

Rapport

Projet Data Visualisation E-Commerce

Master 1 Big Data et Fouilles de Données

Groupe:

Mohamed Amine Kaced

Idris Dada

Enseignante:

Dr. Asma BOUHAFS HAFSIA

1. Introduction	2
2.Description Dataset Schéma des Tables	2 3
3.Les outils utilisés	3
4.Préparation des données	4
Traduction des catégories de produits	4
Suppression des tables inutiles	4
Suppression des valeurs Null	4
Renommage de variable	4
5.Les KPIs mis au point	4
6.Graphiques Et analyse	7
7.Notre travail	22
8.Conclusion et perspectives	22
9.Bibliographie :	23

1. Introduction

Le E-commerce a pris une tournure extraordinaire durant la dernière décennie, il s'est démocratisé dans le monde entier, c'est pour cela que nous avons choisi une dataset E-Commerce d'un magasin au brésil afin de visualiser les données et tirer des remarques et des conclusion. On a utilisé Python pour les prétraitements et Tableau pour la visualisation des KPIs dans différents dashboards.

2.Description Dataset

On a choisi d'utiliser la dataset Brazilian E-Commerce Public Dataset by Olist qui est disponible sur Kaggle, après anonymisation cette dataset est maintenant public regroupant les ventes et les produits de différentes catégories, L'ensemble de données contient des informations sur 100000 commandes de 2016 à 2018 effectuées sur plusieurs marchés au Brésil. Ses fonctionnalités permettent de visualiser une commande à partir de plusieurs dimensions: du statut de la commande, du prix, du paiement et des performances de transport à l'emplacement du client, aux attributs du produit et enfin aux avis rédigés par les clients. Il s'agit de données commerciales réelles, elles ont été anonymisées et les références aux entreprises et partenaires.

Schéma des Tables

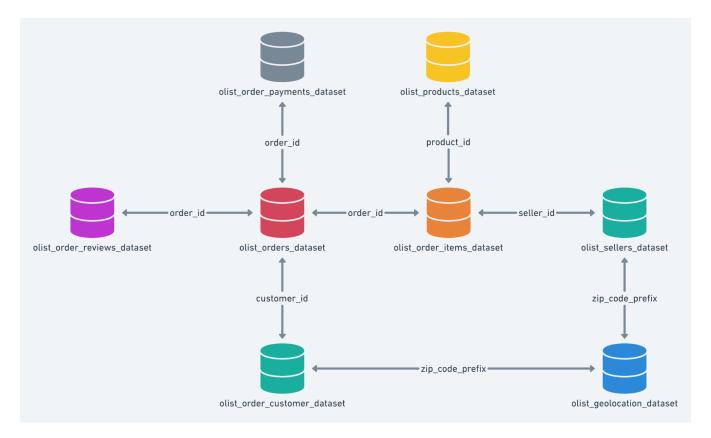


Fig 1: Tables

3.Les outils utilisés

Pour notre projet nous avons utilisé du python pour les prétraitements de la dataset (data cleaning, data preprocessing...) ensuite on a opté pour Tableau afin de réaliser la partie data visualisation (feuille de calculs, graphe et dashboards ..).





4. Préparation des données

Durant cette phase nous avons fait préparé la dataset avant de l'utiliser sur tableau, de ce fait on a réalisé les traitement suivants :

Traduction des catégories de produits

Nous avons traduit les catégories de produits qui étaient à la base en Espagnol pour les rendre en anglais.

Suppression des tables inutiles

Nous avons supprimé la table de géolocalisation, car nous avons trouvé que ça nous était pas nécessaire car tableau reconnaît l'emplacement de chaque ville et département.

Suppression des valeurs Null

Nous avons opté pour supprimer les lignes contenant des valeurs aberrantes ou Null.

Renommage de variable

dans cette étape nous avons renommé différentes colonnes afin d'avoir des appellation plus parlante

5.Les KPIs mis au point

• Le Nombre de commandes

représente le nombre totale des commandes sur le site

• Nombre d'Articles Commandés:

Représente la somme des articles vendu dans toutes les commandes

• Indice de Vente

L'Indice de Vente (IDV) est un indicateur qui mesure le nombre moyen d'articles achetés lors d'une transaction.

Indice De Vente = Nombre d'articles vendus / Nombre de Commandes

• Nombre de produits dans le catalogue

Représente la somme de produit unique disponible dans le catalogue de la boutique.

