Mémoire de Projet de Fin d’année

Pour l’obtention du titre

D’ingénieur d’état en informatique

Option

Systèmes d’information et Big Data

Sujet

…….

…….

Soutenu par : Sous la direction de :

**M. Adnane DRIOUCHE** **M. Youness SKANDRANI (CHRONOPOST)**

***M. Driss ELAMRANI (CHRONOPOST)***

***M. Khalid FAKHAR (ENSAB)***

# Dédicace

Je dédie ce travail,

A mes chers parents

Je ne trouve pas les mots pour rendre hommage à votre

vertu, j’espère que vous trouverez dans ce travail l’expression de ma reconnaissance pour vous.

A ma soeur Hiba

A mes frères Radouane, Hamza et Adam

A mes co-chambres BOUZ, BENHMEMMID, FANGACHI et SELMANI pour leur soutien et leurs conseils prodigieux.

A toute la promo ISI&BD 2018/2021

pour les beaux moments que nous avons vécu ensemble.

Adnane

# Remerciements

Il m’est agréable de m’acquitter d’une dette de reconnaissance auprès de toutes les personnes, dont l’intervention au cours de ce projet a favorisé son aboutissement.

Je tiens à remercier mon encadrant à l’ENSIAS, M. xxxx pour tout le temps qu’il nous a consacré, pour ses conseils pertinents, ses directives précieuses, pour la qualité de son suivi et pour ses réflexions éclairées qui nous ont été salutaires durant toute la période du projet.

Ma reconnaissance s’adresse également à M. xxxx et M. xxxx, Mlle xxxx et M.xxxx, pour leur accueil, leur confiance et leur esprit d’équipe et de partage qui ont rendu ce projet intéressant. Je les remercie également pour leur patience, leur amabilité, les conseils et les formations qu’ils nous ont prodigués durant ce projet.

Que les membres du jury trouvent ici l’expression de ma reconnaissance pour avoir accepté d’évaluer notre travail.

Que tous ceux qui ont contribué de près ou de loin dans notre formation et/ou dans l’accomplissement de ce travail en particulier le corps professoral de ma chère école ENSAB, trouvent ici l’expression de mes remerciements les plus chaleureux.

# Résumé

Le présent document constitue le fruit de mon travail accompli au sein de

CHRONOPOST afin d'analyser et segmenter ses clients en se basant sur

"unsupervised learning algorithm" et précisement K-means Clustering.

Mon projet consiste à créer une application web qui regroupe la totalité d'analyse

après la segmentation des clients en "clusters". L'application est réalisée 100% en

langage python.

# Abstract

# Table des matières

# Introduction générale

# Chapitre 1 : Contexte général du projet

Ce chapitre est un chapitre introductif qui a pour mission la présentation du cadre général du projet en exposant Chronopost où se déroule mon projet de fin d’année, l’état des lieux du projet, la problématique et les objectifs.

En effet, je vais en premier lieu présenter l’organisme l’accueil, ensuite je vais mettre l’accent sur le cadre général du projet, la problématique traitée ainsi que les objectifs visés. Vers la fin de ce chapitre je vais exposer le planning du déroulement de notre stage.

## Présentation des organismes d’accueil :

## Cadre général du projet :

### Présentation du projet :

Le secteur de la livraison connaît des mutations depuis quelques années, notamment sa libération, la forte croissance du marché, la rude concurrence et les nouvelles offres qui se multiplient. Dans ce contexte hostile, l’entreprise doivent être très réactifs face à leurs clients afin d’augmenter le taux de leur satisfaction.

Pour lutter contre cette concurrence rude, la gestion des campagnes marketing doit être beaucoup plus proactive. Au fil de la campagne, les résultats doivent être analysés en se basant sur Machine Learning en tant qu’une forte tendance, afin d'obtenir un meilleur taux de retour et une meilleure rentabilité.

Dans un contexte de forte concurrence, de sur sollicitation, de volatilité des clients et de recherche de leviers de rentabilité, il est impératif de bien identifier les clients à cibler dans une campagne marketing et de proposer les offres adéquates pour fidéliser, acquérir et inciter les abonnés à augmenter leurs consommations.

C’est dans ce contexte que s’inscrit mon projet de fin d’année qui consiste à segmenter les clients à base Machine Learning afin de générer le meilleur retour sur investissement pour chaque client.

### Choix du nom et du Logo :

J’ai pris **DRALYSE** App comme nom de mon application. Le nom est mergé à partir de mon nom ( **DR**IOUCHE ) et ( AN**ALYSE** ).



**Figure 1: Logo de l'application**

### Objectifs du projet :

L’entreprise souhaite améliorer son processus d’analyse des clients par l’acquisition et la mise en place d’une nouvelle solution.

A travers cette nouvelle solution, elle voudra atteindre les objectifs suivants :

* Réalisation d’une application web qui couvre la totalité d’analyse sur les clients et les produits.
* Segmentation des clients à base d’algorithme K-means.
* Application du k-means clustering après la segmentation des clients par ville et par type.

Conclusion

Le premier chapitre positionne le projet dans son contexte général et décrit sa problématique, ses principaux objectifs et le planning de sa concrétisation pour avoir une vision claire sur le projet en général. Dans le chapitre suivant, nous allons exposer maintenant l’étude de l’existant et l’analyse des différents besoins que l’entreprise souhaite implémenter.

# Chapitre 2: Analyse, spécification des besoins et conception

Ce chapitre constitue le cœur du projet. IL présente l’analyse des besoins à satisfaire par la future solution d’analyse. Pour ce faire l’étude de l’existant fera l’objet de la première section de ce chapitre, l’expression des besoins sera abordée dans la deuxième section, les axes d’analyse, les mesures et les spécifications techniques seront présentés dans la troisième section.

## Analyse de l’existant :

La connaissance de l’existant est une phase indispensable qui sert de base pour aider l’entreprise à déterminer ses besoins afin d’améliorer son système tout en gardant les fonctionnalités principales existantes.

Actuellement, l’analyse s’est réalisée d’une façon classique en se basant sur une visualisation classique du data.

D’une part, l’analyse ne sera pas performant si nous la comparons avec l’analyse qui sera basée sur des algorithme du Machine Learning à savoir k-means.

D’autre part, l’analyse de la base des clients sur un contexte général va perturber la prise de la bonne décision envers ce client. C’est pour cela qu’il faut s’appuyé sur la bonne méthode ou plutôt le meilleur algorithme pendant la segmentation.

