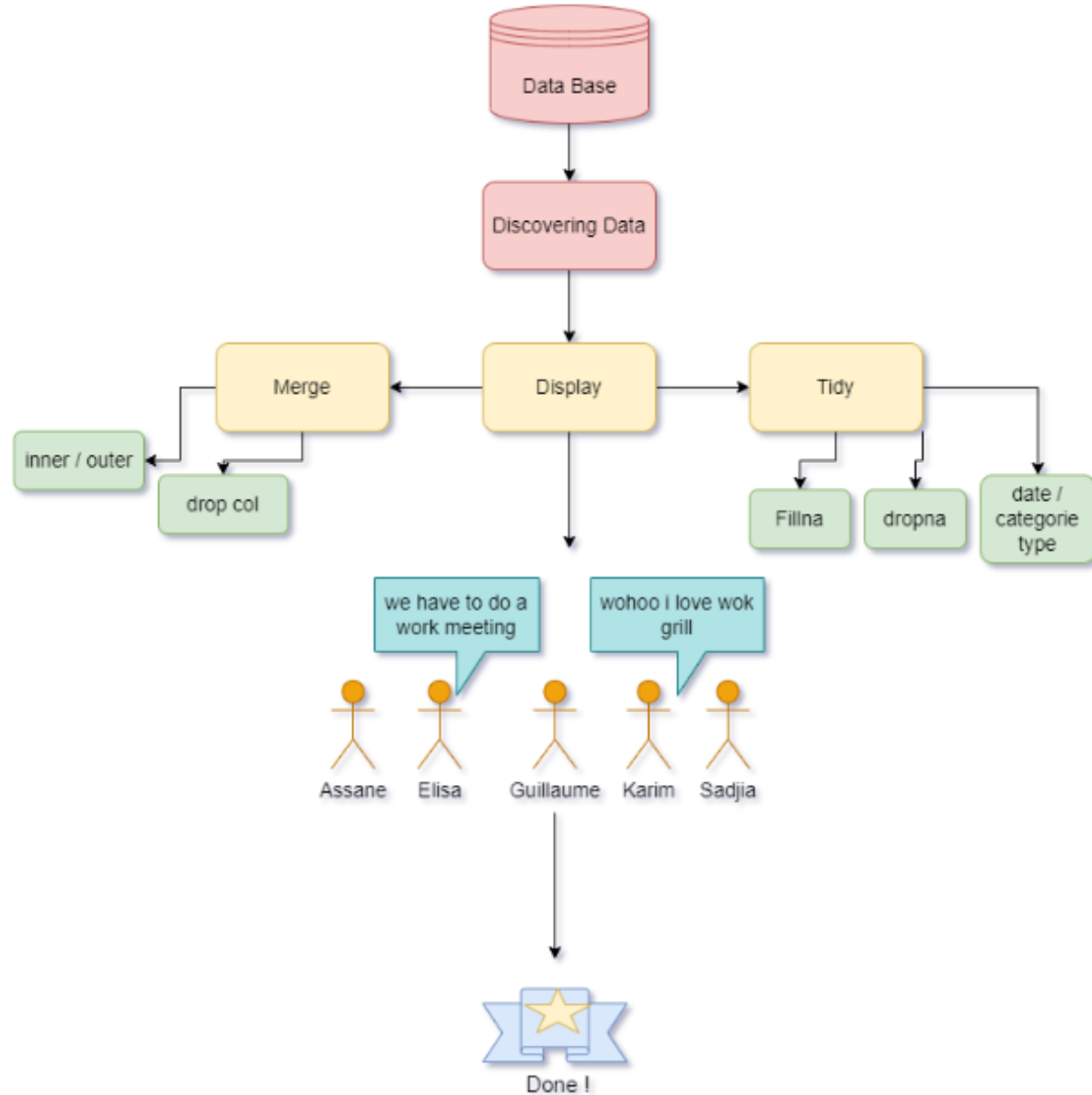


Analyse des données pour l'entreprise Olist Store

Olist project

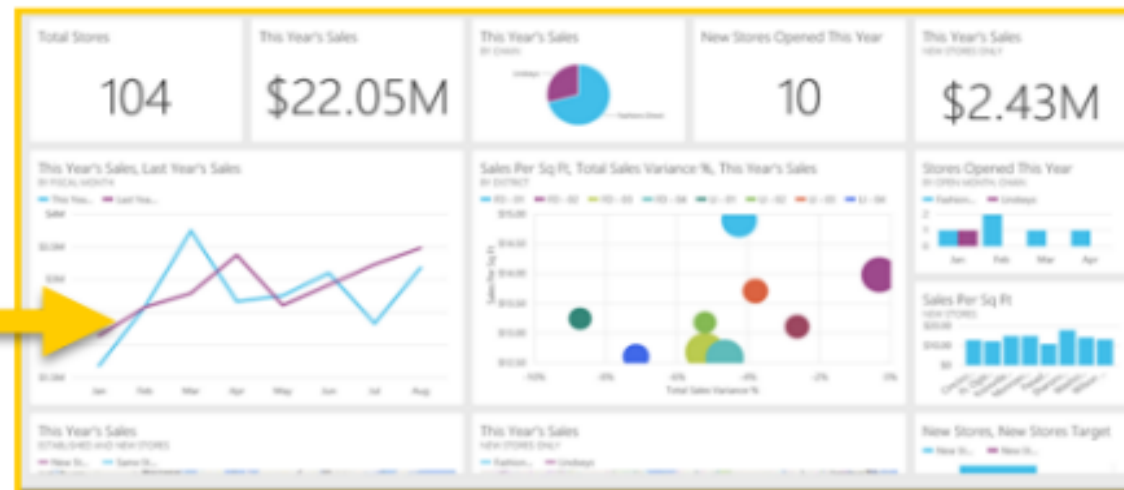


étape de notre analyse :