Panier moyen

Le Panier Moyen est le montant moyen d'une commande sur votre site, sur une période donnée.

<u>Panier Moyen (PM) = Chiffre d'Affaires / Nombre de commandes</u>

Délai de livraison moyen

Représente la durée moyenne pour la finalisation de livraison.

• Chiffre d'affaires

Il représente la somme de toutes les ventes de la boutique.

• Nombre de clients

la somme des clients de toutes les catégories de produits.

• Nombre de transactions

Il est important de mesurer le nombre de transactions, et pas uniquement le chiffre d'affaires

Nombre d'évaluations de produits

représente le nombre d'évaluations laissées par les clients après leurs achats.

Nombre Commandes

Temps moyen de livraison

Nombre d'Articles Commandés

96 461

12,56 DAYS

134 936

Indice de Vente

1,399

Nombre de produits

32 340

Panier moyen

165,96\$

Chiffre d'affaire

16 008 872 \$

Nombre de clients

99 441

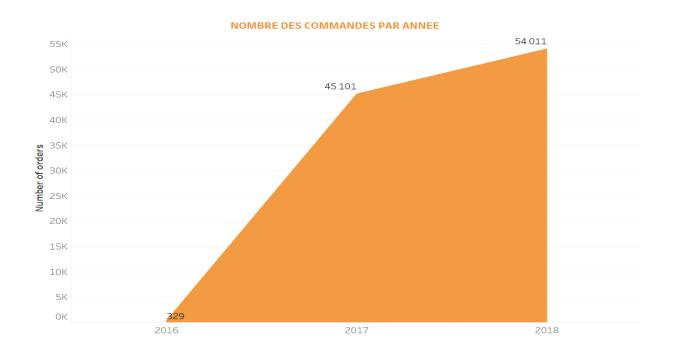
Nombre d'évaluations

9986

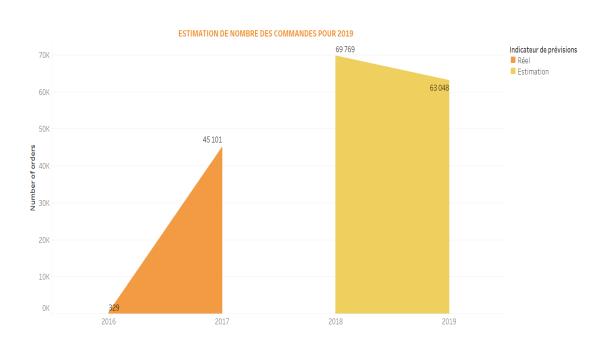
fig 2: KPIs

6.Graphiques Et analyse

1-Commandes:



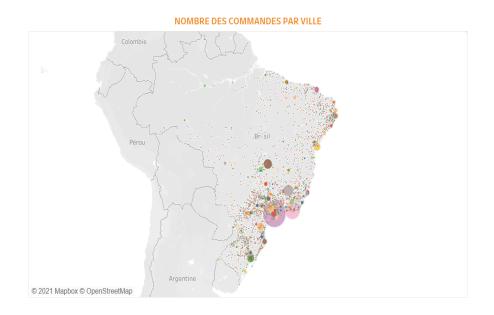
 On voit que le nombre des commandes augmente d'une manière rapide et exponentielle au cours des 3 années.



 D'après le graphe, on peut déduire que le nombre des commandes baissera dans l'année 2019 en se basant sur les années précédentes 2016,2017 et 2018.



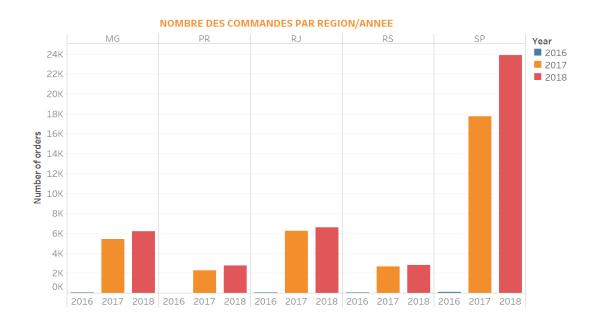
 On voit que le nombre des commandes est très élevé dans les départements qui se situent dans la partie l'est à côté de l'océan, les top 3 départements sont São Paulo,Rio de Janeiro et Minas Gerais respectivement et cela est due à la densité de population élevée dans ces régions.