## Axes d’analyse et indicateurs :

### Axes d’analyse :

En fonction de l’analyse des besoins et en étudiant les sources de données qui font la base des clients, leurs comportements et les données pertinentes requises pour le ciblage des clients.

|  |  |
| --- | --- |
| Indicateurs | Définitions |
| Customer | Il contient des informations décrivant un client *: id, nom du client, city, type du client (A, B, C), adresse e-mail et numéro.* |
| Shipment | Il contient les informations décrivant une expédition : *id, le prix, volume, le poids et la date d’expédition. Ainsi le pays expéditeur et destinataire.* |
| Product | Il contient : *id du produit et son nom.* |
| Country | Il contient **:** *id du pays, son nom et son code alpha* |

**Tableau 1: Dimensions et leurs définitions**

Dimensions et leurs définitions

### Indicateurs:

Les mesures ou les indicateurs recensés pendant la phase d’étude et d’analyse des besoins peuvent être classés selon : expédition, produit et pays.

* **Expédition**

|  |  |
| --- | --- |
| Indicateurs | Descriptions |
| Nombre d’expédition | Nombre d’expédition faites par client |
| Poids d’expédition | Le poids expédié par client |
| Volume d’expédition | Le volume expédié par client |
| Prix d’expédition | Le prix total payé |
| Type d’expédition | ---- |

**Tableau 2: Indicateurs des expéditions et leurs descriptions**

* **Produit**

|  |  |
| --- | --- |
| Indicateurs | Descriptions |
| Vente par produit | Nombre de vente du produit X |
| Poids par produit | Le poids envoyé utilisant un produit X |
| Volume par produit | Le volume envoyé utilisant un produit X |
| Expédition par produit | Nombre d’expédition utilisant un produit X |

**Tableau 3: Indicateurs des produits et leurs descriptions**

* **Pays**

|  |  |
| --- | --- |
| Indicateurs | Descriptions |
| Nombre d’expédition envoyé | Nombre d’expédition envoyée d’une pays X |
| Nombre de destination reçu | Nombre d’expédition reçu d’une pays Y |

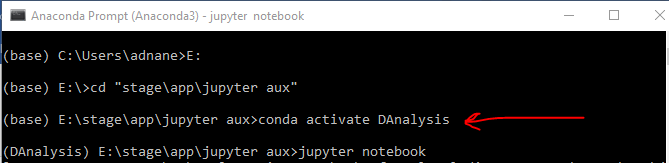
**Tableau 4: Indicateurs des pays et leurs descriptions**

## Spécifications techniques :

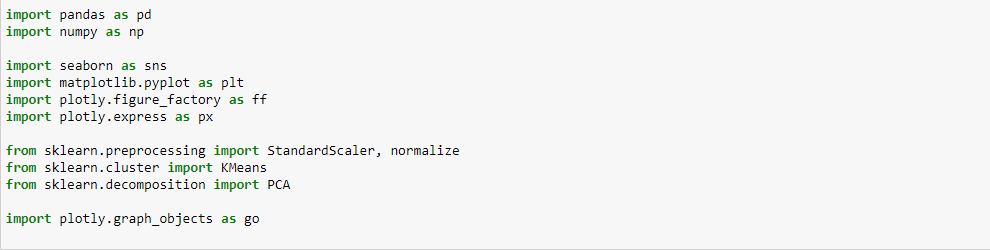
Une fois les spécifications fonctionnelles sont déterminées, je passe aux spécifications techniques en présentant l’environnement de développement, data pre-processing, k-means clustering, pca et la visualisation.

### Environnement de développement :

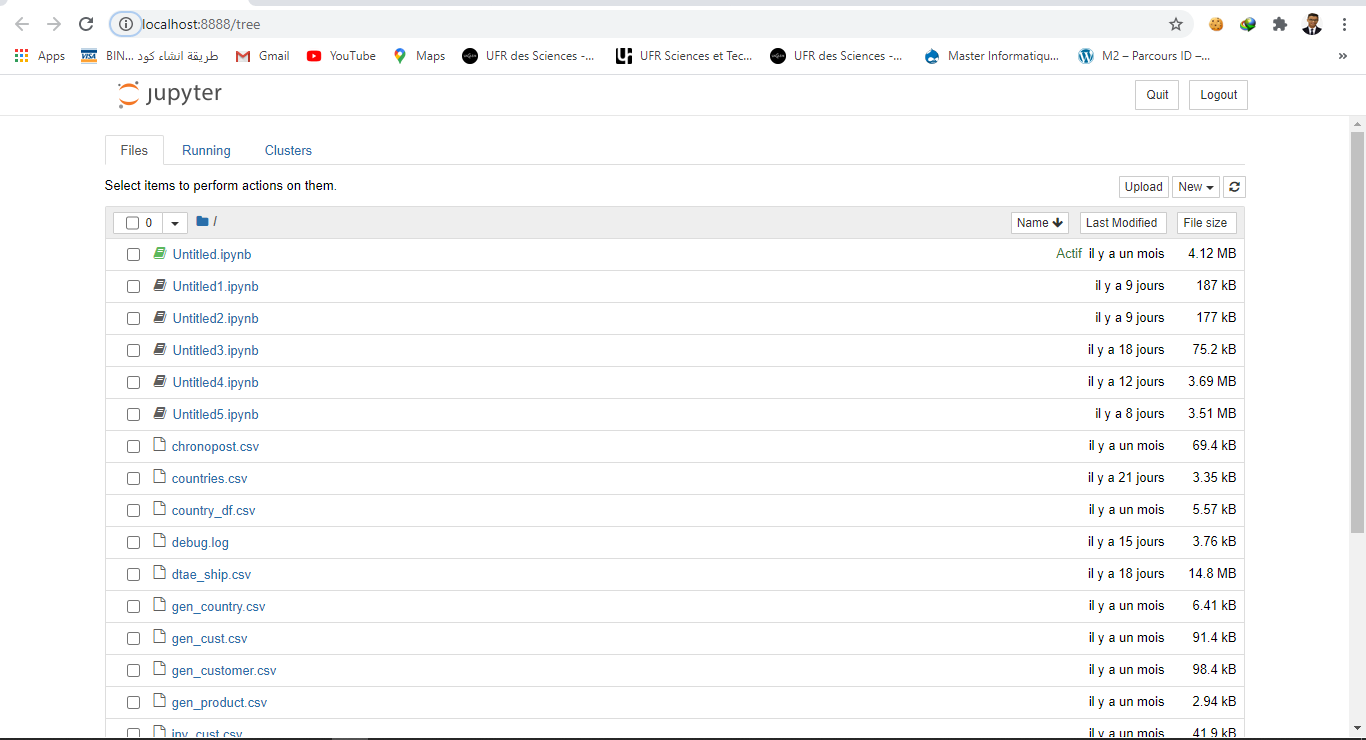
* **Anaconda** : J’ai créé un environnement nommé « *DAnalysis* » sur Anaconda comme étant une distribution libre et open source des langages de programmation Python et R appliqué au développement d'applications dédiées à la science des données et à l'apprentissage automatique.
* **DAnalysis** : cet environnement contient la totalité des bibliothèques que j’en ai eu besoin.