- Olist Store en chiffre.
- Revendeurs , clients , leur chiffre d'affaire et leur localisation.
- délai et retard dû au transport et sa corrélation avec la note
- catégories optimisation et conséquence sur la note.
- articles non livrés, corrélation avec le poids

Projet olist

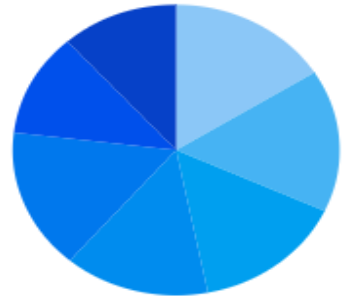
Chapitre I: Dashboard



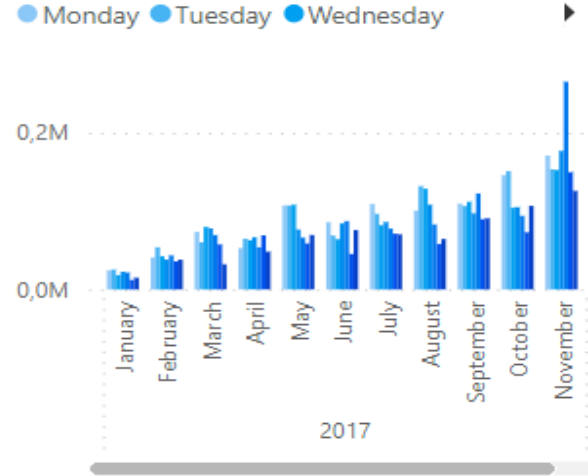
Olist en chiffres

chiffre d'affaire par année, mois et jour:

Monday, Tuesday, Wednesday,
Thursday, Friday, Saturday et Sunday

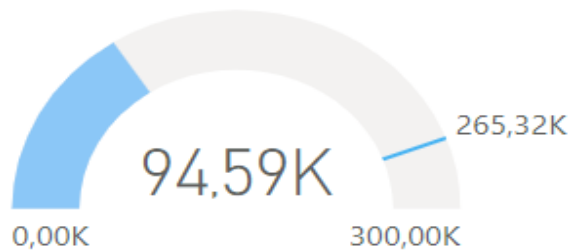


Monday
Tuesday
Wednesd...
Thursday
Friday
Saturday
Sunday

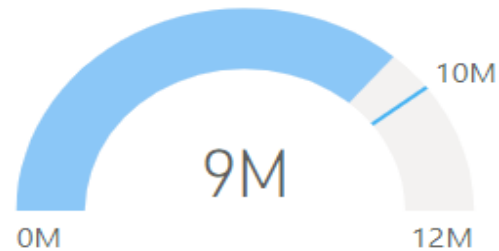


Objectif black Friday et clôture année 2018 :

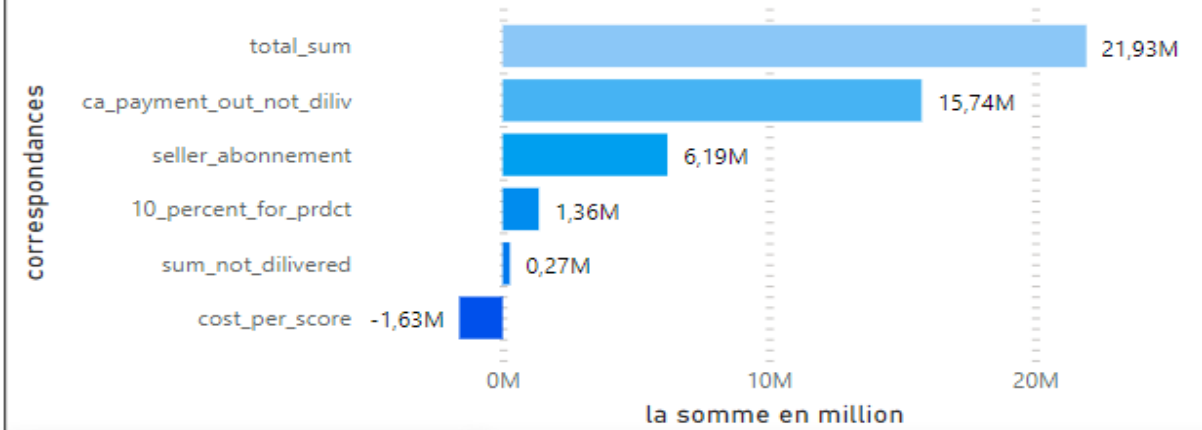
objectif black Friday



objectif fin d'année



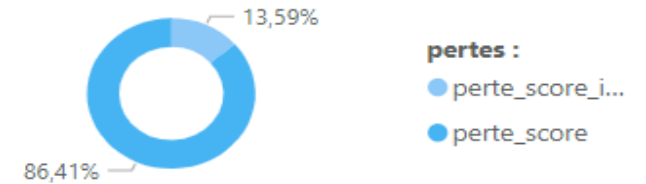
composition de chiffre d'affaire et les pertes associées



