

Université Gustave Eiffel

Master 2 Data Science et Société Numérique

PROJET PYTHON

**ANALYSE DE L'ÉCONOMIE NUMÉRIQUE DU
ROYAUME-UNI EN 2011**

VENTES EN LIGNE

MENARD June Camille

PHAM Dong Pha

Table des matières

I - Problématique

II - Analyse de la base de données

Importer les bibliothèques, la base de données et afficher les colonnes:

Chiffre d’Affaire (CA) mensuel

Visualisation des variations mensuelles du Chiffre d’Affaires

Représentation du taux de croissance sur une base mensuelle

Nombre de clients actifs mensuels

Nombre de commande mensuelle

Chiffre d’Affaires moyen par commande

Chiffre d’Affaires mensuel des catégories Nouveaux et Anciens clients

Représentation graphique des ratios Nouveaux clients et Anciens clients

Examinons le pourcentage de nouveaux clients enregistrés par mois

Taux mensuel de rétention de clients

III - Segmentation de la clientèle

Calcul de la récurrence (R)

Fréquence (F)

Classification des clients selon la fréquence (F)

Valeur monétaire (M)

Le Scoring RFM

IV - Conclusion et recommandation

I - Problématique

Dans ce travail, nous chercherons à extraire des informations économiques des achats en ligne faits au Royaume-Uni en 2010-2011 pour mieux comprendre le comportement de cette population par rapport aux achats en ligne sur toutes les périodes de l'année. Pour ce faire, nous étalerons notre étude sur plusieurs paramètres du dataset afin de déterminer les catégories d'acheteurs, leur périodes d'achats, les fréquences des achats et bien d'autres encore. Et pour finir, proposer des recommandations marketing adaptés aux diverses catégories identifiées.

Avec un intérêt pour les sujets économiques, pour notre analyse, nous avons choisi d'explorer une base de données existante sur un sujet économique.

Titre de la base de données : "Online Retail Data Set"

Nous avons récupéré notre base de données sur le site UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/online+retail>).

II - Analyse de la base de données

Importer les bibliothèques, la base de données et afficher les colonnes:

D'abord, nous importons les bibliothèques:

```
#Importer les bibliothèques
!pip install plotly
!pip install chart-studio
from datetime import datetime, timedelta
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from __future__ import division

import chart_studio.plotly as py
import plotly.offline as pyoff
import plotly.graph_objs as go

#Lancer la bibliothèque de visualisation pour le notebook
pyoff.init_notebook_mode()
```

Nous importons et affichons le base de données dans le Jupyter Notebook:

```
Entrée [428]: df = pd.read_csv('D:/OnlineRetail.csv', encoding= 'unicode_escape')
df.head()
```

Out[428]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom

Nous affichons les noms des colonnes:

```
#Voir les noms des colonnes
df.columns
```

```
Index(['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate',
       'UnitPrice', 'CustomerID', 'Country'],
      dtype='object')
```

Nous calculons Chiffre d'Affaires (CA) mensuel:

	FactureAnneeMois	ChiffreAffaire	CroissanceMensuelle
0	201012	748957.020	NaN
1	201101	560000.260	-0.252293
2	201102	498062.650	-0.110603
3	201103	683267.080	0.371850
4	201104	493207.121	-0.278163
5	201105	723333.510	0.466592
6	201106	691123.120	-0.044530
7	201107	681300.111	-0.014213
8	201108	682680.510	0.002026
9	201109	1019687.622	0.493653
10	201110	1070704.670	0.050032
11	201111	1461756.250	0.365228
12	201112	433668.010	-0.703324

Nous comptons les colonnes "country":

United Kingdom	495478
Germany	9495
France	8557
EIRE	8196
Spain	2533
Netherlands	2371
Belgium	2069
Switzerland	2002
Portugal	1519
Australia	1259
Norway	1086
Italy	803
Channel Islands	758
Finland	695
Cyprus	622
Sweden	462
Unspecified	446
Austria	401
Denmark	389
Japan	358
Poland	341
Israel	297
USA	291
Hong Kong	288
Singapore	229
Iceland	182
Canada	151
Greece	146
Malta	127
United Arab Emirates	68
European Community	61
RSA	58
Lebanon	45
Lithuania	35
Brazil	32
Czech Republic	30
Bahrain	19
Saudi Arabia	10
Name: Country, dtype: int64	

À partir de cette étape, nous décidons de n'utiliser que les données au Royaume-Uni ppays choisi pour l'analyse (notons qu'il représente la majorité des données de commande dans ce dataframe).

Chiffre d'Affaire (CA) mensuel

D'abord, nous convertissons le type de colonne "InvoiceDate " de string à datetime. Puis, nous créons le colonne "FactureAnneeMois" pour faciliter l'analyse et la visualisation. Ensuite, nous calculons le Chiffre d'Affaire pour chaque ligne et créons un nouveau dataframe avec le colonne "FactureAnneeMois" et le colonne "Chiffre d'Affaire"

	FactureAnneeMois	ChiffreAffaire
0	201012	748957.020
1	201101	560000.260
2	201102	498062.650
3	201103	683267.080
4	201104	493207.121
5	201105	723333.510
6	201106	691123.120
7	201107	681300.111
8	201108	682680.510
9	201109	1019687.622
10	201110	1070704.670
11	201111	1461756.250
12	201112	433668.010

Visualisation des variations mensuelles du Chiffre d'Affaires

Nous créons un graphique et nous l'affichons sous forme de graphique linéaire.

Chiffre Affaires (CA) mensuel



Nous constatons une importante augmentation du Chiffre d'Affaires à partir du mois d'Août 2011 avec un pic au mois de novembre suivi d'une chute considérable au mois de décembre. L'hypothèse de la préférence des achats sur place dans l'ambiance des fêtes et parfois de dernières minutes semble la plus plausible.

Représentation du taux de croissance sur une base mensuelle

Nous utilisons de la fonction "pct_change()" pour calculer le pourcentage de variation du chiffre d'affaires et nous affichons un tableau pour montrer le CA et le taux de croissance mensuel.

