de tout devoir implémenter. Une librairie permet, par exemple, de convertir un fichier CSV en JSON. Il aurait été possible de créer une méthode de toutes pièces implémentant cette fonction, mais pourquoi perdre du temps à tout coder alors qu'une librairie existante propose cette fonctionnalité en quelques lignes de codes.

Selon la définition d'une bibliothèque logicielle, trouvée sur le site internet techno-science.net l'intérêt des bibliothèques réside dans le fait qu'elles contiennent du code utile qui n'est pas nécessaire de réécrire chaque fois. Les bibliothèques logicielles sont parfois regroupées en un framework (définition à la page 37), de façon à constituer un ensemble cohérent et complémentaire de bibliothèques (techno-science, 2018).

2.2 JavaScript



Figure 13 - Logo JavaScript (PluralSight, 2018)

JavaScript (JS) est un des premiers langages de programmation orienté Web qui a été inventé par Brendan Eich en 1995 (Pardanaud & Sébastien, 2017). JavaScript programme le comportement des pages web en HTML en exécutant les requêtes (scripts) au niveau du navigateur (client) et non au

niveau du serveur web. C'est-à-dire que l'ordinateur de l'utilisateur recevra les instructions du code JavaScript en chargeant la page web et ceci permettra d'ajouter des petites animations sur le site internet (InfoWebMaster Glossaire, 2018).

Selon w3schools et le site internet de JavaScript, c'est un langage de programmation facile à apprendre (W3Schools, 2018). De plus, JavaScript possède une liste incalculable de librairies, Frameworks et plugins et donc nous trouvons énormément de documentation à propos de ceux-ci.

Lors de l'implémentation d'animations sur un site internet, le CSS propose une syntaxe très simple, mais limitée. Il ne permet pas de combler toutes nos envies. C'est pourquoi il est important d'utiliser JavaScript afin de subvenir à ses besoins. Pour ce projet, nous allons principalement travailler sur la visualisation. Il est donc astucieux d'utiliser du JavaScript.

2.3 Critères de sélection

Grâce à cette analyse, nous avons ressorti les points importants afin de choisir avec quelles librairies nous allons développer cette plateforme. Nous avons listé les aspects qui nous semblent les plus cohérents afin d'avoir un résultat final cohérent.

Les notes sont attribuées pour des personnes ne connaissant pas vraiment le domaine de l'analyse et la visualisation de données. Par exemple, la facilité de mise en œuvre devra être compréhensible naturellement en lisant les instructions claires et précises de la librairie. Si tel est le cas, la note de 3 sera attribuée.

Les points sont attribués de 0 - 3 dans les tableaux de récapitulation :

- 0 = aucune information
- 1 = peu
- 2 = moyen
- 3 = parfait

L'année de création de la librairie est un point intéressant à connaître, car grâce à cela il est possible de savoir si elle est fiable ou non. En effet, si une librairie date de 2018, elle aura encore beaucoup de bugs par rapport à une autre qui date de 2009 et qui aura plus d'expérience.

L'origine est également insérée dans ce tableau afin de savoir d'où vient l'entreprise de la librairie.

Il est important que la librairie soit open-source afin de pouvoir avoir accès au code dans le but de l'utiliser et peut-être de le modifier.

Le prix est aussi un critère de sélection afin de savoir s'il faudra investir de l'argent dans ce projet ou non et si oui le justifier.

Avoir des graphiques prédéfinis est assez intéressant car ainsi il ne faudra pas coder de toutes pièces les visualisations.

Avoir des graphiques interactifs est intéressant à avoir, car ils permettent aux utilisateurs de mieux comprendre les visualisations.

La documentation est importante afin de pouvoir comprendre et utiliser facilement la librairie. (0 = aucune documentation, 3 = énormément de documentation).

Les performances techniques sont importantes afin de savoir jusqu'à quel point la librairie peut nous amener. Certaines ne proposent que quelques points et d'autres beaucoup plus. Par exemple, le fait d'avoir des visualisations responsives, le type de données pouvant être importées et exportées, les navigateurs internet supportant les visualisations, etc. (0 = aucune performance, 3 = énormément de performances techniques)

La facilité de mise en œuvre est pour notre part très importante, car nous n'avons pas énormément de temps pour l'implémentation de la plateforme. Il est donc fondamental de choisir une librairie qui soit facile d'installation et d'apprentissage (0 = impossible à installer, 1 = très dur à installer, 3 = très facile à installer).

La personnalisation est importante afin d'avoir la liberté, par exemple, de changer l'aspect des graphiques proposés et de les personnaliser selon nos besoins. (0 = pas possible de personnaliser, 3 = une grande liberté de personnaliser).

Un tableau récapitulatif avec une note finale sera réalisé lors de la conclusion de cette analyse.

2.3.1 D3.js



Figure 14 - Logo D3.js (D3.js, 2017)

D3.js (ou D3) est une librairie de JavaScript qui a pour but de transformer des données sous forme visualisation. D3 a été créé en 2011 par le successeur de Protovis, qui lui a été créé en 2009 à l'université de Stanford. En quelque sorte, D3 donne vie aux données en utilisant HTML, SVG et CSS afin de créer des visualisations.

La bibliothèque est open source et gratuite. Elle ne fournit pas de graphique prédéfini, mais il existe énormément de documentation sur la création de visualisation avec celle-ci.

Tableau 8 - Résumé D3.js

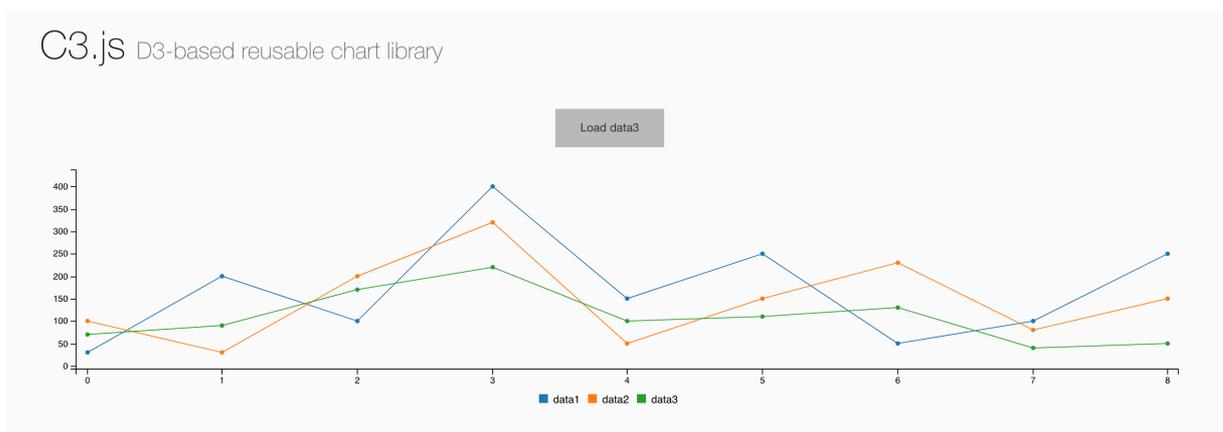
D3.js	
Site internet	https://d3js.org/
Création	2011
Origine	USA
Prix	Gratuit
Open Source	Oui
Documentations	3
Performances techniques	3
Facilité de mise en œuvre	2
Personnalisation	3
Graphiques prédéfinis	Non
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

- a) (D3.js, 2017)
- b) (D3 Github, 2018)
- c) (Heer, 2017)
- d) (Marty, The D3/Vega "stack", 2016)

2.3.2 C3.js

Figure 15 - Exemple de visualisation C3.js



Source : (C3.js, 2014)

C3.js est une librairie basée sur D3.js, et qui facilite son utilisation. En effet, grâce à C3.js ce n'est pas obligatoire d'écrire du code en D3. Cependant, pour aller plus loin, il est possible de créer et modifier des classes. Par contre, vu la simplicité de celle-ci, les fonctionnalités sont moins complètes que D3. Les visualisations étant très simples aussi cette librairie convient pour des

utilisateurs voulant des graphiques classiques et épurés. Et, pour finir, la documentation est très complète.

Tableau 9 - Résumé C3.js

C3.js	
Site internet	http://c3js.org/
Création	2014
Origine	USA
Prix	Gratuit
Open Source	Oui
Documentations	2
Performances techniques	1
Facilité de mise en œuvre	2
Personnalisation	2
Graphiques prédéfinis	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples
 a) (C3.js, 2014)
 b) (C3js Github, 2018)

2.3.3 Plotly



Figure 16 - Logo Plotly (Plotly, 2017)

Plotly est l'une des bibliothèques les plus connues qui sont basées sur D3.js. Depuis novembre 2015, la bibliothèque est open-source. C'est possible de l'utiliser en JavaScript, Python ou encore en R. Le point fort de Plotly est au niveau des points techniques mais les visualisations sont un peu moins *user-friendly* que les autres bibliothèques. La documentation est très complète et plusieurs tutoriels sont proposés pour différents types de graphiques. Le point négatif est que c'est une bibliothèque payante. Voir ci-dessous la liste des prix.

Figure 17 - Prix des abonnements de Plotly

STUDENT	PERSONAL	PROFESSIONAL
\$59	\$396	\$948
per year, per user	per year, per user	per year, per user
SIGN UP	SIGN UP	SIGN UP
Includes:	Includes:	Includes:
2500 Charts	2500 Charts	Unlimited Charts
Private or Public Charts, Dashboards and Slide Decks	Private or Public Charts, Dashboards and Slide Decks	Private or Public Charts, Dashboards and Slide Decks
Community Support for Online Workspace	Chat Support for Online Workspace	Chat and Phone Support for Online Workspace
SVG, EPS, HTML, PDF, PNG and JPEG Export	SVG, EPS, HTML, PDF, PNG and JPEG Export	SVG, EPS, HTML, PDF, PNG and JPEG Export
1000 Chart and Image Saves per day	1000 Chart and Image Saves per day	10,000 Chart and Image Saves per day
Community support for Falcon Sql Client	Chat support for Falcon Sql Client , <i>not including</i> support for persistent connections or non-SQL databases	Chat and phone support for Falcon Sql Client , <i>including</i> support for persistent connections or non-SQL databases

Source : (Plotly, 2017)

Tableau 10 - Résumé Plotly

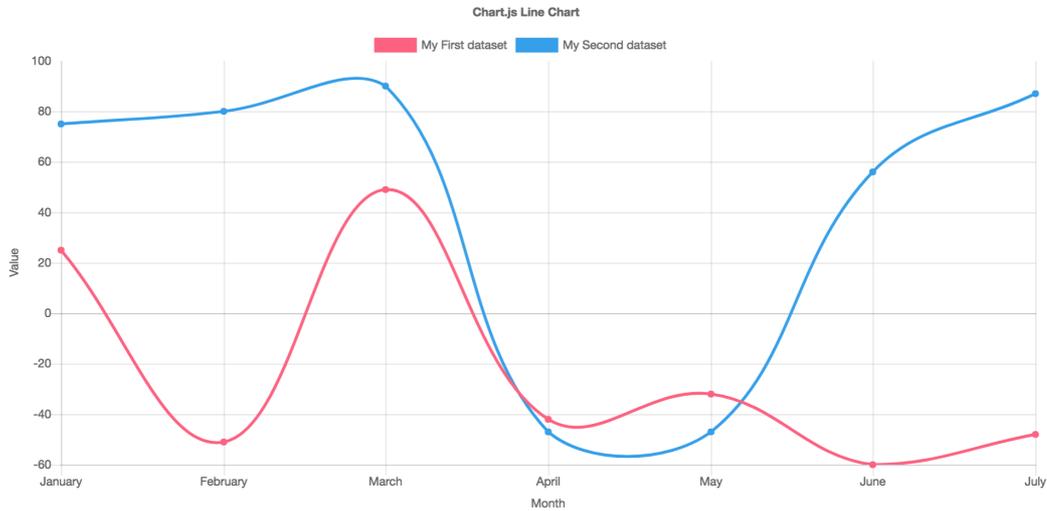
Plotly	
Site internet	https://plot.ly/
Création	2015
Origine	Canada
Prix	Payant
Open Source	Oui
Documentations	3
Performances techniques	3
Facilité de mise en œuvre	1
Personnalisation	3
Graphiques prédéfinis	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

- a) (Plotly, 2017)
- b) (Plotly Github, 2018)

2.3.4 Chart.js

Figure 18 - Exemple de visualisation Chart.js



Source : (Chart.js)

Chart.js est une librairie de JavaScript open-source pour la conception de visualisation. Les graphiques sont responsifs et flat-design. La librairie utilise HTML5. Chart.js propose 8 styles de visualisations qui sont customisables. Un des avantages de celle-ci est qu'il est possible de n'importer que l'un des graphiques. Le programme sera donc moins lourd. En revanche, la documentation peut être améliorée.

Tableau 11 - Résumé Chart.js

Chart.js	
Site internet	https://www.chartjs.org/
Création	2013
Origine	
Prix	Gratuit
Open Source	Oui
Documentations	1
Performances techniques	1
Facilité de mise en œuvre	2
Personnalisation	2
Graphiques prédéfinis	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

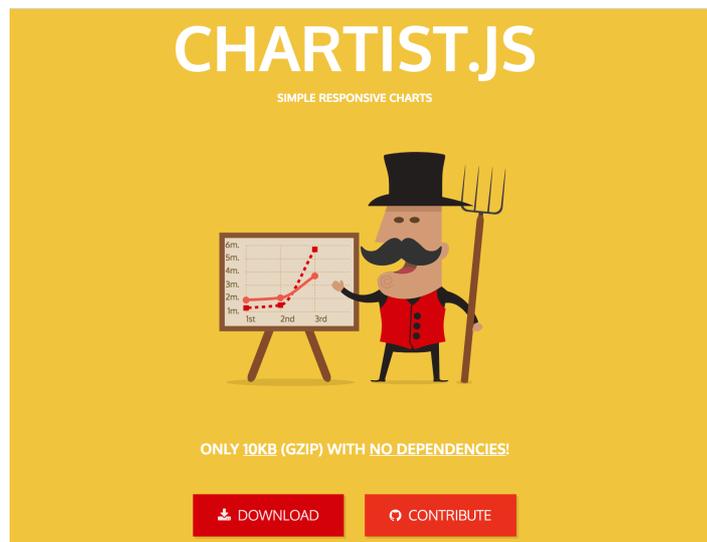
a) (Chart.js, 2018)

b) (Chart.js Github, 2018)

2.3.5 Chartist.js

Chartist est une librairie open-source basée sur JavaScript et sur SVG pour afficher les graphiques. Ils sont donc responsifs. Chartist propose de personnaliser des animations qui fonctionnent souvent sur les navigateurs internet les plus récents. De plus, il est possible d'implémenter des animations plus déterminées avec SMIL. Il est facile de mettre en œuvre la librairie et la documentation fournie est très complète.

Figure 19 - Site internet de Chartist.js



Source : (Chartist.js, s.d.)

Tableau 12 - Résumé Chartist.js

Chartist.js	
Site internet	https://gionkunz.github.io/chartist-js/
Création	
Origine	
Prix	Gratuit
Open Source	Oui
Documentations	2
Performances techniques	2
Facilité de mise en œuvre	2
Personnalisation	2
Graphiques prédéfinis	Oui
Graphiques interactifs	Non

Source : Tableau de l'auteur provenant de sources multiples

- a) (Chartist.js, s.d.)
- b) (Chartist.js Github, 2017)

2.3.6 HighCharts JS



Figure 20 - Logo de HighCharts JS (Highcharts, 2018)

HighCharts JS (ou HighCharts) fait partie de la compagnie Highsoft basée en Norvège avec 20 employés à plein temps et 8 à l'étranger. C'est une librairie open-source basée sur du JavaScript et sur SVG qui a vu le jour en 2009. HighCharts propose la création de graphiques interactifs et responsifs et il est possible de les customiser. Il existe un module d'exportation pour exporter les graphiques au format PNG, JPG, PDF ou SVG. Un avantage de HighCharts est sa compatibilité avec des anciens navigateurs internet tel que Internet Explorer 6. De plus, la documentation proposée sur leur site internet est très complète. Selon leur site, HighCharts est utilisée par plus de 80% des plus grandes entreprises du monde (Highcharts, 2018).

HighCharts est gratuite pour les projets à buts non commerciaux. Par contre elle est payante pour les sites Web ou projets commerciaux (plus d'information sur leur site internet <https://shop.highsoft.com/highcharts>) (Highcharts, 2018).

Tableau 13 - Résumé Highcharts

HighCharts JS	
Site internet	https://www.highcharts.com/
Création	2009
Origine	Norvège
Prix	Gratuit et payant
Open Source	Oui
Documentations	3
Performances techniques	2
Facilité de mise en œuvre	3
Personnalisation	2
Graphiques prédéfinis	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

- a) (Highcharts, 2018)
- b) (Highcharts Github, 2018)

2.3.7 Vega

Selon son site internet, VEGA est une « grammaire de visualisation » : Elle transforme les données en visualisations au format JSON. Puis, il est possible de générer les graphiques en HTML ou en SVG. La librairie fournit le chargement des données, leur transformation (filtre, calcul, etc), transforme les données (chiffre) en visuel (ex. couleur). VEGA est open-source et donc gratuit.

Jeff Heer cite « Vega n'est PAS destiné à remplacer D3. D3 est intentionnellement un système de bas niveau. Lors de la conception initiale de D3, nous l'avons même appelé "noyau de visualisation" plutôt que "toolkit" ou "framework". En plus de la conception sur mesure, D3 est conçu comme couche de support pour les outils de visualisation de niveau supérieur. Vega est l'un de ces outils qui s'appuie fortement sur D3 dans sa mise en œuvre » (Heer, 2017).

Comme dit ci-dessus VEGA transforme les données en JSON. Voici un exemple de la forme :

```
{
  "data": [...],
  "scales": [...],
  "axes": [...],
  "marks": [...]
}
```

- Data : les données à visualiser :
 - Transform : Regroupe, filtre les données.
- Scales : Transforme les données (chiffre) en visuel (ex. couleur).
- Axes : Définition des axes X et Y.
- Marks : Design du graphique et les éléments textes de la visualisation.

Tableau 14 - Résumé de Vega

VEGA	
Site internet	https://vega.github.io/vega/
Création	
Origine	
Prix	Gratuit
Open Source	Oui
Documentations	3
Performances techniques	3
Facilité de mise en œuvre	3

Personnalisation	3
Graphiques prédéfinis	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples

- a) (Vega, s.d.)
- b) (Vega Github, 2018)
- c) (Marty, 2016)
- d) (Satyanarayan, The reactive vega stack: declarative interaction design for data visualization, 2017)

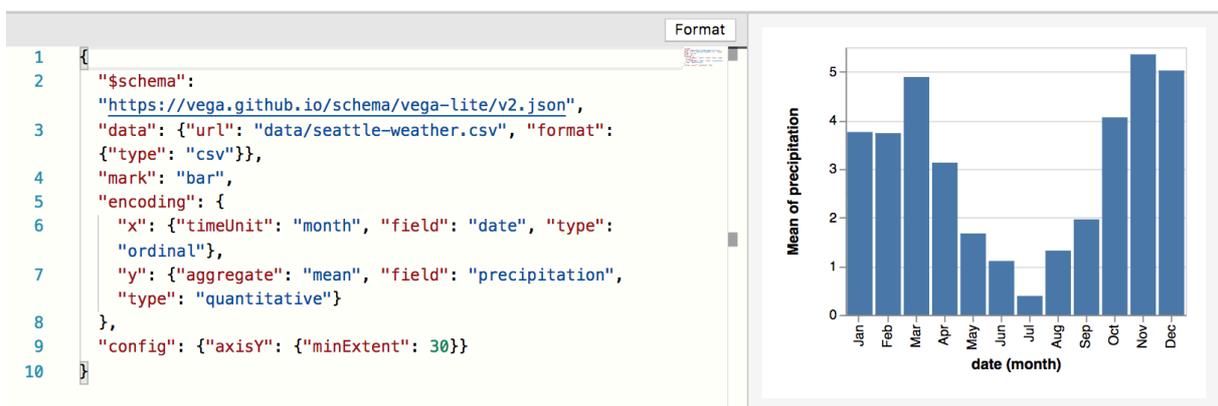
2.3.8 Vega-Lite

Vega-Lite est une librairie JavaScript de niveau supérieur, basée sur Vega et D3 qui permet de créer des visualisations de données interactives. Le format des visualisations est en JSON (Medium, 2016).