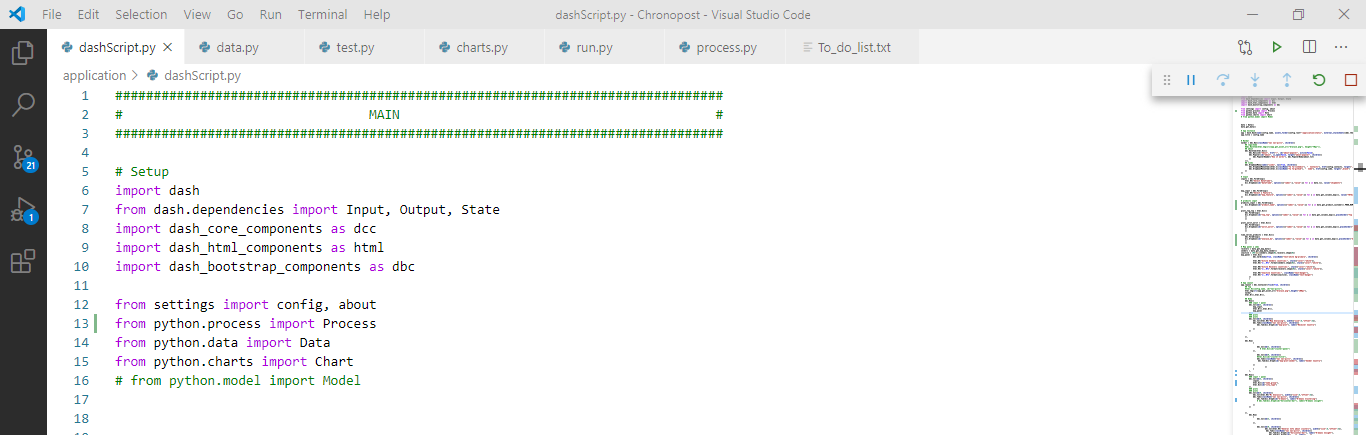
* **Les bibliothèques** : pandas, numpy, matplotlib, seaborn, plotly.express et sklearn.



* **Jupyter Notebook** : Je l’ai utilisé pendant les tests d’analyse.



* **Dash**: Je me suis basé principalement sur la Framework Dash pendant la réalisation de l’application.



### Data pre-processing:

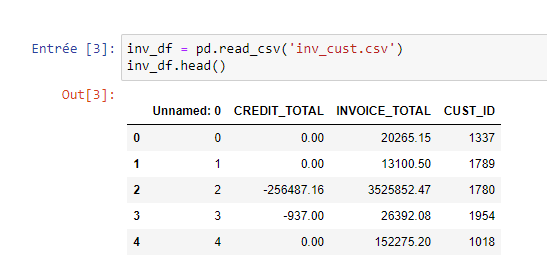
Dans cette partie, j’ai suivi les étapes suivantes :

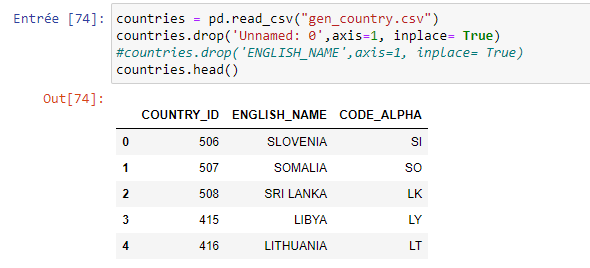
* **Acquire the dataset:**

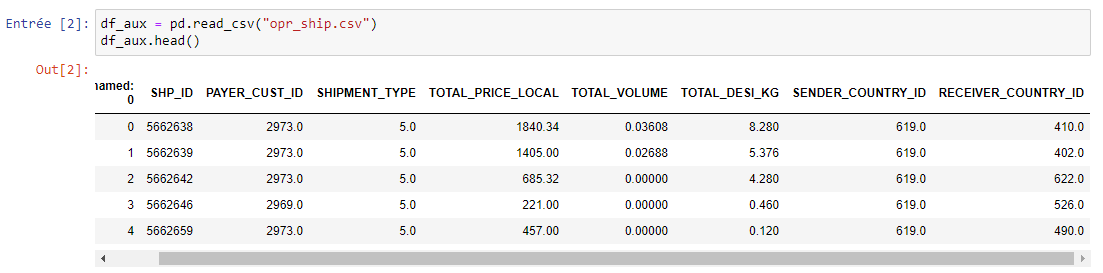
Pour créer et développer des modèles de Machine Learning, J’acquiers d'abord dataset pertinent. Pour ça je me suis connecté à la des données, et transférer les tables au dataframes pour simplifier data pre-processing.



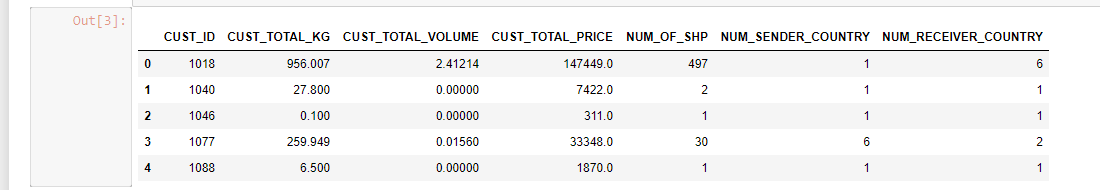
* **Import all the crucial libraries:**
* **Import the dataset :**