	FactureAnneeMois	ChiffreAffaire	CroissanceMensuelle
0	201012	748957.020	NaN
1	201101	560000.260	-0.252293
2	201102	498062.650	-0.110603
3	201103	683267.080	0.371850
4	201104	493207.121	-0.278163

Nous le visualisons par graphique linéaire:

Taux de croissance mensuel



Nous pouvons constater un fort ralentissement en Avril 2011(27%) suivi d'une croissance de 49% en Septembre 2011.

L'hypothèse de l'effet de saison semble la plus probable pour expliquer ces variations. Le plateau entre Juin et Août, période estivale donc de voyages à l'étranger suivi d'une rapide remontée en Septembre, période de la rentrée scolaire et des achats associés, semblent être des pistes confirmant cette hypothèse.

Nous pouvons donc supposer que le ralentissement du mois d'Avril est lié à la fin de toutes les périodes de fêtes et le pic du mois de Novembre 2011 à leur début cumulé aux périodes de congés scolaires particuliers de ce mois(achats de billets de trains, bus et autres).

Nombre de clients actifs mensuels

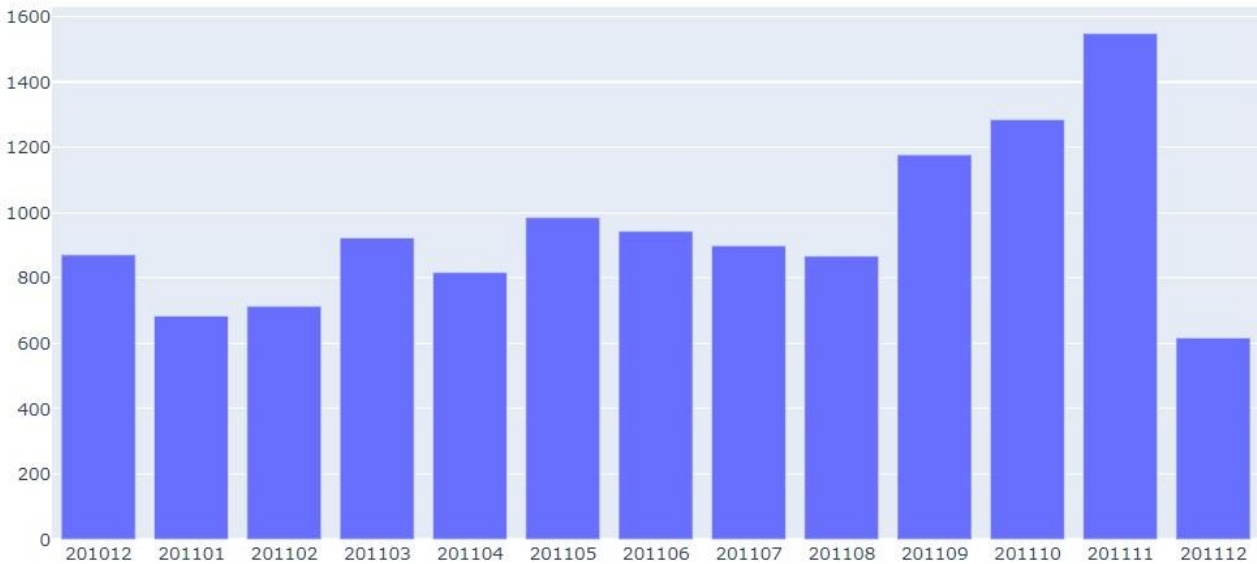
Pour avoir une vue détaillée des clients actifs sur une base mensuelle, nous procédons aux mêmes étapes que le calcul des variations mensuelles du Chiffre d'Affaire. Nous calculons donc le nombre de clients actifs mensuels à partir du nombre de "CustomerID".

Nous créons un nouveau dataframe uniquement pour les clients du Royaume-Uni, et nous créons une base de données mensuelle pour les clients actifs en comptant uniquement la colonne "CustomerID".

	FactureAnneeMois	CustomerID
0	201012	871
1	201101	684
2	201102	714
3	201103	923
4	201104	817
5	201105	985
6	201106	943
7	201107	899
8	201108	867
9	201109	1177
10	201110	1285
11	201111	1548
12	201112	617

Nous affichons le nouveau dataframe par graphique:

Nombre de clients actifs mensuels

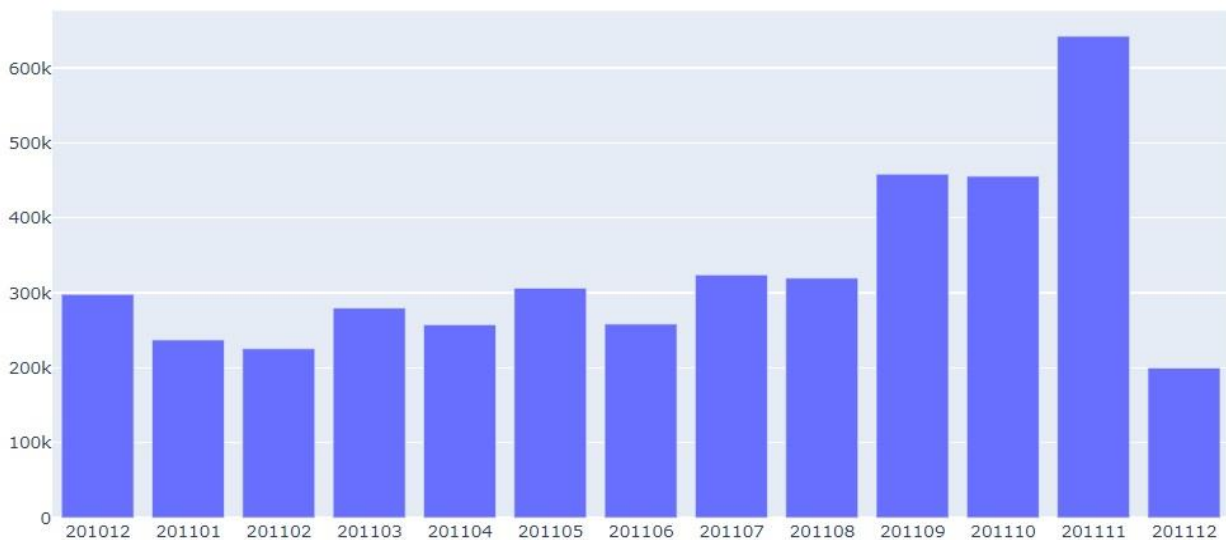


Nous constatons qu'entre Mars et Avril 2011, le nombre de clients actifs est passé de 923 à 817 (soit une diminution de 11,5 %).

