

P5 : SEGMENTEZ DES CLIENTS D'UN SITE E-COMMERCE

September 16, 2021

Carlos SACRISTAN

OpenClassrooms

MISSION OBJECTIVE



- 1 Regrouper ensemble des clients de profils similaires
- 2 Mieux comprendre les différents types d'utilisateurs.

TABLE DES MATIÈRES



- 1 Mission objective**
 - Analyse graphique
 - Test de corrélation
- 2 Data preparation**
 - Data cleansing
 - Combining data
- 3 Data analyses**
 - Preparation data
 - KMeans
- 4 Modeling**
 - Conclusions et perspectives
- 5 Conclusions et perspectives**

INDEX



1 Mission objective

2 Data preparation

- Data cleansing

- Combining data

3 Data analyses

- Analyse graphique

- Test de corrélation

4 Modeling

- Preparation data

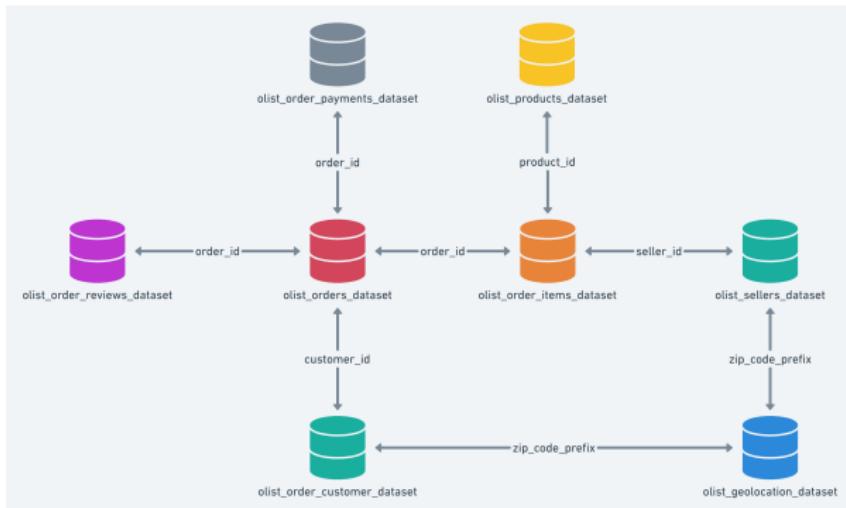
- KMeans

5 Conclusions et perspectives

DATA CLEANSING



- 1 Structure of 8 df
- 2 Basic cleaning
- 3 Outliers
- 4 Features transformations



DATA CLEANSING



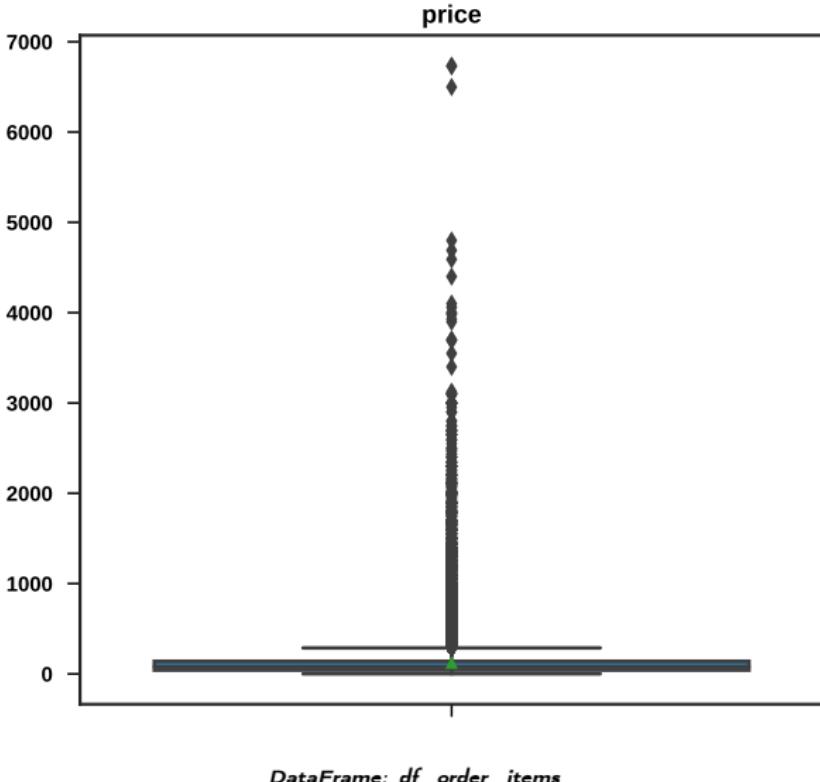
- 1 Structure of 8 df
- 2 **Basic cleaning**
- 3 Outliers
- 4 Features transformations

- Import the 8 csv
- Lowercases of features name
- Convert features to best possible dtypes
- Same features name (geo_code)
- Delete of duplicate values
- Delete of "empty" features : 'review_comment_message', 'review_comment_title' (df_order_reviews)
- Delete of missing values
- Physically impossible values (weight, length, height, width >0)

DATA CLEANSING



- 1 Structure of 8 df
- 2 Basic cleaning
- 3 **Outliers**
- 4 Features transformations



DATA CLEANSING



- 1 Structure of 8 df
- 2 Basic cleaning
- 3 Outliers
- 4 **Features transformations**

The ordinal categorical features are classified in 5 level (percentile):

- Weight, length, height, width, name_length, description_length, photos_qty

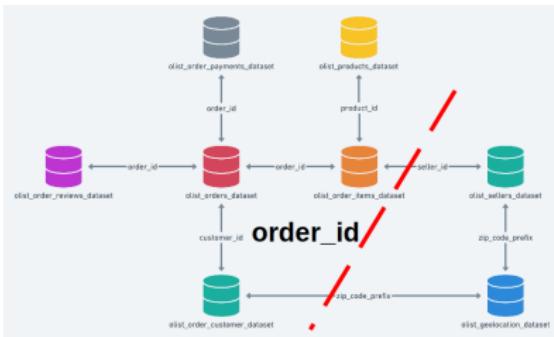
The temporal features :

- Nominal categorical features : Per trimester, Week / Weekend, per hours (7-12h/12-14h/14-18h/18-23h/23-7)

Translation category (Br → Eng) :

- 'product_category_name' ⇔ df_product_category_name_translation

COMBINING DATA

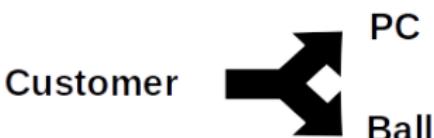


1 Joining 8 df

- 2 Grouping data
- 3 Population

merge \implies dfTOTAL (75286 rows \times 44 columns)

customer_unique_id customer_id



Index \implies 'customer_id'

COMBINING DATA



customer_unique_id customer_id



`groupby('customer_unique_id') ==> dfTOTAL2
(73054 rows × 44 columns)`

- 1 Joining 8 df
- 2 **Grouping data**
- 3 Population

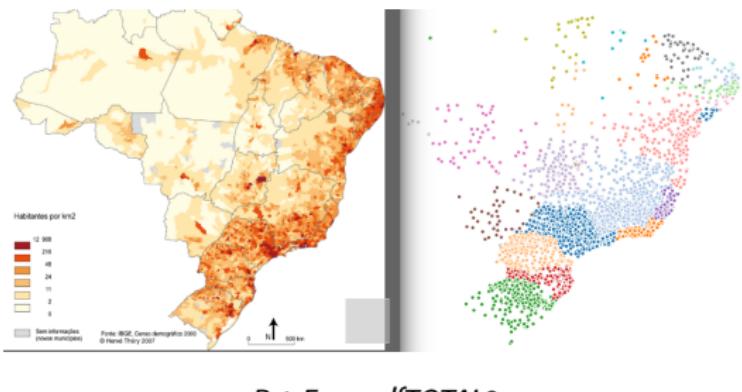
- Last one → 'customer_city', 'customer_state', 'order_purchase', 'lat', 'lon'
 - Mean → 'price', 'freight_value', 'delivered_time', 'review_score', 'name_length', 'description_length'
 - Median → 'photos_qty', 'purchase_M', 'purchase_W', 'purchase_H'
-

- Sum('price') → 'price_sum'
- Count('customer_id') → 'item_number' per customer

COMBINING DATA

olist

- 1 Joining 8 df
- 2 Grouping data
- 3 **Population**



- `groupby('city') → Count('customer_unique_id') → 'population'`
- `groupby('state') → Count('customer_unique_id') → 'population_st'`