****

****

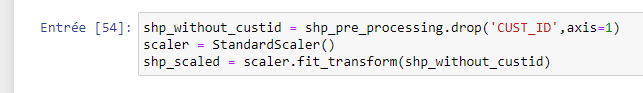
****

* **Identifying and handling the missing values:**

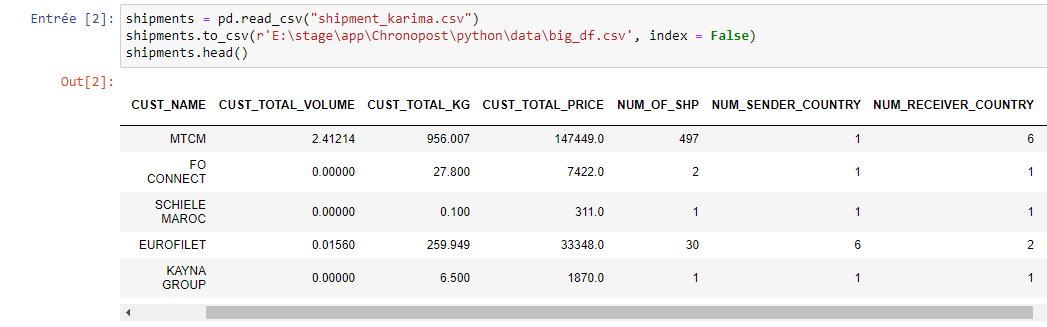
****

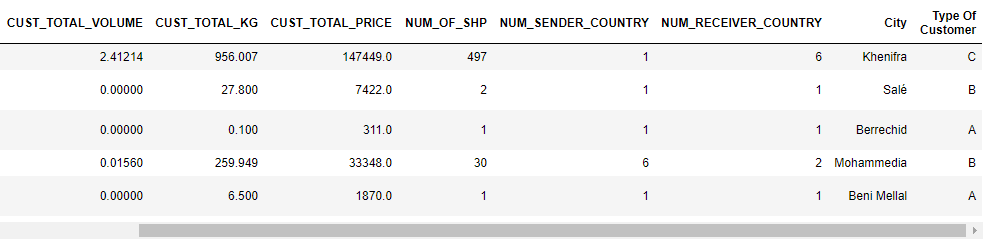
****

* **Feature scaling:**

****

* **Final dataset:**

****

****

### K-means Clustering :

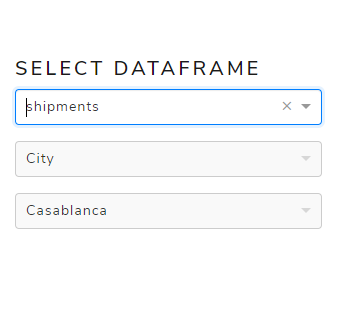
**Customer segmentation** consiste à partitionner une clientèle en groupes d'individus ayant des caractéristiques similaires. C’est une stratégie importante car une entreprise peut cibler ces groupes spécifiques de clients et allouer efficacement des ressources marketing.

Par exemple, un groupe peut contenir des clients à haut profit et à faible risque, c'est-à-dire plus susceptibles d'acheter des produits ou de s'abonner à un service ; un autre groupe pourrait inclure des clients d'organisations à but non lucratif. Une tâche commerciale consiste à fidéliser ces clients.

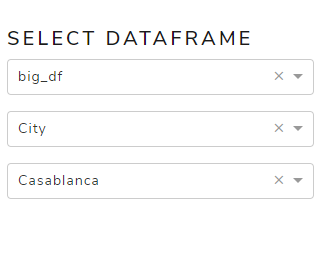
Dans notre cas, mon application consiste à segmenter les clients de deux façons :

**1. Segmentation générale :** consiste à segmenter les clients de façon général,

Puisque je n’ai pas pris en considération le type de client, son pays ou sa ville. Pour segmenter de manière général, l’utilisateur doit choisir *« Shipment dataframe* ».



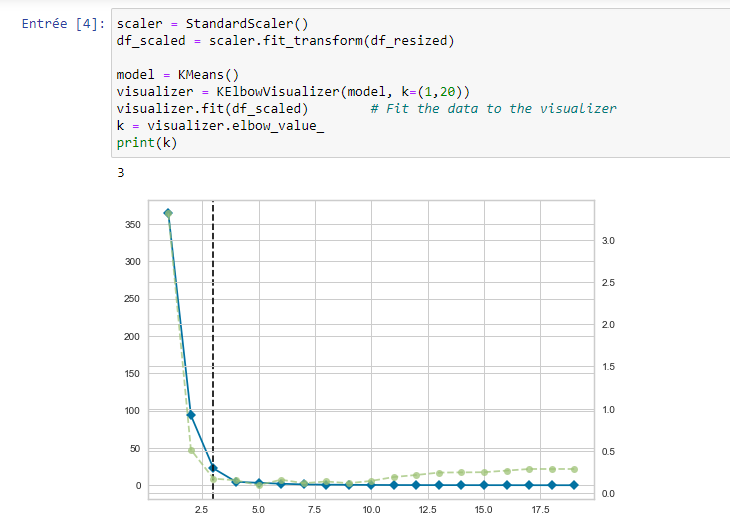
**2. Segmentation après groupement classique :** Dans ce cas, j’ai appliquée k-means après le groupement des clients suivant leurs types ; Il y trois types de client ; ou la ville (30 villes) où ils résident. Pour segmenter de manière général, l’utilisateur doit choisir la 2eme *dataframe.*

****

* **Le choix du K optimal :**

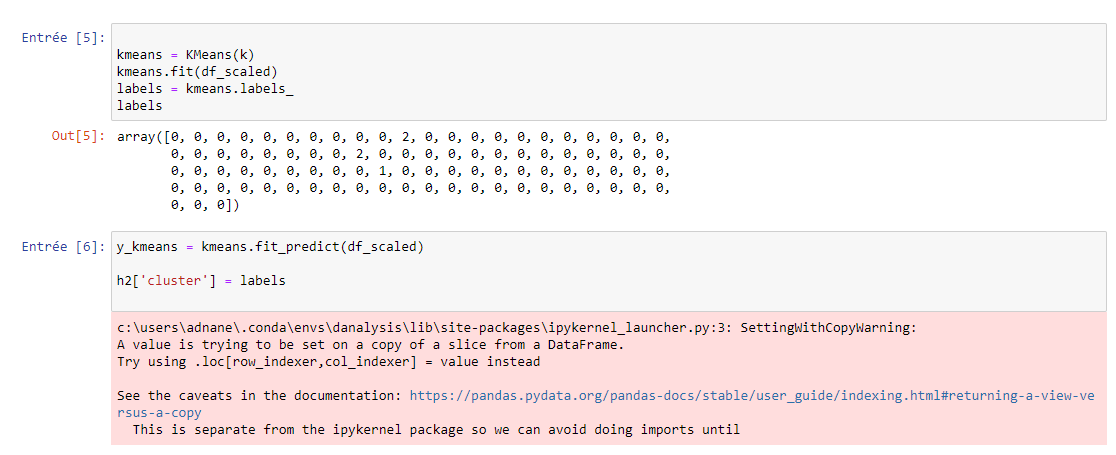
Après data pre-processing, je dois choisir le k optimal (nombre optimal des cluster) pour effectuer le k-means. Pour en faire, je me suis basé sur **Elbow method.**

La méthode consiste àCalculez **Within Cluster Sum of Squared Errors** (WSS) pour différentes valeurs de k, et choisissez le k pour lequel WSS commence à diminuer. Dans le graphique de WSS-versus k, cela est visible comme un coude.