Nombre de commande mensuel

Nous créons un nouveau dataframe du nombre de commande mensuel en utilisant la colonne de quantité en les affichant graphiquement.

Nombre de commande mensuel



Nous affichons un tableau du nouveau dataframe

	FactureAnneeMois	Quantity
0	201012	298101
1	201101	237381
2	201102	225641
3	201103	279843
4	201104	257666
5	201105	306452
6	201106	258522
7	201107	324129
8	201108	319804
9	201109	458490
10	201110	455612
11	201111	642281
12	201112	199907

Le nombre de commandes en avril 2011 a diminué de 8%, passant de 279843 en Mars 2011 à 257666 en Avril 2011 pour atteindre 642281 en novembre 2011. Nous voyons que le nombre de clients actifs affecte directement le nombre de commandes. Ensuite, considérons le Chiffre d'Affaire moyen par commande.

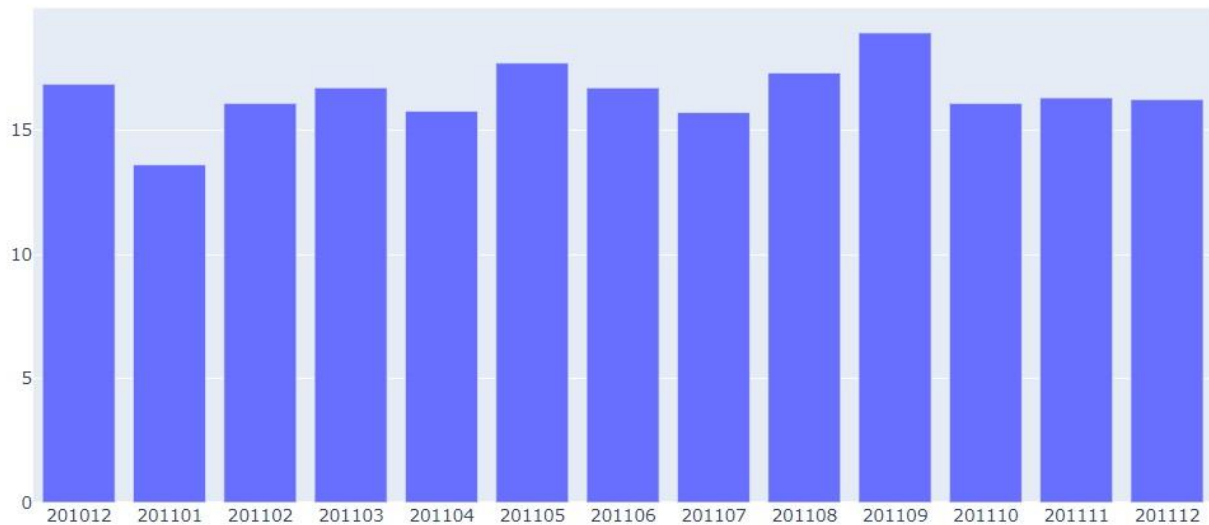
Chiffre d'Affaire moyen par commande

Nous créons un nouveau dataframe pour le chiffre d'affaires mensuels moyen

	FactureAnneeMois	ChiffreAffaire
0	201012	16.865860
1	201101	13.614680
2	201102	16.093027
3	201103	16.716166
4	201104	15.773380
5	201105	17.713823
6	201106	16.714748
7	201107	15.723497
8	201108	17.315899
9	201109	18.931723
10	201110	16.093582
11	201111	16.312383
12	201112	16.247406

Nous l'affichons graphiquement:

Chiffre Affaire moyen mensuelle



Nous constatons une diminution du nombre moyen de commandes mensuelles en Janvier 2011. Mais cela n'affecte pas beaucoup les ventes mensuelles. Ensuite, nous explorerons 2 autres paramètres: le Ratio des nouveaux client et le Taux de rétention.

- Ratio nouveau client: C'est un paramètre qui nous dira si nous perdons d'anciens clients ou si nous ne sommes pas en mesure d'attirer de nouveaux clients.
- Taux de rétention: Il s'agit du taux d'anciens clients qui continuent à utiliser les services pendant un certain temps.

Représentation graphique des ratios Nouveaux clients et Ancients clients

Avec les données concernant uniquement les commandes au Royaume-Uni que nous avons dans les étapes précédentes, nous pouvons supposer que le nouveau client est un client qui achète pour la première fois dans une période prédéfinie. Nous désignerons cette période en un mois. Nous utiliserons la fonction “min()” pour trouver la date du premier achat de chaque client et trouver de nouveaux clients en fonction de cela.

Nous créons un dataframe contenant le colonne "CustomerID" et la première date d'achat. Ensuite, nous fusionnerons la colonne de la première date d'achat dans la base de données principale.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	FactureAnneeMois	ChiffreAffaire	DateAchatMin	DateAchat
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	201012	15.30	2010-12-01 08:26:00	201012
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	201012	20.34	2010-12-01 08:26:00	201012
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	201012	22.00	2010-12-01 08:26:00	201012
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	201012	20.34	2010-12-01 08:26:00	201012
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	201012	20.34	2010-12-01 08:26:00	201012

Nous créons une colonne "Type d'utilisateur" et attribuons la ligne "Existante" sur le graphique (Si l'année du premier achat de l'utilisateur est avant l'année de facturation sélectionnée). Ensuite, nous calculons le chiffre d'affaire par mois pour chaque type d'utilisateur. Nous filtrons les dates et nous traçons le graphique.