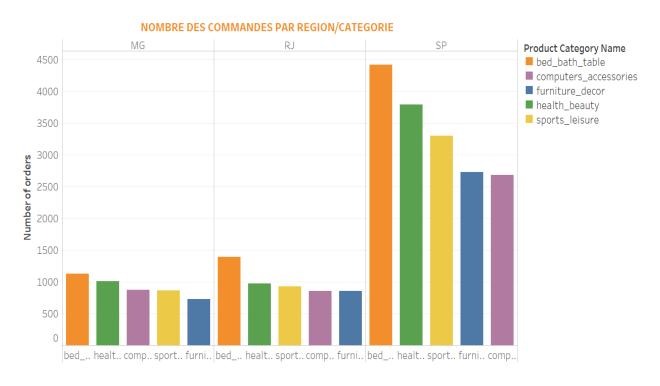
 On remarque que le nombre des commandes est élevé dans les villes qui sont à côté de la côte est du Brésil, notamment à São Paulo(département de São Paulo) ,Rio de Janeiro (département de Rio de Janeiro).



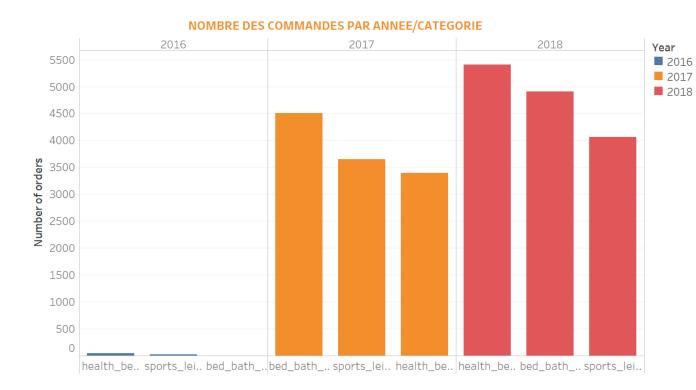
 Les 5 meilleures villes dans les commandes sont : São Paulo,Rio De Janeiro,Belo Horizonte,Brasília,Curitiba respectivement.



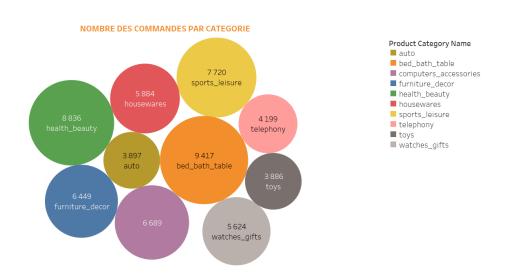
 On voit que le nombre des commandes augmente rapidement au cours des années dans tous les départements brésiliens (on a pris que les top 5).



 On remarque que le bed_bath_table et healthy beauty sont les catégories des produits les plus demandées dans le magasin Olist alors que le reste sont moins demandés dans les 3 départements São Paulo, Rio De Janeiro et Minas Gerais (top 5 categories, top 3 cities)



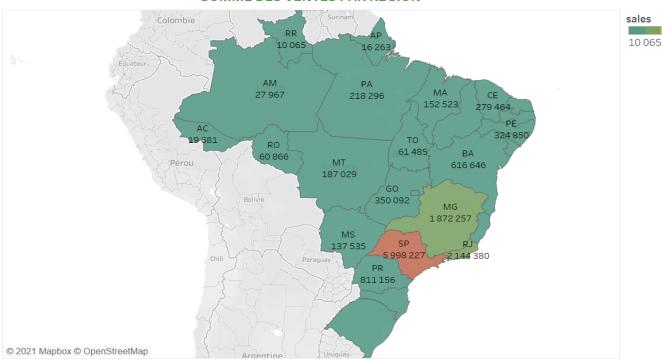
 On remarque que les produits les plus demandés en 2017 appartiennent à la catégorie bed bath table,tandis qu' en 2018 les produits les plus demandés appartiennent à la catégorie health beauty. (top 3 categories)



- Les produits les plus demandés dans le magasin Olist appartiennent aux catégories:
 - 1-bed_bath_table (environ 9500 commandes)
 - 2- health_beauty (environ 900 commandes)
 - 3-sport_leisure (environ 8000 commandes)

2-Ventes:

SOMME DES VENTES PAR REGION



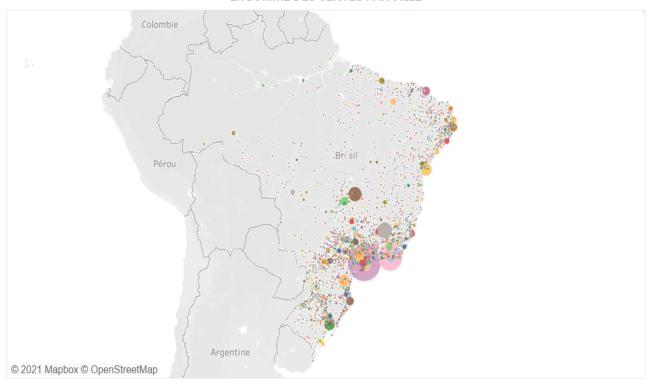
Carte basée sur les Longitude (générée) et Latitude (générée). La couleur met en avant le/la somme de sales. Les repères sont étiquetés par Customer State et somme de sales.

- On remarque que les ventes du magasin Olist est très élevé dans les 3 départements :
 - 1-São Paulo (environs 6 million dollars)
 - 2-Rio de Janeiro (environs 2 million dollars)

6M

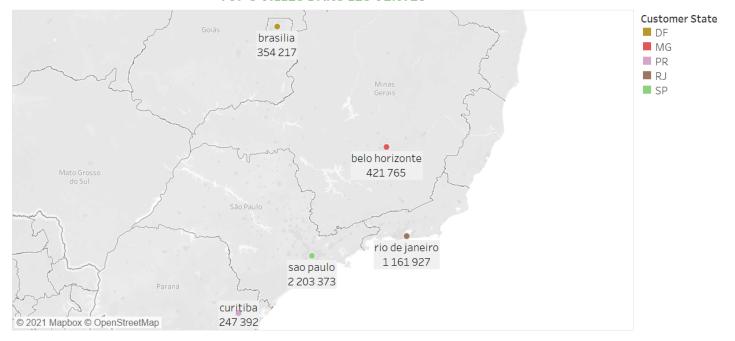
3-Minas Gerais (environs 1 million 900 dollars)

LA SOMME DES VENTES PAR VILLE



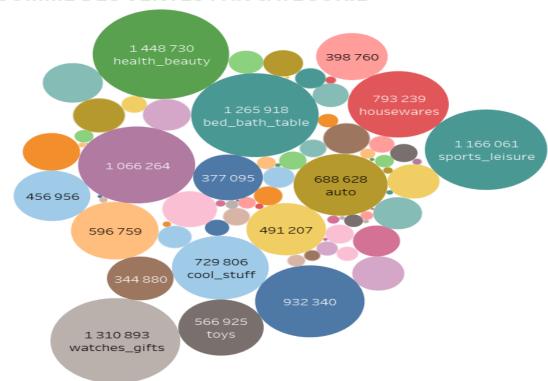
 On remarque que les ventes ont éclaté dans les villes qui ont une grande densité populaire: São Paulo et Rio de Janeiro.



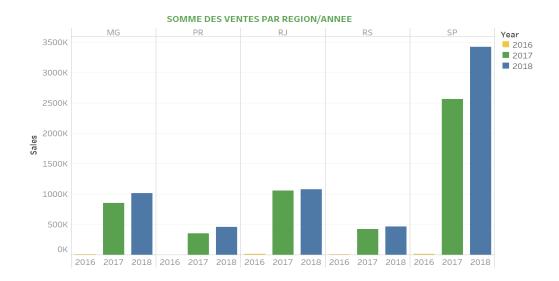


 Les 5 meilleures villes dans les ventes sont : São Paulo,Rio De Janeiro,Belo Horizonte,Brasília,Curitiba respectivement.