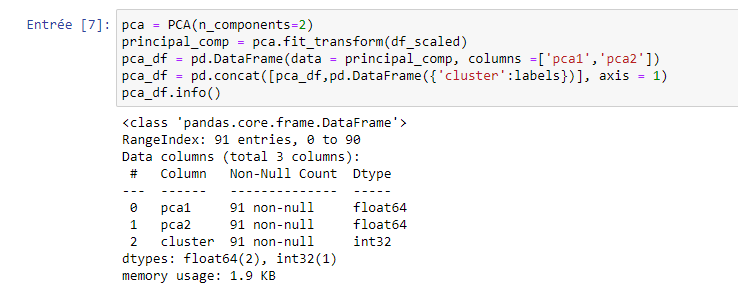
* **Modeling :**

Dans cette étape, je vais appliquer k-means après que j’ai eu mon K optimal.



### PCA (Principal component analysis) :

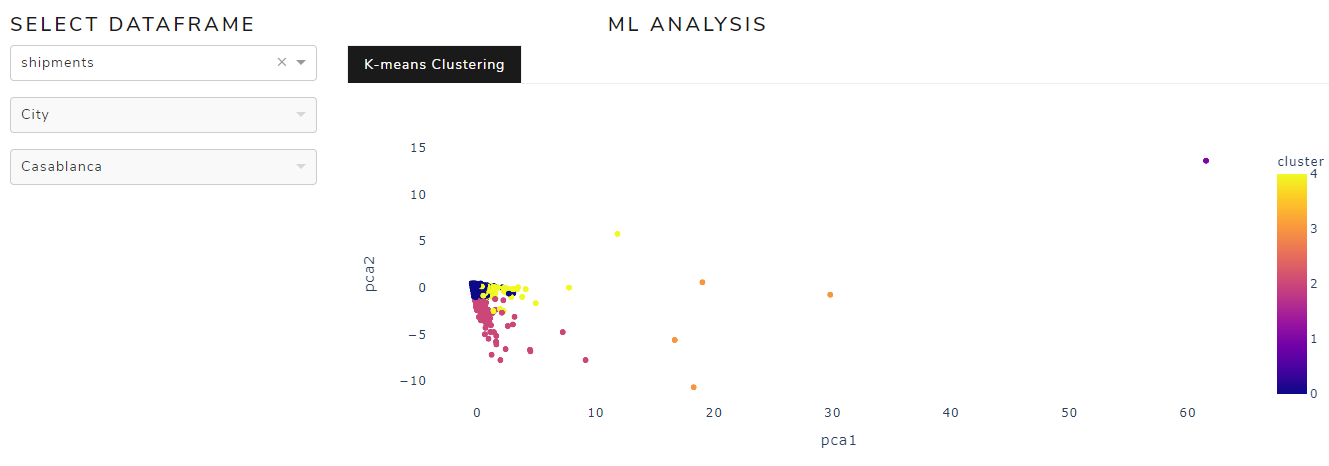
Pour visualiser mon data, j’étais obliger de réduire les dimensions et de choisir le meilleur plan sur lequel je vais projeter mon data. Dans mon cas, PCA est la bonne méthode.



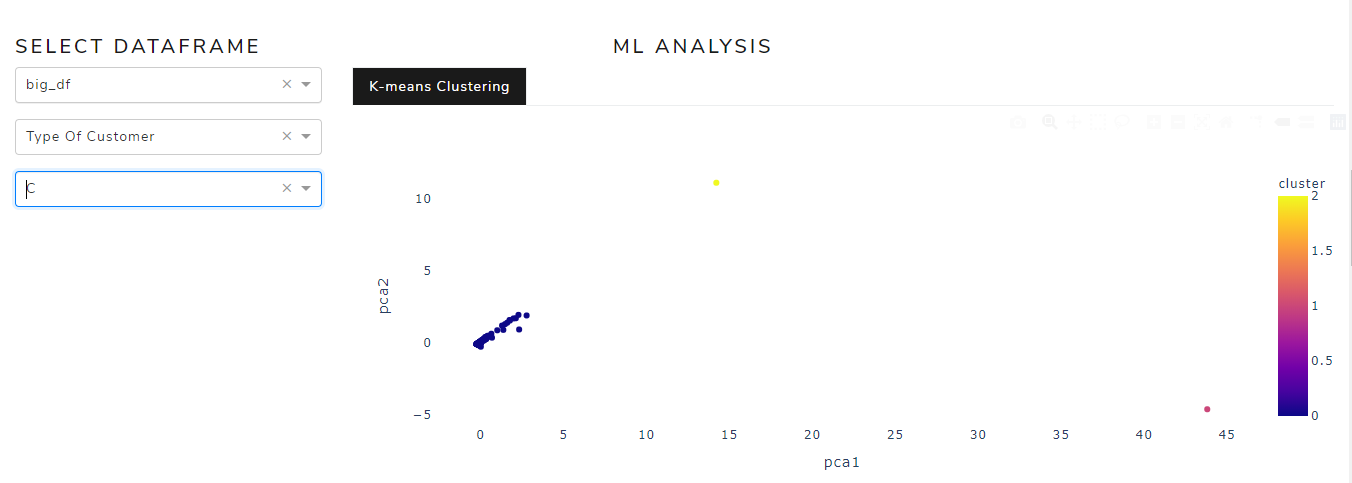
### Visualisation :

L’application peut faire 34 segmentations :