Nouveau vs Existante



Chiffre d'Affaire mensuel des catégories Nouveaux et Anciens clients

	FactureAnneeMois	TypeUtilisateur	ChiffreAffaire
1	201101	Existante	195275.510
2	201101	Nouveau	156705.770
3	201102	Existante	220994.630
4	201102	Nouveau	127859.000
5	201103	Existante	296350.030
6	201103	Nouveau	160567.840
7	201104	Existante	268226.660
8	201104	Nouveau	108517.751
9	201105	Existante	434725.860
10	201105	Nouveau	90847.490
11	201106	Existante	408030.060
12	201106	Nouveau	64479.190
13	201107	Existante	407693.610
14	201107	Nouveau	53453.991
15	201108	Existante	421388.930
16	201108	Nouveau	55619.480
17	201109	Existante	640861.901
18	201109	Nouveau	135667.941
19	201110	Existante	648837.600
20	201110	Nouveau	133940.280
21	201111	Existante	838955.910
22	201111	Nouveau	117153.750

Nous constatons que les ventes ont tendance à augmenter avec les anciens clients, ce qui n'est pas le cas des nouveaux clients. L'explication de la segmentation acheteurs réguliers et acheteurs occasionnels semble être la plus plausible.

Examinons le pourcentage de nouveaux clients enregistrés par mois

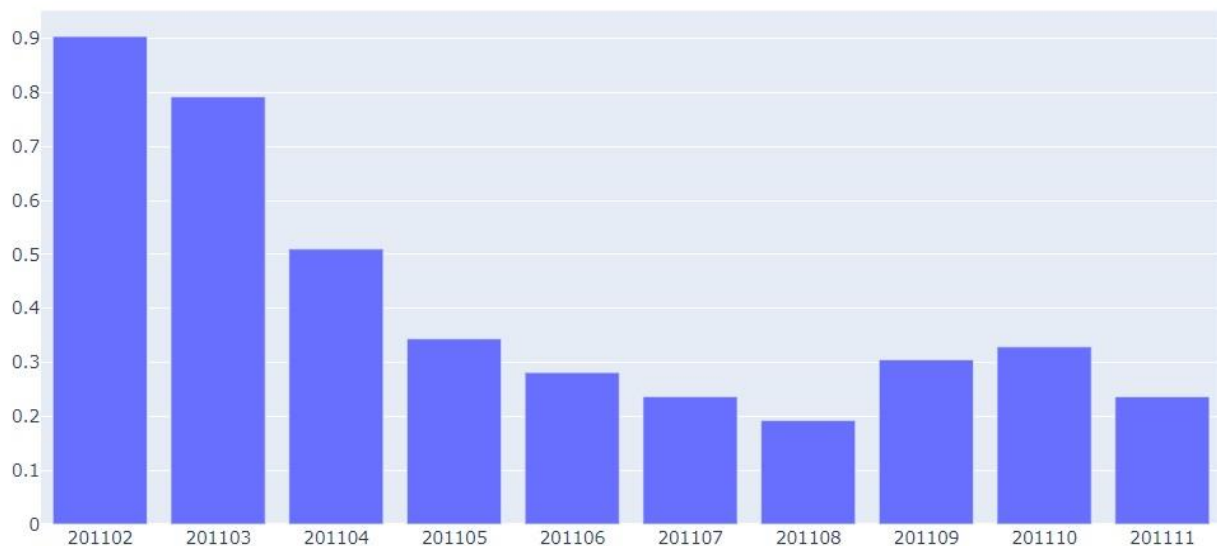
Nous créons un dataframe qui montre le nouveau ratio d'utilisateurs et nous supprimons les valeurs "nans" (NA), le ratio de nouveaux utilisateurs du premier mois est de 0.

Nous affichons le dataframe:

	FactureAnneeMois	CustomerID
1	201101	1.124224
2	201102	0.904000
3	201103	0.792233
4	201104	0.510166
5	201105	0.343793
6	201106	0.281250
7	201107	0.236589
8	201108	0.192572
9	201109	0.304878
10	201110	0.328852
11	201111	0.236422
12	201112	0.058319

Nous traçons le graphique:

Nouveau ratio client



Nous voyons que le ratio des nouveaux clients présente une fluctuation proportionnelle aux résultats précédemment récoltés (diminué comme prévu).

Taux mensuel de rétention de clients

Nous évaluons le taux de rétention des clients pour déterminer le degré d'adéquation au marché des produits et services proposés par ces plateformes.

Monthly Retention Rate = Retained Customers from Previous Month/Active Customers Total

D'abord, nous créons une matrice de rétention avec un tableau croisé. Nous utilisons la fonction `crosstab()` pour calculer le taux de fidélisation de la clientèle.

FactureAnneeMois	CustomerID	201012	201101	201102	201103	201104	201105	201106	201107	201108	201109	201110	201111	201112
0	12346.0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	12747.0	1	1	0	1	0	1	1	0	1	0	1	1	1
2	12748.0	1	1	1	1	1	1	1	1	1	1	1	1	1
3	12749.0	0	0	0	0	0	1	0	0	1	0	0	1	1
4	12820.0	0	1	0	0	0	0	0	0	0	1	1	0	1

Ici, nous avons réalisé une représentation binaire des clients en utilisant d'autres services en ligne au cours du mois (1) et ceux qui n'en utilisent pas (0).

Ensuite, nous regroupons les colonnes et créons un tableau de Chiffre d'Affaire mensuel par client. Puis, nous déterminons quels utilisateurs sont actifs en examinant les chiffres d'affaires mensuels que nous obtenons d'eux.

	CustomerID	FactureAnneeMois	ChiffreAffaire
0	12346.0	201101	0.00
1	12747.0	201012	706.27
2	12747.0	201101	303.04
3	12747.0	201103	310.78
4	12747.0	201105	771.31
...
12325	18283.0	201110	114.65
12326	18283.0	201111	651.56
12327	18283.0	201112	208.00
12328	18287.0	201105	765.28
12329	18287.0	201110	1072.00

12330 rows × 3 columns

D'abord, nous créons un tableau de dictionnaires qui permet de conserver le nombre total d'utilisateurs pour chaque mois. Ensuite, nous convertissons le tableau en dataframe et nous calculons le taux rétention de client mensuel. Enfin, nous traçons le graphique du taux de rétention.

Taux de rétention mensuel



Le résultat graphique fait état d'un taux de rétention mensuel croissant de Juin à Août qui retourne à la normale après cette période.

III - SEGMENTATION DE LA CLIENTELE

A partir des éléments récoltés nous avons la possibilité de classer les clients par catégories afin d'élaborer des stratégies marketing beaucoup plus ciblées.

Il existe de nombreuses méthodes de classement des clients. Dans cette partie, nous classerons les clients en fonction du modèle RFM (Recency - Frequency - Monetary Value), un modèle utilisé pour mesurer les taux d'utilisation des services clients, et évaluer l'adéquation des produits et services des entreprises sur le marché.