Vega et D3 permettent de faire des visualisations plus complexes que Vega-Lite, mais ceci a des conséquences. En effet, pour créer des graphiques avec ces deux librairies, il est nécessaire d'avoir une implémentation complexe, cependant avec Vega-Lite il n'y a que quelques lignes de code.

Vega-Lite propose des exemples graphiques déjà implémentés. L'avantage est qu'ils sont facilement personnalisables en modifiant les quelques lignes de codes, qui contrairement à Vega en a beaucoup. La librairie offre la possibilité de transformer les données, les filtrer, les trier (Satyanarayan, Moritz, Wongsuphasawat, & Heer, 2017). De plus, Vega-Lite propose un éditeur en ligne permettant de créer des visualisations instantanées ce qui est très pratique pour voir rapidement le résultat des graphiques.

Figure 21 - Exemple de visualisation Vega-Lite



Source : (Vega-Lite, s.d.)

Tableau 15 - Résumé Vega-Lite

Vega-Lite	
Site internet	https://vega.github.io/vega-lite/
Création	
Origine	
Prix	Gratuit
Open Source	Oui
Documentations	3
Performances techniques	3
Facilité de mise en œuvre	3
Personnalisation	2
Graphiques prédéfinis	Oui
Graphiques interactifs	Oui

Source : Tableau de l'auteur provenant de sources multiples
a) (Vega-Lite, s.d.)
b) (Vega-Lite Github, 2018)

2.4 Conclusion

Après avoir analysé plusieurs libraires JavaScript, nous avons pu remarquer qu'il existe énormément de librairies sur le marché. Nous en avons choisi quelques-unes afin d'approfondir notre savoir sur celles-ci.

Tableau 16 - Récapitulatif des libraires JavaScript

	D3.js	C3.js	Plotly	Chart.js	Chartist.js	HighCharts	Vega	Vega-Lite
Création	2011	2014	2015	2013		2009		
Origine	USA	USA	Canada			Norvège		
Open Source	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Prix	Gratuit	Gratuit	Payant	Gratuit	Gratuit	Gratuit et payant	Gratuit	Gratuit

Graphiques prédéfinis	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Graphiques interactifs	Oui	Oui	Oui	Oui	Non	Oui	Oui	Oui
Documentations	3	2	3	1	2	3	3	3
Performances techniques	3	1	3	1	2	2	3	3
Facilité de mise en œuvre	2	2	1	2	2	3	3	3
Personnalisation	3	3	3	2	2	2	3	2
TOTAL	11	8	10	6	8	10	12	11

Source : Tableau de l'auteur provenant de sources multiples

- a) (Rahman, 2015)
- b) (SINGHAL, 2015)
- c) (Baaaj, 2017)

Pour conclure, selon l'analyse faite ci-dessus, nous choisissons d'utiliser la librairie Vega-Lite. En effet, c'est celle qui nous paraît être la plus adaptée pour notre plateforme. Lorsque l'on regarde le tableau récapitulatif, ce n'est pas la librairie qui obtient le plus de points. La librairie Vega a un meilleur score. Vega-Lite étant une version simplifiée de Vega nous la choisissons car elle n'utilise que très peu de lignes de code, contrairement à Vega. De plus, après une analyse pointue des deux, nous avons remarqué que Vega-Lite remplissait plus notre objectif d'avoir une structure de l'architecture simple et épurée. Ceci est très important pour la suite, car c'est le début du projet VASCO et les développeurs qui continueront ce travail devront comprendre facilement l'architecture logicielle.

3. Technologies utilisées

3.1 Langages de programmation

Nous allons décrire les différents langages de programmation que nous avons utilisés pour l'implémentation de la plateforme.

3.1.1 HyperText Markup Language

HyperText Markup Language (HTML) ou en français langage de balisage hypertexte, est un langage pour la création de pages web. Son nom est minutieusement choisi car il est structuré avec des balises (ex. `<body></body>`). Elles permettent d'indiquer la façon dont doit être présenté le document et les liens qu'il établit avec d'autres documents (Comment ça marche, 2018).

Les navigateurs n'affichent pas les balises HTML, mais les utilisent pour rendre le contenu de la page (w3schools, 2018).

Le langage HTML a été créé par Tim Berners-Lee en 1989, mais a été réellement considéré comme langage en 1993 et cette première version s'appelait HTML 1.0 (Comment ça marche, 2018). La dernière version, sortie en 2014, et qui est utilisée actuellement, s'appelle HTML5.

Afin d'améliorer ses fonctionnalités HTML fonctionne, la plupart du temps, avec d'autres langages tels que CSS, qui permet de mettre page le contenu d'une page web, JQuery qui est une librairie JavaScript, etc. JavaScript dépend du HTML car sans lui il ne peut fonctionner.

3.1.2 Pug

Pug est un moteur de template de haute performance pour Node.js et qui utilise du JavaScript. Pug fait la même chose que HTML, mais sans les balises et sans utiliser le code en ligne. Le but de ce langage est que l'indentation (l'alignement) soit faite parfaitement.

Ce projet était autrefois connu sous le nom de "Jade". Cependant, il s'est révélé que "Jade" est une marque déposée ; par conséquent un changement de nom était nécessaire. Après quelques discussions entre les responsables, "Pug" a été choisi comme nouveau nom pour ce projet. À partir de la version 2, "pug" est le nom officiel du paquet (Pugjs Github, 2018).

Nous utilisons Pug dans notre projet à la place de HTML, car c'est un langage lié à Node.js. De plus, l'un de nos objectifs principaux est d'apporter un grand soin à l'architecture logicielle, car ce n'est qu'un prototype et, par la suite, des développeurs continueront à l'implémentation du projet. Ils doivent donc comprendre facilement notre code et Pug apporte une clarté sur celui-ci.

3.1.3 Cascading Style Sheets

Cascading Style Sheets (CSS) ou en français feuilles de style en cascade est un langage permettant de mettre en page le contenu des pages internet utilisant du HTML.

Son but principal est de séparer la structure d'un document HTML et sa présentation. En effet, avec HTML, on peut définir à la fois la structure (le contenu et la hiérarchie entre les différentes parties d'un document) et la présentation. Mais cela pose quelques problèmes. Avec le couple HTML/CSS, on peut créer des pages web où la structure du document se trouve dans le fichier HTML tandis que la présentation se situe dans un fichier CSS (Futura Tech, 2018).

3.1.4 Node.js



Figure 22 - Logo Node.js (Node.js, s.d.)

Node.js est un serveur web open source qui permet d'utiliser JavaScript sur le serveur. La plateforme a été créée en 2009 avec le but d'utiliser du JavaScript afin de bénéficier de la rapidité de celui-ci pour le développement des sites web dynamiques.

Pourquoi ne pas utiliser PHP ? Les deux sont pour la programmation *back-end* mais Node.js a l'avantage de faire le lien entre le *back-end* et le *front-end* car les deux utilisent JavaScript. De plus, Node.js a un système très rapide. C'est l'une des raisons pour laquelle nous allons l'utiliser, car la plateforme doit charger des données et les assembler rapidement afin d'afficher des graphiques. Par contre, l'apprentissage de Node.js est un peu plus complexe que PHP, mais dans le monde de l'informatique il est important d'en connaître au moins les principes de base (Chrzanowska, 2017).

3.2 Bibliothèques

Nous allons décrire les différentes bibliothèques que nous avons utilisées pour l'implémentation de la plateforme.

3.2.1 JQuery

JQuery est une bibliothèque de JavaScript qui permet de simplifier celle-ci ainsi que d'avoir une meilleure rapidité. Leur slogan est « Write less, do more » (jQuery, 2018).

jQuery récupère plusieurs lignes de code JavaScript et les résume en très peu de lignes, voire une seule. Ainsi, jQuery améliore la rapidité et la performance d'un programme.

3.2.2 Convert-csv-to-json

Convert-csv-to-json est une librairie JavaScript utilisée avec Node.js. Comme son nom l'indique, elle permet de convertir des fichiers CSV en JSON. Nous avons installé cette librairie, car nous avons utilisé CompassQL qui doit recevoir des données au format JSON.

3.2.3 CompassQL

CompassQL est un langage de requête pour les visualisations de Vega permettant de choisir les graphiques adéquats selon les différents paramètres fournis lors de l'implémentation du code. CompassQL utilise le format JSON afin d'afficher les bons types de graphiques (CompassQL Github, 2018).

Ci-dessous voici la différence entre une structure Vega-Lite et CompassQL. Vous pouvez remarquer que dans CompassQL nous avons mis un point d'interrogation pour le paramètre « mark » qui représente le style de graphique. Par le biais de ce signe, CompassQL saura quel type de visualisation il faudra choisir (Wongsuphasawat, et al., 2016).

Vega-Lite :

```
vlspec = {
  "data": {"url":
"data/exemple.json"},
  "mark": bar,
  "encoding": {
    "x": {
      "field": AxeX,
      "type": ordinal,
    },
    "y": {
      "field": AxeY,
      "type": ordinal
    }
  }
};
```

CompassQL :

```
cqlspec = {
  "spec": {
    "data": {"url":
"data/exemple.json"},
    "mark": "?",
    "encodings": [
      {
        "channel": "x",
        "field": AxeX,
        "type": ordinal
      },{
        "channel": "y",
        "field": AxeY,
        "type": ordinal
      }
    ]
  }
};
```

CompassQL semble un peu plus complexe que Vega-Lite, mais il est très utile pour notre projet. En effet, nous ne pouvons pas prédire le type de données que l'utilisateur va importer, donc la librairie va faire le choix par elle-même afin d'afficher des visualisations les plus adéquates possibles.

3.2.4 Vega-tooltip

Vega-tooltip est un plugin de Vega. Il permet de faire interagir les visualisations en temps réel. Lorsqu'un utilisateur passe la souris sur les données du graphique, Vega-tooltip affiche les informations liées au point sélectionné (Vega-tooltip Github, s.d.).

Nous avons trouvé pertinent d'utiliser cette fonction pour que l'utilisateur comprenne plus facilement la signification des données dans le graphique.

3.2.5 Vega-embed

Vega-Embed est une librairie de Vega. Sur les pages Web elle facilite l'intégration des vues interactives sur Vega-Lite (Vega-embed Github, 2018). Grâce à Vega-embed, il est possible d'ajouter des liens d'exportation en SVG, PNG, JSON et d'ouvrir l'éditeur de Vega. Elle permet aussi de charger les données grâce à du JSON ou d'une URL. De plus, elle permet d'utiliser Vega-tooltip et pour finir, d'utiliser des styles de thèmes tels qu'un fond noir, une visualisation en mode Excel, etc (Vega-themes Github, 2018).

3.2.6 Express.js

Express.js est une infrastructure d'applications Web Node.js minimaliste et flexible qui fournit un ensemble de fonctionnalités robustes pour les applications Web et mobiles (expressjs, 2018). Elle permet de gérer les liens entre les différentes pages d'un site internet.

3.2.7 Multer

Multer est une librairie JavaScript utilisée avec Node.js servant à gérer les données, et surtout le transfert de fichiers.

Nous avons eu besoin d'installer Multer, car la première étape de la plateforme est d'importer un fichier CSV. Grâce à la librairie nous avons pu le transférer à la page suivante.

3.2.8 Express-session

Express-session est un module Node.js qui permet de créer des variables de session unique. Nous avons utilisé cette librairie pour renommer les fichiers uploadés lors de l'importation de données.

3.3 Framework

Un framework ou en français cadre de travail ou encore infrastructure logicielle est une structure logicielle ou conceptuelle qui permet de gérer toutes les couches de la création d'un projet. C'est une base sur laquelle repose le logiciel ou l'application informatique en d'autres mots, c'est le squelette du projet. Il est composé d'un ensemble d'outils et de composants logiciels (JDN, 2018).

Le but principal d'un framework est de faciliter l'implémentation (coder) pour les développeurs en leur proposant une structure « prête à l'emploi » ayant plusieurs composants déjà installés afin d'avoir une bonne base pour commencer le travail.

3.3.1 Materialize

Materialize, créé et conçu par Google, est un framework moderne, basé sur du *front-end* ayant comme avantage d'être responsif. Il s'agit un langage de design qui combine les principes classiques d'un design réussi avec l'innovation et la technologie. L'objectif de Google est de développer un système de conception qui permet une expérience utilisateur unifiée pour tous leurs produits et sur n'importe quelle plateforme (Materialize, 2018).

Nous avons choisi ce Framework, car il est facile d'utilisation et propose une riche documentation.

4. Définition des besoins

Avant d'implémenter la plateforme, il est important de faire une étude des besoins des futurs utilisateurs afin d'être sûr de créer un produit remplissant dûment les attentes de ceux-ci.

Pour cela, nous avons ressorti les *user stories* principales de notre *Product Backlog* afin de voir les différentes fonctionnalités souhaitées pour la création de la plateforme. Pour ce projet, nous avons trois types de personnes : Le développeur, l'administrateur et l'utilisateur.

4.1 Fonctionnalités souhaitées par le développeur

Le développeur est la personne qui va implémenter tout le projet. Il est important que les fonctionnalités qu'il souhaite soient correctement accomplies. Ci-dessous vous trouverez les *user stories* principales pour le développeur :

- En tant que développeur, je veux choisir une technologie adéquate au projet afin de pouvoir implémenter la plateforme avec les technologies les plus adaptées.
- En tant que développeur, je veux installer et configurer mon environnement de travail afin de pouvoir implémenter la plateforme.
- En tant que développeur, je veux semi-automatisée la création de graphiques afin de faciliter la visualisation des données.
- En tant que développeur, je veux avoir une structure de l'architecture correctement implémentée afin que d'autres développeurs comprennent l'architecture facilement.
- En tant que développeur, je veux déployer la plateforme afin que la plateforme devienne publique.

4.2 Fonctionnalités souhaitées par l'administrateur

L'administrateur est la personne aux commandes du travail qui gère tout le projet. Il a de grandes responsabilités. Il est donc très important qu'il atteigne ses objectifs. Ci-dessous vous trouverez les *user stories* principales pour l'administrateur :

- En tant que administrateur, je veux faire un cahier des charges afin d'avoir une visualisation globale sur le travail de Bachelor.
- En tant que administrateur, je veux faire un *Product Backlog* afin d'avoir une visualisation globale des objectifs et besoins du projet.
- En tant que administrateur, je veux faire un état de l'art afin de savoir quelles sont les technologies existantes.

- En tant que administrateur, je veux créer des mockup afin d'avoir une visualisation globale sur la plateforme finale.
- En tant que administrateur, je veux avoir un accès au *repository* du projet afin de pouvoir récupérer tous les fichiers du projet.
- En tant que administrateur, je veux faire tester ma plateforme afin de faire des améliorations et d'éviter les bugs.

4.3 Fonctionnalités souhaitées par les utilisateurs

Il est très important de considérer toutes les fonctionnalités souhaitées par les utilisateurs, car sans eux, la plateforme ne sert à rien, étant donné que s'ils ne l'utilisent pas, le projet n'a plus aucun intérêt.

L'un des objectifs principaux à atteindre est que l'utilisateur puisse comprendre facilement et rapidement la plateforme. Ci-dessous vous trouverez les *user stories* principales pour les utilisateurs :

- En tant que utilisateur, je veux importer des ensembles de données afin d'avoir une vue d'ensemble des données.
- En tant que utilisateur, je veux pouvoir voir plusieurs graphiques afin de sélectionner celui qui me convient le mieux.
- En tant que utilisateur, je veux pouvoir modifier la visualisation afin d'améliorer la compréhension du graphique.
- En tant que utilisateur, je veux pouvoir exporter le graphique afin de pouvoir l'utiliser sans être sur la plateforme.

5. MODÉLISATION ET CRÉATION DE LA PLATEFORME

5.1 MOCKUP

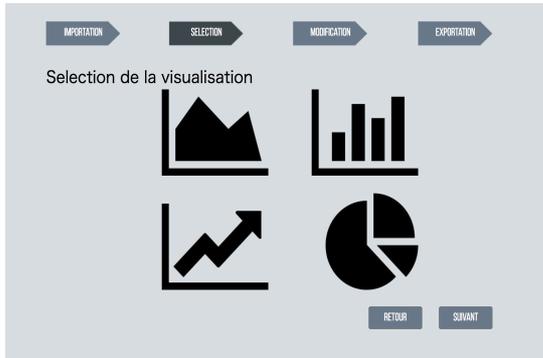
Mockup est un terme utilisé dans le monde du design et de l'informatique représentant le croquis d'un projet. Il est souvent réalisé lors de la phase de conception du travail. Le but est d'avoir un aperçu global sur le produit final avant même d'avoir débuté l'implémentation. De plus, il offre la possibilité d'avoir un exemple de l'emplacement des différents éléments ainsi que l'interaction entre les boutons et les pages. Et pour finir, il permet de considérer la complexité du projet et donc d'estimer le temps nécessaire pour la réalisation de la plateforme (Info Webmaster Glossaire, 2018).

Pour la création des mockup de ce projet, nous avons utilisé le logiciel Axure RP (Axure, 2018). Pour arriver à des mockup satisfaisants, nous avons dû faire plusieurs versions.

Pour commencer, nous avons réalisé une interprétation sur papier afin d'avoir une vue globale sur les différentes pages de la plateforme. Puis, nous avons créé une première version sur Axure RP qui a été discutée avec le professeur. Lors de ce partage, plusieurs points essentiels ont été argumentés tels que la modification de l'étape « Sélection ». Il faut montrer des visualisations grâce à la librairie CompassQL qui génère le type idéal de graphique. L'idéal est d'afficher le plus d'exemples possibles de visualisation. De plus, nous avons trouvé pertinent d'afficher une liste des noms, des dimensions sur le côté avec différentes mesures telles que le type, la somme, la moyenne et un petit visuel. Grâce à cela, l'utilisateur aura plus facilement une vue d'ensemble sur les données qu'il a importées.

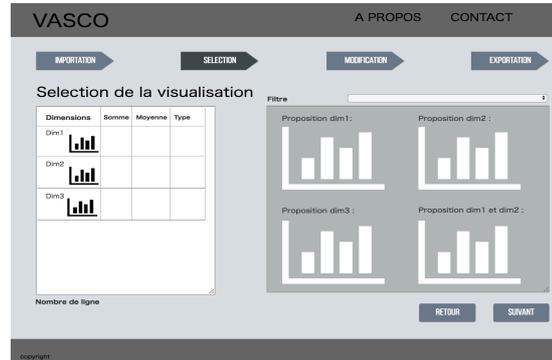
Et pour finir, nous avons modifié l'étape « Modification » afin d'avoir la possibilité de changer le style de graphique. Voici ci-dessous la différence des deux versions Axure RP qui sont susceptibles de changer, car ce ne sont que des croquis :

Figure 23 - Étape de Sélection version 1



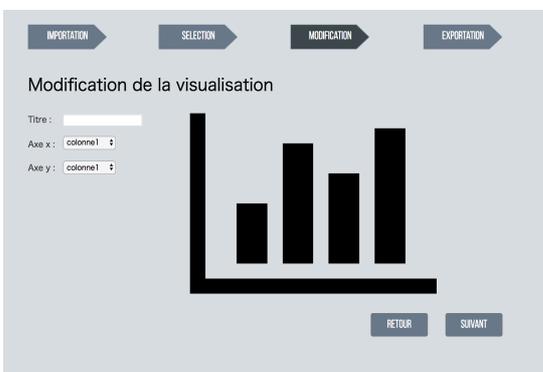
Données de l'auteur

Figure 24 - Étape de Sélection version 2



Données de l'auteur

Figure 25 - Étape de Modification version 1



Données de l'auteur

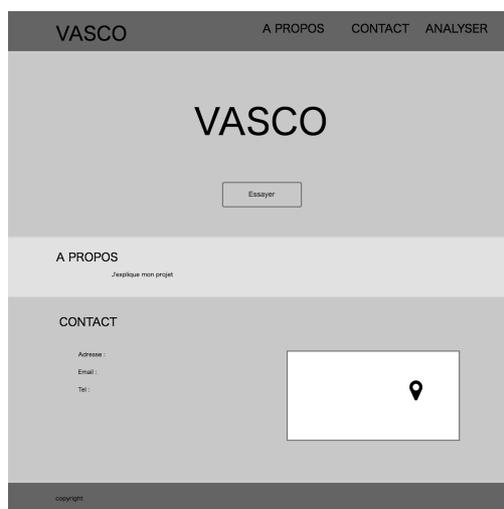
Figure 26 - Étape de Modification version 2



Données de l'auteur

Nous avons aussi créé un mockup pour une page d'accueil vitrine pour que les utilisateurs aient directement un aperçu du projet lorsque qu'ils iront sur la plateforme.