INDEX



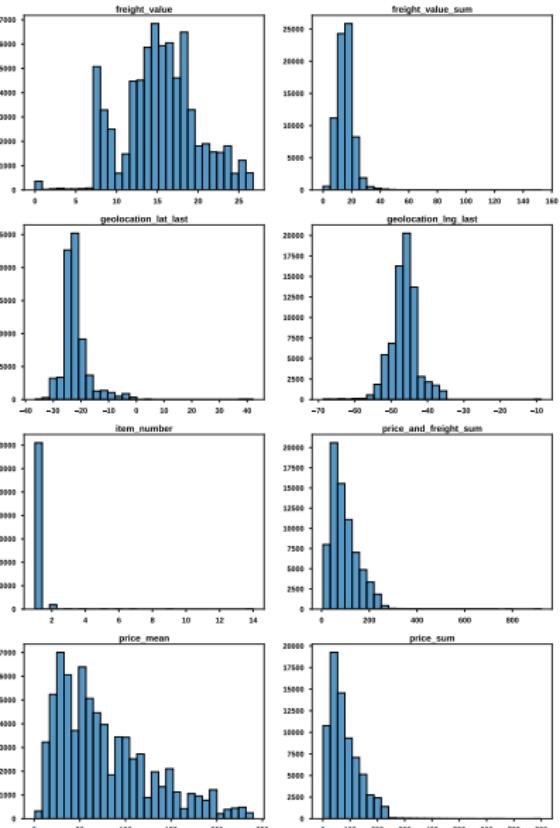
- 1 Mission objective
- 2 Data preparation
 - Data cleansing
 - Combining data
- 3 Data analyses
 - Analyse graphique
 - Test de corrélation
- 4 Modeling
 - Preparation data
 - KMeans
- 5 Conclusions et perspectives

ANALYSE GRAPHIQUE

1 1D Distribution

- Continuos
- Discret

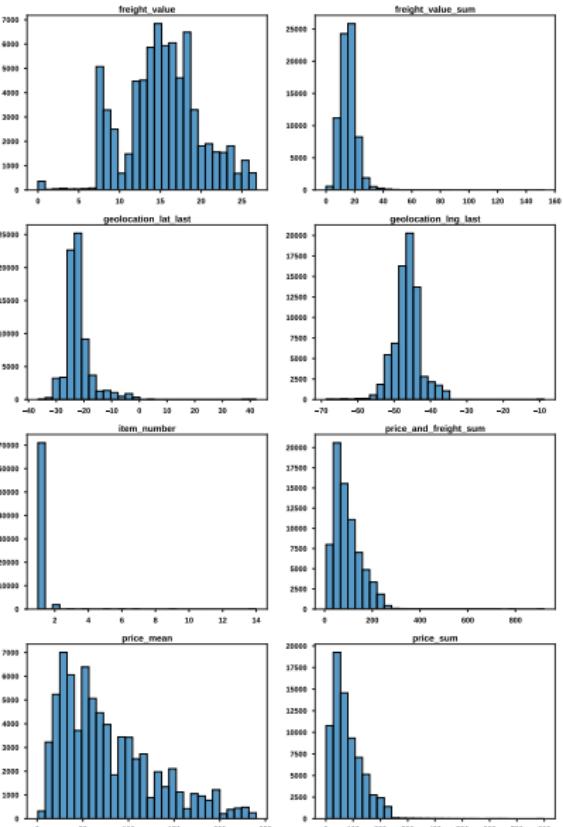
2 Lorenz curve



ANALYSE GRAPHIQUE

- 1** 1D Distribution
 - Continuos
 - Discret

- 2** Lorenz curve



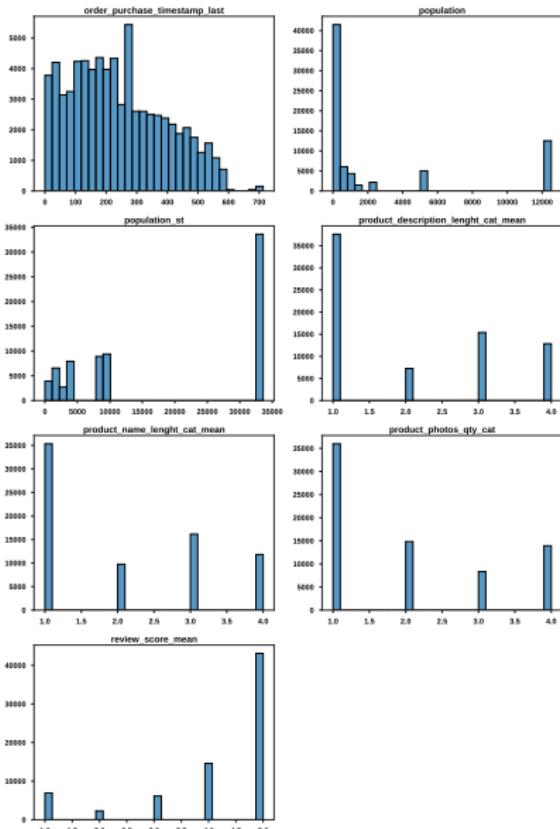
ANALYSE GRAPHIQUE



1 1D Distribution

- Continuos
- Discret

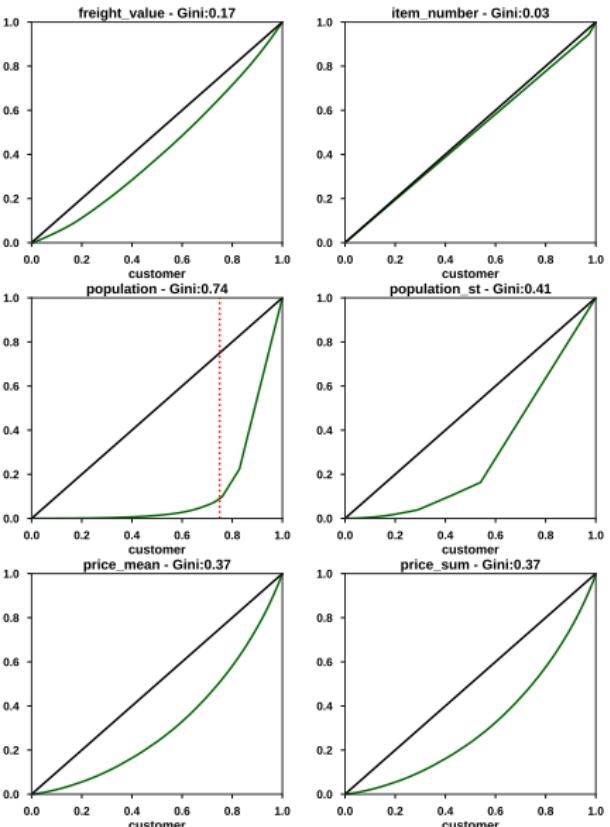
2 Lorenz curve



ANALYSE GRAPHIQUE

- 1 1D Distribution
 - Continuos
 - Discret

- 2 Lorenz curve



DataFrame: dfTOTAL3

TEST DE CORRÉLATION



1 r2 (>0.7)

2 Test de normalité

- Shapiro-Wilk's test
- Visualisation (qqplot)

1 Transformation de Box-Cox

1 Kruskal-Wallis test

1 Chi2

- pval
- chi2

1 Summary

price_mean	-1.0	1.0	0.3	0.2	0.9
price_sum	-1.0	1.0	0.3	0.4	1.0
freight_value	-0.3	0.3	1.0	0.8	0.3
freight_value_sum	-0.2	0.4	0.8	1.0	0.4
price_and_freight_sum	-0.9	1.0	0.3	0.4	1.0

	price_mean	price_sum	freight_value	freight_value_sum	price_and_freight_sum
price_mean	-1.0	1.0	0.3	0.2	0.9
price_sum	-1.0	1.0	0.3	0.4	1.0
freight_value	-0.3	0.3	1.0	0.8	0.3
freight_value_sum	-0.2	0.4	0.8	1.0	0.4
price_and_freight_sum	-0.9	1.0	0.3	0.4	1.0

TEST DE CORRÉLATION



1 r2 (>0.7)

2 Test de normalité

- Shapiro-Wilk's test
- Visualisation (qqplot)

1 Transformation de Box-Cox

1 Kruskal-Wallis test

1 Chi2

- pval
- chi2

1 Summary

Appliquer l'ANOVA -> Vérifier la normalité de la distribution et l'homoscedasticité

L'hypothèse nulle : La population est normalement distribuée.