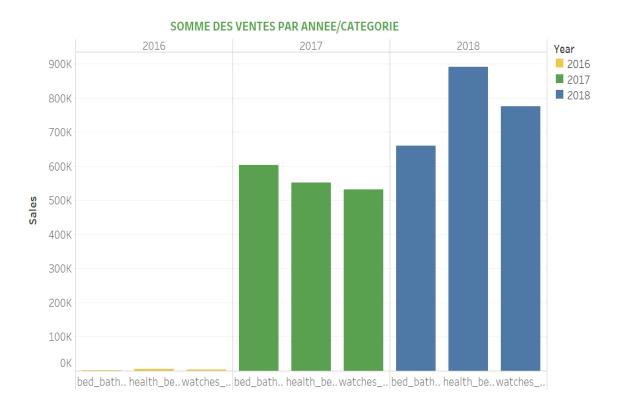
SOMME DES VENTES PAR CATEGORIE



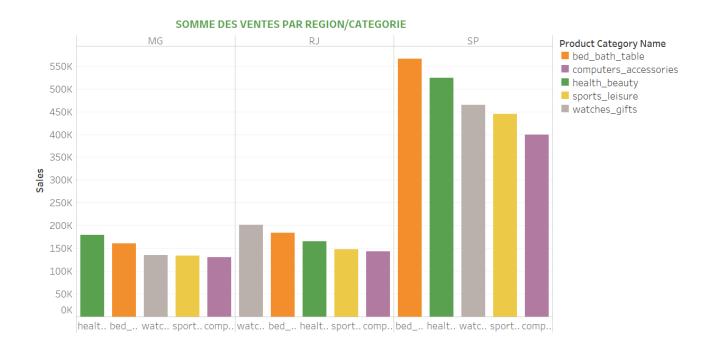
- Le magasin Olist gagne plus d'argent lorsqu' il vend les produits qui appartiennent aux catégories suivants par ordre :
 - 1-health_beauty (environ 1 million 500 \$)
 - 2- watches_gifts (environ 1 million 300 \$)
 - 3-bed_bath_table (environ 1 million 300 \$)
 - 4-sports_leisure (environ 1 million 200 \$)



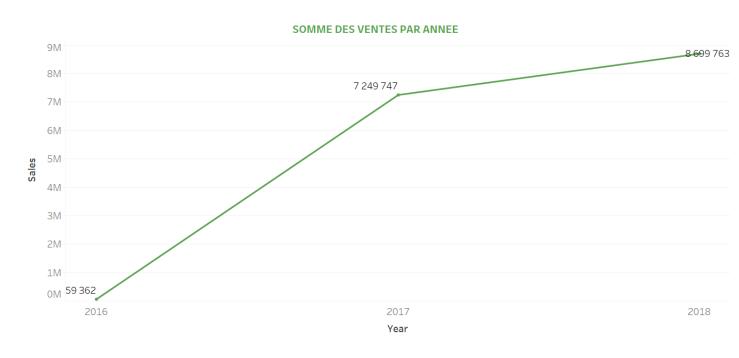
 On voit que les ventes ont explosé au cours des années 2018 et 2019 dans tous les départements brésiliens notamment à São Paulo (on a pris que les top 5).



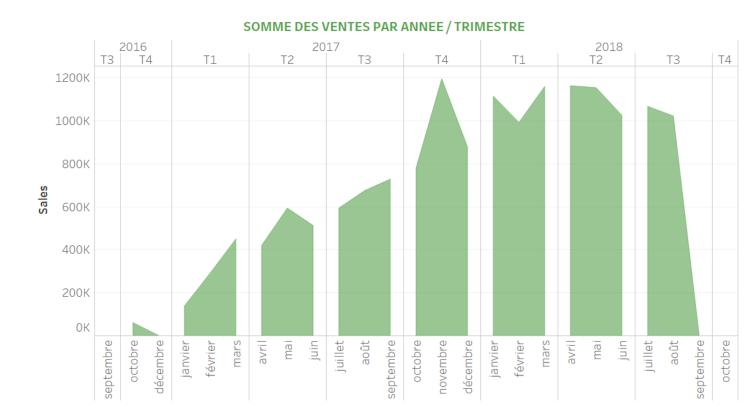
 On remarque qu'en 2017 le magasin Olist a gagné plus d'argent à partir de la vente des produits qui appartiennent à la catégorie bed_bath_table,alors qu' en 2018 ce magasin a gagné plus d'argent à partir de la vente des produits qui appartiennent à la catégorie health_beauty (top 3 catégories).



 On remarque que le bed_bath_table est la catégorie des produits qui fournit plus de gains aux magasin Olist dans le département São Paulo, alors que watches_ gifts représente la catégorie qui fournit plus de bénéfices dans Rio De Janeiro et la catégorie healthy_beaty dans le département Minas Gerais (top 5 categories, top 3 cities).