Figure 27 - Site vitrine VASCO



Données de l'auteur

Après avoir réalisé cette deuxième version de mockup, nous avons commencé à implémenter le projet afin de voir si il était possible de développer ces croquis en pratique.

5.2 Sélection de données

5.2.1 Format CSV

Lors de l'implémentation de la plateforme, nous avons choisi le format de données que nous voulions gérer. Nous avons trouvé pertinent de manipuler les données CSV, car elles sont très compliquées à comprendre en l'état brut. En effet, un fichier CSV est composé de données séparées par des virgules. La première ligne est composée des dimensions (nom des colonnes dans un tableau) puis, en-dessous, chaque ligne représente les données relatives à toutes les dimensions.

Figure 28 - Exemple d'un fichier CSV des meilleures entreprises dans l'informatique entre 2014 et 2015

```
Employer Name,Early Career Median Pay,Mid-Career Median Pay,Median Age,Years of Experience,Years with Company,% Female,% High Job Satisfaction,% High Job Meaning,% High Job Stress
HP,65400,104200,38,7,7,6,3,32,64,41,56
Oracle,71400,126400,37,9,5,3,29,64,32,56
IBM,72700,112700,36,7,4,7,1,25,59,40,62
Microsoft,102500,139700,33,6,4,4,8,27,72,59,58
Qualcomm,94400,132500,33,7,7,4,3,20,69,50,62
Cisco,83800,134700,33,6,7,4,23,76,60,52
eBay,88900,130000,33,6,9,2,7,43,78,55,66
Samsung,65900,143000,33,6,2,3,37,64,36,69
Intel,85800,129400,32,5,5,5,25,72,66,61
Adobe,91900,141900,31,5,2,3,3,28,84,24,47
Apple,91300,145400,31,6,2,7,31,73,74,66
Amazon,81900,130400,30,5,2,2,26,69,42,66
Google,106900,151600,30,5,2,2,30,89,72,53
Salesforce,84800,154600,30,5,3,1,8,30,89,54,55
Tesla,81400,118500,30,5,6,1,6,20,70,89,70
LinkedIn,110800,159600,29,5,0,42,0,0,0
SpaceX,78500,0,29,6,1,2,3,14,73,92,80
Facebook,116800,149300,29,4,3,1,1,32,96,78,44
```

Données de l'auteur

5.2.2 Type de données

Nous avons dû faire un choix sur le type de données que nous voulions manipuler. Nous avons fait le choix de ne gérer que les données quantitatives et qualitatives. Nous avons décidé de ne pas prendre en compte les données relatives au temps (ex. 2018-07-04 13:44) ni celles de type géographique avec les longitudes et latitudes.

Après avoir pris cette décision, nous avons choisi des exemples de fichiers CSV avec des données quantitatives et qualitatives. Nous avons testé la plateforme avec un fichier de données trouvées sur Github à propos de céréales (voir annexes). Dans ce fichier, il y avait le nom des céréales, le nom du fabricant, le nombre de calories, le nombre de protéines, etc (Gupta, 2014).

Nous avons aussi choisi un fichier CSV avec des données à propos des Airbnb à Genève afin de savoir quelle quantité d'appartements étaient loués, à quels prix, quels types de logements (chambre privée, logement entier), etc (Inside Airbnb, s.d.) (voir les annexes).

Et pour finir, nous avons utilisé un fichier CSV avec des données qui indiquent les meilleures compagnies en informatique entre 2014 et 2015 (voir annexes). Le fichier contient le nom des

entreprises, le salaire en début de carrière, le salaire en fin de carrière, l'âge moyen, etc (PayScale, s.d.). Un aperçu des données se trouve dans les annexes.

5.2.3 Données quantitatives

Un type quantitatif représente des données qui peuvent être comptées ou mesurées. En général des données quantitatives sont des nombres (ex. 2 ou 2, 4).

5.2.4 Données qualitatives

Un type qualitatif représente des données qui peuvent être nommées. En général des données qualitatives ne peuvent être exprimées par des nombres (ex. marié, célibataire, veuf).

6. Tests utilisateurs

Le prototype de la plateforme fini, nous avons fait des tests afin de savoir si nous avons atteint notre objectif principal qui est celui de créer une plateforme compréhensible et facile d'utilisation. Le but de ce questionnaire est donc d'analyser la manière dont les personnes réagissent avec la plateforme.

Pour procéder aux tests, nous avons choisi deux types de personnes, celles qui ont un peu d'expérience dans l'analyse de données et celles qui n'en ont pas afin de voir les différentes réactions face à la plateforme.

La phase de test commence par une brève explication de notre part à propos du projet VASCO. Puis, nous donnons aux testeurs un scénario sur papier avec des questions auxquelles ils doivent répondre et si possible sans notre aide. Pendant ce temps nous analysons leur stratégie afin de pouvoir, par la suite, améliorer la plateforme selon les résultats obtenus.

6.1 Ressources nécessaires

Quelques ressources sont nécessaires afin que la phase de tests se déroule correctement. Les utilisateurs doivent par exemple avoir un ordinateur avec du wifi afin de pouvoir aller sur la plateforme. C'est aussi un moyen de savoir si le site internet est responsif sur des machines ayant des autres dimensions que les nôtres.

Les futurs utilisateurs doivent importer deux fichiers CSV différents afin de pouvoir répondre aux questions du formulaire. Elles leur sont envoyés par mail. Ces deux fichiers sont des données avec lesquelles nous avons travaillé lors de l'implémentation du projet.

6.2 Utilisateurs cibles

Nous voulons toucher toutes sortes de personnes. C'est pourquoi notre plateforme doit impérativement être compréhensible et facile d'utilisation. Nous avons donc ciblé deux types d'utilisateurs : Les personnes ayant un peu d'expérience dans l'analyse de données et les personnes non-initiiées à ceci.

La plateforme a été testée par 10 personnes. Nous avons commencé par un utilisateur « beta », c'est-à-dire une personne qui a fait le questionnaire en premier et grâce à qui, selon ses réponses et réactions nous avons pu l'adapter. Puis, nous avons fait faire le questionnaire à 5 personnes, dont 3 qui sont dans le domaine de l'informatique. Après cette première phase, nous avons changé un peu l'interface de la plateforme selon les résultats du questionnaire. Pour finir, 4 autres utilisateurs lambdas, dont une personne dans le domaine de l'analyse, ont testé le site internet modifié.

6.2.1 Utilisateurs dans le domaine de l'analyse de données

Les utilisateurs dans le domaine de l'analyse de données ont une moyenne d'âge de 24 ans. Nous avons eu 3 personnes venant du monde de l'informatique de gestion et 1 venant du monde de la climatologie et de l'océanographie.

6.2.2 Utilisateurs non-initiés

Les utilisateurs non-initiés ont entre 24 et 57 ans. Certains viennent du monde du tourisme et de l'événementiel et d'autres sont dans l'enseignement.

6.3 Type de tâches

Sur la base du document (Jeremy Boy, 2014), nous avons choisi 4 types de tâches pour créer les questions.

- Extrema: C'est une tâche qui permet de trouver un maximum ou un minimum selon des données.
- Variation: C'est une tâche qui permet de détecter des tendances, des similitudes ou des divergences dans les données.
- Average: C'est une tâche qui permet d'estimer une valeur moyenne.
- Comparaison: C'est une tâche qui permet de comparer différentes valeurs ou tendances.

Nous avons mis pour chaque question le type de tâche relative à celle-ci (voir tableau dans les annexes).

6.4 Degré de difficultés

Pour chaque question, nous avons fixé un niveau de difficulté afin d'estimer le temps dont les utilisateurs auront besoin pour y répondre. L'utilisateur bêta a confirmé nos choix dans les degrés de difficulté.

- Facile (moins de 10 secondes pour répondre).
- Moyen (moins de 20 secondes pour répondre).
- Difficile (30 secondes ou plus pour répondre).

6.5 Scénarios

Pour commencer, quelques questions générales sont posées telles que l'âge, le sexe, le métier et l'expérience dans l'analyse de données, ceci afin de cibler le style de la personne interrogée. Puis, nous leur demandons de télécharger sur leur ordinateur deux fichiers CSV afin de faire le scénario

avec ces données. Ensuite, ils doivent aller sur le site internet qui est en ligne et nous analysons la façon dont ils se familiarisent avec la plateforme tout en répondant au questionnaire. Celui-ci est composé de plusieurs parties. La première, citée ci-dessus, a des questions générales. La deuxième se rapporte à un fichier CSV en relation avec des céréales. La troisième porte sur des questions à propos de données concernant les meilleures entreprises informatiques entre 2014 et 2015. Pour finir, la dernière partie a pour but de savoir si cette plateforme leur serait utile ou non.

Le but de ce questionnaire est d'analyser la manière dont ils réagissent avec la plateforme et pour cela, nous les invitons à penser à « haute voix » afin de bien comprendre leurs réactions. Le questionnaire et le document d'analyse se trouvent dans les annexes.

6.6 Résultats de la première phase de test

Cette phase de tests a été très intéressante, car elle a soulevé des petits détails qui paraissent évidents aux utilisateurs réguliers de ce type de plateforme.

La phase de tests s'est déroulée en deux parties. Une première avec des personnes plus jeunes et qui ont un peu plus l'habitude de travailler sur l'ordinateur. Après cette première partie, nous avons fait les changements nécessaires proposés par les testeurs afin que la plateforme soit plus compréhensible. Puis, une deuxième phase de test a été faite avec des personnes travaillant un peu moins sur l'ordinateur afin de savoir si le site internet était devenu plus accessible.

Page d'importation

Sur cette page, nous avons juste ajouté un « placeholder » dans le *input* du fichier d'importation indiquant que c'est ici qu'il faut importer les données.

Figure 29 - placeholder de la page importation



Page de Sélection

Lors de la première phase, plusieurs points similaires d'amélioration sont ressortis par les utilisateurs.

Premièrement, nous avons remarqué que ce n'était pas adéquat de mettre le tableau récapitulatif en haut de la page et mettre la liste des graphiques en bas de la page. En effet, les utilisateurs ne

voyaient pas cette liste de graphiques et ne restaient que sur le tableau récapitulatif. Afin de résoudre ce problème, certaines idées sont ressorties telles que mettre le tableau et la liste l'un à côté de l'autre ou mettre deux onglets contenant chacun les éléments. Nous avons opté pour cette deuxième proposition (voir l'image sous « Changements de la plateforme » n° 1).

Les utilisateurs ne comprenaient et ne voyaient pas bien la ligne en dessous du tableau récapitulatif indiquant le nombre de lignes du fichier importé. Nous avons choisi de changer le texte de « Nombre de lignes » à « Nombre de lignes du fichier importé » et de le mettre au tout début. Ainsi les personnes verraient cette information directement (voir l'image sous « Changements de la plateforme » n° 2).

De plus, ils n'arrivaient pas à lire le petit histogramme inséré dans le tableau récapitulatif. C'est pourquoi nous avons mis une nouvelle fonctionnalité qui est de passer le curseur sur l'aperçu pour qu'il s'agrandisse et pouvoir ainsi le lire plus facilement.

D'autre part, les testeurs ne voyaient pas la liste déroulante pour trier la liste des graphiques et ne comprenaient pas son fonctionnement. Nous avons donc changé le texte de « Afficher les graphiques selon... » par « Je trie les données selon... » (voir l'image sous « Changements de la plateforme » n° 3).

Pour finir, la plupart des utilisateurs avaient un Mac et nous avons constaté que sur ces ordinateurs les « scroll bar » ne se voient pas. Ils avaient donc de la difficulté à voir que le tableau récapitulatif et la liste de graphiques pouvaient être « scrollés ». Pour remédier à ce problème, nous avons choisi d'enlever la « scroll bar » du tableau récapitulatif et de ne mettre que deux visualisations par ligne pour la liste de graphiques (voir l'image sous « Changements de la plateforme » n° 4).

Page de modification

Sur la page de modification, nous avons essayé d'agrandir au maximum le graphique sélectionné afin qu'il soit plus visible. De plus, nous avons ajouté un petit carré explicatif sur des fonctionnalités à propos des graphiques, dont la « toolbox » qui donne des détails sur celui-ci (voir l'image sous « Changements de la plateforme » n° 5).

Pour finir, nous avons remarqué que les utilisateurs ne savaient pas comment exporter. Ils ne voyaient pas le bouton « suivant » pour être redirigés sur la page exportation. Selon leur avis, nous avons donc changé le nom du bouton par « exporter » (voir l'image sous « Changements de la plateforme » n° 6).

Page d'exportation

Nous n'avons pas fait de changement sur cette page.

Vous trouverez ci-dessous un tableau récapitulatif de la première phase de tests.

Tableau 17 - Synthèse des résultats de la première partie des tests

Premiers testeurs						
Testeurs	1- bêta	2	3	4	5	6
Premières impressions	<ul style="list-style-type: none"> Comprend pas l'utilité du tableau récapitulatif et prend trop de place et du coup on ne voit pas la liste des graphiques qui est plus importante. Mettre des onglets pour séparer les deux parties ? Histogrammes illisibles, car trop petit. L'agrandir ? 	<ul style="list-style-type: none"> Interface jolie et compréhensible et assez facile à utiliser expliquer mieux le graphique dans modification 	<ul style="list-style-type: none"> Problèmes avec la scroll bar car sur son Mac elle n'est pas visible. → Mettre une icône avec une flèche pour dire de descendre bouton pour aller à la page d'exportation devrait être plus explicite. Expliquer mieux le graphique sur la page modification Jolie interface 	<ul style="list-style-type: none"> Histogrammes illisibles, car trop petit. Afficher que quelques graph et après avec le tri en afficher plus ? 	<ul style="list-style-type: none"> Mettre exporter sur le bouton à la place de suivant, car ce n'est pas clair Joli le menu avec les étapes 	<ul style="list-style-type: none"> Problèmes avec la scroll bar car sur son mac elle n'est pas visible Interface jolie et compréhensible après ne pense pas très utile
Deuxièmes impressions	<ul style="list-style-type: none"> Plus difficile les questions, mais connaît l'interface donc plus simple Bien qu'il y ait les titres sur les graphiques 	<ul style="list-style-type: none"> Connaissait l'interface donc plus facile. Trouve que l'histogramme est trop petit Long à charger les données 	<ul style="list-style-type: none"> Mettre deux filtres ? pour l'axe x et y L'histogramme trop petit Graphique sur modification est trop petit Mettre numéro avant chaque étape par ex 1 importation > 2 sélections, etc Enlever les titres de chaque page 	<ul style="list-style-type: none"> Plus facile la deuxième partie, car connaît l'interface Trop d'informations sur la page de sélection, il faudrait épurer par exemple avec des onglets 	<ul style="list-style-type: none"> Mettre titre au-dessus du tableau, car comprend pas (ex Synthèse du fichier importé) Trouve que le but de la plateforme est d'analyser et comprendre les données Changer le texte du filtre : Trié les données selon... Aime beaucoup le design de la plateforme 	<ul style="list-style-type: none"> Agrandir l'aperçu de l'histogramme, car illisible Pas facile de voir l'option du filtre et la toolbox User-friendly
Totaux des réponses						
Correct	10	11	10	13	11	11
Faux	1	2	4	2	2	2
Aidé	5	3	2	1	3	3

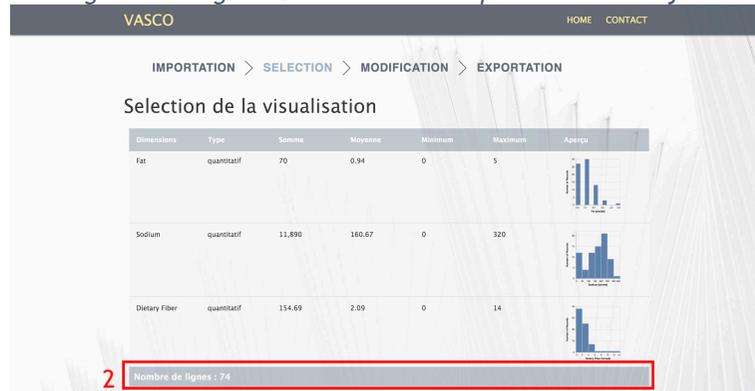
Données de l'auteur

6.7 Modifications apportées à l'interface

Ci-dessous vous pouvez voir par le biais d'images les changements qui ont été fait après la première phase.

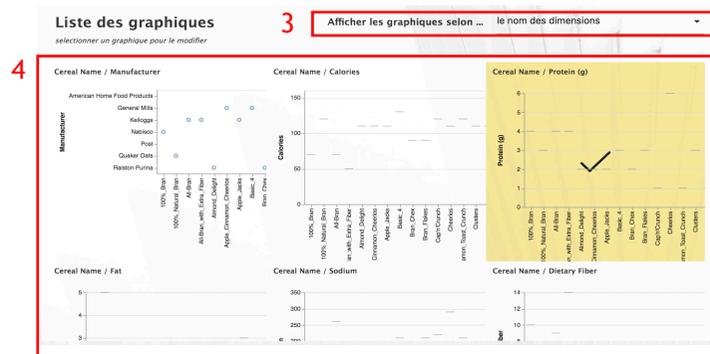
AVANT

Figure 30 - Page de Sélection avant la phase de tests - synthèse



Données de l'auteur

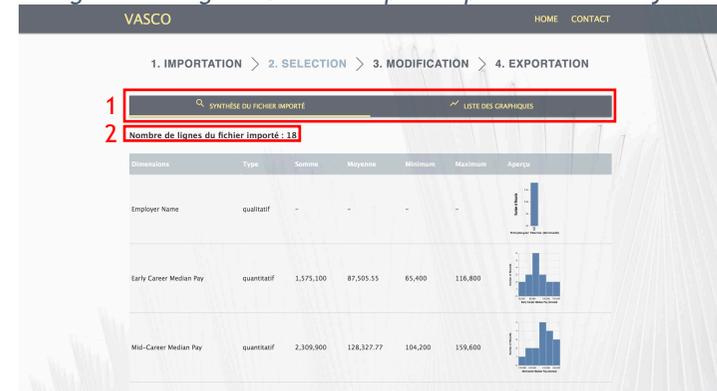
Figure 32 - Page de Sélection avant la phase de tests - liste des graphiques



Données de l'auteur

APRÈS

Figure 31 - Page de Sélection après la phase de tests - synthèse



Données de l'auteur

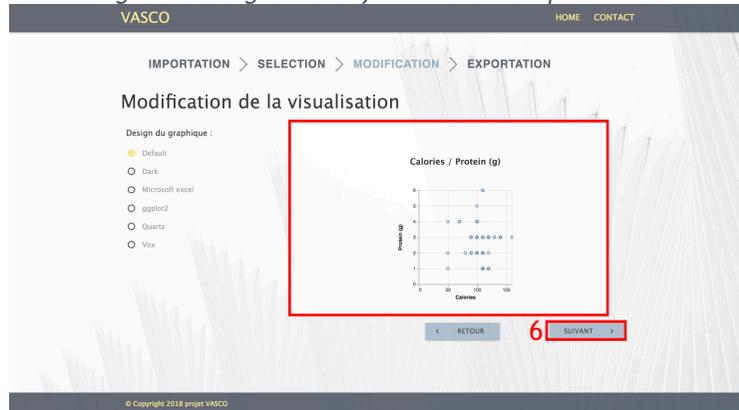
Figure 33 - Page de Sélection après la phase de tests - liste des graphiques



Données de l'auteur

AVANT

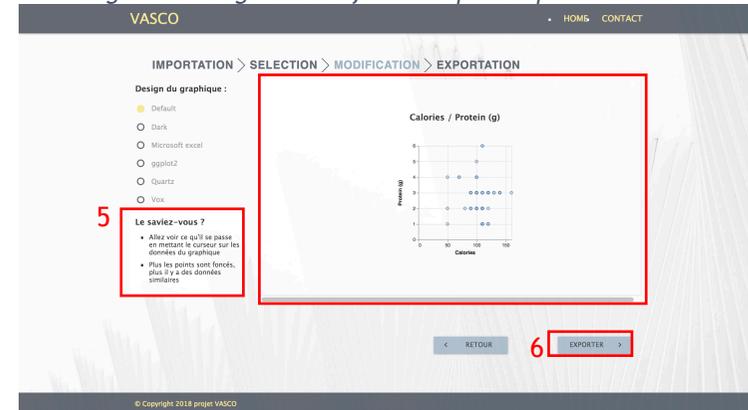
Figure 34 - Page de Modification avant la phase de tests



Données de l'auteur

APRÈS

Figure 35 - Page de Modification après la phase de tests



Données de l'auteur

6.8 Résultats de la deuxième phase de tests

Les changements nécessaires réalisés, nous avons à nouveau fait tester la plateforme à des personnes lambda qui ont appliqué le même scénario que la phase précédente. Nous avons remarqué qu'ils avaient beaucoup plus de facilité à trouver les éléments sur le site internet sans avoir besoin de notre aide. De plus, les utilisateurs de cette phase avaient moins l'habitude de travailler sur les ordinateurs.