Sur la base des informations: dernier achat, nombre d'achats et valeur totale des offres. Nous regrouperons les clients dans les segments suivants:

- Faible valeur : il s'agit d'un groupe de clients qui achètent moins que les autres groupes, utilisent moins de services donc apportent de faibles profits aux entreprises.
- Valeur moyenne: il s'agit d'un groupe de clients avec une utilisation moyenne des paramètres d'achats, d'utilisation des services et donc de profits moyens aux entreprises.
- Valeur élevée : il s'agit du segment de clientèle le plus important, ils apportent des profits élevés à l'entreprise et ils utilisent très fréquemment les services.

Nous calculerons les métriques: Récence, Fréquence et Valeur monétaire et utiliserons des méthodes d'apprentissage non supervisées pour regrouper ces clients.

Pour calculer le score de récence, nous devons connaître la date d'achat la plus récente du client, et leurs jours d'inactivité (compté à partir de la date d'achat la plus récente jusqu'au moment considéré). Lorsque nous avons un nombre de jours inactifs pour chaque client, nous appliquerons l'algorithme de clustering (K-means) pour calculer le score de récence.

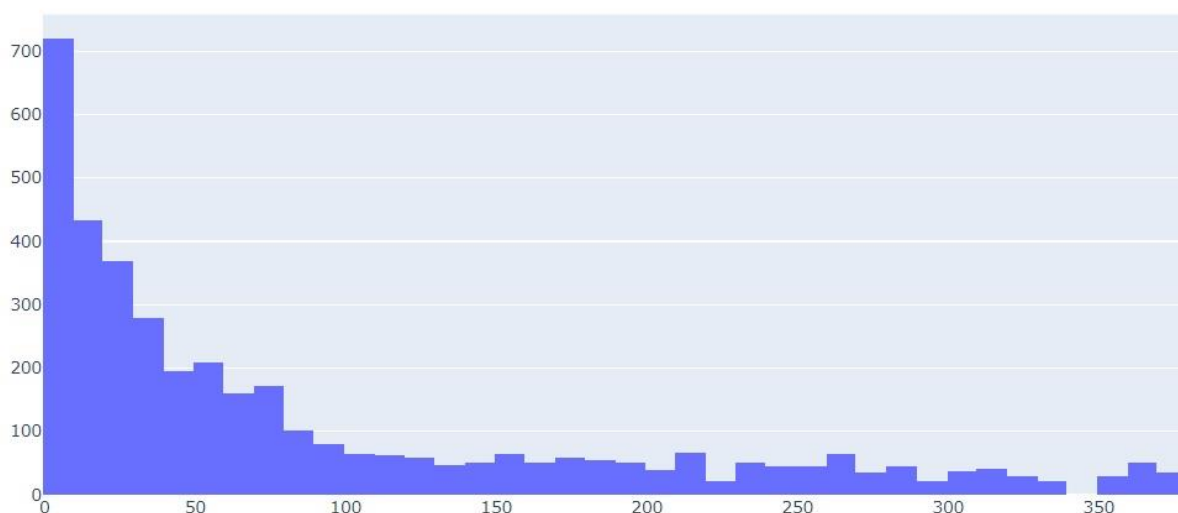
Comme dans les étapes précédentes, nous convertissons le type de string à datetime dans le colonne "InvoiceDate". Nous n'utiliserons que des données au Royaume Uni, et nous affichons le dataframe.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Calcul de la récence (R)

Nous créons un cadre de données d'utilisateurs pour conserver la colonne "CustomerID" et les nouveaux scores de segmentation. Ensuite nous calculons la date d'achat maximale pour chaque client puis nous créons un dataframe. Enfin nous fusionnons ce dataframe avec notre nouveau dataframe utilisateurs et traçons l'histogramme.

Récence



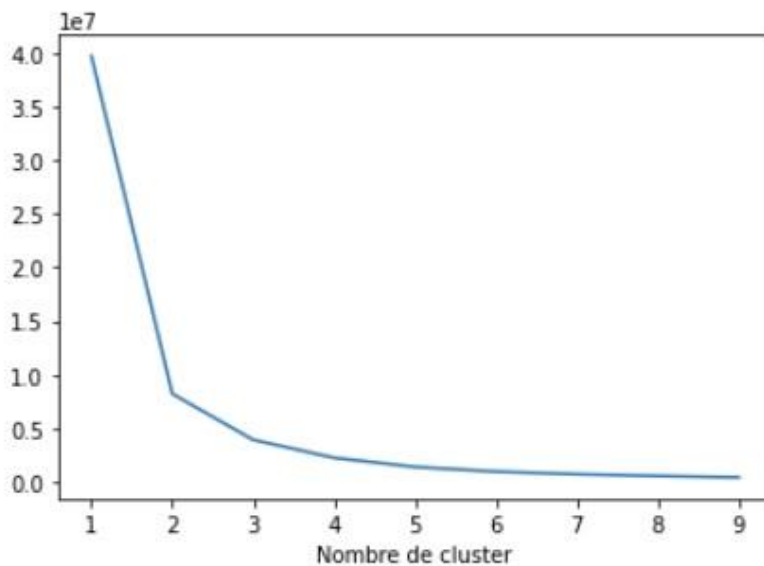
Nous affichons le tableau et examinons la distributions des données de récence (R):

	CustomerID	Recence
0	17850.0	301
1	13047.0	31
2	13748.0	95
3	15100.0	329
4	15291.0	25

```
count    3950.000000
mean      90.778481
std       100.230349
min        0.000000
25%       16.000000
50%       49.000000
75%      142.000000
max       373.000000
Name: Recence, dtype: float64
```

La période moyenne d'achat la plus récente des clients est de 90 jours et 50% des clients ont une période d'achat la plus récente de 49 jours, et de 142 jours pour 70% des clients, et de 16 jours pour 25% des clients.

Nous appliquerons l'algorithme de clustering K-means pour calculer le score de récence (Recency). Nous formerons d'abord l'algorithme pour déterminer le nombre de groupe à attribuer en appliquant la méthode Elbow. Cette méthode nous donnera le nombre optimal de clusters pour l'algorithme.