-1,63M

Somme de valeurs_pertes

pertes dues au score



les vendeurs non bénéfiques à Olist

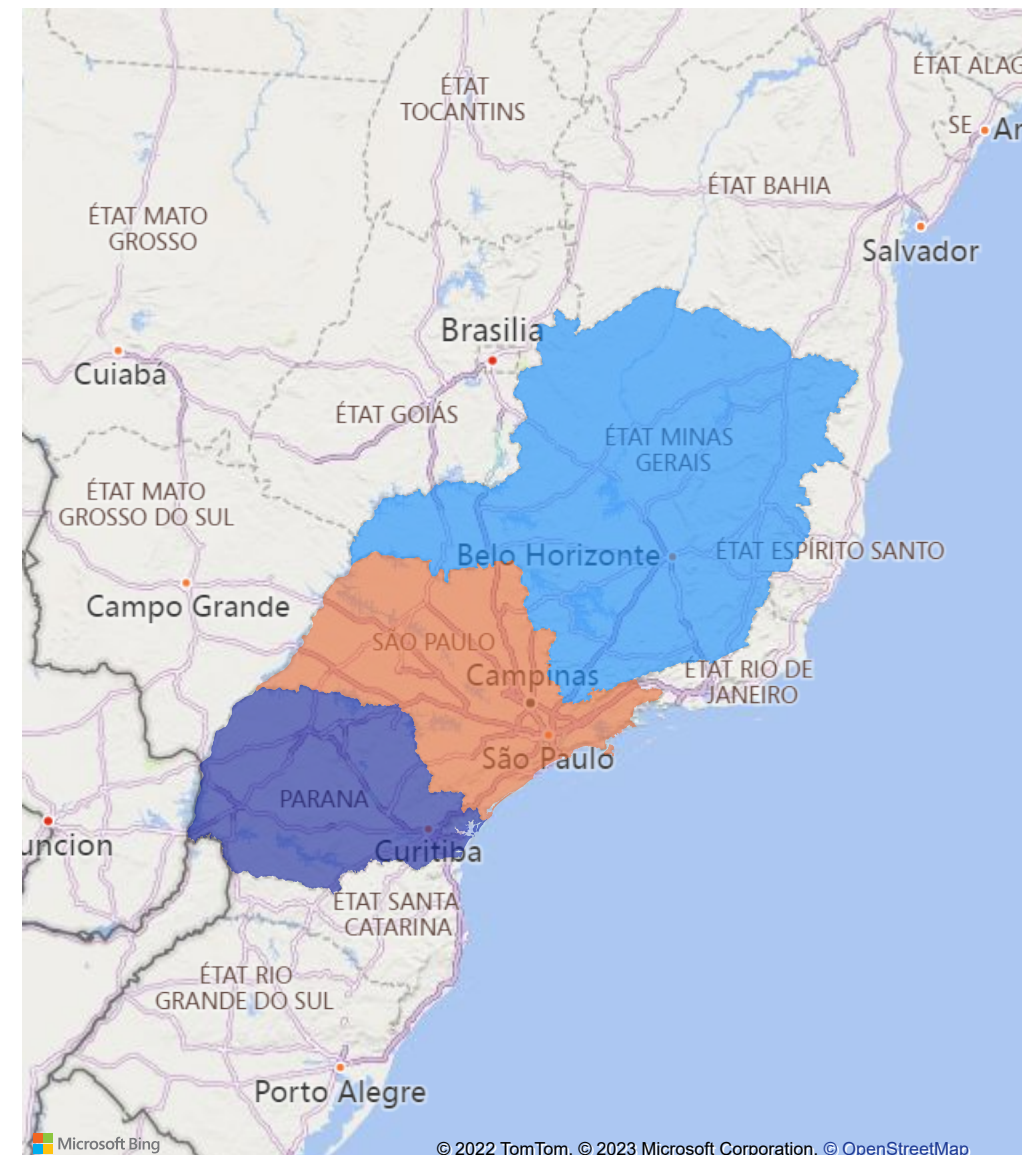
belo horizonte borda da mata ibitinga londrina