Si $pval < \alpha$ choisi, alors l'hypothèse nulle est rejetée et il est prouvé que les données testées ne sont pas normalement distribuées.

		W	pval	normal
	freight_value	0.58	0.00	False
	freight_value_sum	0.60	0.00	False
	item_number	0.15	0.00	False
	price_and_freight_sum	0.50	0.00	False
	price_mean	0.48	0.00	False
	price_sum	0.48	0.00	False

DataFrame: dfTOTAL_cor

Dans notre cas, la probabilité que les variables soient normalement distribuées est de 0.00%.
 $pval < \alpha$ (5%).

TEST DE CORRÉLATION



Appliquer l'ANOVA -> Vérifier la normalité de la distribution et l'homoscedasticité

1 r2 (>0.7)

2 Test de normalité

- Shapiro-Wilk's test
- Visualisation (qqplot)

1 Transformation de Box-Cox

1 Kruskal-Wallis test

1 Chi2

- pval
- chi2

1 Summary

L'hypothèse nulle : La population est normalement distribuée.
Si $pval < \alpha$ choisi, alors l'hypothèse nulle est rejetée et il est prouvé que les données testées ne sont pas normalement distribuées.

		W	pval	normal
1	freight_value	0.58	0.00	False
1	freight_value_sum	0.60	0.00	False
1	item_number	0.15	0.00	False
1	price_and_freight_sum	0.50	0.00	False
1	price_mean	0.48	0.00	False
1	price_sum	0.48	0.00	False

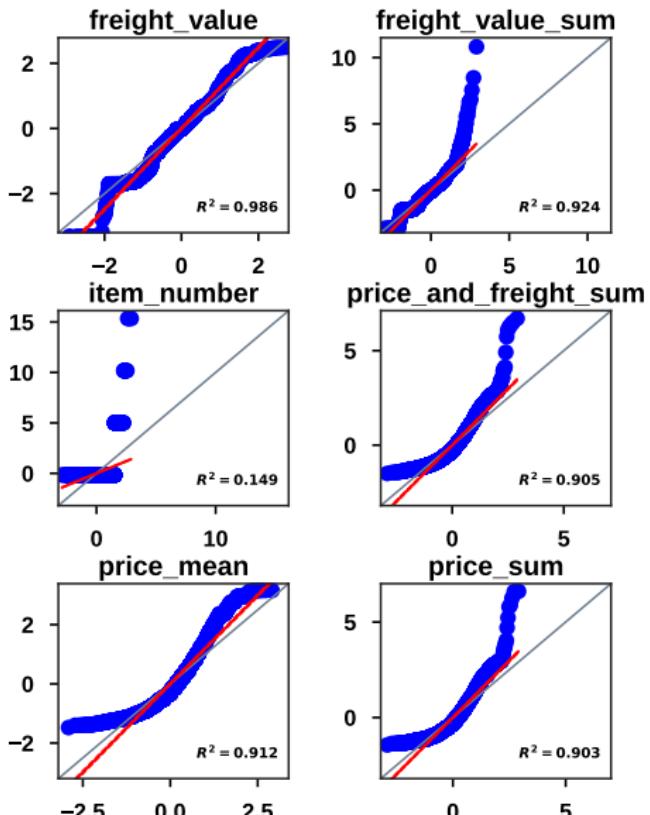
DataFrame: dfTOTAL_cor

Dans notre cas, la probabilité que les variables soient normalement distribuées est de 0.00%.
 $pval < \alpha$ (5%).

TEST DE CORRÉLATION

Appliquer l'ANOVA -> Vérifier la normalité de la distribution et l'homoscedasticité

- 1 r2 (>0.7)
- 2 Test de normalité
 - Shapiro-Wilk's test
 - Visualisation (qqplot)
- 1 Transformation de Box-Cox
- 1 Kruskal-Wallis test
- 1 Chi2
 - pval
 - chi2
- 1 Summary

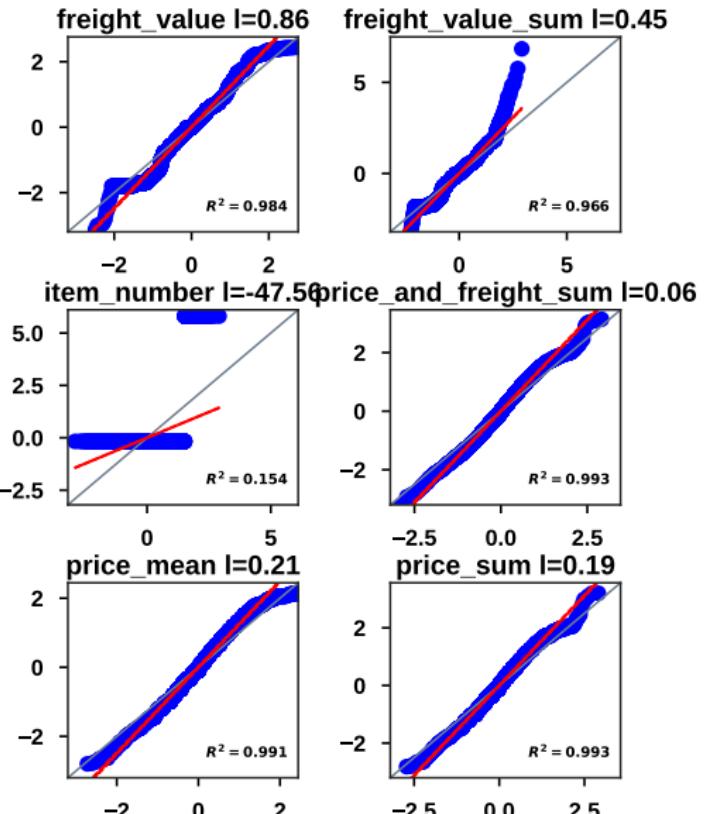


TEST DE CORRÉLATION

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(y_i) & \text{if } \lambda = 0, \end{cases}$$

olist

- 1 r2 (>0.7)
- 2 Test de normalité
 - Shapiro-Wilk's test
 - Visualisation (qqplot)
- 1 Transformation de Box-Cox
- 1 Kruskal-Wallis test
- 1 Chi2
 - pval
 - chi2
- 1 Summary



TEST DE CORRÉLATION



1 r² (>0.7)

2 Test de normalité

- Shapiro-Wilk's test
- Visualisation (qqplot)

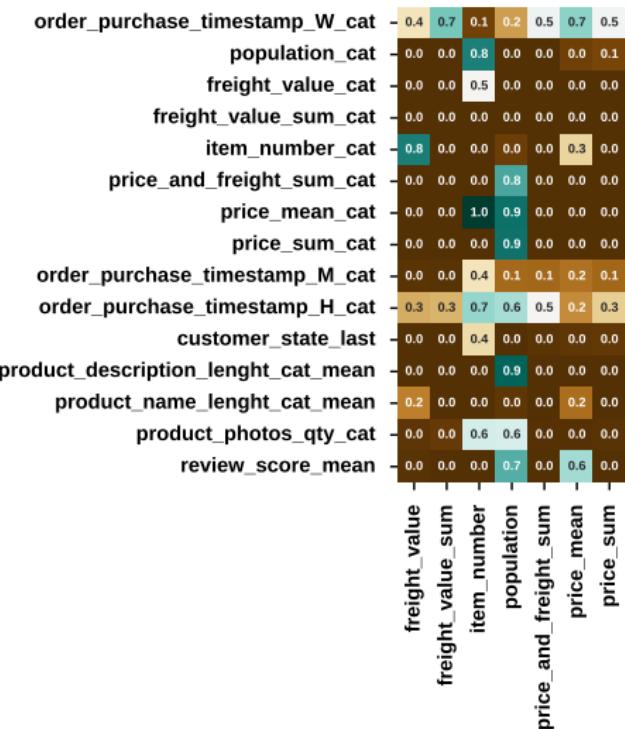
1 Transformation de Box-Cox

1 Kruskal-Wallis test

1 Chi2

- pval
- chi2

1 Summary



DataFrame: dfTOTAL_cor'

TEST DE CORRÉLATION



1 $r^2 (>0.7)$

2 Test de normalité

- Shapiro-Wilk's test
- Visualisation (qqplot)