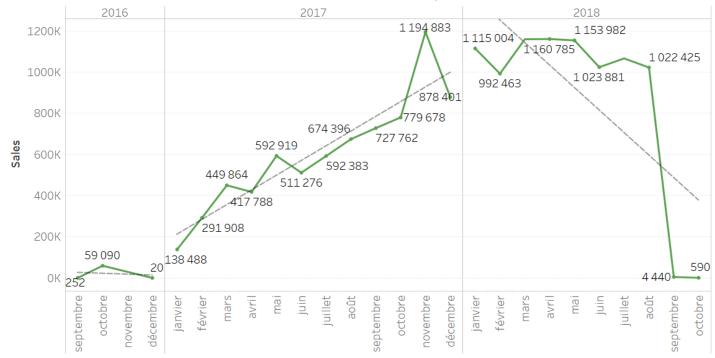


Nous remarquons une explosion des ventes du magasin Olist du 59 000 \$ en
 2016 vers environ les 7 et 9 millions \$ en 2017 et 2018 respectivement.



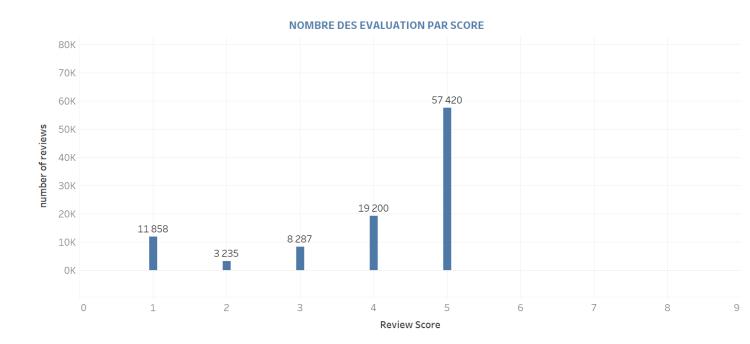
 Nous voyons que les gains de Olist ont augmenté dans le dernier semestre de l'année 2017 et aussi dans les 3 premiers semestres de l'année 2018.Donc,on peut dire que la majorité des clients preferente d'acheter des produits à la fin d'année, pourquoi pas les faire des promotions dans cette période.



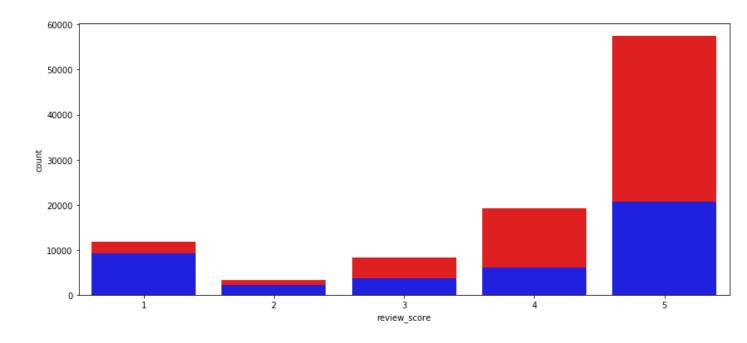


- On voit que : en 2017 le magasin Olist a gagné une somme de presque 2 million \$
 dans le mois de novembre qui est la somme la plus élevée dans toute l'année, elle
 était environ 140 000 \$ en début d'année mois de janvier, d'après la courbe de
 tendance on peut remarquer qu'il ya une grande augmentation des ventes dans
 cette année.
- En 2018 la somme de ventes de ce magasin a atteint le pic en début d'anné le mois de janvier avec presque 1 million \$ et elle a commencé à dégrader jusqu' a les 600 \$ en mois d'octobre

3-Reviews:

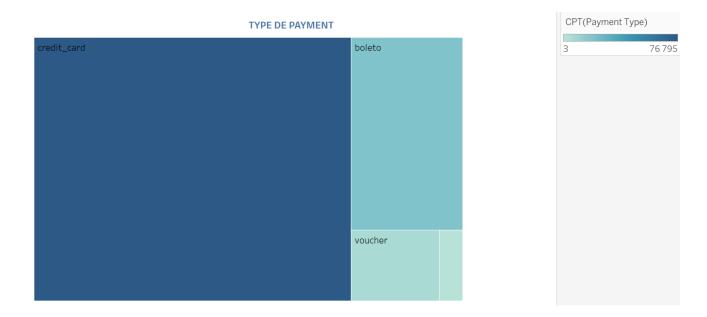


 On voit que la plupart des clients ont noté les produits par 5 étoiles, cela veut dire que presque tous les clients du magasin Olist sont satisfait de la qualité de ces produits.