Toutefois, nous avons encore modifié quelques petits détails tels que la responsivité des pages, un petit texte sur la ligne d'importation du fichier précisant que c'est ici qu'ils doivent mettre le document, la couleur de l'onglet sur la page de sélection et l'emplacement de la liste de recherche.

Vous trouverez ci-dessous un tableau récapitulatif de la deuxième phase de test.

Tableau 18 - Synthèse des résultats de la deuxième partie des tests

Deuxièmes testeurs				
Testeurs	7	8	9	10
Premières impressions	<ul style="list-style-type: none"> Onglets pas visibles, on dirait une barre avec des mots dessus → mettre plus large ? ou écrire étape 1 et étape 2 en dessus ? Prise par les questions donc regarde pas très bien l'interface Peut-être afficher deux filtres, car trop de possibilité lors de la recherche Rapprocher le texte et la dropdown, car ne comprend pas que c'est ensemble 	<ul style="list-style-type: none"> Mettre comme titre : « Liste complète des graphiques » Mettre icône descendre et monter pour montrer qu'il y a plusieurs graphiques dans la liste 	<ul style="list-style-type: none"> Comprend pas comment importer Sur son ordi le bouton « j'essaye » ne se voit pas très bien 	<ul style="list-style-type: none"> Afficher que l'onglet au début et lorsqu'on clique, les infos s'affichent pour elle très compréhensible, car l'habitude de travailler sur l'ordinateur
Deuxièmes impressions	<ul style="list-style-type: none"> Plus facile, car connaît comment l'interface fonctionne et ce n'est pas compliqué c'est juste la première mise en main qui est un peu difficile comme pour n'importe quel site internet 	<ul style="list-style-type: none"> Le texte est trop loin par rapport à la liste et n'est pas assez visible Changer la couleur des onglets afin de voir que c'est des onglets et non une barre avec des titres 	<ul style="list-style-type: none"> Interface lisible, mais ne travaille pas souvent sur l'ordinateur donc un peu de peine au début pour comprendre 	<ul style="list-style-type: none"> Bonne plateforme compréhensible Serait bien qu'elle accepte d'autres types de fichiers (temps, coordonnées) pour que ça soit utile pour elle. Long le chargement de donnée
Totaux des réponses				
Correct	12	11	15	16
Faux	1	1	0	0
Aidé	3	4	1	0

Données de l'auteur

6.9 Résultats finaux

Nous remarquons une différence entre les deux phases grâce aux changements apportés à l'interface. En effet, les totaux des tableaux récapitulatifs sont plus satisfaisants dans la deuxième partie.

Par le biais de ces tests, nous avons pu apporter des modifications nécessaires et ainsi réussir à remplir correctement notre objectif qui est de faire une plateforme compréhensible et facile d'utilisation.

7. Difficultés rencontrées

Lors de l'implémentation de la plateforme, nous avons été confrontés à plusieurs difficultés et problèmes qui ont ralenti le planning de notre projet. Nous avons dû faire des heures supplémentaires afin de résoudre ces problèmes et fournir le travail dans les temps.

7.1 Manque de documentation

Après avoir fait l'état de l'art et après discussions avec le professeur, nous avons choisi d'utiliser la librairie JavaScript Vega-lite qui nous paraissait la plus pertinente selon nos besoins. Elle a pu remplir tous nos objectifs, mais de longues heures ont été nécessaires afin de pouvoir comprendre cette librairie. En effet, bien que cette bibliothèque propose de très bons services, c'est encore une toute nouvelle librairie donc il n'y a pas beaucoup de documentations à part la leur. Lorsque nous rencontrions des problèmes, nous ne pouvions consulter que leur documentation, très bien faite, mais qui manque de temps en temps un peu de précision. Donc, pour l'implémentation, nous avons perdu beaucoup de temps à essayer de résoudre des problèmes sur lesquels nous ne pensions pas devoir passer autant de temps.

Pour la librairie CompassQL, nous avons eu le même problème du manque de documentation. En effet, nous n'avons pas trouvé d'autre documentation à part la leur et nous avons donc dû passer beaucoup de temps à comprendre son fonctionnement. Heureusement, nous avons reçu de l'aide de camarades qui avaient plus d'expérience dans l'implémentation. Ils ont pu nous assister et nous ont permis de mieux comprendre la librairie et la mettre en place.

7.2 Taille des fichiers

Pour créer les visualisations, nous avons utilisé la librairie Vega-lite et, en complément, pour afficher le style de graphique, la librairie CompassQL a été appliquée.

Le chargement est long, car nous avons fait en sorte de sélectionner toutes les possibilités des graphiques. C'est pourquoi nous avons dû restreindre une valeur maximum aux fichiers importés à 20 Ko. Pour l'instant ce n'est qu'un prototype mais si elle reste ainsi, la plateforme ne touchera qu'un petit marché de clients et ce n'est pas l'objectif du projet. Nous pensons donc que pour la suite de ce projet, ce problème doit être géré rapidement afin que la plateforme soit fonctionnelle et surtout utile pour nos futurs utilisateurs. Une des solutions pourrait être de mettre en mémoire cache les graphiques.

7.3 Données sélectionnées

Pour tester la plateforme, nous avons choisi trois fichiers CSV qui nous intéressaient. Lors de l'implémentation, nous avons choisi de ne gérer que les données qualitatives et quantitatives. C'est pourquoi nous nous sommes permis « d'épurer » un peu les fichiers. En effet, nous avons supprimé les colonnes ayant des données géographiques et temporelles ainsi que les colonnes ayant des identifiants car elles nous paraissaient insignifiantes pour ces graphiques. Pour finir, nous avons aussi supprimé des lignes du fichier Airbnb afin que le programme se charge plus rapidement. Il est important que pour la suite de ce projet qu'il n'y ait plus besoin de supprimer des données ainsi que de devoir réduire la taille des fichiers.

7.4 Valeurs manquantes

Dans les fichiers CSV, il y a souvent quelques valeurs manquantes. Pour l'instant nous avons un peu « triché » en remplaçant cette valeur absente par un 0. Nous avons dû changer cette valeur nulle, car, si on ne le faisait pas, notre tableau récapitulatif avait des problèmes à calculer la somme, moyenne, minimum et maximum. Pour la suite de ce projet, nous pensons qu'il est pertinent de gérer ces valeurs manquantes. Elles peuvent être sous différentes formes : Nulles, non définies ou vides. Il serait intéressant de sortir ces valeurs des fichiers importés afin qu'elles ne soient pas incluses dans les graphiques. De plus, il serait pertinent de les afficher à côté des graphiques sous forme de listes afin que l'utilisateur puisse avoir conscience que ces valeurs manquantes existent. Grâce à cela, les statistiques des visualisations seront beaucoup plus adéquates.

7.5 Variables de session

Pour commencer, lors de l'importation de fichiers, nous avons renommé en dur le nom du fichier CSV afin de le récupérer plus facilement. Mais, lors du déploiement de la plateforme, nous avons remarqué que ce n'était pas possible de le faire comme ainsi. En effet, plusieurs utilisateurs ne pouvaient pas utiliser la plateforme en même temps, car s'ils le faisaient, ils écrasaient chaque fois le fichier de l'autre utilisateur en important leur propre fichier. Nous avons géré ce problème avec des variables de session. Le but de celles-ci est de générer un identifiant unique pour chaque utilisateur lors de l'importation de fichiers et grâce à cela les anciens fichiers ne sont plus écrasés par le nouveau importé. Après de grandes modifications dans le code, nous avons réussi à gérer ce problème et maintenant il est possible que plusieurs utilisateurs utilisent le site internet en même temps sans perdre leurs données.

CONCLUSION

L'objectif principal de ce travail était de créer un prototype d'une plateforme publique de visualisation de données compréhensible et facile d'utilisation. Cet objectif a été atteint en mettant le site sur un serveur de la HES-SO. Mais surtout, des utilisateurs lambdas ont testé la plateforme et ont confirmé la simplicité de celle-ci.

Ce travail a été divisé en deux grandes parties principales : La phase d'analyses et la phase de développement. La première comprenait un état de l'art technique qui avait pour but d'analyser les plateformes et programmes similaires à notre projet. Il en est ressorti que nous voulions créer une plateforme qui proposait aux utilisateurs de choisir parmi plusieurs propositions graphiques générées automatiquement et éviter qu'ils doivent créer eux-mêmes les visualisations. Pour réaliser ceci, nous avons choisi de travailler avec la librairie Vega-lite qui produit des graphiques avec très peu de lignes de code en JSON. Nous avons également fait une analyse des besoins et pour finir nous avons créé des mockup. La deuxième phase se composait de l'implémentation du prototype selon les résultats obtenus lors de la première partie. Tout au long de cette phase, nous avons passé beaucoup de temps à apprendre les différents langages de programmation. Pour finir, une phase de tests de la plateforme a été effectuée avec des utilisateurs lambdas.

Il est important de se rappeler que le but de ce travail était de réaliser un *proof of concept* et non une plateforme finie de toutes pièces.

AVIS TECHNIQUE

Ce travail m'a permis de connaître des librairies dans l'analyse et la création de visualisation. Je pense que sans ce projet, je ne me serais pas penchée sur ce type de librairies. Je suis heureuse d'avoir réussi à créer un prototype avec Vega-Lite. En effet, cette bibliothèque est très intéressante mais il est difficile de trouver de la documentation en dehors de leur site internet et de leur forum car c'est une librairie toute nouvelle mais néanmoins, très complète. Pour une personne comme moi qui n'a pas beaucoup d'expérience au niveau technique, c'était assez compliqué de l'utiliser et de la comprendre. Pour finir, j'ai été impressionnée des visualisations réalisables en seulement quelques lignes de code en JSON.

De plus, comme je n'avais jamais travaillé avec Node.js, je suis donc heureuse d'avoir appris à manier ce langage même si ce n'est qu'une infime partie de son potentiel.

AVIS PERSONNEL

Durant la formation à la HES-SO, plusieurs projets ont dû être réalisés et la plupart du temps j'étais chargée de m'occuper du *front-end*. Ce fut un grand défi pour moi de réaliser ce travail de bachelor qui consistait à implémenter énormément du côté *back-end*. J'ai passé des heures/jours entiers à essayer de comprendre des problèmes, ce qui m'a fait faire beaucoup d'heures supplémentaires, mais c'était passionnant. Maintenant, je suis fière du résultat et ce travail m'a permis de prendre confiance en moi au niveau *back-end*.

AMÉLIORATIONS FUTURES

Les objectifs principaux ont été atteints, mais ce travail n'est encore qu'un prototype, c'est pourquoi il y a encore des points à améliorer notamment sur la partie pratique qui est d'approfondir les possibilités avec Vega-Lite et d'optimiser le code. De plus, la plateforme n'accepte que des fichiers de petites tailles avec deux types de données (quantitatives et qualitatives) et avec un format CSV. Par la suite, il serait préférable que les utilisateurs puissent importer de grandes tailles de fichier avec tout type de données et sous plusieurs formes (Excel, JSON, TSV, etc).

Lors de l'état de l'art, nous avons trouvé intéressant que les plateformes proposent des exemples de données à importer et nous avons donc voulu faire la même chose. Mais après plusieurs jours d'essai et d'aide par des camarades, nous n'avons, malheureusement, pas réussi à créer cette fonctionnalité qui serait judicieux de pouvoir implémenter dans le futur.

Le projet télécharge à chaque chargement des pages, les données pour créer les visualisations. Ceci est un point à améliorer, car l'utilisateur n'aime pas attendre. Une solution serait de mettre les graphiques dans la mémoire cache.

Nous avons passé beaucoup de temps sur la phase de sélection, mais nous n'avons pas en passer autant sur la partie de modifications. Pour la suite de ce projet, il serait intéressant d'ajouter d'autres fonctionnalités telles que le changement de types de graphiques, changer les couleurs des axes et des données ainsi que de pouvoir les commenter et lui donner un titre.

Enfin, nous avons mis de côté certaines user stories qui étaient en « SHOULD » par manque de temps. Pour les personnes qui reprendront ce projet, il serait préférable de commencer par terminer ces petites fonctionnalités.

RÉFÉRENCES

- Axure. (2018). *DESIGN THE RIGHT SOLUTION*. Récupéré sur Axure: <https://www.axure.com/>
- Baaj, A. (2017, Juin). *Compare the Best Javascript Chart Libraries*. Récupéré sur Sicara: <https://blog.sicara.com/compare-best-javascript-chart-libraries-2017-89fbe8cb112d>
- C3.js. (2014). *C3.js D3-based reusable chart library*. Récupéré sur C3.js: <https://c3js.org/>
- C3js Github. (2018). *C3.js A D3-based reusable chart library*. Récupéré sur Github: <https://github.com/c3js/c3>
- Caviglia, C. G., Mauri, M., Uboldi, G., & Azzi, M. (2017). *RAWGraphs app*. Récupéré sur RAWGraphs: <http://app.rawgraphs.io/>
- Chart.js. (2018). *Chart.js*. Récupéré sur Chart.js: <https://www.chartjs.org/>
- Chart.js. (s.d.). *Chart.js Line chart*. Récupéré sur Chart.js: <http://www.chartjs.org/samples/latest/charts/line/basic.html>
- Chart.js Github. (2018). *Chart.js Simple HTML5 Charts using the <canvas> tag*. Récupéré sur Github: <https://github.com/chartjs/Chart.js>
- Chartist.js. (s.d.). *Chartist.js*. Récupéré sur Chartist.js: <https://gionkunz.github.io/chartist-js/>
- Chartist.js Github. (2017). *Chartist.js Simple responsive charts*. Récupéré sur Github: <https://github.com/gionkunz/chartist-js>
- Chrzanowska, N. (2017, Septembre). *Node.js vs. PHP: Which Environment To Choose For Your Next Project?* Récupéré sur Netguru: <https://www.netguru.co/blog/nodejs-vs-php>
- Comment ça marche. (2018, Juillet). *HTML - Langage*. Récupéré sur CCC: <https://www.commentcamarche.com/contents/498-html-langage>
- CompassQL Github. (2018). *CompassQL Query Language for visualization recommendation*. Récupéré sur Github: <https://github.com/vega/compassql>
- D3 Github. (2018). *D3.js Bring data to life with SVG, Canvas and HTML*. Récupéré sur Github: <https://github.com/d3/d3>
- D3.js. (2017). *D3 Data-Driven Documents*. Récupéré sur D3.js: <https://d3js.org/>

- Data Illustrator. (2018, Mai). *Data Illustrator*. Récupéré sur Data Illustrator: <http://data-illustrator.com/about.php>
- Datavoyager. (2015). *Datavoyager*. Washington, USA.
- Datawrapper. (2018). *Datawrapper*. Récupéré sur Datawrapper: <https://www.datawrapper.de/>
- Datawrapper Github. (2018). *Datawrapper*. Récupéré sur Github: <https://github.com/datawrapper/datawrapper>
- expressjs. (2018, Juillet). *Express Infrastructure Web minimaliste, souple et rapide pour Node.js*. Récupéré sur expressjs: <http://expressjs.com/fr/>
- Florent, L. (s.d.). *Guide de démarrage Scrum*. Récupéré sur L'Agiliste: <https://www.agiliste.fr/guide-de-demarrage-scrum/#Les-R-les-en-bref>
- Futura Tech. (2018, Juillet). *CSS*. Récupéré sur Futura Tech: <https://www.futura-sciences.com/tech/definitions/internet-css-4050/>
- Gupta, S. (2014). *cereal.csv*. Récupéré sur Github: <https://gist.github.com/SinnerShanky/925f08febd10b40b8b5e>
- Heer, J. (2017, Mars). *Vega and D3*. Récupéré sur Github: <https://github.com/vega/vega/wiki/Vega-and-D3>
- Highcharts. (2018). *Highcharts*. Récupéré sur Highcharts: <https://www.highcharts.com/products/highcharts/>
- Highcharts. (2018). *Highcharts Licenses*. Récupéré sur Highcharts: <https://shop.highsoft.com/highcharts>
- Highcharts. (2018). *Make your data come alive*. Récupéré sur Highcharts: <https://www.highcharts.com/>
- Highcharts Github. (2018). *Highcharts JS, the JavaScript charting framework*. Récupéré sur Github: <https://github.com/highcharts/highcharts>
- IDC. (2017, April). *Data Age 2025: The Evolution of Data to Life-Critical*. Retrieved from Seagate: <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>

- Info Webmaster Glossaire. (2018). *Mockup*. Récupéré sur Info Webmaster Glossaire: <http://glossaire.infowebmaster.fr/mockup/>
- InfoWebMaster Glossaire. (2018, Juillet). *Javascript*. Récupéré sur Informations et ressources pour webmasters : <http://glossaire.infowebmaster.fr/javascript/>
- Inside Airbnb. (s.d.). *Get the Data*. Récupéré sur Inside Airbnb Adding data to the debate: <http://insideairbnb.com/get-the-data.html>
- JDN. (2018, Avril). *Framework ou infrastructure logicielle : définition et traduction*. Récupéré sur Journal dunet: <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203355-framework/>
- Jeremy Boy, R. A.-D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 10. Récupéré sur <https://hal.inria.fr/hal-01027582/document>
- jQuery. (2018, Juillet). *jQuery*. Récupéré sur jQuery: <https://jquery.com/>
- Les Echos. (2017, Avril). *L'explosion du volume de données fait saliver l'industrie du stockage*. Récupéré sur LesEchos.fr: https://www.lesechos.fr/04/04/2017/lesechos.fr/0211941994405_l-explosion-du-volume-de-donnees-fait-saliver-l-industrie-du-stockage.htm
- Littlefield, A. (2017, Juin 5). *Gestion de projets: méthode Agile et Scrum pour les nuls*. Récupéré sur Trello: <https://blog.trello.com/fr/methode-agile-scrum-gestion-projet>
- Liu, Z., Thompson, J., Wilson, A., Dontcheva, M., Delorey, J., Grigg, S., . . . Stasko, J. (2018, Avril). *Data Illustrator: Augmenting Vector Design Tools with Lazy Data Binding for Expressive Visualization Authoring*. Récupéré sur Data Illustrator: <http://data-illustrator.com/papers/DataIllustratorCHI18.pdf>
- Lorenz, M. (s.d.). *How Datawrapper helps journalists to be first with their stories*. Récupéré sur Datawrapper: <https://handsontable.com/static/resources/ModCustomers/datawrapper/Datawrapper-Customer-Story.pdf>
- Marty, É. (2016, Février). *The D3/Vega "stack"*. Récupéré sur Eric Marty: <https://blog.ericmarty.com/the-d3-/-vega-stack>