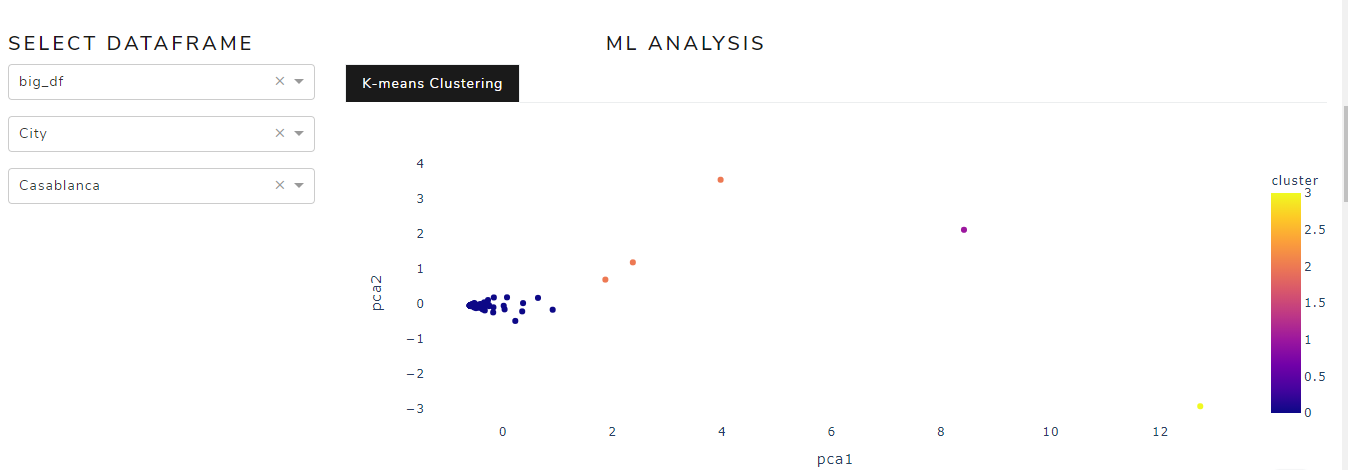
* Une segmentation générale



* 3 segmentations après le groupement par type de client



* 30 segmentations après le groupement des clients par ville



### Analyse des « clusters » :

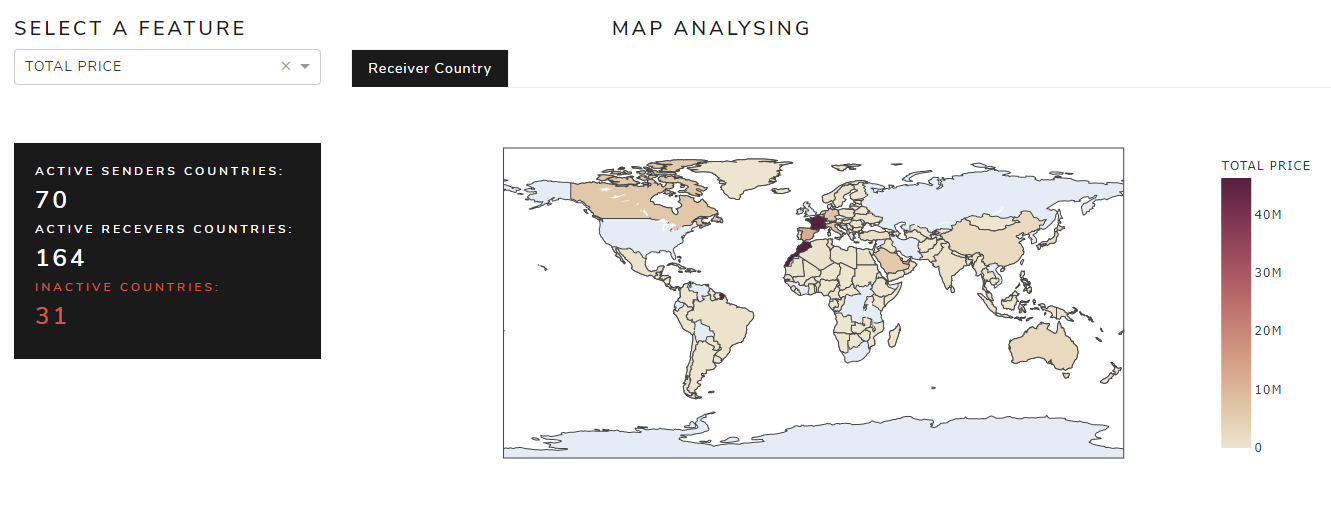
Après la segmentation, J’ai fait des analyses à chaque cluster et client pour que je puisse savoir comment je peux transférer un client d’un cluster X à un cluster Y.

L’analyse est basée sur les nombres d’expédition, le type de produit, le prix payé, le poids et le volume expédié.

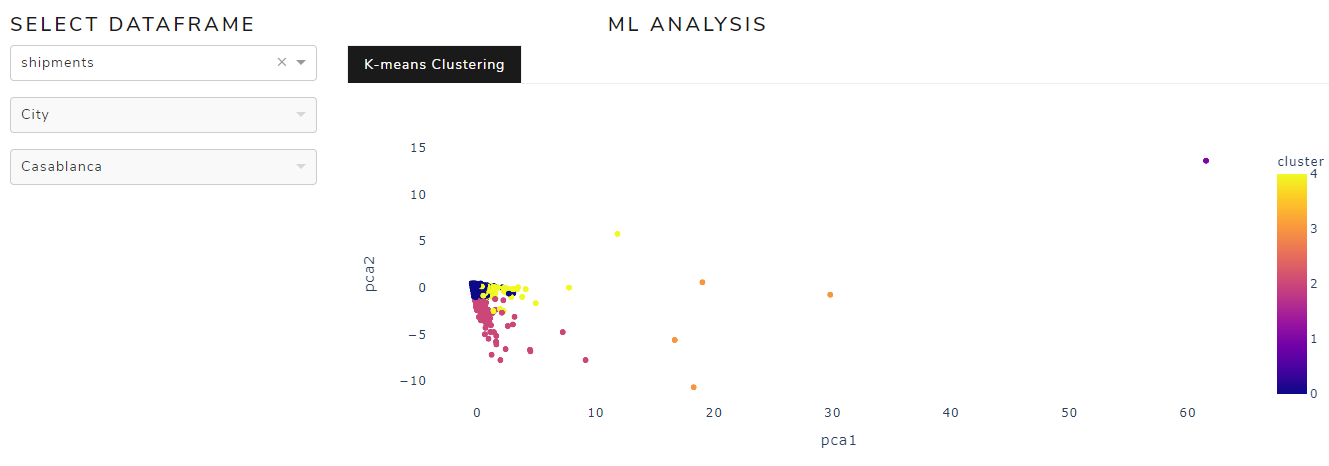
# Chapitre 4: Mise en œuvre

Ce chapitre a pour mission de présenter la phase de réalisation qui concrétise la conception exposée lors du chapitre précédent. Il expose le contact avec l’application.