1 Transformation de Box-Cox

1 Kruskal-Wallis test

1 Chi2

- pval
- chi2

1 Summary

	order_purchase_timestamp_W_cat	population_cat	freight_value_cat	freight_value_sum_cat	item_number_cat	price_and_freight_sum_cat	price_mean_cat	price_sum_cat	order_purchase_timestamp_M_cat	order_purchase_timestamp_H_cat	customer_state_last	product_description_lenght_cat_mean	product_name_lenght_cat_mean	product_photos_qty_cat	review_score_mean	pval
order_purchase_timestamp_W_cat	-	1.0	0.2	0.5	0.0	0.3	0.2	0.1	0.2	0.0	0.1	0.4	0.8	0.5	0.2	
population_cat	1.0	-	0.0	0.0	0.0	0.3	0.3	0.5	0.5	0.1	0.0	0.2	0.2	0.4	0.1	
freight_value_cat	0.2	0.0	-	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.4	0.0	0.0
freight_value_sum_cat	-0.5	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
item_number_cat	0.0	0.0	0.3	0.0	-	0.0	0.6	0.0	0.4	0.2	0.7	0.0	0.0	0.0	0.0	0.0
price_and_freight_sum_cat	-0.3	0.3	0.0	0.0	0.0	-	0.0	0.0	0.8	0.4	0.4	0.0	0.0	0.0	0.0	0.0
price_mean_cat	-0.2	0.3	0.0	0.0	0.6	0.0	-	0.0	0.7	0.2	0.6	0.0	0.5	0.0	0.5	0.8
price_sum_cat	-0.1	0.5	0.0	0.0	0.0	0.0	0.0	-	0.7	0.6	0.6	0.0	0.0	0.0	0.0	0.0
order_purchase_timestamp_M_cat	-0.2	0.5	0.0	0.0	0.4	0.8	0.7	0.7	-	0.9	0.1	0.5	0.2	0.3	0.0	
order_purchase_timestamp_H_cat	-0.0	0.1	0.1	0.2	0.2	0.4	0.2	0.6	0.6	-	0.1	0.1	0.4	0.2	0.4	
customer_state_last	-0.1	0.0	0.0	0.0	0.7	0.4	0.6	0.6	0.1	0.1	-	1.0	0.0	0.6	0.0	
product_description_lenght_cat_mean	-0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.1	1.0	-	0.0	0.0	0.0	
product_name_lenght_cat_mean	-0.8	0.2	0.4	0.0	0.0	0.0	0.5	0.0	0.2	0.4	0.0	-	0.0	0.0	0.0	
product_photos_qty_cat	-0.5	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.6	0.0	-	0.0	0.9	
review_score_mean	-0.2	0.1	0.0	0.0	0.0	0.8	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.9	-	

TEST DE CORRÉLATION



1 $r^2 (>0.7)$

2 Test de normalité

- Shapiro-Wilk's test
- Visualisation (qqplot)

1 Transformation de Box-Cox

1 Kruskal-Wallis test

1 Chi2

- pval
- chi2

1 Summary

	order_purchase_timestamp_W_cat	population_cat	freight_value_cat	freight_value_sum_cat	item_number_cat	price_and_freight_sum_cat	price_mean_cat	price_sum_cat	order_purchase_timestamp_M_cat	order_purchase_timestamp_H_cat	customer_state_last	product_description_lenght_cat_mean	product_name_lenght_cat_mean	product_photos_qty_cat	review_score_mean	pval
order_purchase_timestamp_W_cat	1.0	0.2	0.5	0.0	0.3	0.2	0.1	0.2	0.0	0.1	0.4	0.8	0.5	0.2	-	-
population_cat	-1.0	1.0	0.0	0.0	0.0	0.3	0.3	0.5	0.5	0.1	0.0	0.2	0.2	0.4	0.1	-
freight_value_cat	-0.2	0.0	1.0	0.0	0.3	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.4	0.0	0.0	-
freight_value_sum_cat	-0.5	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	-
item_number_cat	-0.0	0.0	0.3	0.0	1.0	0.0	0.6	0.0	0.4	0.2	0.7	0.0	0.0	0.0	0.0	-
price_and_freight_sum_cat	-0.3	0.3	0.0	0.0	0.0	1.0	0.0	0.0	0.8	0.4	0.4	0.0	0.0	0.0	0.0	-
price_mean_cat	-0.2	0.3	0.0	0.0	0.6	0.0	1.0	0.0	0.7	0.2	0.6	0.0	0.5	0.0	0.5	-
price_sum_cat	-0.1	0.5	0.0	0.0	0.0	0.0	0.0	1.0	0.7	0.6	0.6	0.0	0.0	0.0	0.0	-
order_purchase_timestamp_M_cat	-0.2	0.5	0.0	0.0	0.4	0.8	0.7	0.7	1.0	0.1	0.5	0.2	0.3	0.0	-	-
order_purchase_timestamp_H_cat	-0.0	0.1	0.1	0.2	0.2	0.4	0.2	0.6	0.9	1.0	0.1	0.1	0.4	0.2	0.4	-
customer_state_last	-0.1	0.0	0.0	0.0	0.7	0.4	0.6	0.6	0.1	0.1	1.0	0.0	0.6	0.0	0.0	-
product_description_lenght_cat_mean	-0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.1	1.0	0.0	0.0	0.0	0.0	-
product_name_lenght_cat_mean	-0.8	0.2	0.4	0.0	0.0	0.0	0.5	0.0	0.2	0.4	0.0	0.0	0.0	0.0	0.0	-
product_photos_qty_cat	-0.5	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.6	0.0	0.0	0.0	0.9	-
review_score_mean	-0.2	0.1	0.0	0.0	0.0	0.8	0.0	0.0	0.4	0.0	0.0	0.0	0.9	0.0	0.9	-

TEST DE CORRÉLATION



1 $r^2 (>0.7)$

2 Test de normalité

- Shapiro-Wilk's test
- Visualisation (qqplot)

1 Transformation de Box-Cox

1 Kruskal-Wallis test

1 Chi2

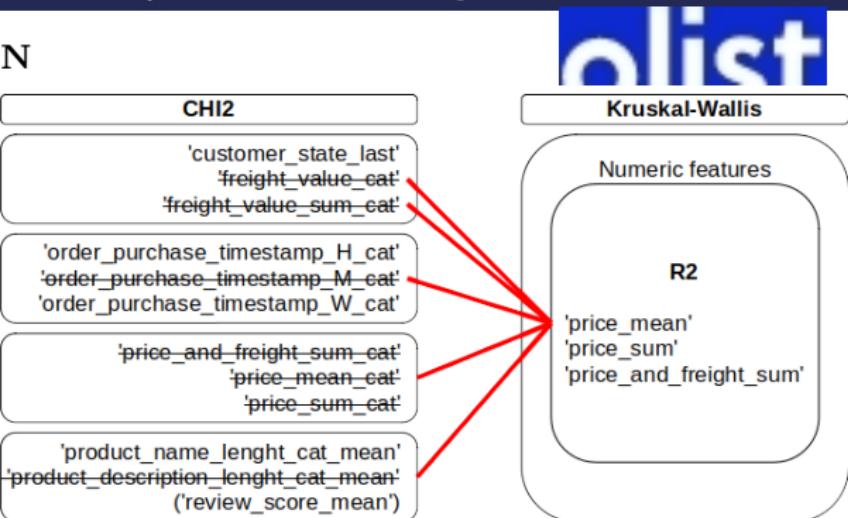
- pval
- chi2

1 Summary

order_purchase_timestamp_W_cat	0	3	1	8	2	4	5	5	33	34	9	6	3	13
population_cat	0	125	99	32	3	3	1	2	8	339	13	13	3	15
freight_value_cat	3	125	7648	5	291	208	186	150	15	1972	52	22	23	46
freight_value_sum_cat	1	99	7648	42	500	246	337	92	14	1717	160	160	14	76
item_number_cat	8	32	5	42	19	3	14	6	14	43	102	97	22	57
price_and_freight_sum_cat	2	3	291	500	19	7976	8092	3	10	52	203	116	31	55
price_mean_cat	4	3	208	246	3	7976	8248	4	13	48	111	19	47	15
price_sum_cat	5	1	186	337	14	8092	8248	4	8	47	176	99	44	46
order_purchase_timestamp_M_cat	5	2	150	92	6	3	4	4	8	92	26	35	11	60
order_purchase_timestamp_H_cat	33	8	15	14	14	10	13	8	8	150	56	52	19	52
customer_state_last	34	339	1972	1717	43	52	48	47	92	150	116	291	71	1864
product_description_lenght_cat_mean	9	13	52	160	102	203	111	176	26	56	116	8328	82	3748
product_name_lenght_cat_mean	6	13	22	160	97	116	19	99	35	52	291	8328	172	4308
product_photos_qty_cat	3	3	23	14	22	31	47	44	11	19	71	82	172	20
review_score_mean	13	15	46	76	57	55	15	46	60	52	1864	3748	4308	20
order_purchase_timestamp_W_cat														
population_cat														
freight_value_cat														
freight_value_sum_cat														
item_number_cat														
price_and_freight_sum_cat														
price_mean_cat														
price_sum_cat														
customer_state_last														
product_description_lenght_cat_mean														
product_name_lenght_cat_mean														
product_photos_qty_cat														
review_score_mean														