- Marty, É. (2016, Février). *Vega Visualization Grammar*. Récupéré sur Eric Marty: <https://blog.ericmarty.com/vega-visualization-grammar>
- Materialize. (2018, Juillet). *Materialize*. Récupéré sur Materialize css: <https://materializecss.com/>
- Mauri, M. (2013). *RAWGraphs*. Récupéré sur Density Design: <https://densitydesign.org/research/raw/>
- Médiacademie. (2013, Octobre). *Datawrapper nouvelle version, pour des visualisations de données faciles et en temps réel*. Récupéré sur Médiacademie: <http://www.mediacademie.org/130/datawrapper-nouvelle-version-pour-des-visualisations-faciles-et-temps-reel>
- Medium. (2016, Février). *Data visualization and interactive analysis research at the University of Washington*. Récupéré sur Medium: <https://medium.com/hci-design-at-uw/introducing-vega-lite-438f9215f09e>
- Node.js. (s.d.). *Node.js*. Récupéré sur Node.js: <https://nodejs.org/en/>
- Opensource. (2007, Mars). *The Open Source Definition (Annotated)*. Récupéré sur Open Source Initiative: <https://opensource.org/osd>
- Pardanaud, J., & Sébastien. (2017, Septembre). *Dynamisez vos sites web avec JavaScript !* Récupéré sur Openclassrooms: <https://openclassrooms.com/courses/1916641-dynamisez-vos-sites-web-avec-javascript/1916776-introduction-au-javascript>
- PayScale. (s.d.). *By the Numbers: Comparing Tech Employee Salary, Age, Stress and More*. Récupéré sur PayScale: <https://www.payscale.com/data-packages/top-tech-companies-compared/tech-salaries>
- Plotly. (2017). *Modern visualization for the data era*. Récupéré sur Plotly: <https://plot.ly/products/cloud/>
- Plotly. (2017). *Modern Visualization for the Data Era*. Récupéré sur Plotly: <https://plot.ly/>
- Plotly Github. (2018). *Plotly.js The open source JavaScript graphing library that powers plotly*. Récupéré sur Github: <https://github.com/plotly/plotly.js>
- PluralSight. (2018). *JavaScript*. Récupéré sur JavaScript: <https://www.javascript.com/>
- Predictive Analytics Today Review Desk. (2017, Septembre). *DATAWRAPPER*. Récupéré sur Pat research: <https://www.predictiveanalyticstoday.com/datawrapper/>

- Predictive Analytics Today Review Desk. (2017). *TABLEAU PUBLIC*. Récupéré sur Pat Research: <https://www.predictiveanalyticstoday.com/tableau-public/>
- Predictive Analytics Today Review Desk. (2018, Janvier). *RAWGRAPHS*. Récupéré sur Pat Research: <https://www.predictiveanalyticstoday.com/rawgraphs/>
- Pugjs Github. (2018, Juin). *Pug - robust, elegant, feature rich template engine for Node.js*. Récupéré sur Github: <https://github.com/pugjs/pug>
- Rahman, S. F. (2015, Juillet). *16 JavaScript Libraries For Creating Beautiful Charts*. Récupéré sur Sitepoint: <https://www.sitepoint.com/15-best-javascript-charting-libraries/>
- RAWGraphs. (2018). *RAWGraphs*. Récupéré sur RAWGraphs: <https://rawgraphs.io/>
- Satyanarayan, A. (2017). *The reactive vega stack: declarative interaction design for data visualization*. Stanford University. Stanford: Creative Commons.
- Satyanarayan, A., Moritz, D., Wongsuphasawat, K., & Heer, J. (2017). *Vega-Lite: A Grammar of Interactive Graphics*. Récupéré sur UW Interactive Data Lab: <http://idl.cs.washington.edu/files/2017-VegaLite-InfoVis.pdf>
- SINGHAL, V. (2015, Juin). *20 best JavaScript charting libraries*. Récupéré sur The next web: <https://thenextweb.com/dd/2015/06/12/20-best-javascript-chart-libraries/>
- Tableau. (2018). *Découvrez Tableau Server avec une version d'évaluation*. Récupéré sur Tableau: <https://www.tableau.com/fr-fr/products/server/options#options-explained>
- Tableau. (2018). *Des offres adaptées à tous les besoins*. Récupéré sur Tableau: <https://www.tableau.com/fr-fr/pricing/teams-orgs>
- Tableau. (2018). *Tableau*. Récupéré sur Tableau: <https://www.tableau.com/>
- Tableau. (2018). *Tableau Desktop : profitez d'un essai gratuit de 14 jours*. Récupéré sur Tableau: <https://www.tableau.com/fr-fr/products/trial>
- Tableau. (2018). *Tableau Software vous aide à voir et à comprendre vos données*. Récupéré sur Tableau: <https://www.tableau.com/fr-fr/about/mission>
- techno-science. (2018, Juillet). *Bibliothèque logicielle*. Récupéré sur techno-science.net: <https://www.techno-science.net/?onglet=glossaire&definition=1470>

- Vallandingham, J. (2015, Novembre). *OUR WORK WITH DATA VOYAGER: DESIGNING FOR FAST DATA EXPLORATION*. Récupéré sur bocoup: <https://bocoup.com/blog/our-work-with-data-voyager>
- Vega Github. (2018). *Vega A visualization grammar*. Récupéré sur Github: <https://github.com/vega/vega>
- Vega. (s.d.). *Vega - A Visualization Grammar*. Récupéré sur Vega: <https://vega.github.io/vega/>
- Vega-embed Github. (2018, Juillet). *Publish Vega visualizations as embedded web components with interactive parameters*. Récupéré sur Github: <https://github.com/vega/vega-embed>
- Vega-Lite Github. (2018). *Vega-Lite A concise grammar of interactive graphics, built on Vega*. Récupéré sur Github: <https://github.com/vega/vega-lite>
- Vega-Lite. (s.d.). *Type*. Récupéré sur Vega-Lite: <https://vega.github.io/vega-lite/docs/type.html>
- Vega-Lite. (s.d.). *Vega-Lite - A Grammar of Interactive Graphics*. Récupéré sur Vega-Lite: <https://vega.github.io/vega-lite/>
- Vega-themes Github. (2018, Mai). *Themes for stylized Vega and Vega-Lite visualizations*. Récupéré sur Github: <https://github.com/vega/vega-themes>
- Vega-tooltip Github. (s.d.). *Tooltip Plugin for Vega-Lite*. Récupéré sur Github: <https://github.com/vega/vega-tooltip/>
- Voyager 2 Github. (2018). *Vega Recommendation-Powered Visualization Tool for Data Exploration*. Récupéré sur Github: <https://github.com/vega/voyager>
- w3schools. (2018, Juillet). *HTML Introduction*. Récupéré sur w3schools.com: https://www.w3schools.com/html/html_intro.asp
- W3Schools. (2018). *JavaScript Tutorial*. Récupéré sur W3Schools.com: <https://www.w3schools.com/js/default.asp>
- Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., & Heer, J. (2016, Juin). *Towards A General-Purpose Query Language for Visualization Recommendation*. San Francisco: ACM SIGMOD. Récupéré sur UW Interactive Data Lab: <https://idl.cs.washington.edu/files/2016-CompassQL-HILDA.pdf>
- Wongsuphasawat, K., Qu, Z., Moritz, D., Chang, R., Ouk, F., Anand, A., . . . Heer, J. (2017, Mai). *Voyager 2: Augmenting Visual Analysis with Partial View Specifications*. Denver: ACM Human

Factors in Computing Systems. Récupéré sur Voyager 2:
<http://idl.cs.washington.edu/files/2017-Voyager2-CHI.pdf>

ANNEXES

ANNEXE I : CAHIER DES CHARGES

1 Contexte

Dans le cadre du travail de Bachelor réalisé durant le 6ème semestre en Informatique de Gestion à la Haute École Spécialisée de Suisse occidentale (HES-SO) Valais-Wallis. Le professeur Florian Evéquoz a proposé le sujet suivant ; « Plateforme publique de visualisation de données - mise en place d'une infrastructure et POC ». L'étudiant a reçu la tâche de développer cette plateforme publique de visualisation de données.

Ce travail a pour but de créer une plateforme qui est « la première brique » du projet VASCO (Assisted Visualisation Generation Platform for AnalySing Data COLlaboratively) développé à l'institut Informatique de Gestion de la HES-SO Valais-Wallis.

Le but de ce travail de Bachelor est dans un premier temps, d'analyser les outils existants afin de cibler les objectifs principaux de la future plateforme. Puis dans un deuxième temps, implémenter une plateforme facile d'utilisation avec l'aide de bibliothèques existantes.

2 Description

« Le Big Data », de nos jours, ces trois petits mots ressortent très souvent dans nos discussions, à la télévision et dans les journaux. Nous avons de plus en plus de données provenant de diverses ressources telles que les téléphones portables, les réseaux sociaux, les objets connectés qui nous suivent dans notre quotidien. Ces données sont pour la plupart des personnes incompréhensibles. Cependant, il existe déjà des outils tels que D3.js, HighCharts, Plotly, Vega, etc qui permettent de visualiser des données par le biais de graphiques, mais ils sont parfois très complexes à utiliser. Cette problématique nous amène à réaliser ce projet qui a pour but de créer une plateforme publique de visualisation de données qui est facile d'utilisation et de compréhension pour des personnes n'ayant pas de grandes connaissances dans l'analyse de données.

Le but de ce projet est de créer une plateforme publique de visualisation de données qui est facile d'accès, d'utilisation et de compréhension pour des personnes n'ayant pas de grandes connaissances dans l'analyse de données.

L'utilisateur peut importer des fichiers au format CSV. Les données sont chargées et automatiquement une liste de graphiques est affichée. De plus, il est possible d'exporter les graphiques en SVG, PNG et en JSON.

3 Objectifs

Durant ce travail de Bachelor, l'étudiant va être amené à atteindre plusieurs objectifs :

- Comprendre comment faire un état de l'art
- Comprendre les architectures supportant la mise en place d'applications web
- Réaliser une architecture logicielle robuste pour une application distribuée
- Comprendre et appliquer les principes d'ingénierie de la visualisation de données
- Réaliser un cas pratique de visualisation de données au moyen de la plateforme mise en place
- Création d'un Product Backlog
- Apporter un grand soin à l'architecture logicielle et à la documentation

4 Phases du travail

4.1 État de l'art technique des librairies

Pour commencer, l'étudiant va faire une analyse des outils existants afin de cibler plus précisément ses objectifs et de lui donner de nouvelles idées. Par le biais de cet état de l'art, il comprendra mieux les difficultés à la réalisation d'une plateforme de visualisation de données.

4.2 Définition du concept et de l'architecture de la solution

L'étudiant va choisir les librairies et le langage de programmation grâce à son état de l'art fait à l'étape précédente. De plus, il fera une étude de besoins et créera des mock-up user-friendly. Cette dernière étape est très importante pour avoir une idée concrète du résultat final afin d'éviter de partir dans de mauvaises directions.

De plus, il sera important de définir 2-3 Use Case pour cibler plus précisément le style de données que la plateforme pourra gérer afin de faire une implémentation correcte dès le début du projet.

D'autre part, l'étudiant va mettre en place toute l'architecture informatique afin de pouvoir tout de suite commencer, à l'étape suivante, l'implémentation de la plateforme.

4.3 Installation de l'infrastructure backend (intégration des librairies)

Dans cette partie, l'étudiant va implémenter toute la partie backend pour le bon fonctionnement de la plateforme. Cette partie servira à charger les données et les transformer en graphique.

4.4 Développement du frontend

Puis l'étape du développement frontend sera réalisée avec précaution. En effet, elle est très importante, car le but du projet est de créer une plateforme publique pour des personnes non-professionnelles dans le domaine de l'informatique. Il sera donc très important de faire une plateforme user-friendly.

4.5 Réalisation d'un proof-of-concept

La dernière étape sera de déployer la plateforme sur un serveur afin qu'elle soit accessible au public.

Puis, des utilisateurs lambda testeront la plateforme en donnant leurs avis afin de pouvoir améliorer les points faibles de celle-ci.

4.6 Documentation

Tout le long du projet l'étudiant devra apporter un grand soin à l'architecture logicielle et à la documentation. En effet, ce projet s'insère comme « la première brique » du projet VASCO. Il est donc impératif que l'étudiant s'applique sur la documentation du projet et la lisibilité du code car d'autres personnes poursuivront le développement de ce travail.

5 Délais

Sprint	Début	Fin	Durée	Objectifs principaux
0	30.04.18	14.05.18	2 semaines	Création du Product Backlog et cahier des charges
1	14.05.18	04.06.18	2 semaines	État de l'art et choix technologiques
2	04.06.18	18.06.18	3 semaines, car une en Finlande pour les cours	Mock-up et choix de data set + préparation de l'environnement
3	18.06.18	02.07.18	2 semaines	Implémentation du Back-end
4	02.07.18	16.07.18	2 semaines	Continuer l'implémentation du Back-end + front-end + rapport
5	16.07.18	30.07.18	2 semaines	Déploiement + tests utilisateurs + rapport

ANNEXE II : PRODUCT BACKLOG

Issue ID	Theme	User Stories			Acceptance Criteria	Priority (1-100)	Status	Story Points (Fibonacci)	DONE	MOSCOW	Sprint
		As an/a ...	I want to ...	so that ...							
1	Implémentation	Utilisateur	importer des ensembles de données	d'avoir une vue d'ensemble des données	utiliser la librairie Multer afficher un bouton pour uploader des fichiers pouvoir uploader que des fichiers csv convertir le fichier en JSON afin d'utiliser CompassQL récupérer les fichiers importés dans un fichier "uploads" avoir les données intacts dans le dossier uploads	65	●	55	DONE	MUST	3
2	Implémentation	Utilisateur	pouvoir voir plusieurs graphiques	de sélectionner celui qui me convient le mieux	Afficher un "hover" lorsque qu'on est sur le graphique pour voir le quel on va sélectionner Lorsque sélection du graphique envoyer les informations sur la page d'après	50	●	34	DONE	MUST	4
3	Implémentation	Utilisateur	pouvoir modifier la visualisation	d'améliorer la compréhension du graphique.	Utiliser vega-tooltip Afficher différent style de design	7	●	5	DONE	SHOULD	4
4	Préparation	Administrateur	avoir un accès au repository du projet	de pouvoir récupérer tous les fichiers du projet	Le repository a bien été crée Les informations concernant les accès au repository se trouvent dans le guide technique La visibilité du repository est uniquement accessibles aux membres du projet	80	●	1	DONE	MUST	2
5	Implémentation	Développeur	deployer la plateforme	que la plateforme devienne publique	Afficher une plateforme en ligne sur un serveur de la hes Afficher une page d'accueil avec une description et un page de contact	20	●	21	DONE	MUST	5
6	Préparation	Développeur	choisir une technologie adéquate au projet	afin de pouvoir implémenter la plateforme avec les technologies les plus adaptées	Choix d'une technologie tester les différentes librairies	93	●	21	DONE	MUST	1
7	Préparation	Développeur	installer et configurer mon environnement de travail	afin de pouvoir implémenter la plateforme	Installation des outils nécessaires Créer un lien github	85	●	3	DONE	MUST	2
8	Implémentation	Développeur	afficher toutes les possibilités des graphiques	pouvoir avoir une visualisation globale des données	Afficher différentes style de graphique selon CompassQL Prendre toutes les possibilités des dimensions et les afficher avec une scrollbar UX plateforme afficher dans tableau résumé : type, aperçu, somme, moyenne, min, max afficher le total des lignes proposer une dropdown list pour filtrer les graphiques afficher bouton pour revenir en arrière et aller à la page suivante	55	●	55	DONE	MUST	3 mais validé au 4

9	Implémentation	Développeur	semi-automatiser la création de graphique	facilité la visualisation des données	utilisé la librairie CompassQL afin de choisir le style de graphique le plus adapté récupérer le fichier uploadé dans le dossier "uploads" afin de créer les graphiques convertir le csv en JSON pour utiliser CompassQL	63	●	34	DONE	MUST	3
10	Préparation	Administrateur	créer des mockup	d'avoir une visualisation globale sur la plateforme finale	Utiliser Axure mockup minutieusement créés validation par le professeur	90	●	21	DONE	MUST	2
11	Préparation	Administrateur	faire un état de l'art	de savoir quels sont les technologies existantes	création d'un état de l'art technique validation par le professeur	95	●	21	DONE	MUST	1
12	Préparation	Administrateur	faire un cahier des charges	d'avoir une visualisation globale sur le travail de Bachelor	création du cahier des charges validation par le professeur	100	●	5	DONE	MUST	0
13	Implémentation	Utilisateur	pouvoir mettre des annotations et des commentaires sur les graphiques	faciliter le partage de connaissances	afficher un éditeur de commentaires modification du graphique	2	●	5	NOT DONE	SHOULD	
14	Finalisation	Administrateur	faire tester ma plateforme	faire des améliorations et d'éviter les bugs	tester à des utilisateurs lambdas améliorer la plateforme régler les bugs	18	●	21	DONE	MUST	5
15	Implémentation	Utilisateur	pouvoir exporter le graphique	pouvoir l'utiliser sans être sur la plateforme	Utiliser la librairie vega-embed exporter en PNG, SVG et json	40	●	5	DONE	MUST	4
16	Préparation	Administrateur	faire un product Backlog	d'avoir une visualisation globale des objectifs et besoins du projet	création du product backlog validation par le professeur	98	●	8	DONE	MUST	0
17	Maintenance	Développeur	avoir une structure de l'architecture correctement implémentée	que d'autres développeurs comprennent l'architecture facilement	Code correctement implémenté commentaires faire le guide technique	25	●	13	DONE	MUST	4
18	Implémentation	Développeur	Afficher une carte avec l'adresse	l'utilisateur voit où on se situe	Afficher une map dans la section contact	4	●	5	NOT DONE	SHOULD	
19	Implémentation	Développeur	Envoyer un mail depuis le formulaire de contact	l'utilisateur peut nous contacter	Implémenter un template de mail envoyer le mail à l'administrateur	1	●	8	NOT DONE	SHOULD	
20	Implémentation	Utilisateur	avoir des exemples de data set	d'avoir des exemples	Proposer des exemples de data set sur la page d'importation afficher un bouton pour choisir	10	●	34	NOT DONE	SOULD	

ANNEXE III : GUIDE DE L'UTILISATEUR

1. Présentation

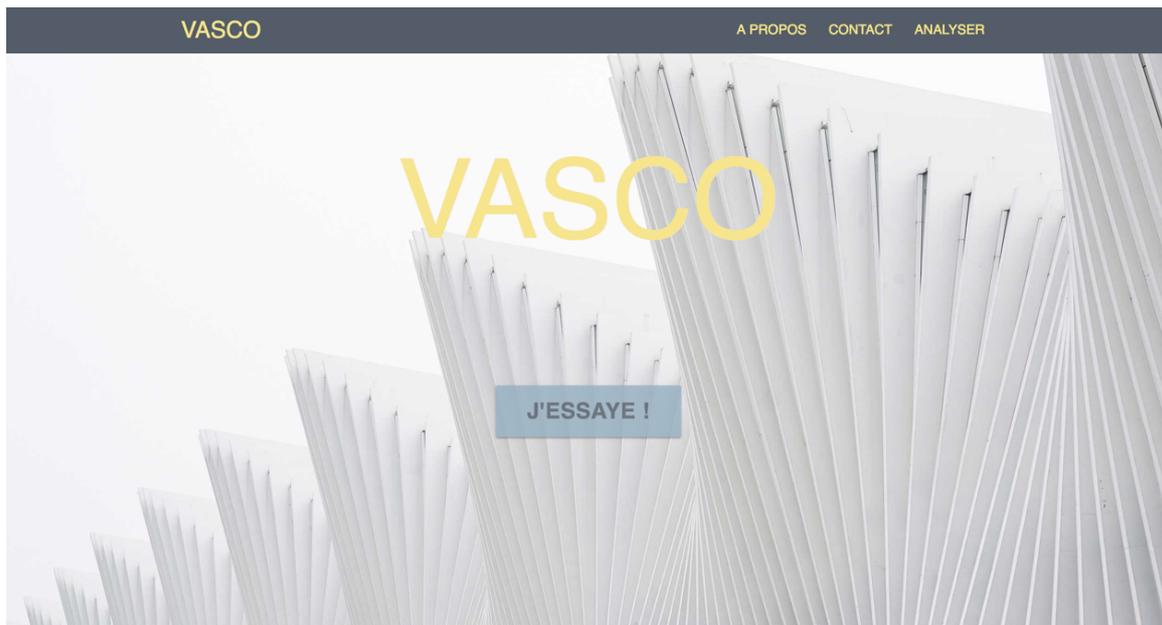
VASCO est une plateforme publique de visualisation de données qui permet aux utilisateurs d'importer des fichiers au format CSV. Les données sont chargées et automatiquement une liste de graphiques est affichée. De plus, il est possible d'exporter les graphiques en SVG, PNG et en JSON.