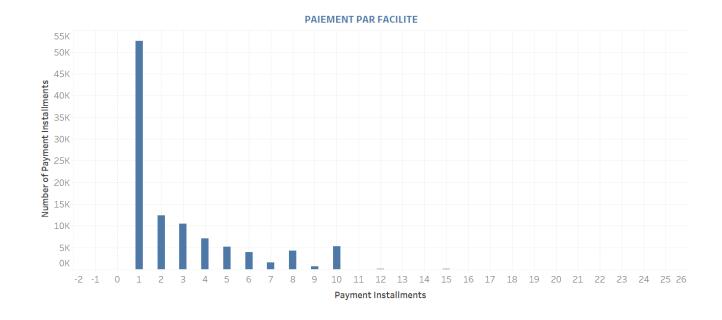


 Nous pouvons remarquer que pour les clients qui ont donné 5 étoiles comme score juste une petite partie qui ont mis des commentaires tandis que les clients qui ont évalué les produits par 1 étoile la majorité ont commenté sur les produits.Donc,un client est plus susceptible de donner des avis lorsqu'il est mécontent.

4-Paiement:



 On peut déduire que la majorité des clients préfèrent payer par carte bancaire par rapport au paiement par billet ou le bon.



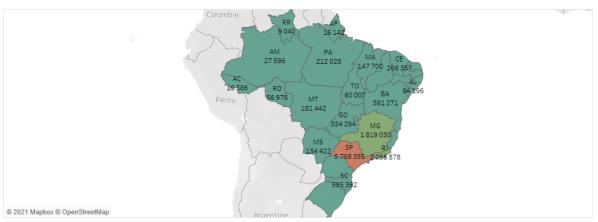
 On peut voir que la plupart des clients de ce magasin préfèrent payer en une seule tranche au lieu de payer en plusieurs fois.