### Visualisation par pays :



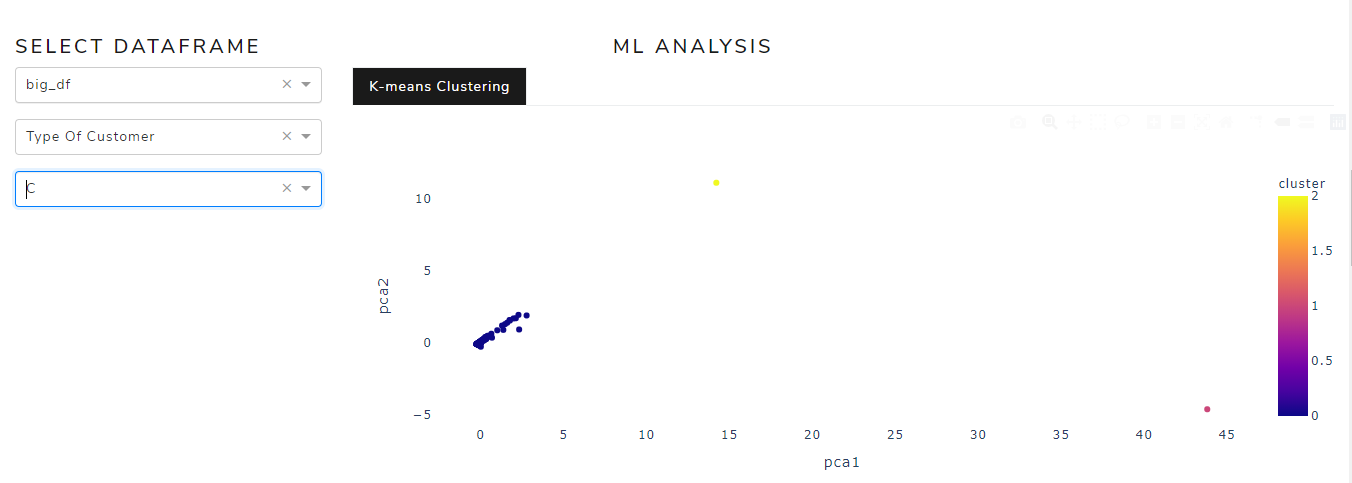
### Segmentation générale :



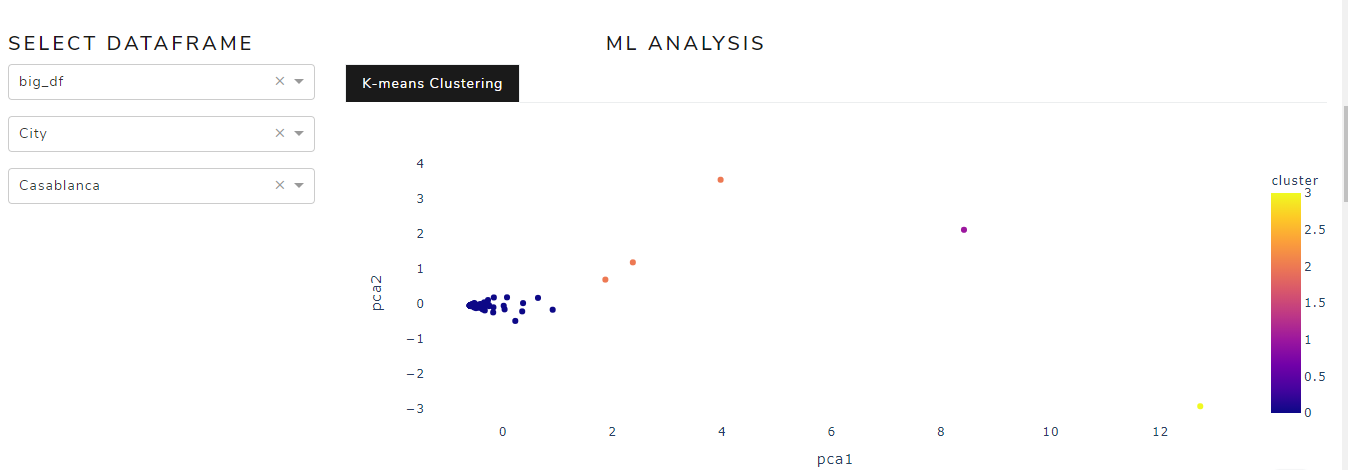
### Segmentation après groupement :

Dans ce cas, on va segmenter les clients après le regroupement selon l’indicateur choisi et puis on va faire le même analyse effectué dans la section précédente.