Le but de cette plateforme est qu'elle soit compréhensible et facile d'utilisation pour analyser des visualisations.

2. Manuel d'utilisation

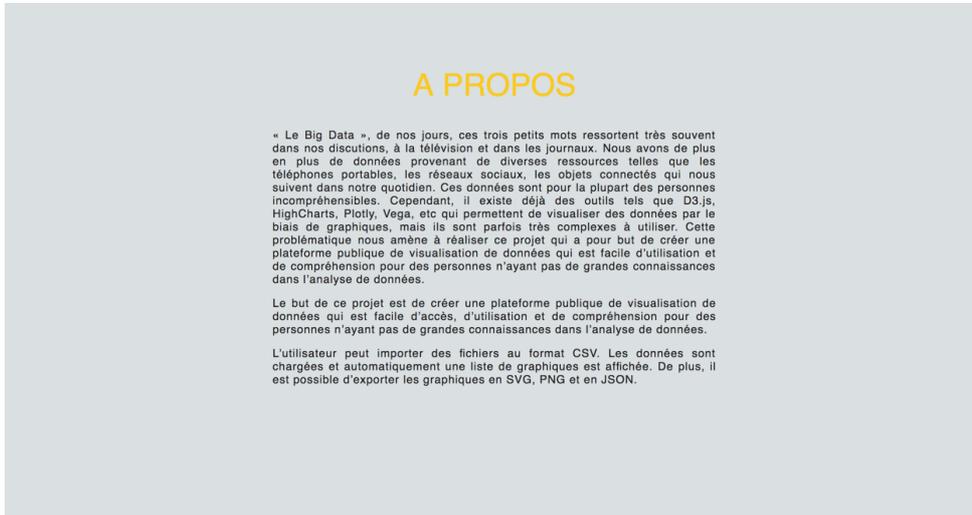
2.1. Page d'accueil

Allez sur <https://vasco.tb.hevs.ch/> pour accéder à la plateforme.



Sur cette page vous avez accès à :

- À propos, un descriptif de la plateforme



- Contacte (comment prendre contact avec nous, par téléphone, par e-mail, par courrier ou directement avec le formulaire)

CONTACT

Nom & Prénom

Email

Sujet

Envoyez-nous un message

Rue Technopole 3

3960 Sierre

022 000 00 00

info@vasco.ch

Pour commencer l'analyse de donnée, vous pouvez cliquer soit sur l'onglet « ANALYSER » dans la barre de navigation soit sur le bouton « J'ESSAYE ! ». Après avoir appuyé sur l'un des deux, vous êtes redirigé sur la page d'importation.

2.2 Importation

Sur cette page vous pouvez importer un fichier CSV de taille maximum de 20Ko. De plus, les données du fichier doivent être que quantitatives et/ou qualitatives.

Si vous n'avez pas de fichier CSV, vous pouvez cliquer sur le bouton « EXEMPLE CSV » et télécharger le fichier sur Github.

Lorsque vous avez importé un fichier, cliquez sur le bouton « SUIVANT »

Lien externe pour télécharger un fichier CSV :

Cereal Name	Manufacturer	Type	Calories	Protein (g)	Fat	Sodium	Dietary Fiber	Carbs	Sugars	Display Shelf
100%Bran	Nabisco	C	70	4	1	130	10	5	6	3
100%_Natural_Bran	Quaker Oats	C	120	3	5	15	2	8	8	3
All-Bran	Kelloggs	C	70	4	1	260	9	7	5	3
All-Bran_with_Extra_Fiber	Kelloggs	C	50	4	0	140	14	8	0	3
Almond_Delight	Ralston Purina	C	110	2	2	200	1	14	8	3
Apple_Cinnamon_Cheerios	General Mills	C	110	2	2	180	1.5	10.5	10	1
Apple_Jacks	Kelloggs	C	110	2	0	125	1	11	14	2
Basic_4	General Mills	C	130	3	2	210	2	18	8	3
Bran_Chech	Ralston Purina	C	90	2	1	200	4	15	6	1
Bran_Flakes	Post	C	90	3	0	210	5	13	5	3
Cap'n_Crunch	Quaker Oats	C	120	1	2	220	0	12	12	2
Cheerios	General Mills	C	110	6	2	290	2	17	1	1
Cinnamon_Toast_Crunch	General Mills	C	120	1	3	210	0	13	9	2
Clusters	General Mills	C	110	3	2	140	2	13	7	3
Cocoa_Puffs	General Mills	C	110	1	1	180	0	12	13	2
Corn_Chech	Ralston Purina	C	110	2	0	280	0	22	3	1
Corn_Flakes	Kelloggs	C	100	2	0	290	1	21	2	1

2.3 Sélection

La page est en premier gelée, car il faut attendre un petit moment afin que les graphiques se téléchargent et s'affichent.

Dimensions	Type	Somme	Moyenne	Minimum	Maximum	Aperçu
Cereal Name	qualitatif	-	-	-	-	
Manufacturer	qualitatif	-	-	-	-	
Calories	quantitatif	7,810	105.54	50	160	
Protein (g)	quantitatif	185	2.5	1	6	
Fat	quantitatif	70	0.94	0	5	
Sodium	quantitatif	11,890	160.67	0	320	
Dietary Fiber	quantitatif	154.69	2.09	0	14	
Carbs	quantitatif	1,080	14.59	-1	23	
Sugars	quantitatif	501	6.77	-1	15	
Display Shelf	quantitatif	161	2.17	1	3	

La page sélection est séparée par deux onglets « Synthèse du fichier importé » et « Liste des graphiques ».

2.3.1 Synthèse du fichier importé

Dans cet onglet, vous avez un tableau récapitulatif du fichier avec toutes les dimensions, leur type, la somme, la moyenne, le minimum, le maximum et un aperçu sous forme d'histogramme.

Cet histogramme est un peu petit donc vous pouvez passer votre curseur dessus afin qu'il s'agrandisse.

Dimensions	Type	Somme	Moyenne	Minimum	Maximum	Aperçu
Cereal Name	qualitatif	-	-	-	-	
Manufacturer	qualitatif	-	-	-	-	
Calories	quantitatif	7,810	105.54	50	160	

2.3.1 Liste des graphiques

Dans cet onglet, vous pouvez apercevoir la liste des graphiques de votre fichier importé précédemment.

Vous avez la possibilité de trier les graphiques selon le nom des dimensions.

Lorsque vous passez le curseur sur les visualisations un « hover » jaune s'affiche afin de vous montrer plus précisément le graphique que vous allez sélectionner en cliquant dessus. Lorsque vous cliquez sur une visualisation, vous êtes redirigé sur la page de modification.

The screenshot shows the VASCO application interface. At the top, there is a navigation bar with 'HOME' and 'CONTACT'. Below it, a breadcrumb trail reads 'IMPORTATION > SELECTION > MODIFICATION > EXPORTATION'. A search bar contains 'SYNTHÈSE DU FICHIER IMPORTÉ' and a button labeled 'LISTE DES GRAPHIQUES' is highlighted. A dropdown menu is open, displaying 'Je trie les données selon ...' and 'le nom des dimensions (nom des colonnes)'. Below the menu, two charts are shown. The first chart, 'Cereal Name / Manufacturer', is a scatter plot with 'Manufacturer' on the y-axis and 'Cereal Name' on the x-axis. The second chart, 'Cereal Name / Calories', is a bar chart with 'Calories' on the y-axis and 'Cereal Name' on the x-axis. A yellow highlight is visible over the 'Cereal Name / Calories' chart, and a checkmark is present next to the 'Basilic 4' category on its x-axis.

2.4 Modification

Sur cette page, vous pouvez apercevoir le graphique que vous avez sélectionné précédemment.

Il est possible de changer la couleur de fond du graphique en appuyant sur l'une des options sous « Design du graphique ».

Lorsque vous avez fini, vous pouvez appuyer sur le bouton « exporter » afin d'exporter le graphique.

VASCO HOME CONTACT

IMPORTATION > SELECTION > MODIFICATION > EXPORTATION

Design du graphique :

- Default
- Dark
- Microsoft excel
- ggplot2
- Quartz
- Vox

Le saviez-vous ?

- Placez votre curseur sur les données du graphique pour avoir plus de détails
- Plus les points sont foncés, plus il y a des données similaires

Calories / Protein (g)

Protein (g)

Calories

< RETOUR EXPORTER >

© Copyright 2018 projet VASCO

De plus, lorsque vous placez le curseur sur les données du graphique vous pouvez apercevoir un petit rectangle indiquant des informations sur celles-ci.

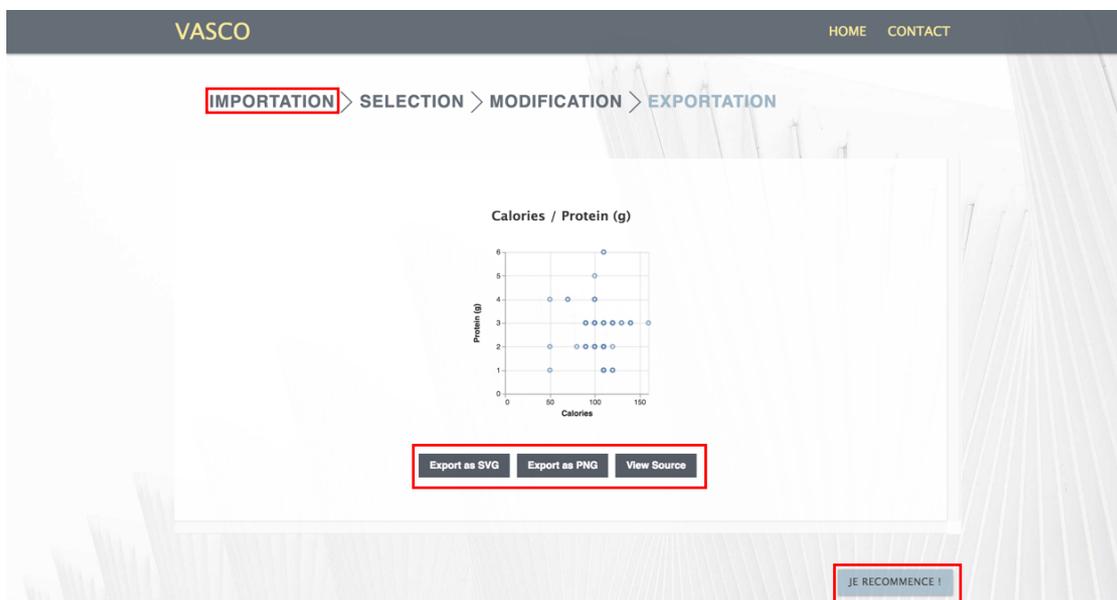


2.5 Exportation

Sur cette page, vous pouvez exporter votre graphique en appuyant sur les boutons « Export as SVG », « Export as PNG » et « View Source ».

- « Export as SVG » pour exporter en SVG
- « Export as PNG » pour exporter en PNG
- « View Source » afin de voir le code en JSON du graphique

Après vous pouvez recommencer soit en appuyant sur le bouton « JE RECOMMENCE ! » soit en appuyant sur l'onglet « Importation ».



ANNEXE IV : GUIDE TECHNIQUE

1. Informations techniques

1.1. Software & Hardware

- Hardware
 - Hébergement Web sur les serveurs de la HES-SO
 - Nom de domaine vasco.tb.hevs.ch
 - Ordinateurs des développeurs
 - 1 ordinateur sous MacOS
- Software
 - Webstorm version 2018.1
 - GitLab Web Plateforme - Outil collaboratif

1.2. Architecture & Librairies

Le projet a été développé en utilisant l'architecture et les librairies suivantes :

- **Couche de présentation**
 - Materialize.css version 1.0.0 - Framework CSS
 - JQuery version 2.2.4
 - Vega version 3
 - Vega-lite version 2.4.3
 - Vega-embed version 3.14.0
- **Couche de traitement**
 - NodeJS version 8.11.2
 - Compassql version 0.20.0
 - Convert-csv-to-json version 0.0.13
 - Express version 4.16.0
 - Express-session version 1.15.6
 - Multer version 1.3.0
 - Pug version 2.0.0
 - Rimraf version 2.6.2
 - Vega-embed version 3.15.0
 - Vega-tooltip version 0.11.0

1.3 Identifiant

Plateforme web : `vasco.tb.hevs.ch`

SSH :

- demander les identifiants à Christophe Pignat (`christophe.pignat@hevs.ch`)

2 Prérequis

2.1 Installation

```
git clone https://github.com/audreycelia/VASCO.git
```

2.2 GitLab Repository

```
https://github.com/audreycelia/VASCO
```

3 Structure du projet

```
VASCO
├── bin
├── node_modules
├── public
├── routes
├── uploads
└── views
```

- **bin** : contient le launcher du site internet
- **nodes_modules** : contient tous les modules NodeJS
- **public** : contient toutes les fichiers de ressources : CSS / Images /JS
 - **css** : contient tous les fichiers css
 - **images** : contient toutes les ressources médias
 - **js** : contient tous les fichiers javascript (client)
- **routes** : contient les contrôleurs
- **uploads** : contient les fichiers importés par l'utilisateur
- **views** : contient toutes les vues du site internet

ANNEXE V : DONNÉES SÉLECTIONNÉES - CÉRÉALES

Cereal Name	Manufacturer	Calories	Protein (g)	Fat	Sodium	Dietary Fiber	Carbs	Sugars	Display Shelf	Potassium	Vitamins and Minerals	Serving Size Weight	Cups per Serving
100%_Bran	Nabisco	70	4	1	130	10	5	6	3	280	25	1	0.33
100%_Natural_Bran	Quaker Oats	120	3	5	15	2	8	8	3	135	0	1	-1
All-Bran	Kelloggs	70	4	1	260	9	7	5	3	320	25	1	0.33
All-Bran_with_Extra_Fiber	Kelloggs	50	4	0	140	14	8	0	3	330	25	1	0.5
Almond_Delight	Ralston Purina	110	2	2	200	1	14	8	3	-1	25	1	0.75
Apple_Jacks	Kelloggs	110	2	0	125	1	11	14	2	30	25	1	1
Bran_Chex	Ralston Purina	90	2	1	200	4	15	6	1	125	25	1	0.67
Bran_Flakes	Post	90	3	0	210	5	13	5	3	190	25	1	0.67
Cap'n'Crunch	Quaker Oats	120	1	2	220	0	12	12	2	35	25	1	0.75

ANNEXE VI : DONNÉES SÉLECTIONNÉES - AIRBNB À GENÈVE

name	host_name	neighbourhood	room_type	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
Geneva center - Hosting females	Noelle	Commune de Genève	Private room	99	3	29	0.38	1	333
Perfect Cozy Bedroom Close to Lake & Transport	Rita	Commune de Genève	Private room	90	1	41	0.46	1	164
GardenPavilion (with TransportPass)	Maya	Plan-les-Ouates	Entire home/apt	120	3	5	0.12	1	210
Studio in a busy area *	Ladyl'	Commune de Genève	Entire home/apt	75	4	24	0.31	1	0
Central apartment by the lake	Virginia	Commune de Genève	Entire home/apt	135	7	62	0.84	1	268
Top location studio 1 block from lake!	Mike	Commune de Genève	Entire home/apt	165	3	45	0.71	1	192
Fabulous apartment in Champel - Geneva	Maria	Commune de Genève	Entire home/apt	118	6	35	0.47	1	234
Modern & cosy loft in Geneva center	Ahone	Commune de Genève	Entire home/apt	90	3	13	0.19	2	0
B&B one in Geneva City - Airport	Éric	Meyrin	Private room	75	2	18	0.25	2	309
B&B two in Geneva City - Airport	Eric	Meyrin	Private room	75	2	32	0.46	2	312
Stylish loft in Paquis, next to UN	Virna	Commune de Genève	Entire home/apt	150	2	69	0.99	1	322
1-bed apartment near Geneva Airport	L+N	Vernier	Entire home/apt	140	2	41	0.57	1	52
Top of Paquis (Downtown Geneva)	Julik	Commune de Genève	Private room	125	1	35	0.77	1	170
Pools, 2 Saunas, & Mountain Views!	Daniel	Chv™ne-Bougeries	Entire home/apt	932	1	9	0.13	1	336
Beautiful Apm't in central Geneva (NB Guest Room)	Markus	Commune de Genève	Private room	250	1	22	0.39	1	363
cosy&quiet in super strategic central location	Lisa	Commune de Genève	Entire home/apt	90	2	10	0.35	1	345
Great View Apartment in Geneva / 20min city center	Jean-Pierre	ThvYnex	Entire home/apt	118	14	19	0.30	1	63
HOMESWISS HOME	Franzy	Carouge	Private room	80	1	42	0.89	1	329

ANNEXE VII : DONNÉES SÉLECTIONNÉES - MEILLEURES COMPAGNIES EN INFORMATIQUE ENTRE 2014 ET 2015

Employer Name	Early Career Median Pay	Mid-Career Median Pay	Median Age	Years of Experience	Years with Company	% Female	% High Job Satisfaction	% High Job Meaning	% High Job Stress
HP	65400	104200	38	7,7	6,3	32	64	41	56
Oracle	71400	126400	37	9	5,3	29	64	32	56
IBM	72700	112700	36	7,4	7,1	25	59	40	62
Microsoft	102500	139700	33	6,4	4,8	27	72	59	58
Qualcomm	94400	132500	33	7,7	4,3	20	69	50	62
Cisco	83800	134700	33	6,7	4	23	76	60	52
eBay	88900	136000	33	6,9	2,7	43	78	55	66
Samsung	65900	143000	33	6	2,3	37	64	36	69
Intel	85800	129400	32	5,5	5	25	72	66	61
Adobe	91900	141900	31	5,2	3,3	28	84	24	47
Apple	91300	145400	31	6	2,7	31	73	74	66
Amazon	81900	130400	30	5,2	2	26	69	42	66
Google	106900	151600	30	5,2	2	30	89	72	53
Salesforce	84800	154600	30	5,3	1,8	30	89	54	55
Tesla	81400	118500	30	5,6	1,6	20	70	89	70
LinkedIn	110800	159600	29	5		42	0	0	0
SpaceX	78500	NA	29	6,1	2,3	14	73	92	88
Facebook	116800	149300	29	4,3	1,1	32	96	78	44

ANNEXE VIII : QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO

Questionnaire pour tester la plateforme VASCO

Introduction

Nous allons vous demander de tester la plateforme de visualisation de données. Pour commencer, nous allons vous poser quelques questions générales. Puis, nous allons vous demander de télécharger sur votre ordinateur les fichiers `cereal.csv` et `TopTechCompanies2014_2015.csv`. Ensuite, vous irez sur le lien du site internet (<https://vasco.tb.hevs.ch/>) et nous allons regarder comment vous vous familiarisez avec la plateforme afin de savoir si elle est compréhensible. Pour finir, nous vous laisserons répondre aux questions.

Durant ces étapes, nous allons observer votre stratégie/méthode afin d'améliorer la plateforme. Attention, le temps est compté pour chaque question, mais ce n'est pas grave si vous n'arrivez pas à y répondre. Nous ne nous intéressons pas à votre performance individuelle, mais nous voulons observer à quel point la plateforme est compréhensible et facile à utiliser.

Le but de ce questionnaire est d'analyser la manière dont vous réagissez avec la plateforme et pour cela, nous vous invitons à penser à « haute voix » afin que l'on puisse bien comprendre vos réactions.

Partie 1

Êtes-vous une femme ou un homme ?

Quel est votre âge ?

Quel est votre métier ?

Connaissez-vous les fichiers `.csv` ?

Avez-vous déjà utilisé Excel pour créer des graphiques ?

Avez-vous déjà utilisé le logiciel Tableau ?

Avez-vous déjà utilisé le logiciel RAWGraphs ?

Avez-vous déjà utilisé le logiciel data-Illustrator ?

Avez-vous déjà utilisé le logiciel datawrapper ?

Avez-vous déjà utilisé le logiciel datavoyager ?

Autres logiciels ou plateforme de visualisation non mentionnés ci-dessus ?