TEST DE CORRÉLATION

- 1 r2 (>0.7)
- 2 Test de normalité
 - Shapiro-Wilk's test
 - Visualisation (qqplot)
- 1 Transformation de Box-Cox
- 1 Kruskal-Wallis test
- 1 Chi2
 - pval
 - chi2
- 1 Summary



6 QUANTITATIVE FEATURES : 'freight_value_sum',
 'population_st', 'price_mean', 'price_sum_z-score',
 'estimated_delivered_time_mean',
 'order_purchase_timestamp_last'

7 QUALITATIVE FEATURES :
 'order_purchase_timestamp_W_cat',
 'product_photos_qty_cat', 'review_score_mean',
 'customer_state_last',
 'o_purchase_time_H_cat', 'product_category_name',
 'product_name_lenght_cat_mean'

INDEX



1 Mission objective

2 Data preparation

- Data cleansing

- Combining data

3 Data analyses

- Analyse graphique

- Test de corrélation

4 Modeling

- Preparation data

- KMeans

5 Conclusions et perspectives

PREPARATION DATA



- 1 Selected features
- 2 Encoding, Scaling
- 3 PCA

4 QUANTITATIVE FEATURES : 'freight_value_sum',
'population_st', 'price_mean', 'price_sum_z-score',
'estimated_delivered_time_mean', 'order_purchase_timestamp_last'
3 QUALITATIVE FEATURES : 'order_purchase_timestamp_W_cat',
'product_photos_qty_cat', 'review_score_mean',
'customer_state_last',
'o_purchase_time_H_cat', 'product_category_name',
'product_name_length_cat_mean'

PREPARATION DATA



- 1 Selected features
- 2 Encoding,
Scaling
- 3 PCA

OrdinalEncoder:

`product_photos_qty_cat', 'review_score_mean'`

OneHotEncoder

`'order_purchase_timestamp_W_cat'`

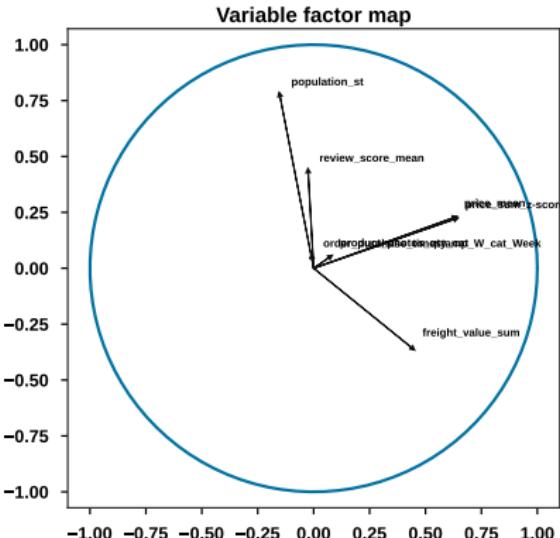
MinMaxScaler

Scale change without distance deformation

PREPARATION DATA

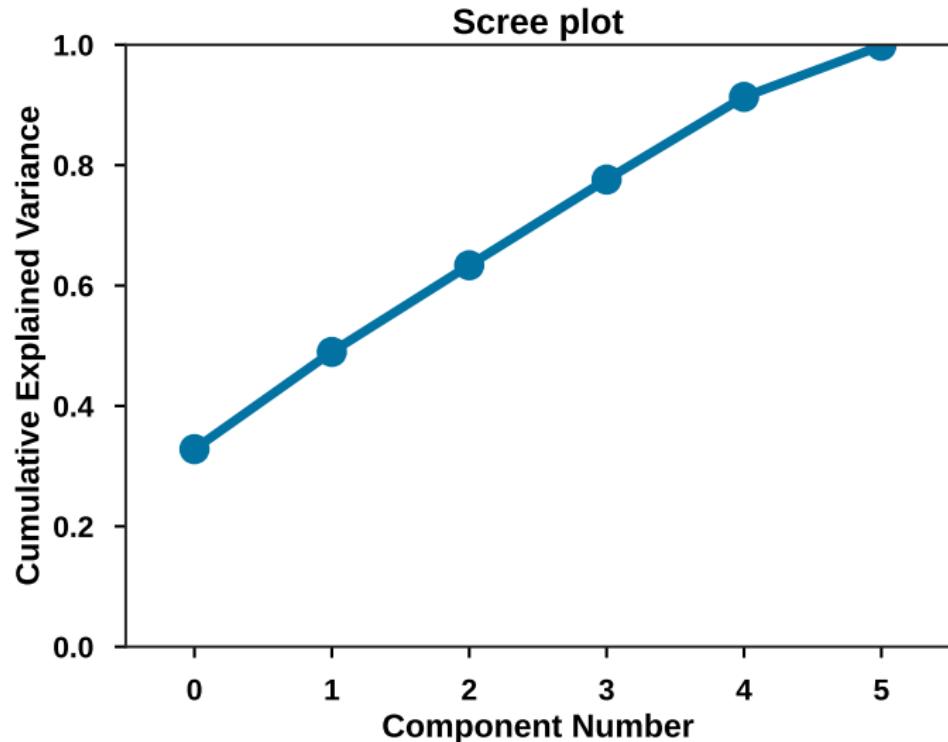


- 1 Selected features
- 2 Encoding, Scaling
- 3 PCA



	PC1	PC2	PC3	PC4	PC5	PC6
price_sum_z-score	63%	22%	7%	5%	6%	-20%
price_mean	62%	23%	7%	5%	6%	-24%
freight_value_sum	44%	-35%	-4%	-3%	-19%	80%
product_photos_qty_cat	7%	5%	-74%	-47%	48%	4%
order_purchase_timestamp_W_cat_Week	-1%	5%	55%	-83%	3%	2%
review_score_mean	-2%	43%	-34%	-24%	-80%	-2%
population_st	-15%	77%	15%	17%	29%	50%

PREPARATION DATA

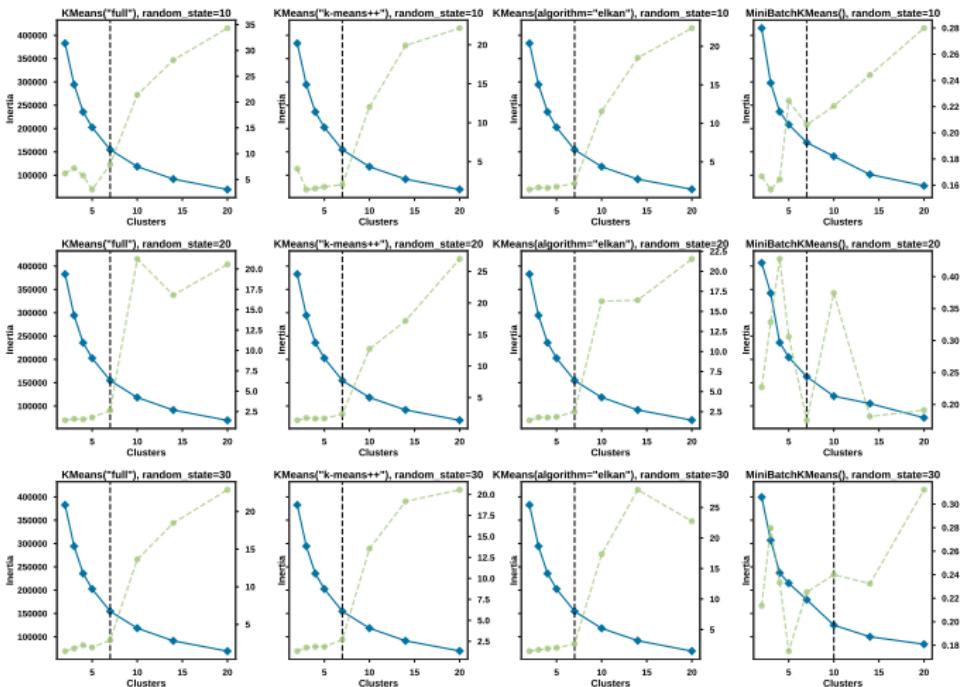


DF of 4 PCs used

KMEANS



- 1 N-Clusters
 - Inertia
 - Silhouette
- 2 LollipopChart
- 3 StackBar
- 4 Prix moyen
- 5 T-SNE
- 6 GaussMixture