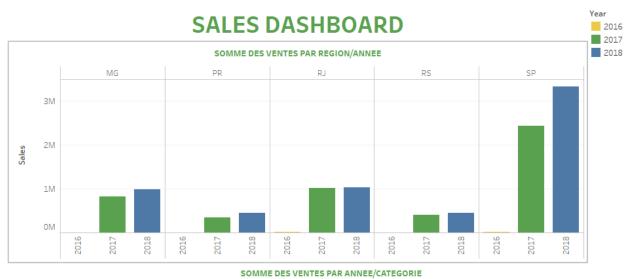
7.Dashboards

SALES DASHBOARD



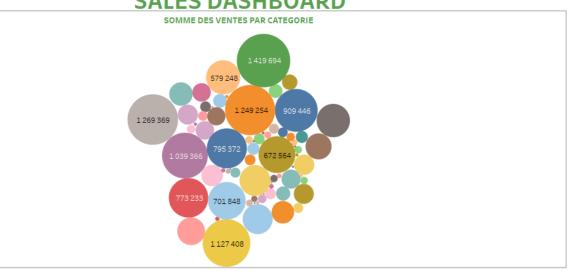
SOMME DES VENTES PAR REGION



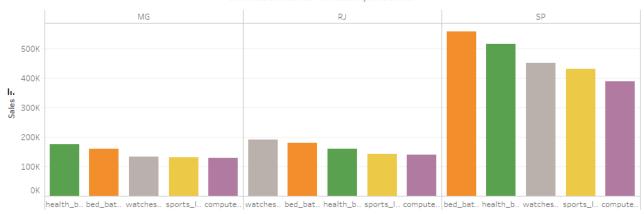




SALES DASHBOARD



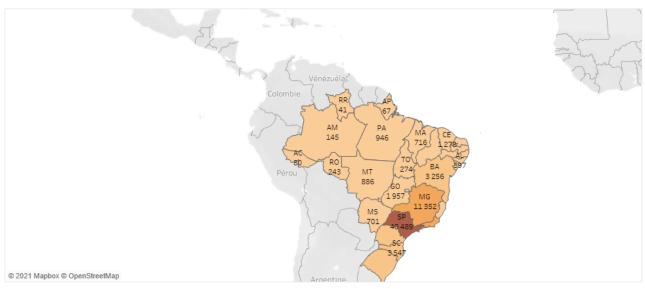
SOMME DES VENTES PAR REGION/CATEGORIE

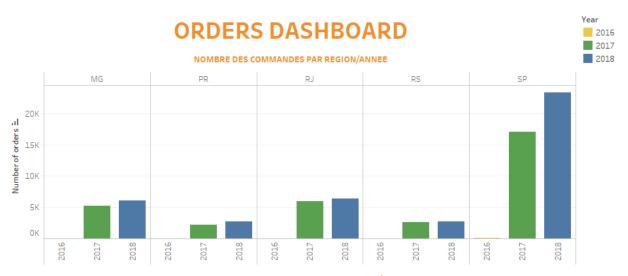


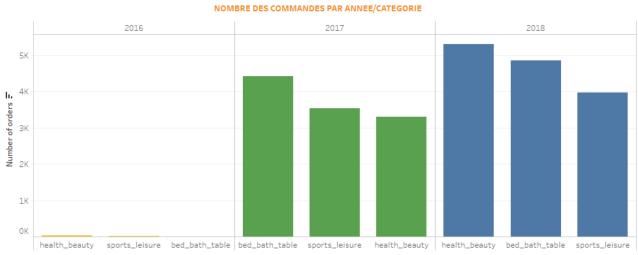
ORDERS DASHBOARD

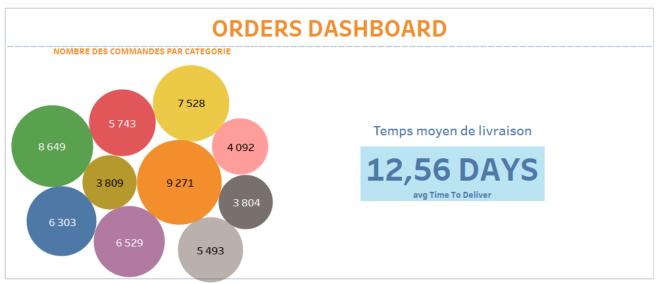


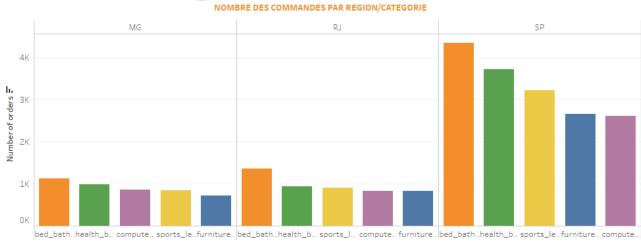
NOMBRE DES COMMANDES PAR REGION











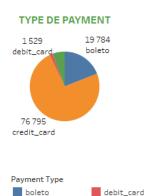
Nombre de clients

99 441

Panier moyen

165,96\$



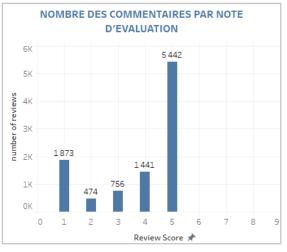


voucher

credit_card

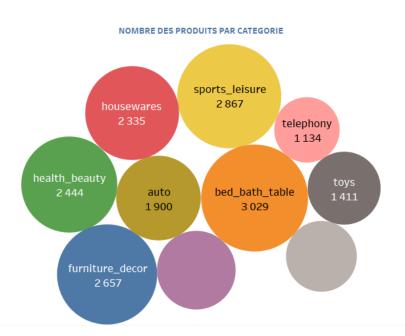
Nombre d'évaluations

9 986



Product Dashboard

Nombre de produits
32 340



8. Notre travail

Nous avons déposé tous notre travail sur un repository Github, dont voici le lien :

https://github.com/idrimadrid/Data-Visualisation-Ecommerce

8. Conclusion et perspectives

Ce projet nous a permis de découvrir et essayer plusieurs méthodes de visualisation et traitement de données, on aurait tout de même voulu explorer d'autres KPI, faute de données qui ne sont pas très enrichies, on aurait cependant aimer travailler sur une autre base de donne plus complet que celle ci avec des discounts, refunds et des returns, afin de relever d'autre tendances.

9.Bibliographie:

- <u>5 Ways to Measure Customer Lifetime Value in Ecommerce (marsello.com)</u>
- 20 KPI e-commerce pour votre tableau de bord Web Analytics Digital Analytics
 Blog (atinternet.com)
- Gross Sales Definition (investopedia.com)
- Top 5 e-commerce KPIs for your Tableau dashboard (alphaservesp.com)
- LTV/CAC Ratio Important eCommerce and SaaS Metrics (corporatefinanceinstitute.com)