.....

Partie 2

Voici un petit aperçu du fichier cereal.csv sous forme de tableau Excel. Chaque ligne contient une sorte de céréale avec les informations nutritionnelles sur celle-ci. Nous pouvons voir que dans la première colonne il y a le nom des céréales puis, dans la deuxième, le nom des fabricants. Toutes les colonnes qui suivent sont les informations nutritionnelles relatives à celles-ci. En d'autres termes, c'est ce que vous pouvez observer à l'arrière des emballages des aliments.

Cereal Name	Manufacturer	Calories	Protein (g)	Fat	Sodium	Dietary Fiber	Carbs	Sugars	Display Shelf	Potassium	Vitamins and Minerals	Serving Size Weight	Cups per Serving
100%_Bran	Nabisco	70	4	1	130	10	5	6	3	280	25	1	0.33
100%_Natural_Bran	Quaker Oats	120	3	5	15	2	8	8	3	135	0	1	-1
All-Bran	Kelloggs	70	4	1	260	9	7	5	3	320	25	1	0.33
All-Bran_with_Extra_Fiber	Kelloggs	50	4	0	140	14	8	0	3	330	25	1	0.5
Almond_Delight	Ralston Purina	110	2	2	200	1	14	8	3	-1	25	1	0.75
Apple_Jacks	Kelloggs	110	2	0	125	1	11	14	2	30	25	1	1
Bran_CheX	Ralston Purina	90	2	1	200	4	15	6	1	125	25	1	0.67
Bran_Flakes	Post	90	3	0	210	5	13	5	3	190	25	1	0.67
Cap'nCrunch	Quaker Oats	120	1	2	220	0	12	12	2	35	25	1	0.75

Aperçu du fichier csv

Cereal Name,Manufacturer,Calories,Protein (g),Fat,Sodium,Dietary Fiber,Carbs,Sugars,Display Shelf,Potassium,Vitamins and Minerals,Serving Size Weight,Cups per Serving
100%_Bran,Nabisco,70,4,1,130,10,5,6,3,280,25,1,0.33
100%_Natural_Bran,Quaker Oats,120,3,5,15,2,8,8,3,135,0,1,-1
All-Bran,Kelloggs,70,4,1,260,9,7,5,3,320,25,1,0.33
All-Bran_with_Extra_Fiber,Kelloggs,50,4,0,140,14,8,0,3,330,25,1,0.5
Almond_Delight,Ralston Purina,110,2,2,200,1,14,8,3,-1,25,1,0.75
Apple_Cinnamon_Cheerios,General Mills,110,2,2,180,1.5,10.5,10,1,70,25,1,0.75
Apple_Jacks,Kelloggs,110,2,0,125,1,11,14,2,30,25,1,1

Téléchargez le fichier cereal.csv et aller sur le site <https://vasco.tb.hevs.ch/> et importez-le

- Combien y a-t-il de lignes dans le fichier ?
- Quelle est la valeur maximale de gras (Fat) ?
- Quelle est la valeur minimale de fabricant (Manufacturer) ?.....
- Quelle est la moyenne des calories ?
- Pouvez-vous sélectionner le graphique qui représente les calories et le sucre ?
 - Que représentent les points ?.....
.....
 - Pourquoi y a-t-il des points plus foncés que d'autres ?
- Pouvez-vous modifier le graphique avec un fond foncé ?
- Pouvez-vous exporter le graphique en SVG ?

Partie 3

Voici un aperçu du fichier TopTechCompanies2014_2015.csv sous forme de tableau Excel. Chaque ligne représente une entreprise dans l'informatique avec des informations relatives à celle-ci. Voici un petit descriptif de chaque colonne :

- Employer Name : le nom de l'entreprise informatique
- Early Career Median Pay : une moyenne du salaire par année dans les premières années de travail
- Mid-Career Median Pay : une moyenne du salaire par année après quelques années d'expérience
- Median Age : l'âge moyen
- Years of Experience : le nombre d'années d'expérience dans le monde du travail
- Years with company : le nombre d'années au sein de l'entreprise
- % Female : le pourcentage de femme dans l'entreprise
- % High Job Satisfaction : le pourcentage de satisfaction à propos du travail
- % High Job Meaning : le pourcentage du sens élevé du travail
- % High Job Stress : le pourcentage de stress au travail

Employer Name	Early Career Median Pay	Mid-Career Median Pay	Median Age	Years of Experience	Years with Company	% Female	% High Job Satisfaction	% High Job Meaning	% High Job Stress
HP	65400	104200	38	7,7	6,3	32	64	41	56
Oracle	71400	126400	37	9	5,3	29	64	32	56
IBM	72700	112700	36	7,4	7,1	25	59	40	62
Microsoft	102500	139700	33	6,4	4,8	27	72	59	58
Qualcomm	94400	132500	33	7,7	4,3	20	69	50	62
Cisco	83800	134700	33	6,7	4	23	76	60	52
eBay	88900	136000	33	6,9	2,7	43	78	55	66
Samsung	65900	143000	33	6	2,3	37	64	36	69
Intel	85800	129400	32	5,5	5	25	72	66	61
Adobe	91900	141900	31	5,2	3,3	28	84	24	47
Apple	91300	145400	31	6	2,7	31	73	74	66
Amazon	81900	130400	30	5,2	2	26	69	42	66
Google	106900	151600	30	5,2	2	30	89	72	53
Salesforce	84800	154600	30	5,3	1,8	30	89	54	55
Tesla	81400	118500	30	5,6	1,6	20	70	89	70
LinkedIn	110800	159600	29	5		42	0	0	0
SpaceX	78500	NA	29	6,1	2,3	14	73	92	88
Facebook	116800	149300	29	4,3	1,1	32	96	78	44

Téléchargez le fichier TopTechCompanies2014_2015.csv et importez-le

- Pourquoi « Employer Name » (nom de l'entreprise) n'a pas de somme, moyenne, min et max ?
- Quelle est la moyenne des années d'expérience (Years of Experience) ?
- Quelle est la tranche d'âge moyenne (Median Age) la plus représentée ?
 - 30-31 ans
 - 33-34 ans
 - 36-37 ans

11. Dans quelle compagnie y a-t-il le plus de femmes (female) ?
- a. Et quel est le pourcentage ?
12. À 30 ans, quelles sont (environ) les années d'expérience (Years of Experience) ?
.....
13. Pour le graphique "Years of Experience / % High Job Satisfaction", qu'observez-vous ?
- b. Plus les années d'expérience passent plus la satisfaction augmente
- c. Plus les années d'expérience passent plus la satisfaction diminue
- d. Rien de spécial ne se passe
- e. Aucune idée

Partie 4

- Est-ce qu'une plateforme comme celle-ci est utile selon vous?
-
- Si oui, pour quel type de personne?
- Si non, pourquoi pas?
- Est-ce que vous seriez prêt à utiliser cette plateforme ?
- Si oui, pour quel type de tâche?
- Si non, quels seraient vos conseils d'amélioration ?
-

Merci de votre participation !

ANNEXE IX : ANALYSE DU QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO

Analyse du questionnaire pour tester la plateforme VASCO

Type de tâche :

- Extrema: trouver un maximum ou un minimum de points de données
- Variation: détecter des tendances, des similitudes ou des divergences dans les données
- Average: estimation d'une valeur moyenne
- Comparaison: comparer différentes valeurs ou tendances

Niveau de difficulté:

- Facile (moins de 10 secondes pour répondre)
- Moyen (moins de 20 secondes pour répondre)
- Difficile (30 secondes ou plus pour répondre)

N°	Type de tâches	Difficulté	Question	Réponse	Observation	Analyse	Réponse correcte ?
1		Facile	Combien y-a-t-il de lignes dans le fichier ?	74	Où va-t-elle regarder ? ----- En combien de temps ?		
2	Extrema	Facile	Quelle est la valeur maximale de gras (Fat) ?	5	Où va-t-elle regarder ? ----- En combien de temps ?		
3	Extrema	Facile	Quelle est la valeur minimale de fabricant (Manufacturer) ?	Aucune, c'est une valeur qualitative	Où va-t-elle regarder ? ----- En combien de temps ?		
4	Average	Facile	Quelle est la moyenne des calories ?	105,54	Où va-t-elle regarder ? ----- En combien de temps ?		

5		Difficile	Pouvez-vous sélectionner le graphique qui représente les calories et le sucre ?		Elle tri avec les filtres si oui : elle choisit les calories ou le sucre ?		
					En combien de temps ?		
5a	Variation	Moyen	Que représentent les points ?	Les valeurs du fichier csv	Combien de temps de réflexion ?		
5b	Variation	Difficile	Pourquoi y-a-t-il des points plus foncés que d'autres ?	beaucoup de valeurs similaires	Combien de temps de réflexion ?		
6		Facile	Pouvez-vous modifier le graphique avec un fond foncé ?	oui	La personne clique directement sur le radio button « dark » ?		
					Et après, elle clique directement sur le bouton « suivant » ?		
					En combien de temps ?		
7		Facile	Pouvez-vous exporter le graphique en SVG ?	Oui	Voir si elle clique directe sur le bouton « exporter en SVG »		
					En combien de temps ?		
Deuxième partie finie : <ul style="list-style-type: none"> • demander au testeur ses premières impressions • noter les facilités • noter les difficultés • regarder avec le testeur s'il a bien compris l'interface et si non, bien lui expliquer 							

		Facile	Maintenant, il faut importer le fichier TopTechCompanies2014_2015.csv		La personne appuie sur : <ul style="list-style-type: none"> • le bouton « je recommence » • l'onglet « importation » • sur VASCO 		
8	Extrema et average	Moyen	Pourquoi « Employer Name » (nom de l'entreprise) n'a pas de somme, moyenne, min et max ?	Car valeur de type qualitative	Combien de temps de réflexion ?		
9	Average	Facile	Quelle est la moyenne d'âge (Median Age) ?	32 ans	Où va-t-elle regarder ? ----- En combien de temps ?		
10	Average	Difficile	Quelle est la tranche d'âge moyenne (Median Age) la plus représentée ?	33-34 ans	Où va-t-elle regarder ? <ul style="list-style-type: none"> • tableau récapitulatif • directement dans les graphiques → le mieux est de regarder dans le tableau sur l'histogramme ----- En combien de temps ?		
11	Variation	Difficile	Dans quelle compagnie y-a-t-il le plus de femmes (female) ?	Ebay	Où va-t-elle regarder ? → le mieux est de sélectionner le graphique en allant sur la page modification pour mieux voir ----- En combien de temps ?		
11a	Variation	Difficile	Et quel est le pourcentage exact ?	43%	Est-ce qu'elle va utiliser la souris sur le graphique et voir le toolbox ? ----- La personne doit revenir en arrière sur la page « Sélection » L'a-t-on aidé ? ----- En combien de temps ?		

12	Comparison	Difficile	À 30 ans, quelles sont (environ) les années d'expérience (Years of Experience) ?	environ 4 à 6 ans	Est-ce que la personne a filtré pour regarder		
					Est-ce qu'elle regarde le graphique Median Age / Years of Experience		
13	Comparison	Difficile	Pour le graphique "Years of Experience / % High Job Satisfaction", qu'observez-vous ?	Plus les années d'expérience passent plus la satisfaction diminue	Où va-t-elle regarder ?		
					Combien de temps de réflexion ?		
Troisième partie finie :							
<ul style="list-style-type: none"> demander au testeur ses deuxièmes impressions (si plus facile ou non que la première) noter les facilités noter les difficultés 							

ANNEXE X : RÉPONSE DU QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO AVANT MODIFICATION

Réponse du questionnaire pour tester la plateforme VASCO avant modification

Premiers testeurs						
Questions	1- beta	2	3	4	5	6
Partie 1 - questions générales						
1. Êtes-vous une femme ou un homme ?	homme	Femme	femme	femme	homme	femme
2. Quel est votre âge ?	24 ans	28 ans	24 ans	24 ans	24 ans	24 ans
3. Quel est votre métier ?	Étudiant en tourisme	Chargée d'accueil, événements & animations au château d'aigle	Assistante de recherche IIG	Informaticienne	Étudiant en informatique	Co-fondatrice start-up
4. Connaissez-vous les fichiers CSV ?	Aucune idée	non	oui	oui	oui	non
5. Avez-vous déjà utilisé Excel pour créer des graphiques ?	oui	oui	oui	oui	oui	oui
6. Avez-vous déjà utilisé le logiciel Tableau ?	non	non	oui	non	oui	non
7. Avez-vous déjà utilisé le logiciel RAWGraphs ?	non	non	non	non	non	non
8. Avez-vous déjà utilisé le logiciel data-illustrator ?	non	non	non	non	non	non
9. Avez-vous déjà utilisé le logiciel datawrapper ?	non	non	non	non	non	non
10. Avez-vous déjà utilisé le logiciel datavoyager ?	non	non	non	non	non	non
11. Autres logiciels ou plateforme de visualisation non mentionnés ci-dessus ?	non	non	ARCGIS, knime	Power Pivot, knime	Power Pivot et knime	Le Sphinx

Type de tâches	Difficulté	Partie 2 - première approche avec la plateforme avec des questions en relation avec un fichier cereal.csv						
	Facile	12. Combien y a-t-il de lignes dans le fichier ?	13	74	74	74	74	74
		a) Où va-t-elle regarder ?	Compter le nombre de lignes dans le tableau résumé Il a dit 14 lignes (les dimensions)	Tableau puis scroll puis regarde	Tableau	tableau	Tableau	tableau
		b) En combien de temps ?	5 sec	5 sec	3 sec	3 sec	7 sec	10 sec
		Réponse correcte ?	✗	✓	✓	✓	✓	✓
Extrema	Facile	13. Quelle est la valeur maximale de gras (Fat) ?	5	5	5	5	5	5
		a) Où va-t-elle regarder ?	Tableau résumé	Scroll dans la liste des graphiques puis revient dans le tableau	tableau	tableau	Regarde la liste puis revient sur le tableau	Tableau et liste puis revient à tableau
		b) En combien de temps ?	5 sec	40 sec	3 sec	3 sec	40 sec	50 sec
		Réponse correcte ?	✓	✓	✓	✓	🔊)	✓
Extrema	Facile	14. Quelle est la valeur minimale de fabricant (Manufacturer) ?	aucune	aucune	aucune	qualitatif	Pas possible	aucune
		a) Où va-t-elle regarder ?	Tableau résumé	tableau	tableau	tableau	Tableau	tableau
		b) En combien de temps ?	7 sec	20 sec	5 sec	3 sec	10 sec	2 sec
		Réponse correcte ?	✓	✓	✓	✓	✓	✓
Average	Facile	15. Quelle est la moyenne des calories ?	105,54	105,54	105,54	105,54	105,54	105,54
		a) Où va-t-elle regarder ?	Tableau résumé	Tableau	Tableau	tableau	tableau	tableau

		b) En combien de temps ?	3 sec	3 sec	3 sec	3 sec	3 secondes	2 sec
		Réponse correcte ?						
	Difficile	16. Pouvez-vous sélectionner le graphique qui représente les calories et le sucre ?	oui	oui	oui	oui	oui	oui
		a) Elle tri avec les filtres si oui : elle choisit les calories ou le sucre ?	Pas vu le tri en premier mais après vu du coup utilise et choisi calories	Trie sur les calories	Liste et scoll pour trouver	Liste et scroller puis trie calories	Trie en sélectionnant calorie et sélectionne le graph	Pas de tri, elle scroll dans la liste pour trouver
		b) En combien de temps ?	50 sec	15 sec	20 sec	25 sec	30 sec	1m30
		Réponse correcte ?	🔊	🔊	🔊	✓	✓	🔊
Variation	Moyen	17. Que représentent les points ?	Les valeurs	données	Le nombre de sucres par calorie	céréales	Calories par rapport au sucre	Le nombre de calories/sucre
		a) Combien de temps de réflexion ?	30 sec	20 sec	5 sec	30 sec	5 sec	5 sec
		Réponse correcte ?	🔊	✓	✓	✓	✓	✓
Variation	Difficile	18. Pourquoi y a-t-il des points plus foncés que d'autres ?	Plusieurs céréales	?	?	?	?	?
		a) Combien de temps de réflexion ?	5 sec	Ne sait pas	Ne sait pas	Ne sait pas	Ne sait pas	Ne sait pas
		Réponse correcte ?	✓	✗	✗	✗	✗	✗
	Facile	19. Pouvez-vous modifier le graphique avec un fond foncé ?	oui	oui	oui	oui	oui	oui
		a) La personne clique directement sur le radio bouton « dark » ?	Oui directement	oui	oui	Oui	oui	oui
		b) Et après, elle clique directement sur le bouton ?	Reste en premier en essayant le clic droit	D'abord essaye d'appuyer sur « exportation » puis sur btn	Non	oui	Oui mais trouve pas clair le btn suivant	oui
		c) En combien de temps ?	9sec	5 sec	2 sec	10 sec	3 sec	5 sec

		Réponse correcte ?	✓	✓	✓	✓	✓	✓
	Facile	20. Pouvez-vous exporter le graphique en SVG ?	oui	oui	Oui en png avec un clic droit	oui	oui	oui
		a) Voir si elle clique directe sur le bouton « exporter en SVG »	Oui mais essaye en premier clic droit pour exporter	oui	Elle a essayé un clic droit sur la page modification	oui	oui	oui
		b) En combien de temps ?	1 sec	1 sec	10 sec	3 sec	1 sec	2 sec
		Réponse correcte ?	✓	✓	✗	✓	✓	✓
		<ul style="list-style-type: none"> • Premières impressions • Facilités ? • Difficultés ? regarder avec le testeur s'il a bien compris l'interface et si non, bien lui expliquer	Mettre le tableau résumé plus petit, car ne voit pas les graphiques en dessous ou mettre les deux à côté ou mettre des onglets → trouve pas forcément utile le tableau résumé → trouve que la colonne aperçu ne sert à rien ou si on laisse il faudrait pouvoir cliquer dessus afin de voir les informations en plus grandes Il faudrait mettre un titre avant le tableau résumé pour savoir ce que c'est	Jolie interface et compréhensible et assez facile Btn suivant il faudrait écrire exporter Dire pourquoi c'est plus foncé les points sur le graphique	Scroll bar pas visible donc on ne sait pas qu'il y a d'autres informations. L'exportation n'est pas forcément explicite Dire pourquoi des points plus foncés dans le graphique Jolie interface	Pouvoir voir mieux l'histogramme Pourquoi pas afficher que quelques graph et après avec le tri en afficher plus ?	Mettre exporter sur le btn à la place de suivant, car ce n'est pas clair Mais joli le menu avec les étapes	Problème de la scroll bar Mais interface jolie et compréhensible après ne pense pas très utile

Partie 3 - deuxième approche avec la plateforme avec des questions en relation avec un fichier TopTechCompanies2014_2015.csv								
	Facile	Importer le fichier TopTechCompanies2014_2015.csv a) La personne appuie sur : • btn « je recommence » • onglet « importation » • tab « VASCO »	btn « Je recommence »	btn « Je recommence »	btn « Je recommence »	btn « Je recommence »	btn « Je recommence »	btn « Je recommence »
Extrema et average	moyen	21. Pourquoi « Employer Name » (nom de l'entreprise) n'a pas de somme, moyenne, min et max ?	Données qualitatives	qualitatives	Parce que c'est du texte	qualitatif	Car c'est des noms, ce n'est pas quantitatif	Pas possible
		a) Combien de temps de réflexion ?	17 sec	15 sec	3 sec	1 sec	2 sec	15 sec
		Réponse correcte ?	✓	✓	✓	✓	✓	✓
Average	Facile	22. Quelle est la moyenne des années d'expérience (Years of Experience) ?	6,17	6,17	6,17	6,17	6,17	6,17
		a) Où va-t-elle regarder ?	Tableau résumé	tableau	tableau	Tableau	Tableau	Tableau
		b) En combien de temps ?	15 sec	7 sec	3sec	4 sec	3 sec	4 sec
		Réponse correcte ?	✓	✓	✓	✓	✓	✓
Average	Difficile	23. Quelle est la tranche d'âge médiane (Median Age) la plus représentée ?	33-34 ans	33-34 ans	33-34 ans	33-34 ans	33-34 ans	33-34 ans
		a) Où va-t-elle regarder ? • tableau récapitulatif puis histogramme • directement dans les graphiques	À deviner grâce à la moyenne et les propositions proposées en dessous	Tableau et sur l'histogramme	Regarde dans le tableau et aperçu mais dit que c'est trop petit	N'as pas compris la question donc je lui ai expliqué Et après zoom avec son trackpad sur l'aperçu	En premier va regarder dans la liste puis revient sur le tableau et sur l'aperçu et zoom avec son trackpad	Tableau et regarde l'histogramme
		b) En combien de temps ?	34 sec	15 sec	17 sec	30 sec	35 sec	1 min