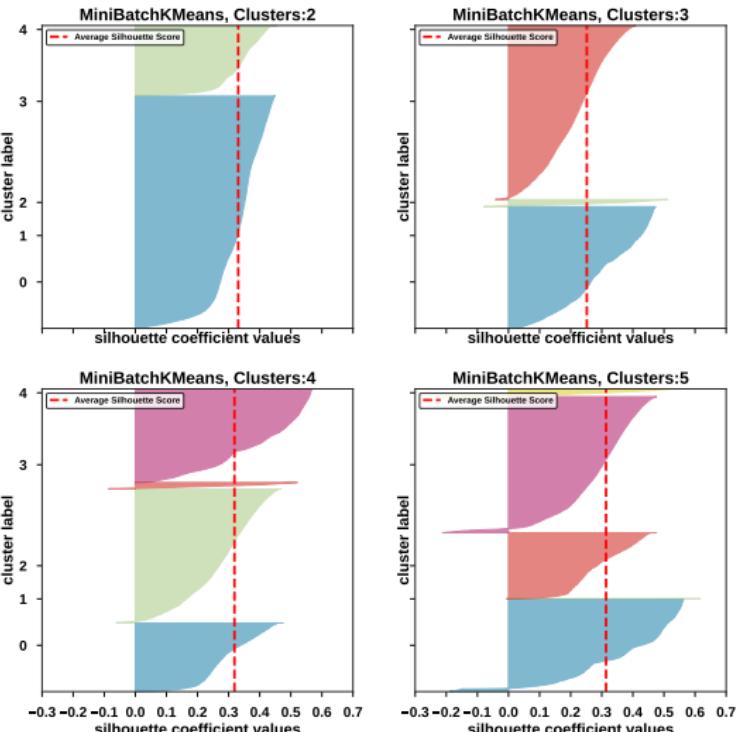


MiniBatchKMeans (init=10) : ↘inertia & ↑faster

KMEANS

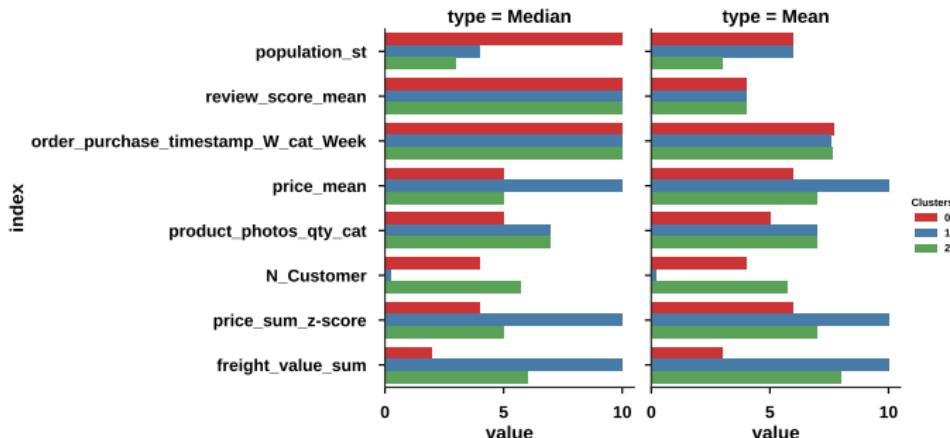
olist

- 1 N-Clusters
 - Inertia
 - Silhouette
- 2 LollipopChart
- 3 StackBar
- 4 Prix moyen
- 5 T-SNE
- 6 GaussMixture



Clusters: 3

KMEANS



1 N-Clusters

- Inertia
- Silhouette

2 LollipopChart

3 StackBar

4 Prix moyen

5 T-SNE

6 GaussMixture

EN GENERAL: Les 3 clusters représentent 90% des clients. Les clients viennent d'une zone rurale, ils achètent pendant la semaine et ils donnent la note maximale

CLUSTER 0 : Ils achètent des produits d'un prix moyen, leurs frais de livraison sont les plus faibles et ils viennent des états les plus peuplés

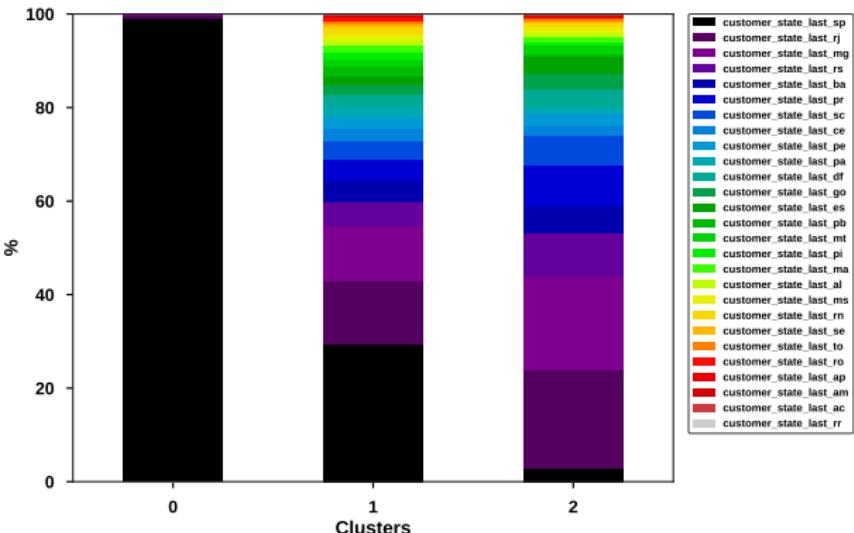
CLUSTER 1 : Ils achètent les produits le plus cher, leurs frais de livraison sont les plus élevés aussi et ils viennent des états moyennes

CLUSTER 2 : Ils achètent des produits d'un prix moyen, leurs frais de livraison sont dans la moyenne et ils viennent des états peu peuplés

KMEANS

olist

- 1 N-Clusters
 - Inertia
 - Silhouette
- 2 LollipopChart
- 3 StackBar
- 4 Prix moyen
- 5 T-SNE
- 6 GaussMixture

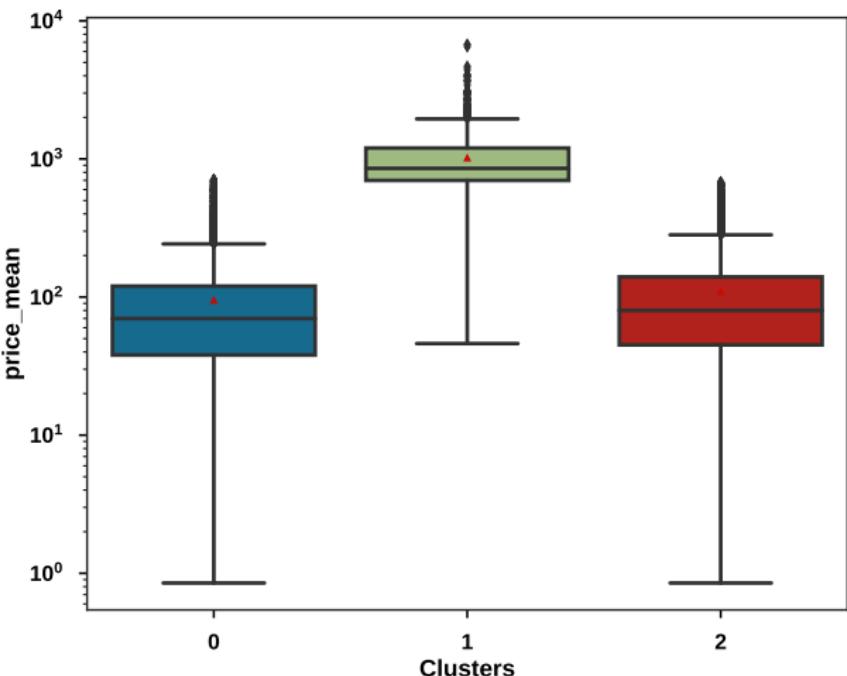


La plus parts des clients vient de l'état SP

KMEANS

olist

- 1 N-Clusters
 - Inertia
 - Silhouette
- 2 LollipopChart
- 3 StackBar
- 4 Prix moyen
- 5 T-SNE
- 6 GaussMixture