Nous voyons que le nombre optimal de clusters est 3. Mais nous pouvons choisir un nombre de clusters inférieur ou supérieur au nombre optimal en fonction des besoins des entreprises. Nous choisissons donc 4 clusters pour ce cas.

Nous créons donc 4 clusters pour les visites récentes de clients en les ajoutant au dataframe. Nous classons des clusters et nous affichons le tableau.

	count	mean	std	min	25%	50%	75%	max
RecenceCluster								
0	478.0	304.393305	41.183489	245.0	266.25	300.0	336.0	373.0
1	570.0	184.436842	31.856230	131.0	156.00	184.0	211.0	244.0
2	952.0	77.567227	22.743569	48.0	59.00	72.0	93.0	130.0
3	1950.0	17.488205	13.237058	0.0	6.00	16.0	28.0	47.0

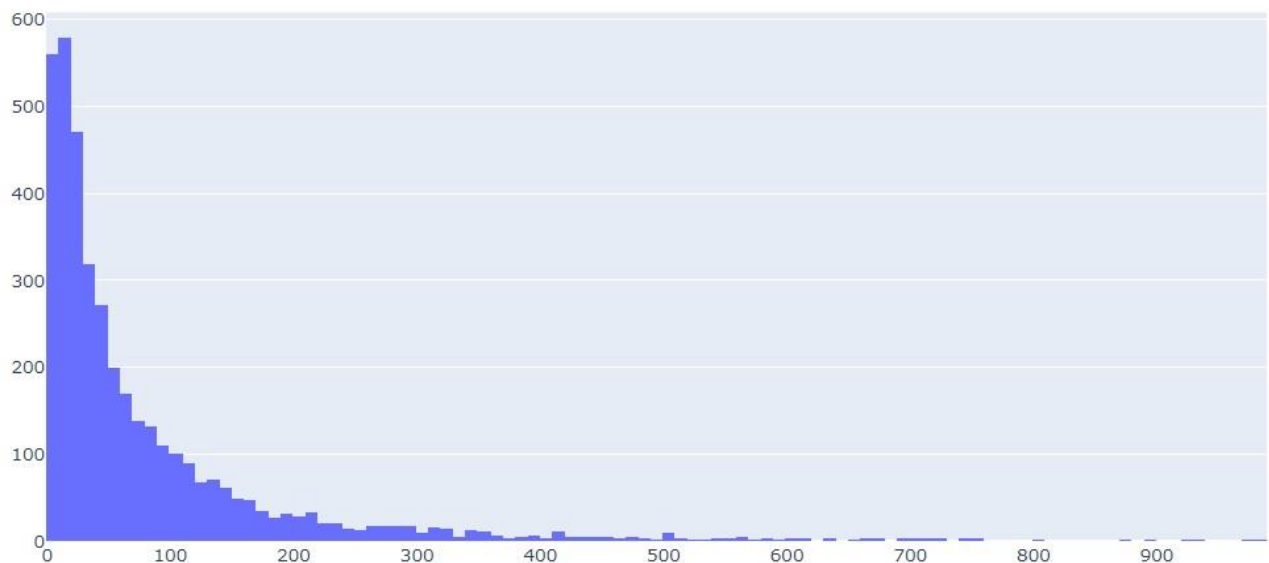
Selon ce résultat, le cluster numéro 3 est le cluster qui a la meilleure récence soit (17 jours). Il est également le cluster ayant le plus de clients (1950).

Frequence (F)

Calculons la fréquence (F) et la valeur monétaire. Pour regrouper le client par fréquence, nous devons trouver le nombre total de commandes par client.

Trouvons le nombre de commandes pour chaque utilisateur ensuite créons un dataframe. Ajoutons ensuite ces données à notre base de données principale puis élaborons le graphique.

Frequence



Classification des clients selon la fréquence (F)

En utilisant K-means, créons le cluster de fréquences et affichons les détails de chaque cluster.

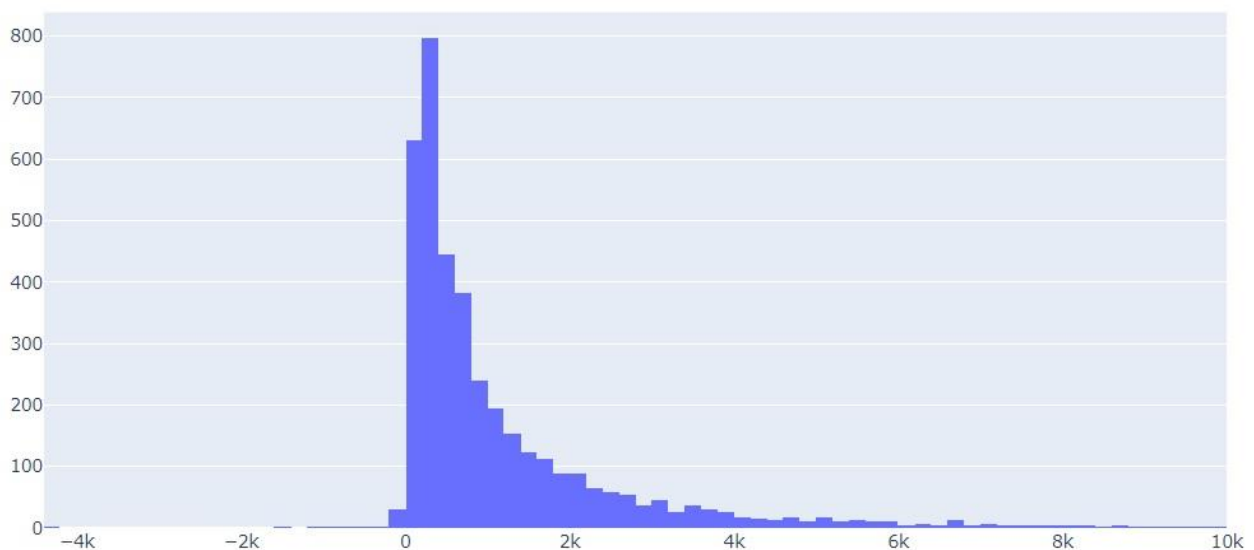
	count	mean	std	min	25%	50%	75%	max
FréquenceCluster								
0	3496.0	49.525744	44.954212	1.0	15.0	33.0	73.0	190.0
1	429.0	331.221445	133.856510	191.0	228.0	287.0	399.0	803.0
2	22.0	1313.136364	505.934524	872.0	988.5	1140.0	1452.0	2782.0
3	3.0	5917.666667	1805.062418	4642.0	4885.0	5128.0	6555.5	7983.0

Selon ces résultats, nous voyons que plus F est grand mieux c'est pour l'entreprise, car les clients passent plus de commandes, donc ils génèrent plus de profits.