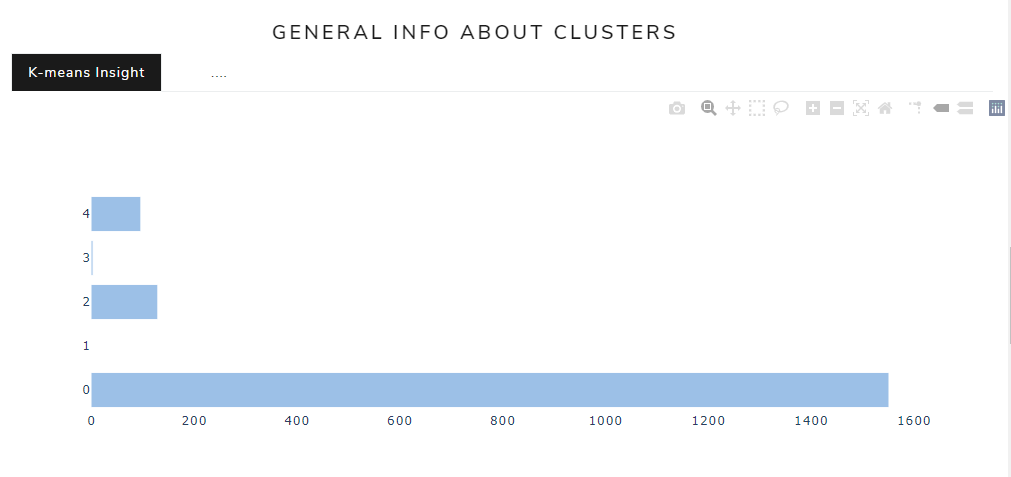
* **Par type de client :**

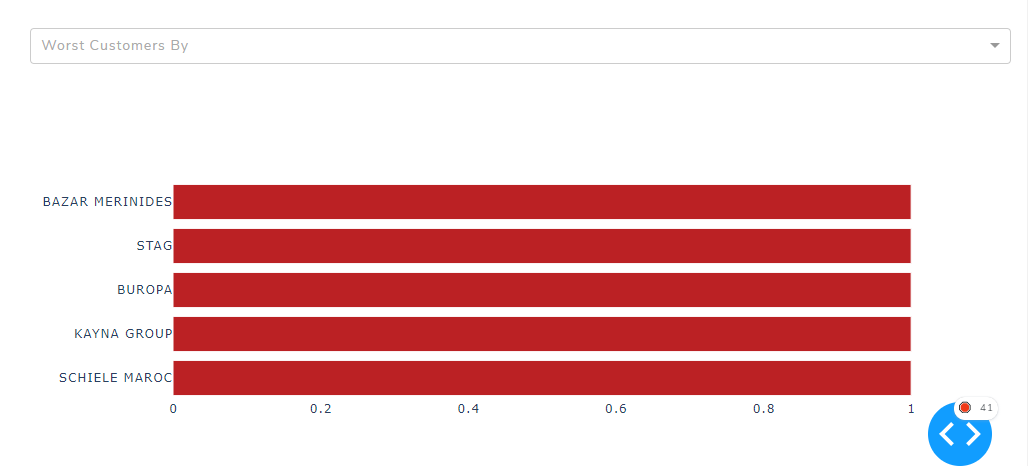


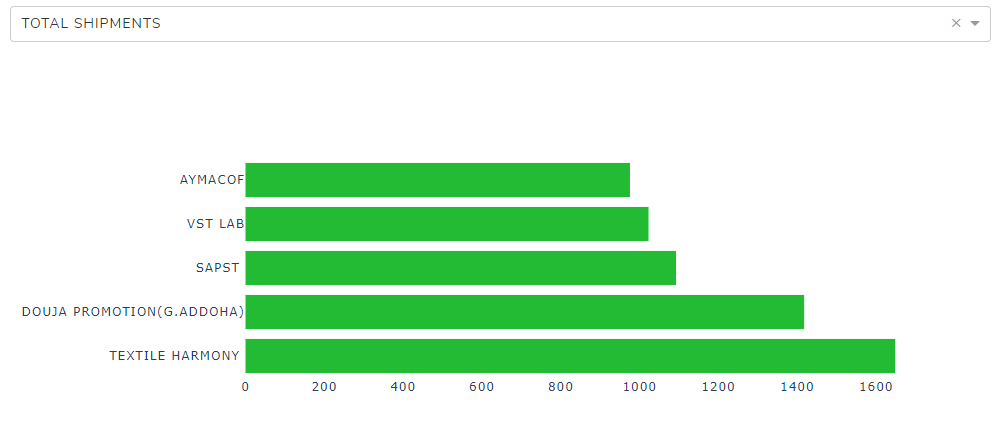
* **Par ville :**



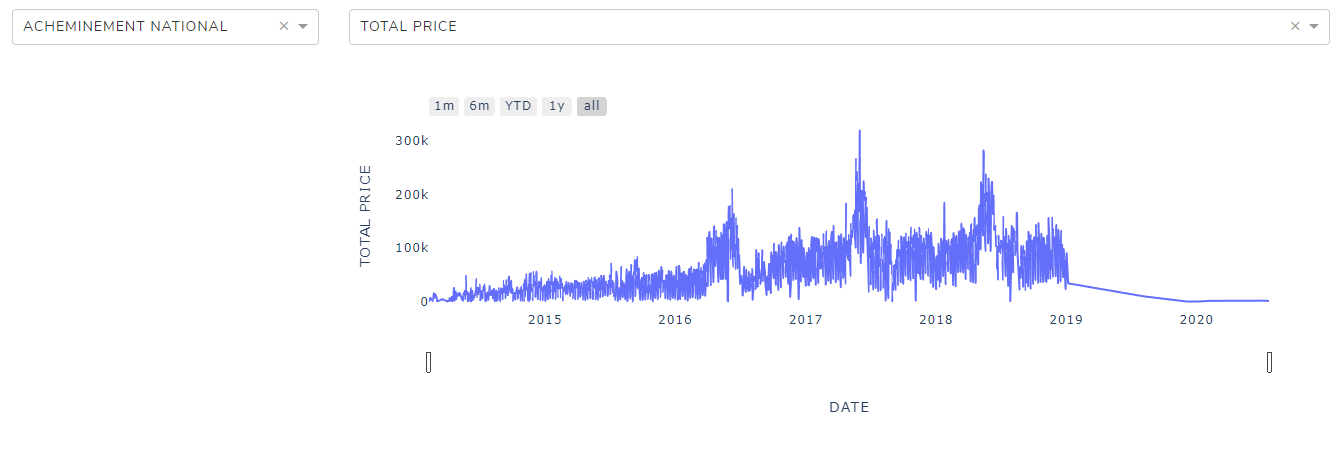
### Analyse par cluster :

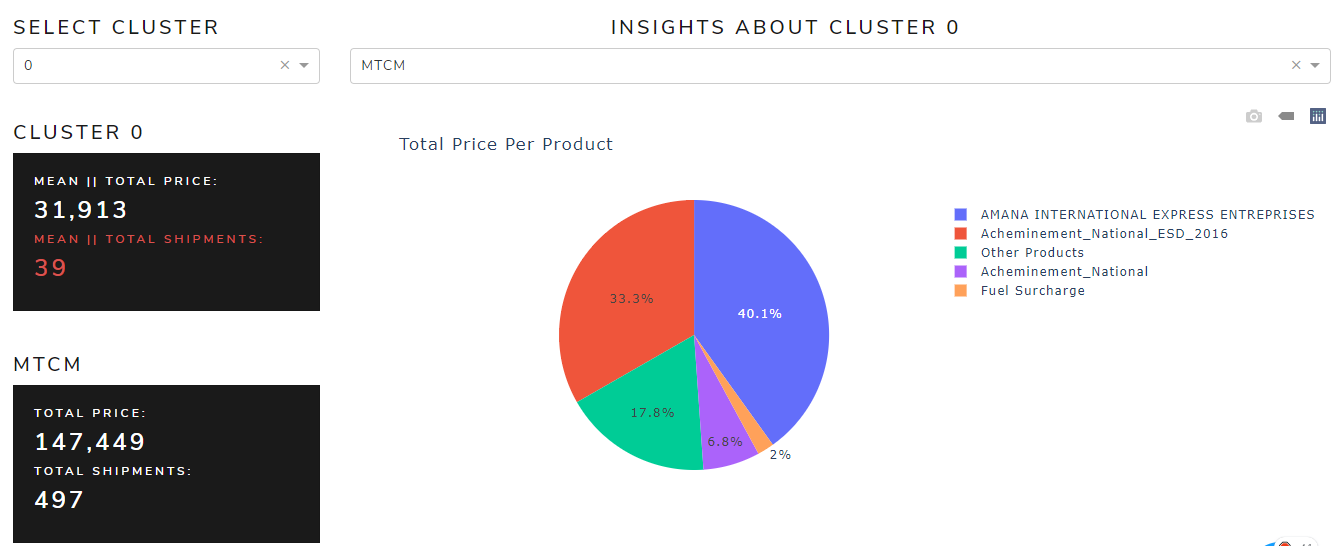






### Analyse par produit :





# Conclusion

# Bibliographie

# Annexes