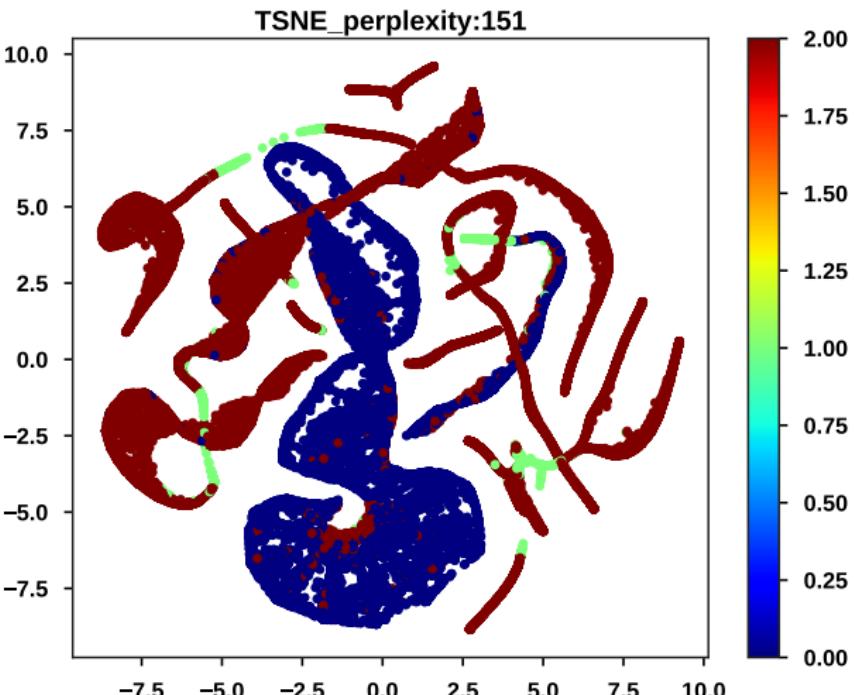


Le prix moyen du cluster est notamment plus élevé

KMEANS

olist

- 1 N-Clusters
 - Inertia
 - Silhouette
- 2 LollipopChart
- 3 StackBar
- 4 Prix moyen
- 5 T-SNE
- 6 GaussMixture



Les 3 clusters sont bien séparés

KMEANS



1 N-Clusters

- Inertia
- Silhouette

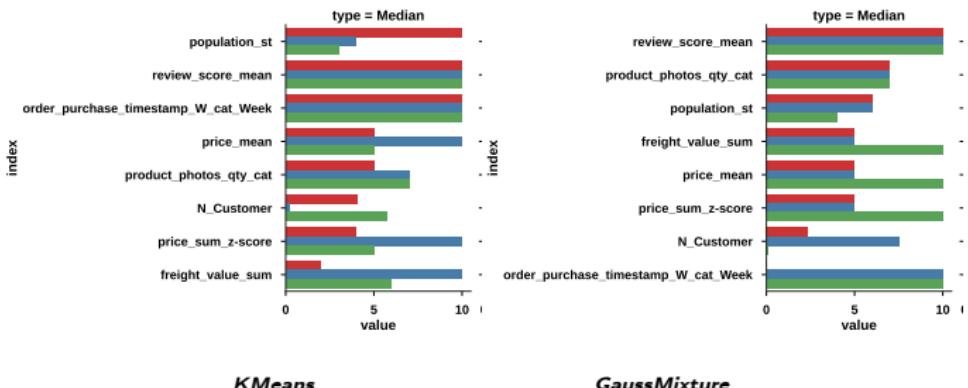
2 LollipopChart

3 StackBar

4 Prix moyen

5 T-SNE

6 GaussMixture



INDEX



- 1 Mission objective
- 2 Data preparation
 - Data cleansing
 - Combining data
- 3 Data analyses
- 4 Modeling
 - Analyse graphique
 - Test de corrélation
- 5 Conclusions et perspectives

CONCLUSIONS - GENERAL



1 Features selection:

CONCLUSIONS - GENERAL



1 Features selection:

- **4 QUANTITATIVE FEATURES** : 'freight_value_sum', 'population_st', 'price_mean',
'price_sum_z-score'

CONCLUSIONS - GENERAL



1 Features selection:

- **4 QUANTITATIVE FEATURES** : 'freight_value_sum', 'population_st', 'price_mean',
'price_sum_z-score'
- **3 QUALITATIVE FEATURES** : 'order_purchase_timestamp_W_cat',
'product_photos_qty_cat', 'review_score_mean'

CONCLUSIONS - GENERAL



1 Features selection:

- **4 QUANTITATIVE FEATURES** : 'freight_value_sum', 'population_st', 'price_mean',
'price_sum_z-score'
- **3 QUALITATIVE FEATURES** : 'order_purchase_timestamp_W_cat',
'product_photos_qty_cat', 'review_score_mean'

2 Models:

CONCLUSIONS - GENERAL



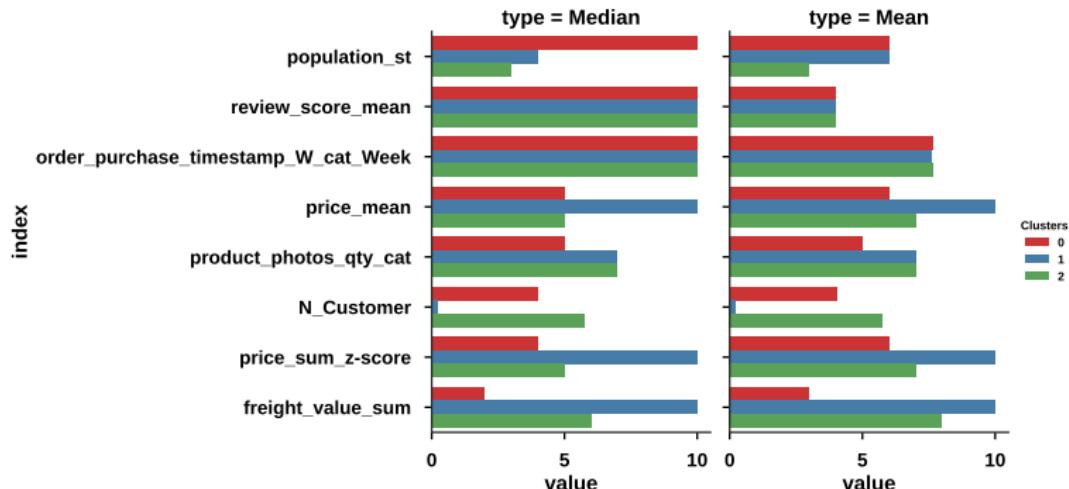
1 Features selection:

- **4 QUANTITATIVE FEATURES** : 'freight_value_sum', 'population_st', 'price_mean',
'price_sum_z-score'
- **3 QUALITATIVE FEATURES** : 'order_purchase_timestamp_W_cat',
'product_photos_qty_cat', 'review_score_mean'

2 Models:

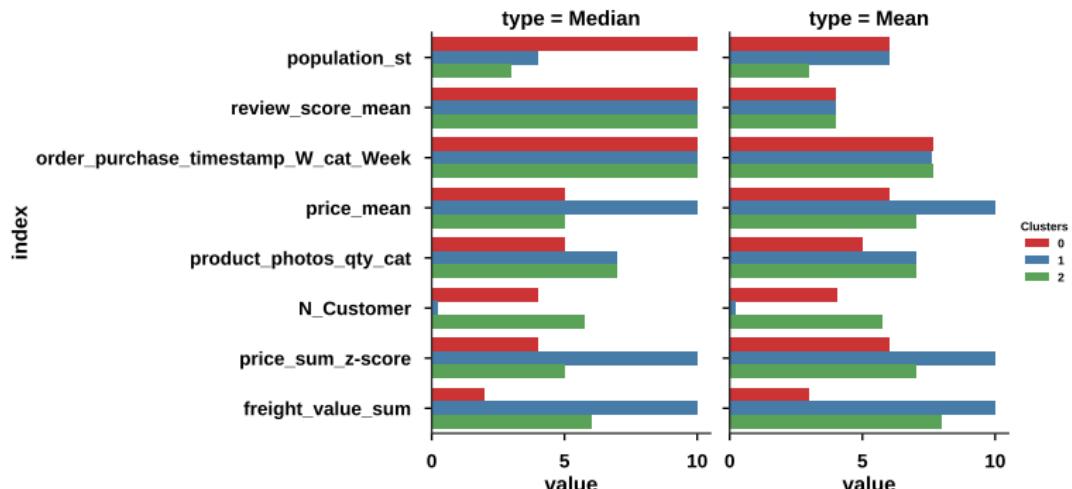
- KMeans classification is better than GaussMixture classification