		Réponse correcte ?	✓	🔊	🔊	✗	🔊	🔊
Variation	Difficile	24. Dans quelle compagnie y a-t-il le plus de femmes (female) ?	ebay	Linkedin	Linked-in	ebay	ebay	ebay
		a) Où va-t-elle regarder ? → le mieux est de sélectionner le graphique en allant sur la page modification pour mieux voir	J'ai dit qu'il fallait sélectionner un graphique	Tableau puis liste et trie %female et regarde mais ne clique pas dessus donc ne trouve pas la bonne réponse	Dans la liste mais n'appuie pas donc réponse fausse	En premier tableau puis liste et trie sur % female Ne clique pas donc je lui ai dit	Sur la liste et trie sur % femme puis sélectionne le graphique	Ne trie pas cherche en scrollant
		b) En combien de temps ?	40 sec	30 sec	7 sec	40 sec	20 sec	30 sec
		Réponse correcte ?	🔊	✗	✗	🔊	✓	✓
Variation	Difficile	25. Et quel est le pourcentage exact ?	43%	?	Environ 40%	43%	47,5%	47%
		a) Est-ce qu'elle va utiliser la souris sur le graphique et voir le toolbox ?	Oui mais j'ai aidé	Je lui ai dit qu'il fallait cliquer et qu'il y avait le toolbox	estime mais ne clique pas pour voir mieux	Oui	Oui après lui avoir expliqué	non
		b) La personne doit revenir en arrière sur la page « Sélection »	Clique sur le bouton retour	Clique sur le bouton retour	Clique sur le bouton retour	Clique sur btn retour	Clique sur btn retour	Clique sur btn retour
		c) En combien de temps ?	30 sec	10 sec	2 sec	20 sec	20 sec	1 sec
		Réponse correcte ?	🔊	🔊	✗	✓	🔊	✗
Comparaison	Difficile	26. À 30 ans, quelles sont (environ) les années d'expérience (Years of Experience) ?	5,6 ans	6 ans	4-6 ans	5	5,5-6ans	5,5 / 4 ans
		a) Est-ce que la personne a filtré pour regarder	Oui choisi years of experience	Trie sur Years of experience	Filtre sur years of experience	Trie sur les Years of experience	Ne comprend pas la question	Trie, car on lui a dit qu'il y avait cette fonction puis sélectionne le graphique
		b) Est-ce qu'elle regarde le graphique Median Age / Years of Experience	Oui et utilise la toolbox	Oui et se sert de la toolbox	oui	Ne sélectionne pas	ne sélectionne pas	Oui

		c) En combien de temps ?	1 min	45 sec	20 sec	25 sec	1 min	1m35
		Réponse correcte ?	✓	✓	✓	✓	✗	🔊)
Comparaison	Difficile	27. Pour le graphique "Years of Experience / % High Job Satisfaction", qu'observez-vous ?	Plus les années passent plus la satisfaction diminue	Plus les années passent plus la satisfaction diminue	Plus les années passent plus la satisfaction diminue	Plus les années passent plus la satisfaction diminue	Plus les années passent plus la satisfaction diminue	Plus les années passent plus la satisfaction diminue
		a) Où va-t-elle regarder ?	Trie et sélectionne	Dans la liste et trie sur years of experience et clic puis utilise la toolbox	Sur la liste puis trie	Sur la liste et sélectionne	Trie et sélectionne le graphique	Trie en sélectionnant Years of experience
		b) Combien de temps de réflexion ?	1m30	1m20	10 sec	20 sec	15 sec	35 sec
		Réponse correcte ?	🔊)	✓	✓	✓	✓	✓
		<ul style="list-style-type: none"> • Deuxièmes impressions (si plus facile ou non que la première) • Facilités ? • Difficultés ? • Amélioration ? 	Plus difficile les questions mais plus compréhensible les informations Doit bien chercher les graphiques pour trouver le bon heureusement qu'il y a les titres	Elle connaissait l'interface donc plus facile. Trouve que l'aperçu est trop petit et que c'est long à charger les données	Pourquoi ne pas mettre deux filtres ? pour l'axe x et y L'affichage de l'histogramme trop petit Le graphique sur modification est trop petit, il faudrait l'agrandir Mettre un numéro avant chaque étape par ex 1 importation > 2 sélections, etc Enelever les titres de chaque page, inutile Problème avec la scroll bar car sur mac on ne la voit pas	Plus facile la deuxième partie, car connaît l'interface Trouve qu'il y a beaucoup d'information sur la page de sélection, il faudrait épurer par exemple avec des onglets Trouve qu'on pourrait mettre moins de propositions de graphiques dans la liste et mettre que deux graph par ligne comme ça plus visible	Mettre un titre au-dessus du tableau, car on ne comprend pas ce que c'est (ex Synthèse du fichier importé) puis afficher le nombre de lignes et après le tableau Selon lui le but de la plateforme n'est pas d'exporter un graphique mais d'analyser et comprendre les données importées Changer le texte du filtre : Trié les données selon... Aime beaucoup le design de la plateforme	Il faudrait agrandir l'aperçu de l'histogramme, car illisible N'a pas vu l'option du filtre et la toolbox Mais sinon bien user-friendly

								pourquoi pas mettre une icône avec une flèche pour dire de descendre			
Total correct :			10	11	10	13	11	11			
Total faux :			1	2	4	2	2	2			
Total aidé :			5	3	2	1	3	3			
Partie 4 - questions générales											
28. Est-ce qu'une plateforme comme celle-ci selon vous ?			Oui pour les personnes qui utilisent ce genre de données	Oui	oui	Oui	oui	non			
29. Si oui, pour quel type de personne ?			Des gens qui font des statistiques	Chefs d'entreprise pour les comparaisons et l'analyse	Pour des étudiants/entrprises qui veulent des informations avec des graphiques rapidement et facilement	Preneur de décision	Employeurs et pour les RH				
30. Si non, pourquoi pas ?								Difficile de cerner l'utilité			
31. Est-ce que vous seriez prêt à utiliser cette plateforme ?			Pas le besoin	non	Non, j'utilise knime	Oui si besoin mais en version finale	Pourquoi pas	Pas l'utilité			
32. Si oui, pour quel type de tâche ?						Prédiction, analyse rapide, remplacer un peu Excel					
33. Si non, quels seraient vos conseils d'amélioration ?				Faire de la « pub » auprès des gens qui utilisent les CSV	J'utilise knime qui est super performant mais qui est compliqué mais vu que je l'utilise souvent j'ai l'expérience avec.	Super début de prototype	Jolie plateforme	Mieux définir l'utilité exacte			

ANNEXE XI : RÉPONSE AU QUESTIONNAIRE POUR TESTER LA PLATEFORME VASCO APRÈS MODIFICATION

Réponse au questionnaire pour tester la plateforme VASCO après modification

Deuxièmes testeurs				
Questions	7	8	9	10
Partie 1 - questions générales				
1. Êtes-vous une femme ou un homme ?	Femme	Homme	femme	femme
2. Quel est votre âge ?	55 ans	57 ans	27 ans	26 ans
3. Quel est votre métier ?	Productrice de spectacle	Enseignant spécialisé	Enseignante	Étudiante en master climatologie et océanographie
4. Connaissez-vous les fichiers CSV ?	non	Déjà entendu parlé	non	oui
5. Avez-vous déjà utilisé Excel pour créer des graphiques ?	oui	Oui il y a longtemps	oui	oui
6. Avez-vous déjà utilisé le logiciel Tableau ?	non	non	non	non
7. Avez-vous déjà utilisé le logiciel RAWGraphs ?	non	non	non	non
8. Avez-vous déjà utilisé le logiciel data-illustrator ?	non	non	non	non
9. Avez-vous déjà utilisé le logiciel datawrapper ?	non	non	non	non
10. Avez-vous déjà utilisé le logiciel datavoyager ?	non	non	non	non
11. Autres logiciels ou plateforme de visualisation non mentionnés ci-dessus ?	Pages, word	non	Word, Pages, Keynote, Powerpoint	Matlab, R

Type de tâches	Difficulté	Partie 2 - première approche avec la plateforme avec des questions en relation avec un fichier cereal.csv				
	Facile	12. Combien y a-t-il de lignes dans le fichier ?	74	14	74	74
		a) Où va-t-elle regarder ?	tableau	Compte dans le tableau les lignes	Tableau, en premier compte et après vois qu'elle s'est trompée et note la bonne réponse	tableau
		b) En combien de temps ?	10 sec	15 sec	30 sec	3 sec
		Réponse correcte ?	✓	✗	✓	✓
Extrema	Facile	13. Quelle est la valeur maximale de gras (Fat) ?	5	5	5	5
		a) Où va-t-elle regarder ?	Tableau résumé	Tableau résumé	tableau	tableau
		b) En combien de temps ?	10 sec	10 sec	15 sec	3 sec
		Réponse correcte ?	✓	✓	✓	✓
Extrema	Facile	14. Quelle est la valeur minimale de fabricant (Manufacturer) ?	aucune	aucune	aucune	aucune
		a) Où va-t-elle regarder ?	Tableau résumé	tableau	tableau	tableau
		b) En combien de temps ?	20 sec	20 sec	15 sec	3 sec
		Réponse correcte ?	✓	✓	✓	✓
Average	Facile	15. Quelle est la moyenne des calories ?	105,54	105,54	105,54	105,54
		a) Où va-t-elle regarder ?	Tableau résumé	Tableau	Tableau	tableau
		b) En combien de temps ?	3 sec	3 sec	8 sec	2 sec
		Réponse correcte ?	✓	✓	✓	✓
	Difficile	16. Pouvez-vous sélectionner le graphique qui représente les calories et le sucre ?	oui	oui	oui	oui

		a) Elle tri avec les filtres si oui : elle choisit les calories ou le sucre ?	Ne trie pas et je l'aide pour lui montrer les filtres	J'aide pour trouver la liste des graphiques, car ne voit pas les onglets, car même couleur	Va sur la liste, mais sélection le mauvais et remarque que c'est pas le bon, mais comprend pas donc je l'aide et lui dis qu'il faut le graph qui représente CALORIE et SUCRE et trouve	Trie sur le sucre
		b) En combien de temps ?	2m30	1m30	3min	40 sec
		Réponse correcte ?	🔊	🔊	🔊	✓
Variation	Moyen	17. Que représentent les points ?	Quantité de sucre/calories	Un produit en fonction des calories et du sucre	Le nombre de calories pour le nombre de sucres	Les calories et les sucres
		a) Combien de temps de réflexion ?	15 sec	15 sec	30 sec	10 sec
		Réponse correcte ?	✓	✓	✓	✓
Variation	Difficile	18. Pourquoi y a-t-il des points plus foncés que d'autres ?	Données similaires	?	Il s'agit de données similaires	Points foncés = données similaires
		a) Combien de temps de réflexion ?	30 sec	Ne sait pas et ne regarde pas tout le site	10 sec	10 sec
		Réponse correcte ?	✓	🔊	✓	✓
	Facile	19. Pouvez-vous modifier le graphique avec un fond foncé ?	oui	oui	oui	oui
		a) La personne clique directement sur le radio button « dark » ?	Oui	oui	oui	Clique en premier sur le graph puis sur le radio button
		b) Et après, elle clique directement sur le bouton ?	Pas directement cherche avant de voir le btn exporter, car pas bien responsif	oui	oui	Appuis sur l'onglet « selection »
		c) En combien de temps ?	25 sec	20 sec	20 sec	5 sec
		Réponse correcte ?	✓	✓	✓	✓
	Facile	20. Pouvez-vous exporter le graphique en SVG ?	oui	oui	Oui	oui
		a) Voir si elle clique directe sur le bouton « exporter en SVG »	Oui, mais ne connaît pas SVG du coup cherche un peu partout	oui	oui	Oui

		b) En combien de temps ?	30 sec	2 sec	5 sec	2 sec
		Réponse correcte ?	✓	✓	✓	✓
		<ul style="list-style-type: none"> Premières impressions Facilités ? Difficultés ? regarder avec le testeur s'il a bien compris l'interface et si non, bien lui expliquer	Les onglets ne sont pas visibles, on dirait une barre avec des mots dessus -> mettre en plus gros la largeur ou écrire étape 1 et étape 2 en dessus Elle était prise par les questions du coup ne regardait pas très bien l'interface Peut-être afficher deux filtres, car encore trop de possibilité lors de la recherche Rapprocher le texte et la dropdown	Ecrire par exemple liste complète des graph Mettre icône descendre et monter pour montrer qu'il y a plusieurs graph dans la liste	Ne sait pas comment importer j'ai dû lui dire d'appuyer sur le btn Sur son ordi le btn « j'essaye » ne se voit pas très bien	p-e afficher que l'onglet au début et rien d'autre puis lorsqu'on clic, les infos s'affichent pour elle très compréhensible, car l'habitude de travailler sur l'ordinateur
Partie 3 - deuxième approche avec la plateforme avec des questions en relation avec un fichier TopTechCompanies2014_2015.csv						
	Facile	Importer le fichier TopTechCompanies2014_2015.csv a) La personne appuie sur : <ul style="list-style-type: none"> btn « je recommande » onglet « importation » tab « VASCO » 	btn « Je recommande »	btn « Je recommande »	btn « Je recommande »	onglet « importation »
Extrema et average	moyen	21. Pourquoi « Employer Name » (nom de l'entreprise) n'a pas de somme, moyenne, min et max ?	Parce que c'est le nom de l'entreprise	C'est le nom de l'entreprise	Parce que c'est des noms	Valeurs qualitatives
		a) Combien de temps de réflexion ?	17 sec	3 sec	30 sec	3 sec
		Réponse correcte ?	✓	✓	✓	✓
Average	Facile	22. Quelle est la moyenne des années d'expérience (Years of Experience) ?	6,17	6,17	6,17	6,17
		a) Où va-t-elle regarder ?	Tableau résumé	tableau	tableau	Tableau
		b) En combien de temps ?	12 sec	7 sec	10 sec	3 sec
		Réponse correcte ?	✓	✓	✓	✓
Average	Difficile	23. Quelle est la tranche d'âge médiane (Median Age) la plus représentée ?	33-34 ans	33-34 ans	33-34 ans	33-34 ans

		a) Où va-t-elle regarder ? • tableau récapitulatif puis histogramme • directement dans les graphiques	Liste des graphiques puis j'aide pour lui dire que c'est dans tableau puis regarde sur l'histogramme	Tableau récapitulatif puis histogramme	Tableau puis regarde l'histogramme	Tableau puis histogramme
		b) En combien de temps ?	4 min	10 sec	20 sec	10 sec
		Réponse correcte ?	🔊	✓	✓	✓
Variation	Difficile	24. Dans quelle compagnie y a-t-il le plus de femmes (female) ?	ebay	ebay	ebay	ebay
		a) Où va-t-elle regarder ? → le mieux est de sélectionner le graphique en allant sur la page modification pour mieux voir	Liste et trie sur % Female	Aidé pour dire qu'il faut chercher dans la liste et non dans le tableau et après il a trié et trouvé	Trie sur % female et sélectionne le graph puis utilise la toolbox	Tableau puis liste et trie % female puis sélectionne le graph
		b) En combien de temps ?	40 sec	3 min	40 sec	35 sec
		Réponse correcte ?	✓	🔊	✓	✓
		25. Et quel est le pourcentage exact ?	43%	Environ 42%	43%	43%
Variation	Difficile	a) Est-ce qu'elle va utiliser la souris sur le graphique et voir le toolbox ?	Ne regarde pas tout le site se focus sur le graphique à chercher et ne lis pas tout	Ne lis pas tout le site se focus sur le graphique en zoomant -> je dis de tout regardé le site	oui	Utilise toolbox
		b) La personne doit revenir en arrière sur la page « Sélection »	Clique sur le bouton retour	Clique sur le bouton retour	Appuis sur son moteur de recherche précédent	onglet « selection »
		c) En combien de temps ?	2 min	35 sec	20 sec	20 sec
		Réponse correcte ?	🔊	🔊	✓	✓

Comparaison	Difficile	26. À 30 ans, quelles sont (environ) les années d'expérience (Years of Experience) ?	5,3ans	5 ans	Environ 5 ans	5,6
		a) Est-ce que la personne a filtré pour regarder	Trie sur Years of Experience	Trie sur Years of Experience	Oui puis sélectionne et utilise la toolbox	Trie sur Years of experience et sélectionne
		b) Est-ce qu'elle regarde le graphique Median Age / Years of Experience	Oui, clique sur le graph et utilise la toolbox	Oui	oui	Oui
		c) En combien de temps ?	1m15	50 sec	40 sec	15 sec
		Réponse correcte ?	✓	✓	✓	✓
Comparaison	Difficile	27. Pour le graphique "Years of Experience / % High Job Satisfaction", qu'observez-vous ?	Rien de spécial ne se passe	Plus les années passent plus la satisfaction diminue	Rien de spécial ne se passe	Plus les années passent plus la satisfaction diminue
		a) Où va-t-elle regarder ?	Trie sur %High job Satisfaction et sélectionne, mais trouve que rien ne change, car la majorité des points sont au même endroit	Liste et trie sur Years of experience Puis sélectionne	Liste et trie sur % high job satisfaction et sélectionne	Trie % high job satisfaction
		b) Combien de temps de réflexion ?	4m	50 sec	40 sec	15 sec
		Réponse correcte ?	✗	✓	✓	✓
		<ul style="list-style-type: none"> • Deuxièmes impressions (si plus facile ou non que la première) • Facilités ? • Difficultés ? • Amélioration ? 	Plus facile, car elle connaît comment l'interface fonctionne et en soit ce n'est pas compliqué c'est juste la première mise en mains, mais qui est un peu difficile comme pour n'importe quel site internet	La phrase « je trie selon » est trop loin par rapport au dropdown et la dropdown pas assez visible Changer la couleur des onglets afin de voir que c'est des onglets et non une barre avec des titres	Bien lisible, mais ne travaille pas souvent sur l'ordinateur donc un peu de peine au début pour comprendre	Bonne plateforme compréhensible, mais ça serait bien qu'elle accepte d'autres types de fichiers (temps, coordonnés) pour que ça soit utile pour elle. Long le chargement de donnée

Total correct :	12	11	15	16
Total faux :	1	1	0	0
Total aidé :	3	4	1	0
Partie 4 - questions d'avenir				
28. Est-ce qu'une plateforme comme celle-ci selon vous ?	Oui	Ça dépend pour qui	oui	Oui
29. Si oui, pour quel type de personne ?	Des spécialistes en relation avec les données	Managers, RH	Entreprises multinationales, gestion, RH, statistique cantonale	Dans le domaine de recherche où beaucoup de données sont traitées en même temps
30. Si non, pourquoi pas ?				
31. Est-ce que vous seriez prêt à utiliser cette plateforme ?	Peut-être pour mon travail	non	Non, je n'en aurai pas l'utilité	Oui
32. Si oui, pour quel type de tâche ?	Pour des statistiques pour mon travail			Traitement de données (géographiques et océanographique)
33. Si non, quels seraient vos conseils d'amélioration ?	Pour d'autres types de données			

DÉCLARATION DE L'AUTEUR

Je déclare, par ce document, que j'ai effectué le travail de Bachelor ci-annexé seule, sans autre aide que celles dûment signalées dans les références, et que je n'ai utilisé que les sources expressément mentionnées. Je ne donnerai aucune copie de ce rapport à un tiers sans l'autorisation conjointe du Responsable de Filière et du Professeur chargés du suivi du travail de Bachelor, y compris au partenaire de recherche appliquée avec lequel j'ai collaboré, à l'exception des personnes qui m'ont fourni les principales informations nécessaires à la rédaction de ce travail et que je cite ci-après : Monsieur Florian Evéquoz.

Sierre, le 30 juillet 2018



Audrey MICHEL