La répartition des vendeurs et clients

Sellers_Customers_CA/States

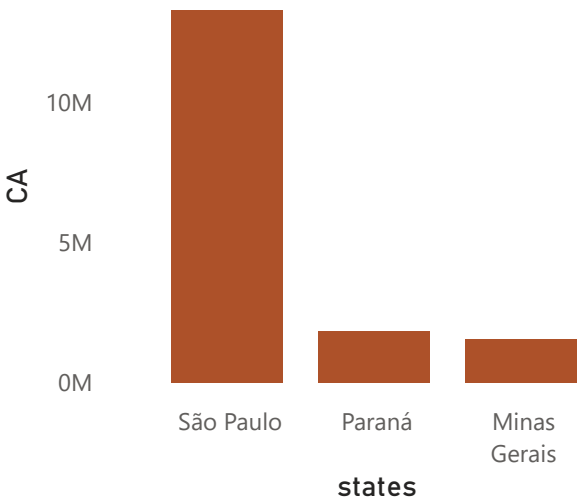
nb_customers ●5045 ●11635 ●41746



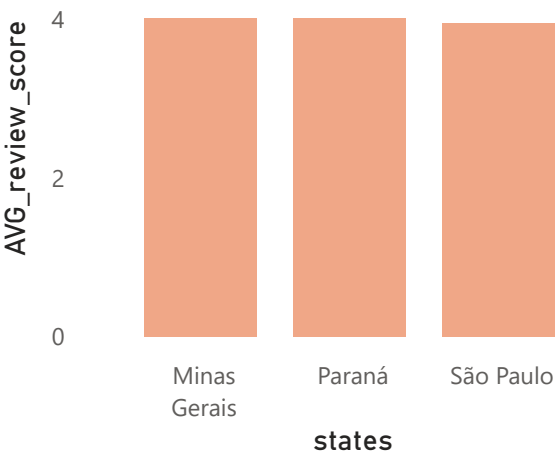
1228899

13286547

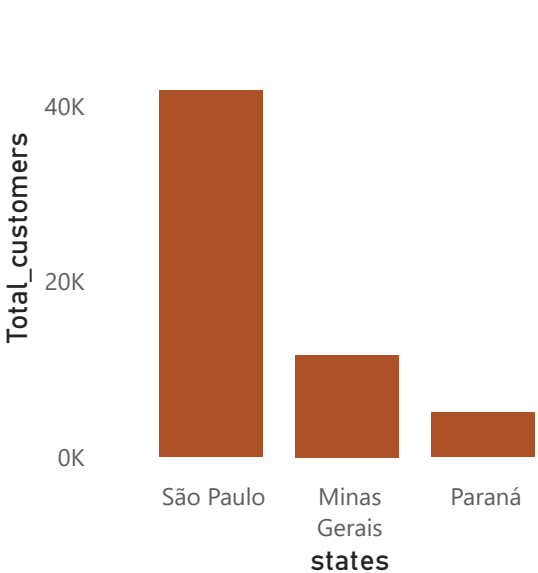
CA par states



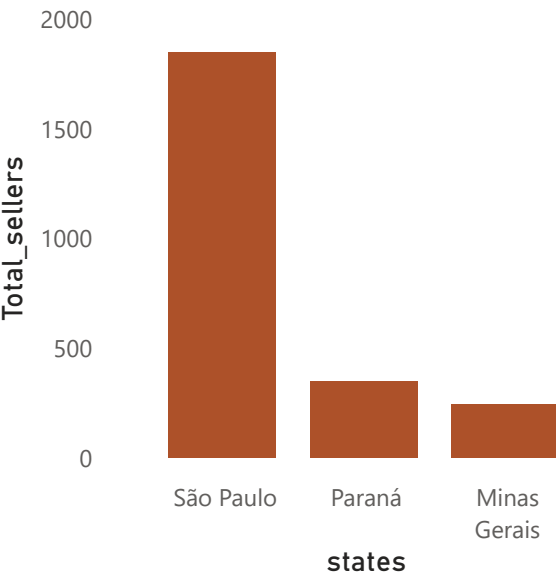
avg_review_score par states



Total_customers par states

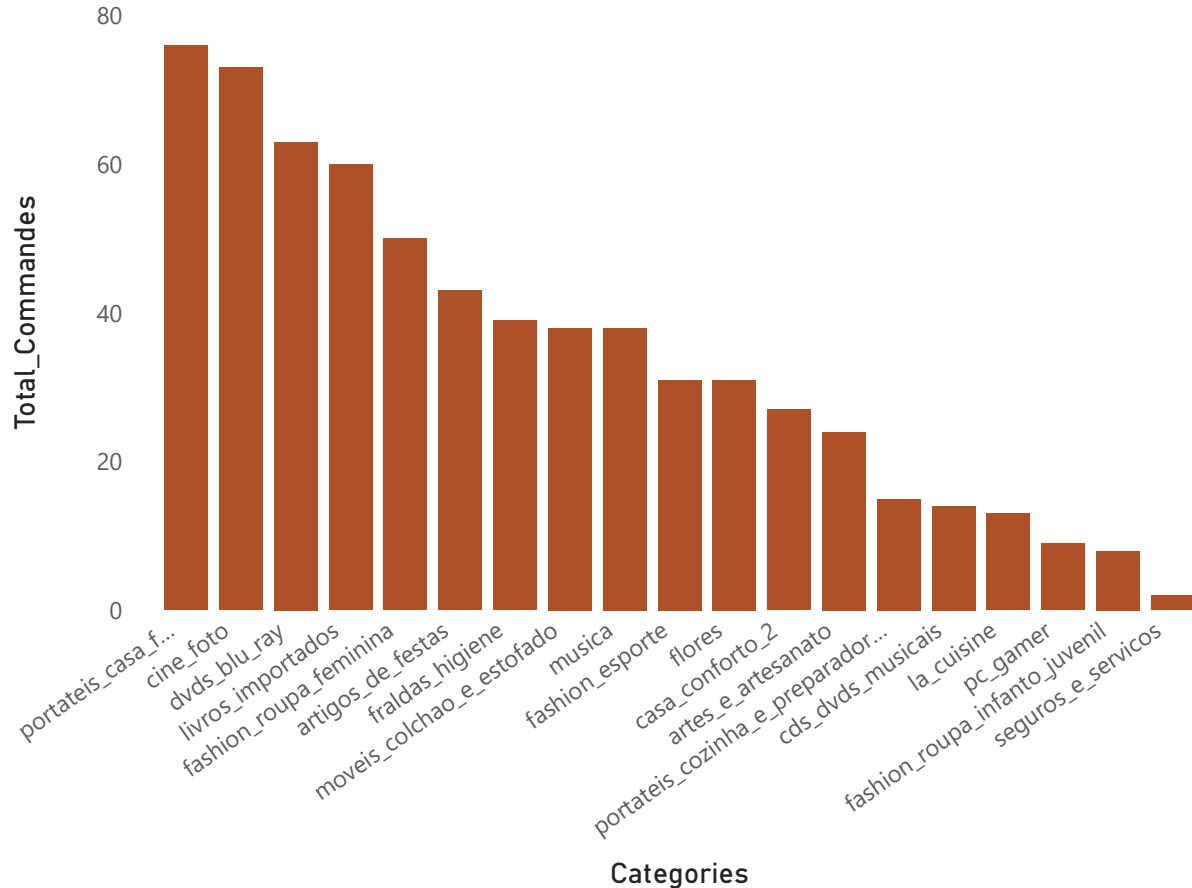


Total_sellers par states

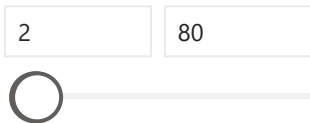
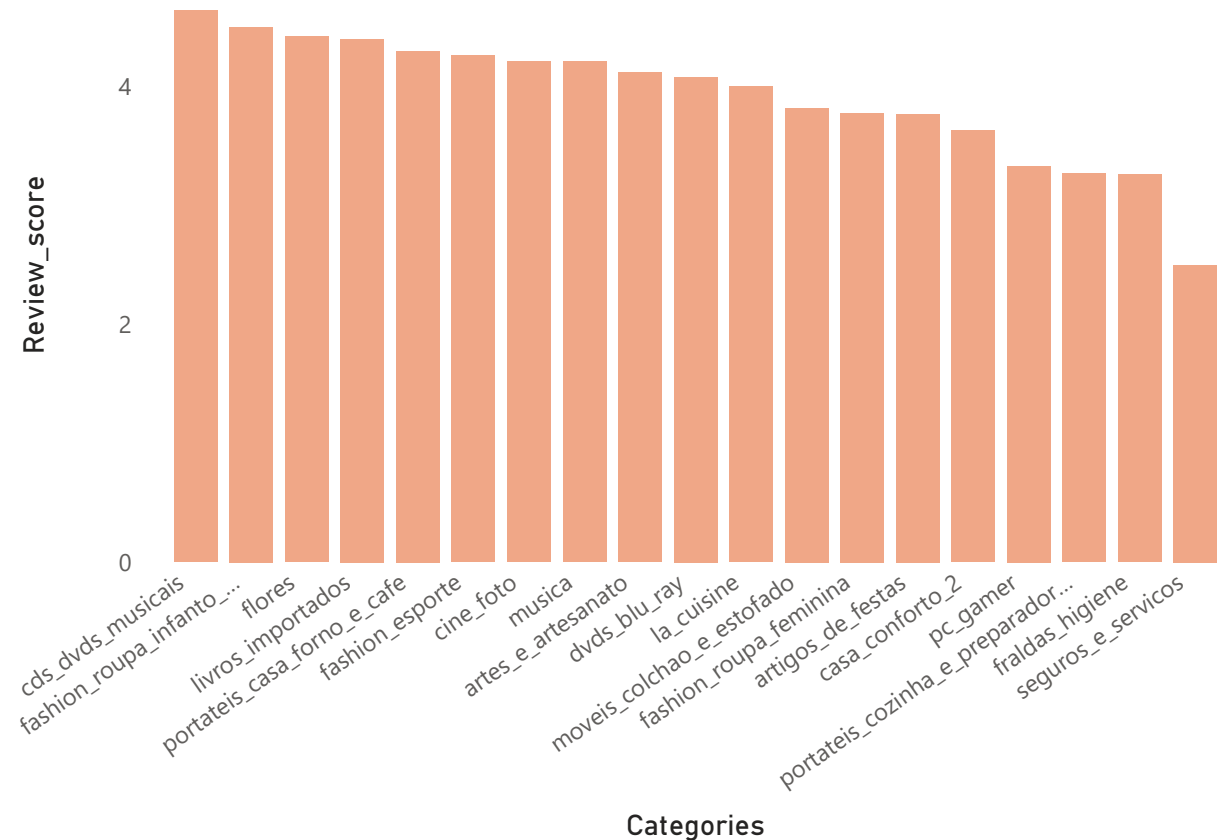


Le total des commandes par catégorie

TotalCommandes_categories



Moyenne_notes_category_name

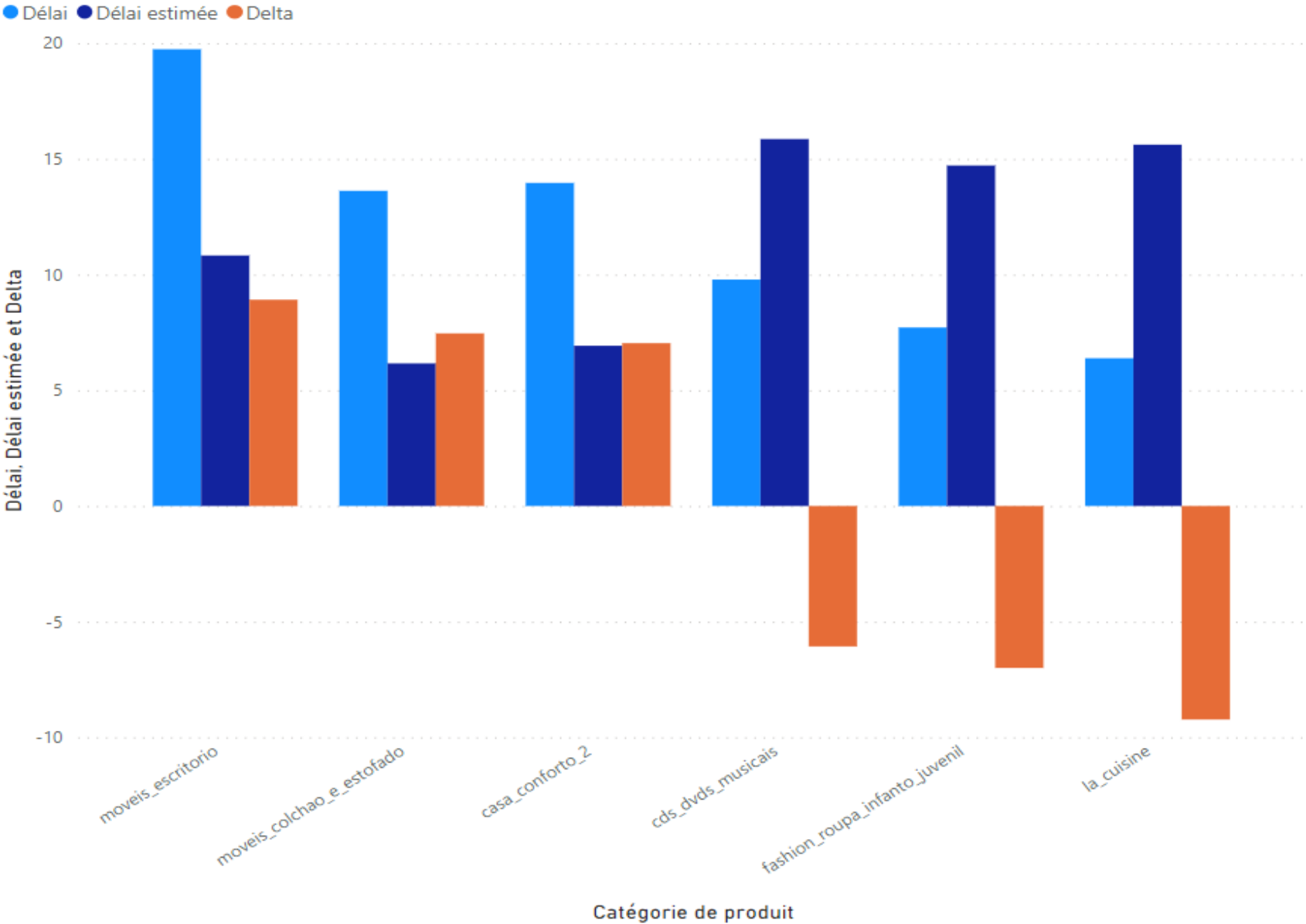


Recommendations:

Il faudrait regrouper les catégories et les renommer car on observe une dispersion des notes ce qui influe sur la moyenne de ces dernières.

Délai de livraison des produits

les 3 produits aux délais de livraison les plus surestimés, et les plus sous-estimés



Catégorie de produits

casa_conforto_2

cds_dvds_musicaux

fashion_roupa_infanto_juvenil

la_cuisine

moveis_colchao_e_estofado

moveis_escritorio

11,87

Délai de livraison moyen

11,68

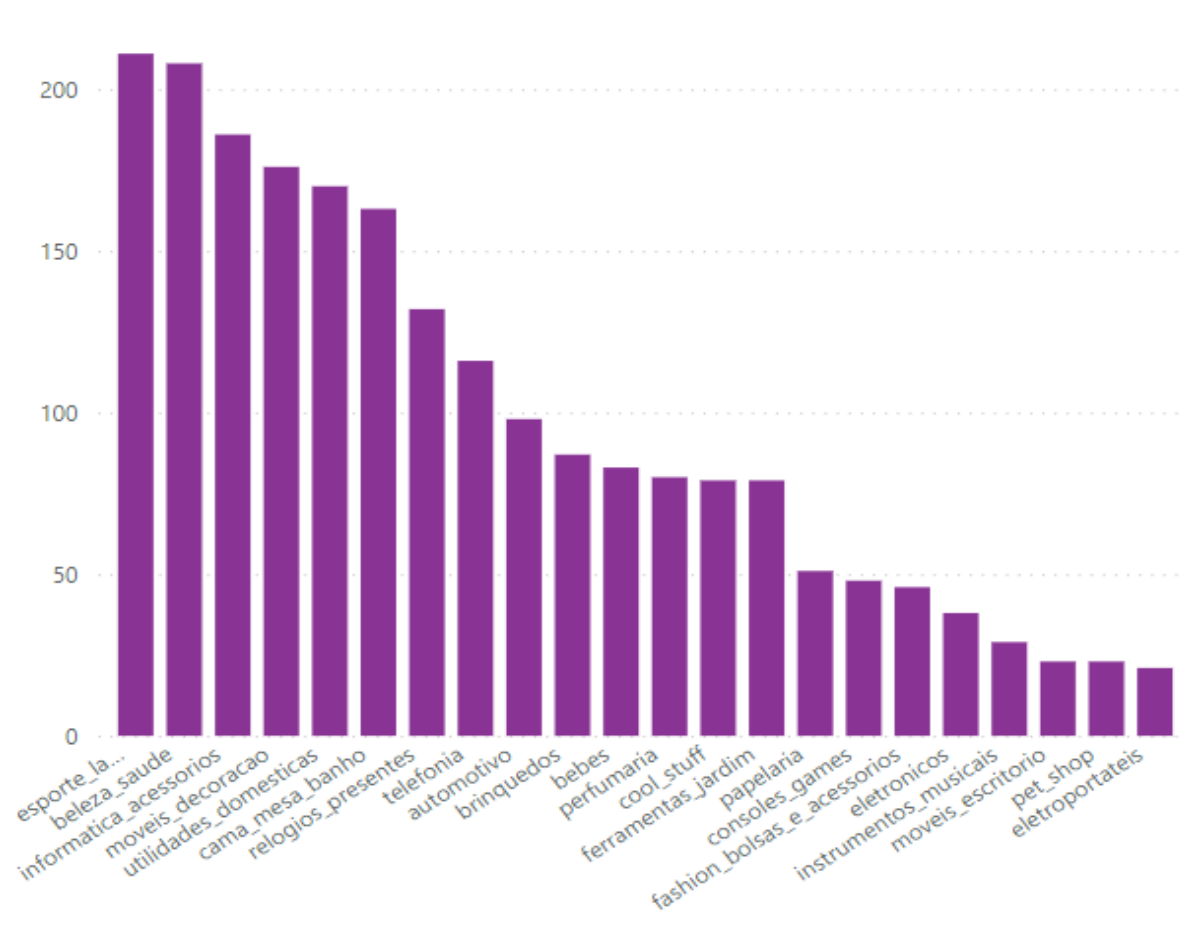
Délai de livraison estimée

0,18

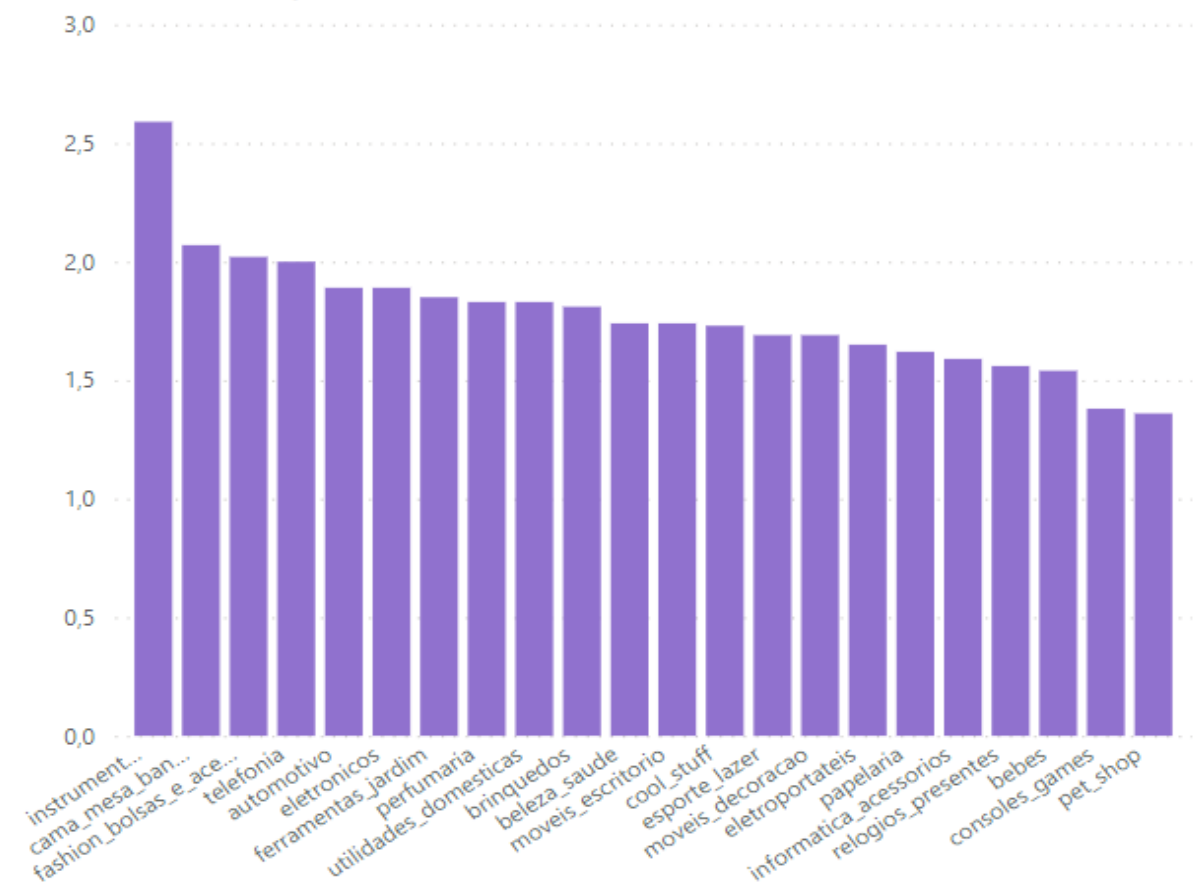
Delta

Les commandes non livrées

Total des commandes non livrées



Moyenne des notes des commandes non livrées



Total des commandes

21

211



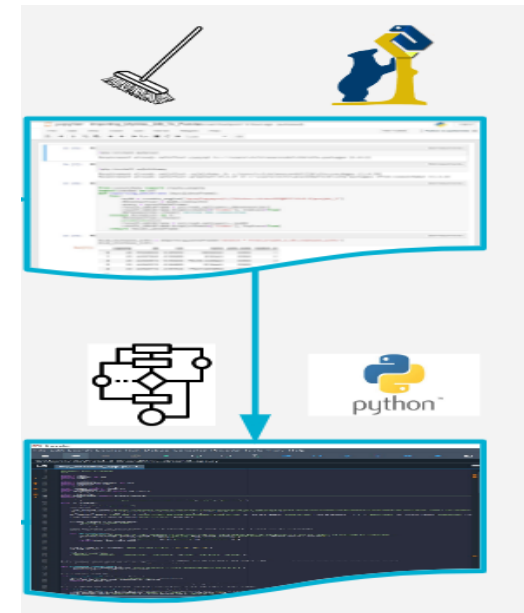
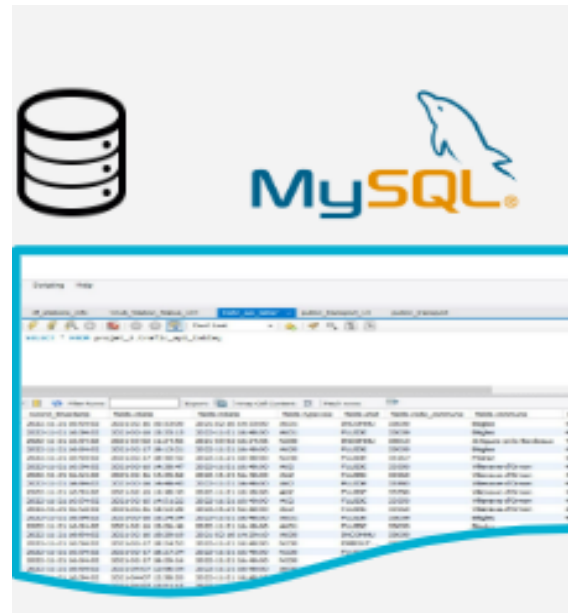
Bien entendu, lorsque les clients ne reçoivent pas leurs commandes, ils sont mécontents d'où les mauvaises notes. Pour y remédier nous proposons :

Préconisations :

- lorsqu'une commande dépasse le délai estimé de trois jours ouvrables pour les petits colis et cinq pour les plus gros , contacter les clients en proposant une offre commerciale (selon une grille tarifaire interne),

Projet Olist

Chapitre II: astuces techniques



Technique N°1: Pivot table

Tri des gains par années , mois , et jours bien ordonné

Trier les gains par années , mois , et jours

```
# calculer la somme des ventes par mois et par jour:
ca_per_year_month_day = orders_payments.pivot_table(index= ['year','month'],
                                                    columns='day',
                                                    values='payment_value',
                                                    fill_value=0,
                                                    aggfunc={'payment_value': 'sum'})

ca_per_year_month_day.reset_index(inplace=True)
ca_per_year_month_day
```

day	year	month	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
0	2016	December	19.62	0.00	0.00	0.00	0.00	0.00	0.00
1	2016	October	8255.19	5270.24	9778.12	4261.45	9398.61	11914.65	10212.22
2	2016	September	0.00	75.06	0.00	136.23	0.00	40.95	0.00
3	2017	April	53785.51	52746.14	69097.32	48107.15	66590.79	64641.29	62819.83
4	2017	August	82887.08	100577.77	57914.50	64352.82	108173.24	131931.68	128559.23
5	2017	December	151319.40	142201.20	114293.33	103845.17	103758.80	129369.90	133613.68
6	2017	February	43389.86	40712.16	35830.02	37896.68	37896.29	53852.27	42330.73

```
# ajouter les années, mois, et jours dans un tableau pour ordonner avec :
yearOrder= ['2016','2017','2018']
monthOrder = ['January', 'February', 'March', 'April','May','June','July','August',
              'September', 'October','November','December']
dayOrder = ['Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday']

ca_per_year_month_day = orders_payments.pivot_table(
    index=['year','month'],
    columns='day',
    values='payment_value',
    fill_value=0,
    aggfunc= {'payment_value': 'sum'}).loc[(yearOrder,monthOrder), dayOrder].round(0)

ca_per_year_month_day.reset_index(inplace=True)
ca_per_year_month_day
```

day	year	month	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	2016	September	75.0	41.0	0.0	0.0	0.0	0.0	136.0
1	2016	October	5270.0	11915.0	10212.0	9399.0	8255.0	9778.0	4261.0
2	2016	December	0.0	0.0	0.0	0.0	20.0	0.0	0.0
3	2017	January	24229.0	25281.0	18180.0	22542.0	21494.0	11960.0	14802.0
4	2017	February	40712.0	53852.0	42331.0	37896.0	43390.0	35830.0	37897.0
5	2017	March	73385.0	60174.0	79657.0	77977.0	69425.0	57232.0	32012.0
6	2017	April	52746.0	64641.0	62820.0	66591.0	53786.0	69097.0	48107.0
7	2017	May	107103.0	107151.0	108062.0	76388.0	66105.0	58601.0	69509.0

Technique N°2: Conversion des dates

```
[ ] df_4['délai'] = (df_4['order_delivered_customer_date'] - df_4['order_approved_at']).dt.days
```

```
df_4['délai_estimée'] = (df_4['order_estimated_delivery_date'] - df_4['order_delivered_customer_date']).dt.days
df_4['délai_estimée']
```

```
0      7.0
1      9.0
2      7.0
3      5.0
4      5.0
...
112367  7.0
112368  7.0
112369 -9.0
112370 10.0
112371 17.0
Name: délai_estimée, Length: 112372, dtype: float64
```

Technique N°3: Groupe by

```
# deuxième data frame
df_produits_orders_no_cat_2 = (df_produits_orders_no.groupby('product_category_name') \
                               .agg(total_commandes_non_livrees = ('order_id', 'count'), \
                                    moyenne_notes = ('review_score', 'mean')).round(2)).reset_index().sort_values(by='total_commandes_non_livrees', ascending = False)

#df_produits_orders_no_cat_2.to_csv('tableau3.csv')
df_produits_orders_no_cat_2
```

	product_category_name	total_commandes_non_livrees	moyenne_notes
28	esporte_lazer	211	1.69
10	beleza_saude	208	1.74
39	informatica_acessorios	186	1.59
48	moveis_decoracao	176	1.69

Technique N°4: scraping

```
[ ] brasil_state= pd.read_html('https://en.wikipedia.org/wiki/ISO_3166-2:BR')
brasil_state[0]['Code'] = brasil_state[0]['Code'].str.split('-').str[1]
state_df = brasil_state[0]
state_df.columns = ['abrv', 'states', 'category']
state_df=state_df.sort_values(by='abrv')
state_df=state_df.loc[:,['abrv', 'states']]
state_df
```

	abrv	states
0	AC	Acre
1	AL	Alagoas
3	AM	Amazonas
2	AP	Amapá
4	BA	Bahia
5	CE	Ceará
6	DF	Distrito Federal
7	ES	Espírito Santo

```
[ ] a['delta'] = a['délai']-a['délai_estimée']
```

```
[ ] a = a.sort_values(by = 'delta', ascending = False )
```

```
[ ] a_top3 = a.head()
```

```
[ ] a_last3 = a.tail()
```

```
[ ] a_top3
```

	délai	délai_estimée	delta
product_category_name			
moveis_escritorio	19.739784	10.825721	8.914062
moveis_colchao_e_estofado	13.621622	6.162162	7.459459
casa_conforto_2	13.962963	6.925926	7.037037
casa_conforto	12.567442	8.858140	3.709302
audio	12.389972	9.091922	3.298050

Technique N°5: Merge

```
[ ] sellers_dataset_sql=pd.read_sql("select * from sellers_dataset", conn)
sellers_dataset=sellers_dataset_sql.sort_values(by='seller_state')
sellers_dataset=sellers_dataset.loc[:,['seller_id','seller_state']]
sellers_dataset=pd.merge(sellers_dataset,state_df, how='left', left_on='seller_state', right_on='abrv')
sellers_dataset=sellers_dataset.groupby(['states']) \
.agg(nb_sellers = ('seller_id', 'count')).reset_index()
sellers_dataset
```

	states	nb_sellers
0	Acre	1
1	Amazonas	1
2	Bahia	19
3	Ceará	13
4	Distrito Federal	30
5	Espírito Santo	23
6	Goiás	40
7	Maranhão	1

Merci pour votre attention