Valeur monétaire (M)

Nous regroupons les clients en fonction de la valeur monétaire (M). Nous calculons d'abord M pour chaque client, ensuite nous les regroupons et puis les traçons.

Valeur monétaire



Nous constatons qu'il y a quelques clients qui rapportent des profits négatifs aux entreprises. Cela implique que nous devons continuer le clustering.

Ensuite, nous appliquons le clustering puis créons des numéros de cluster. Enfin, nous affichons les détails du dataframe.

	count	mean	std	min	25%	50%	75%	max
ChiffreAffaireCluster								
0	3686.0	906.329979	920.325222	-4287.63	262.8975	572.5055	1257.5925	4301.22
1	235.0	7746.035787	3636.348298	4314.72	5152.9650	6530.0400	9116.7900	21535.90
2	27.0	43070.445185	15939.249588	25748.35	28865.4900	36351.4200	53489.7900	88125.38
3	2.0	221960.330000	48759.481478	187482.17	204721.2500	221960.3300	239199.4100	256438.49

Le Scoring RFM

Le scoring consiste à attribuer des points au client en faisant la somme de ses variables. La méthode RFM est la plus connue des méthodes de segmentation par scoring. Elle est employée principalement par les sociétés de vente à distance.

Ici, nous avons déjà des scores RFM (en cluster). Nous avons déjà des scores RFM (en cluster). Nous calculons le score global et nous utilisons la fonction "mean()" pour avoir les détails

	Recence	Frequence	ChiffreAffaire
ScoreGlobal			
0	304.584388	21.995781	303.339705
1	185.170213	32.569149	499.035215
2	78.870229	47.044711	864.548497
3	20.729318	68.304434	1092.971126
4	14.929766	271.481605	3601.606990
5	9.662162	373.290541	9136.946014
6	7.740741	876.037037	22777.914815
7	1.857143	1272.714286	103954.025714
8	1.333333	5917.666667	42177.930000

A partir de nos résultats, nous pouvons dire que, le client ayant un score de 8 est le meilleur client. Pour simplifier, nous allons les diviser en trois groupes comme suit:

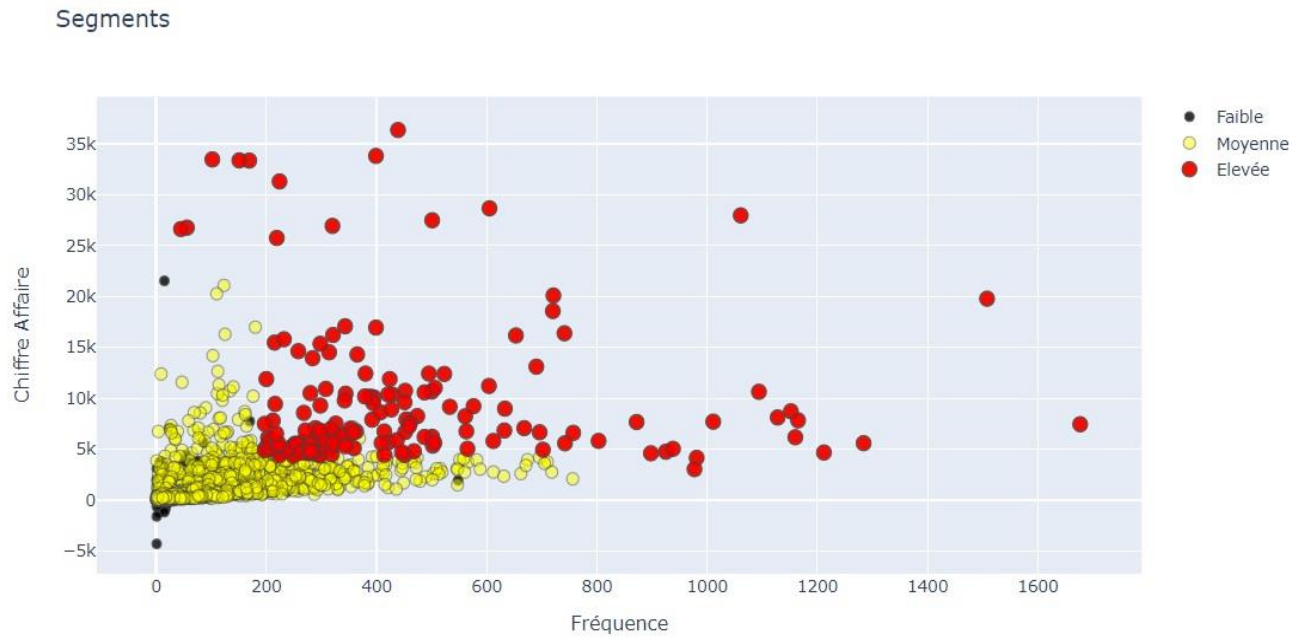
0 à 2: valeur faible

3 à 4: valeur moyenne

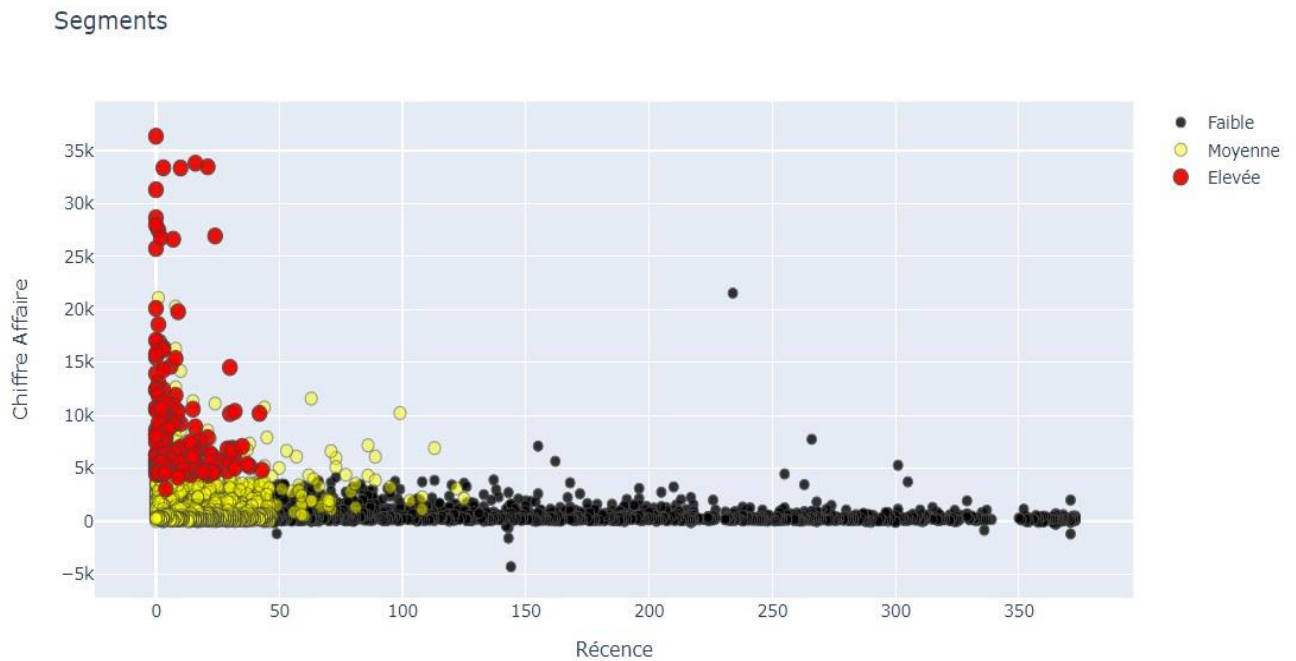
5+: valeur élevée

Dessignons maintenant une figure pour avoir en visuel la répartition de la population de clients.

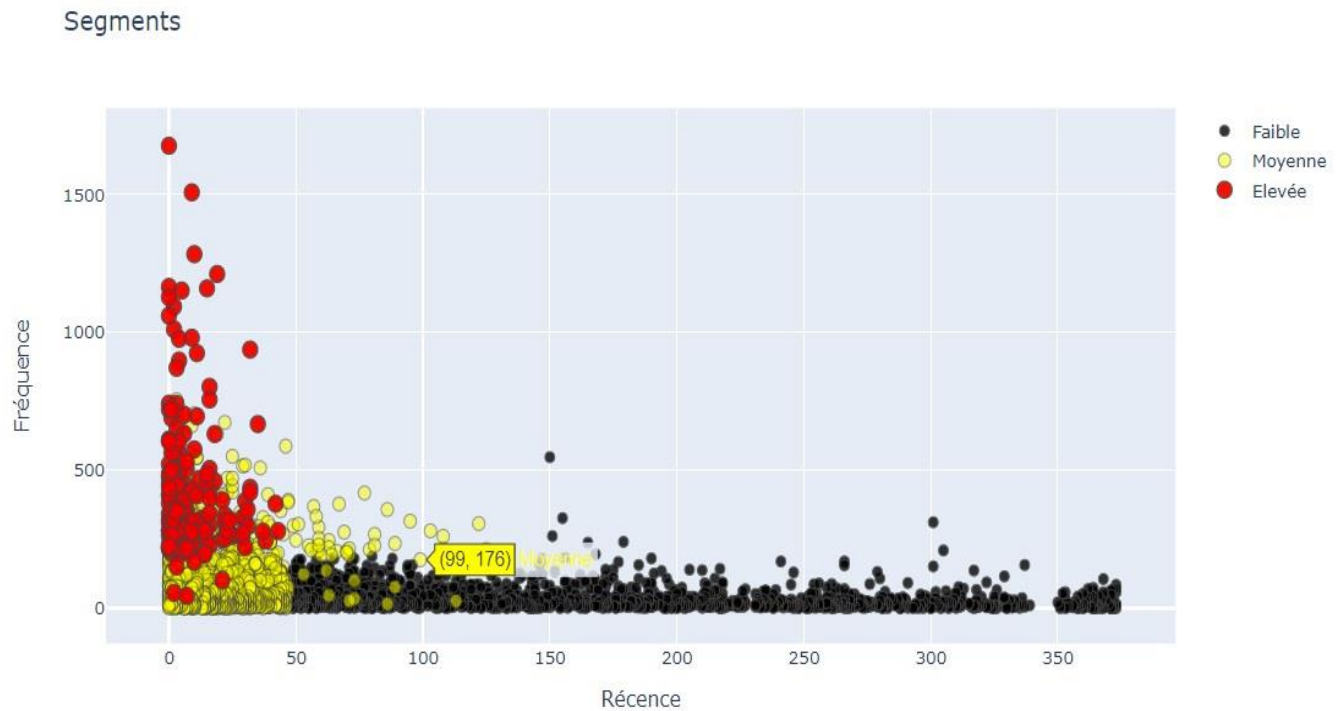
- Chiffre d'Affaires et la fréquence:



- Chiffre d'Affaire et la récence:



- La récence et la fréquence:



IV - CONCLUSION ET RECOMMANDATIONS

Pour conclure, nous pouvons dire qu'en plus des informations économiques que nous avons pu extraire de la base de données, les informations sociologiques y sont très parlantes. Car nous avons pu analyser les comportements de la population du Royaume Uni via leurs achats en ligne. Nous avons également pu retracer des variations périodiques conjuguées sur les plans économique et social avec une grande finesse.

A partir des résultats récoltés, nous recommandons les stratégies suivantes pour les 3 groupes de clients :

Valeur élevée: Améliorer la rétention par des offres de fidélisation (Cartes cadeau, remises spéciales, ect...)

Valeur moyenne: Améliorer la rétention tout en augmentant la fréquence (Nouvelles méthodes publicitaires, bons de réduction ciblés, ect...)

Valeur basse: Augmenter la fréquence par des stratégies de marketing à plus grande échelle, offre de nouveaux produits pour mieux identifier les goûts et les méthodes d'achats de cette catégorie.