CONCLUSIONS - CLASSIFICATION



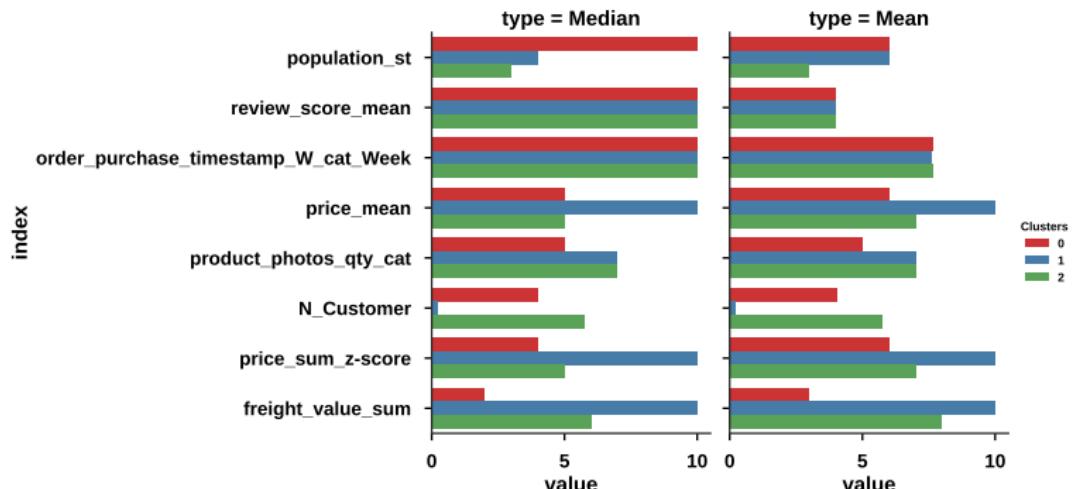
- EN GENERAL: Les 3 clusters représentent 90% des clients. Les clients viennent d'une zone rural et notamment de l'état SP. Ils achètent pendant la semaine et ils donnent la note maximal

CONCLUSIONS - CLASSIFICATION



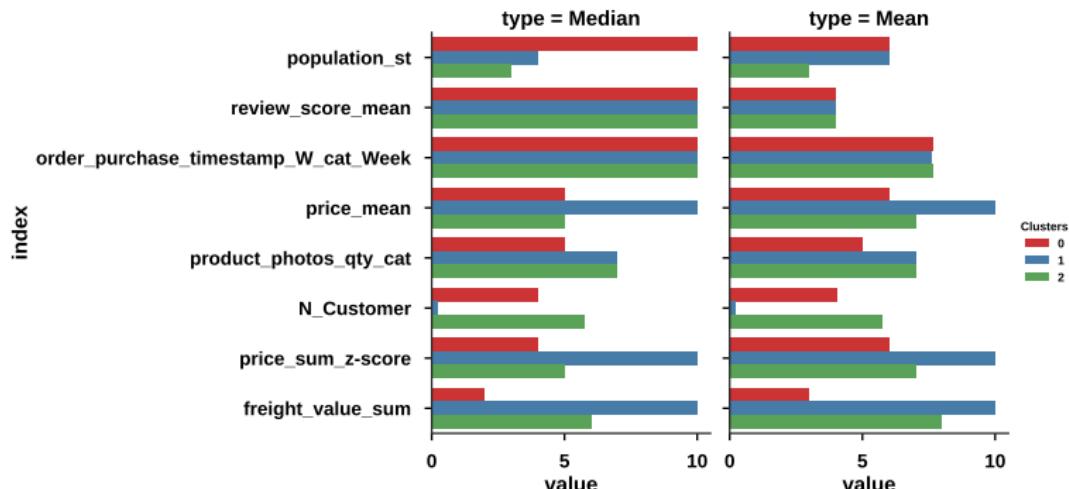
- EN GENERAL: Les 3 clusters représentent 90% des clients. Les clients viennent d'une zone rurale et notamment de l'état SP. Ils achètent pendant la semaine et ils donnent la note maximale
- CLUSTER 0 : prix moyen, frais de livraison ↓ et états ↑

CONCLUSIONS - CLASSIFICATION



- EN GENERAL: Les 3 clusters représentent 90% des clients. Les clients viennent d'une zone rural et notamment de l'état SP. Ils achètent pendant la semaine et ils donnent la note maximale
- CLUSTER 0 : prix moyenne, frais de livraison ↓ et états ↑
- CLUSTER 1 : prix ↑, frais de livraison ↑ et états moyennes

CONCLUSIONS - CLASSIFICATION

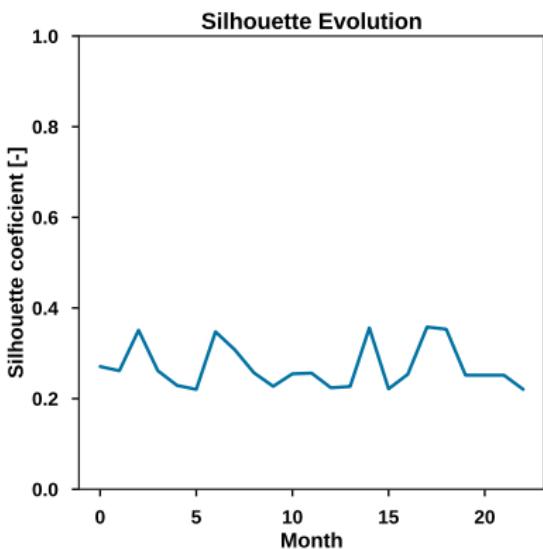


- EN GENERAL: Les 3 clusters représentent 90% des clients. Les clients viennent d'une zone rurale et notamment de l'état SP. Ils achètent pendant la semaine et ils donnent la note maximale
- CLUSTER 0 : prix moyenne, frais de livraison ↓ et états ↑
- CLUSTER 1 : prix ↑, frais de livraison ↑ et états moyennes
- CLUSTER 2 : prix moyenne, frais de livraison moyenne et états ↘

PERSPECTIVES - MAINTENANCE



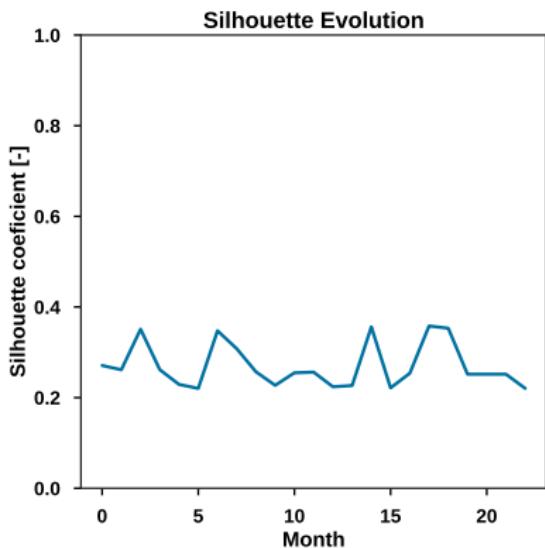
1 New features



PERSPECTIVES - MAINTENANCE



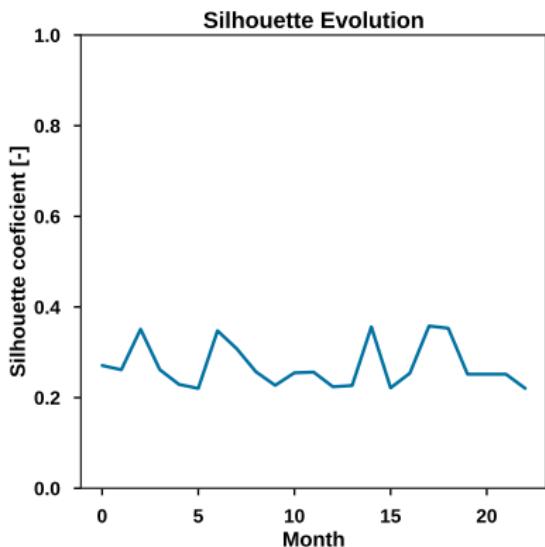
- 1 New features
- 2 Before/after a peak of purchases



PERSPECTIVES - MAINTENANCE



- 1 New features
- 2 Before/after a peak of purchases
- 3 Measure of stability (Silhouette)



PERSPECTIVES - MAINTENANCE



- 1 New features
- 2 Before/after a peak of purchases
- 3 Measure of stability (Silhouette)
 - If the silhouette drop below 0.2:

