

```
[1]: import pandas as pd
import numpy as np
import sqlite3 as sql
import matplotlib.pyplot as plt
from ydata_profiling import ProfileReport
import seaborn as sns

[2]: olist_customers_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_customers_dataset.csv")
olist_geolocation_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_geolocation_dataset.csv")
olist_order_items_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_order_items_dataset.csv")
olist_order_payments_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_order_payments_dataset.csv")
olist_order_reviews_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_order_reviews_dataset.csv")
olist_orders_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_orders_dataset.csv")
olist_products_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_products_dataset.csv")
olist_sellers_dataset = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_sellers_dataset.csv")
product_category_name_translation = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\product_category_name_translation.csv")

[3]: db = sql.connect("BEPO.db")

[7]: print("Number of rows in olist_customers_dataset: ", olist_customers_dataset.to_sql("olist_customers_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in olist_geolocation_dataset: ", olist_geolocation_dataset.to_sql("olist_geolocation_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in olist_order_items_dataset: ", olist_order_items_dataset.to_sql("olist_order_items_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in olist_order_payments_dataset: ", olist_order_payments_dataset.to_sql("olist_order_payments_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in olist_order_reviews_dataset: ", olist_order_reviews_dataset.to_sql("olist_order_reviews_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in olist_orders_dataset: ", olist_orders_dataset.to_sql("olist_orders_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in olist_products_dataset: ", olist_products_dataset.to_sql("olist_products_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in olist_sellers_dataset: ", olist_sellers_dataset.to_sql("olist_sellers_dataset", db, if_exists = 'replace', index = 'False'))
print("Number of rows in product_category_name_translation: ", product_category_name_translation.to_sql("product_category_name_translation", db, if_exists = 'replace', index = 'False'))

[9]: q1 = """
select * from olist_customers_dataset
"""
pd.read_sql(q1,db)

[9]:
```

	False	customer_id	customer_unique_id	customer_zip_code_prefix	customer_city	customer_state
0	0	06b8999e2fba1a1fbcc8172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0	14409	franca	SP
1	1	18955e83d337fd6b2def6b18a428ac77	290c77bc529b7ac935b93aa66c33dc3	9790	sao bernardo do campo	SP
2	2	4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e	1151	sao paulo	SP
3	3	b2b6027bc5c5109e529d4dc6358b12c3	259dac757896d24d7702b9acbbff3f3c	8775	mogi das cruzes	SP
4	4	4f2d8ab171c80ec8364f7c12e35b23ad	345ecd01c38d18a9036ed96c73b8d066	13056	campinas	SP
...
99436	99436	17ddf5dd5d51696bb3d7c6291687be6f	1a29b476fee25c95fbafc67c5ac95cf8	3937	sao paulo	SP
99437	99437	e7b71a9017aa05c9a7fd292d714858e8	d52a67c98be1cf6a5c84435bd38d095d	6764	taboao da serra	SP
99438	99438	5e28dfe12db7fb50a4b2f691faece5e	e9f50caf99f032f0bf3c55141f019d99	60115	fortaleza	CE
99439	99439	56b18e2166679b8a959d72dd06da27f9	73c2643a0a458b49f58cea58833b192e	92120	canaoas	RS
99440	99440	274fa6071e5e17fe303b9748641082c8	84732c5050c01db9b23e19ba39899398	6703	cotia	SP

99441 rows × 6 columns

```
[11]: olist_customers_dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99441 entries, 0 to 99440
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   customer_id      99441 non-null   object 
 1   customer_unique_id 99441 non-null   object 
 2   customer_zip_code_prefix 99441 non-null   int64  
 3   customer_city     99441 non-null   object 
 4   customer_state    99441 non-null   object 
```

```

dtypes: int64(1), object(4)
memory usage: 3.8+ MB

[13]: olist_customers_dataset.describe().round(2)

[13]:   customer_zip_code_prefix
      count    99441.00
      mean     35137.47
      std      29797.94
      min      1003.00
      25%     11347.00
      50%     24416.00
      75%     58900.00
      max     99990.00

[209]: olist_customers_dataset.nunique()

[209]:   customer_id          99441
       customer_unique_id    96096
       customer_zip_code_prefix 14994
       customer_city           4119
       customer_state            27
       dtype: int64

[15]: olist_customers_dataset['customer_city'].nunique()

[15]: 4119

[17]: olist_customers_dataset['customer_state'].nunique()

[17]: 27

[19]: olist_customers_dataset['customer_state'].unique()

[19]: array(['SP', 'SC', 'MG', 'PR', 'RJ', 'RS', 'PA', 'GO', 'ES', 'BA', 'MA',
       'MS', 'CE', 'DF', 'RN', 'PE', 'MT', 'AM', 'AP', 'AL', 'RO', 'PB',
       'TO', 'PI', 'AC', 'SE', 'RR'], dtype=object)

[21]: olist_customers_dataset.isna().sum()

[21]:   customer_id      0
       customer_unique_id 0
       customer_zip_code_prefix 0
       customer_city        0
       customer_state        0
       dtype: int64

```

```
[23]: # Generate a profile report for the DataFrame df
profile = ProfileReport(olist_customers_dataset)
```

```
# Display the report in the notebook
profile
```

Error displaying widget

Error displaying widget

Error displaying widget

Pandas Profiling Report

Overview Variables Interactions Correlations Missing values Sample

Overview

Brought to you by YData

Overview

Alerts 3

Reproduction

Dataset statistics

Number of variables	5
Number of observations	99441
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%

Variable types

Text	3
Numeric	1
Categorical	1

Total size in memory	3.8 MiB
Average record size in memory	40.0 B

Variables

Select Columns ▾

[23]:

[25]: olist_customers_dataset.value_counts()

```
[25]: customer_id          customer_unique_id      customer_zip_code_prefix  customer_city    customer_state
00012a2ce6f8dcda20d059ce98491703 248ffe10d632bebe4f7267f1f44844c9 6273           osasco         SP             1
aa62a16b10e3fb24ecdefd7e4dc3fa65 6526a724d515750e89a291c3d0176923 25010          duque de caxias  RJ             1
aa738cd72af39e38482fa25b992c67c 79f6bb4f638b4214ad8d62d60148a280 23055          rio de janeiro  RJ             1
aa7210736aad65a84e964cfa8d64aa66 ec2b8deff9fbfb59af09425a12597a3c 7936           francisco morato  SP             1
aa72002446dfbac0691304465c1b982d 7302f7368abf7ec6cb2e16fb005c4785 13660          porto ferreira   SP             1
Name: count, Length: 99441, dtype: int64
```

[27]: olist_customers_dataset.shape

[27]: (99441, 5)

```
[29]: q1 = """
select * from olist_customers_dataset
"""
pd.read_sql(q1, db)
```

	False	customer_id	customer_unique_id	customer_zip_code_prefix	customer_city	customer_state
0	0	06b8999e2fba1a1fb88172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0		14409	franca SP
1	1	18955e83d337fd6b2def6b18a428ac77	290c77bc529b7ac935b93aa66c333dc3		9790	sao bernardo do campo SP
2	2	4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e		1151	sao paulo SP
3	3	b2b6027bc5c5109e529d4dc6358b12c3	259dac757896d24d7702b9acbbff3f3c		8775	mogi das cruzes SP
4	4	4f2d8ab171c80ec8364f7c12e35b23ad	345ecd01c38d18a9036ed96c73b8d066		13056	campinas SP
...
99436	99436	17ddf5dd5d51696bb3d7c6291687be6f	1a29b476fee25c95fbafc67c5ac95cf8		3937	sao paulo SP
99437	99437	e7b71a9017aa05c9a7fd292d714858e8	d52a67c98be1cf6a5c84435bd38d095d		6764	taboao da serra SP
99438	99438	5e28dfe12db7fb50a4b2f691faece5e	e9f50caf99f032f0bf3c55141f019d99		60115	fortaleza CE
99439	99439	56b18e2166679b8a959d72dd06da27f9	73c2643a0a458b49f58cea58833b192e		92120	canoas RS
99440	99440	274fa6071e5e17fe303b9748641082c8	84732c5050c01db9b23e19ba39899398		6703	cotia SP

99441 rows × 6 columns

[31]: olist_geolocation_dataset.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000163 entries, 0 to 1000162
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   geolocation_zip_code_prefix  1000163 non-null  int64  
 1   geolocation_lat            1000163 non-null  float64 
 2   geolocation_lng            1000163 non-null  float64 
 3   geolocation_city           1000163 non-null  object  
 4   geolocation_state          1000163 non-null  object  
dtypes: float64(2), int64(1), object(2)
memory usage: 38.2+ MB
```

[33]: olist_geolocation_dataset.describe().round(2)

	geolocation_zip_code_prefix	geolocation_lat	geolocation_lng
count	1000163.00	1000163.00	1000163.00
mean	36574.17	-21.18	-46.39
std	30549.34	5.72	4.27
min	1001.00	-36.61	-101.47

25%	11075.00	-23.60	-48.57
50%	26530.00	-22.92	-46.64
75%	63504.00	-19.98	-43.77
max	99990.00	45.07	121.11

```
[35]: olist_geolocation_dataset.isna().sum()
```

```
[35]: geolocation_zip_code_prefix      0
geolocation_lat                      0
geolocation_lng                      0
geolocation_city                     0
geolocation_state                    0
dtype: int64
```

```
[205]: olist_geolocation_dataset.nunique()
```

```
[205]: geolocation_zip_code_prefix      19015
geolocation_lat                      717360
geolocation_lng                      717613
geolocation_city                     8011
geolocation_state                    27
dtype: int64
```

```
[207]: olist_geolocation_dataset['geolocation_state'].value_counts()
```

```
[207]: geolocation_state
SP    404268
MG    126336
RJ    121169
RS    61851
PR    57859
SC    38328
BA    36045
GO    20139
ES    16748
PE    16432
DF    12986
MT    12031
CE    11674
PA    10853
MS    10431
MA    7853
PB    5538
RN    5041
PI    4549
AL    4183
TO    3576
SE    3563
RO    3478
AM    2432
AC    1301
AP    853
RR    646
Name: count, dtype: int64
```

```
[37]: olist_geolocation_dataset.head()
```

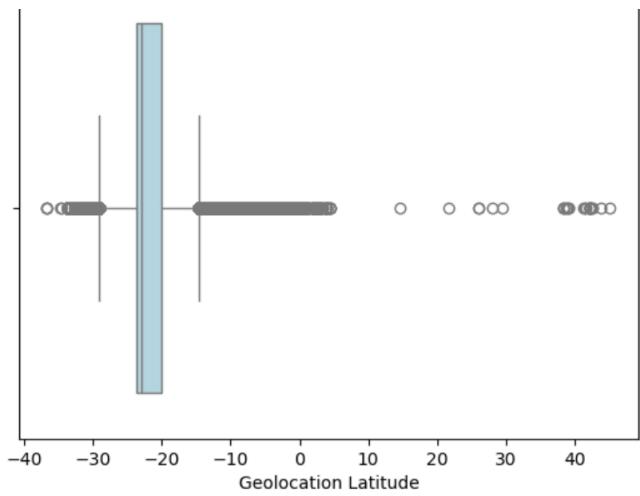
	geolocation_zip_code_prefix	geolocation_lat	geolocation_lng	geolocation_city	geolocation_state
0	1037	-23.545621	-46.639292	sao paulo	SP
1	1046	-23.546081	-46.644820	sao paulo	SP
2	1046	-23.546129	-46.642951	sao paulo	SP
3	1041	-23.544392	-46.639499	sao paulo	SP
4	1035	-23.541578	-46.641607	sao paulo	SP

```
[39]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# لتفعيل عرض الرسوم داخل Jupyter
%matplotlib inline
# قراءة البيانات (نأخذ من المسار المحدد)
data = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\Brazilian E-Commerce Sales.csv")

# إنشاء Box Plot باستخدام Seaborn
sns.boxplot(x=data["geolocation_lat"], color="lightblue")

# إضافة عنوان ونطاق
plt.title("Geolocation Sales Box Plot")
plt.xlabel("Geolocation Latitude")
plt.show()
```

Geolocation Sales Box Plot



```
[41]: # Generate a profile report for the DataFrame df
profile = ProfileReport(olist_geolocation_dataset)

# Display the report in the notebook
profile
```

Error displaying widget
Error displaying widget
Error displaying widget

Pandas Profiling Report

Overview Variables Interactions Correlations Missing values Sample Duplicate rows

Overview

Brought to you by [YData](#)

[Overview](#) [Alerts 4](#) [Reproduction](#)

Dataset statistics

Number of variables	5
Number of observations	1000163
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	128174
Duplicate rows (%)	12.8%
Total size in memory	38.2 MiB
Average record size in memory	40.0 B

Variable types

Numeric	3
Text	1
Categorical	1

Variables

Select Columns ▾

[41]:

```
[43]: olist_geolocation_dataset.shape
```

```
[43]: (1000163, 5)
```

```
[45]: q1 = """
select * from olist_order_items_dataset
"""
pd.read_sql(q1,db)
```

[45]:

False

order_id order_item_id

product_id

seller_id shipping_limit_date price ...

		order_id	order_item_id	product_id	seller_id	shipping_limit_date	price
0	0	00010242fe8c5a6d1ba2dd792cb16214	1	4244733e06e7ecb4970a6e2683c13e61	48436dade18ac8b2bce089ec2a041202	2017-09-19 09:45:35	58.90
1	1	00018f77f2f0320c557190d7a144bdd3	1	e5f2d52b802189ee658865ca93d83a8f	dd7ddc04e1b6c2c614352b383efe2d36	2017-05-03 11:05:13	239.90
2	2	000229ec398224ef6ca0657da4fc703e	1	c777355d18b72b67abbeef9df44fd0fd	5b51032eddd242adc84c38acab88f23d	2018-01-18 14:48:30	199.00
3	3	00024acbcdf0a6daa1e931b038114c75	1	7634da152a4610f1595efa32f14722fc	9d7a1d34a5052409006425275ba1c2b4	2018-08-15 10:10:18	12.99
4	4	00042b26cf59d7ce69dfabb4e55b4fd9	1	ac6c3623068f30de03045865e4e10089	df560393f3a51e74553ab94004ba5c87	2017-02-13 13:57:51	199.90
...
112645	112645	fffc94f6ce00a00581880bf54a75a037	1	4aa6014eceb682077f9dc4bffebc05b0	b8bc237ba3788b23da09c0f1f3a3288c	2018-05-02 04:11:01	299.99
112646	112646	ffffcd46ef2263f404302a634eb57f7eb	1	32e07fd915822b0765e448c4dd74c828	f3c38ab652836d21de61fb8314b69182	2018-07-20 04:31:48	350.00
112647	112647	fffe4705a9662cd70adb13d4a31832d	1	72a30483855e2eafc67aaee5dc2560482	c3cfdc648177fdbbbb35635a37472c53	2017-10-30 17:14:25	99.90
112648	112648	ffe18544ffabc95dfada21779c9644f	1	9c422a519119dcad7575db5af1ba540e	2b3e4a2a3ea8e01938cabda2a3e5cc79	2017-08-21 00:04:32	55.99
112649	112649	ffffe41c64501cc87c801fd61db3f6244	1	350688d9dc1e75ff97be326363655e01	f7ccf836d21b2fb1de37564105216cc1	2018-06-12 17:10:13	43.00

112650 rows × 8 columns

```
[47]: olist_order_items_dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 112650 entries, 0 to 112649
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   order_id         112650 non-null   object 
 1   order_item_id    112650 non-null   int64  
 2   product_id       112650 non-null   object 
 3   seller_id        112650 non-null   object 
 4   shipping_limit_date 112650 non-null   object 
 5   price            112650 non-null   float64
 6   freight_value    112650 non-null   float64
dtypes: float64(2), int64(1), object(4)
memory usage: 6.0+ MB
```

```
[49]: olist_order_items_dataset.describe().round(2)
```

	order_item_id	price	freight_value
count	112650.00	112650.00	112650.00
mean	1.20	120.65	19.99
std	0.71	183.63	15.81
min	1.00	0.85	0.00
25%	1.00	39.90	13.08
50%	1.00	74.99	16.26
75%	1.00	134.90	21.15
max	21.00	6735.00	409.68

```
[51]: olist_order_items_dataset.isna().sum()
```

```
order_id          0
order_item_id     0
product_id        0
seller_id         0
shipping_limit_date 0
price             0
freight_value     0
dtype: int64
```

```
[203]: olist_order_items_dataset.nunique()
```

```
order_id          98666
order_item_id     21
product_id        32951
seller_id         3095
shipping_limit_date 93318
price             5968
freight_value     6999
dtype: int64
```

```
[53]: # Generate a profile report for the DataFrame df
profile = ProfileReport(olist_order_items_dataset)
```

```
# Display the report in the notebook
profile
```

Error displaying widget
Error displaying widget
Error displaying widget

Overview

Brought to you by YData

Overview	Reproduction
Dataset statistics	
Number of variables	7
Number of observations	112650
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	6.0 MiB
Average record size in memory	56.0 B
Variable types	
Text	3
Numeric	3
DateTime	1

Variables

Select Columns ▾

[53] :

```
[55]: olist_order_items_dataset.shape
```

[55]: (112650, 7)

```
[57]: q1 = """
    select * from olist_order_payments_dataset
"""
pd.read_sql(q1, db)
```

[57]:		False	order_id	payment_sequential	payment_type	payment_installments	payment_value
0	0	b81ef226f3fe1789b1e8b2acac839d17		1	credit_card	8	99.33
1	1	a9810da82917af2d9aefcd1278f1dcfa0		1	credit_card	1	24.39
2	2	25e8ea4e93396b6fa0d3dd708e76c1bd		1	credit_card	1	65.71
3	3	ba78997921bbcdc1373bb41e913ab953		1	credit_card	8	107.78
4	4	42fdf880ba16b47b59251dd489d4441a		1	credit_card	2	128.45
...
103881	103881	0406037ad97740d563a178ecc7a2075c		1	boleto	1	363.31
103882	103882	7b905861d7c825891d6347454ea7863f		1	credit_card	2	96.80
103883	103883	32609bbb3dd69b3c066a6860554a77bf		1	credit_card	1	47.77
103884	103884	b8b61059626efa996a60be9bb9320e10		1	credit_card	5	369.54
103885	103885	28bbae6599b09d39ca406b747b6632b1		1	boleto	1	191.58

103886 rows × 6 columns

```
[59]: olist.order_payments.dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103886 entries, 0 to 103885
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   order_id         103886 non-null   object 
 1   payment_sequential 103886 non-null   int64  
 2   payment_type      103886 non-null   object 
 3   payment_installments 103886 non-null   int64  
 4   payment_value     103886 non-null   float64 
dtypes: float64(1), int64(1), object(2)
```

```
types: float64(4), int64(2), object(2)
memory usage: 4.0+ MB
```

```
[61]: olist_order_payments_dataset.describe().round(2)
```

	payment_sequential	payment_installments	payment_value
count	103886.00	103886.00	103886.00
mean	1.09	2.85	154.10
std	0.71	2.69	217.49
min	1.00	0.00	0.00
25%	1.00	1.00	56.79
50%	1.00	1.00	100.00
75%	1.00	4.00	171.84
max	29.00	24.00	13664.08

```
[63]: olist_order_payments_dataset.isna().sum()
```

```
[63]: order_id          0
payment_sequential    0
payment_type          0
payment_installments  0
payment_value          0
dtype: int64
```

```
[201]: olist_order_payments_dataset.nunique()
```

```
[201]: order_id          99440
payment_sequential    29
payment_type          5
payment_installments  24
payment_value          29077
dtype: int64
```

```
[65]: olist_order_payments_dataset['payment_type'].nunique()
```

```
[65]: 5
```

```
[67]: olist_order_payments_dataset['payment_type'].unique()
```

```
[67]: array(['credit_card', 'boleto', 'voucher', 'debit_card', 'not_defined'],
      dtype=object)
```

```
[69]: olist_order_payments_dataset['payment_type'].value_counts()
```

```
[69]: payment_type
credit_card    76795
boleto        19784
voucher        5775
debit_card     1529
not_defined      3
Name: count, dtype: int64
```

```
[71]: # Generate a profile report for the DataFrame df
profile = ProfileReport(olist_order_payments_dataset)
```

```
# Display the report in the notebook
profile
```

Error displaying widget
Error displaying widget
Error displaying widget

Pandas Profiling Report

Overview Variables Interactions Correlations Missing values Sample

Overview

Brought to you by YData

Overview

Alerts 1

Reproduction

Dataset statistics

Number of variables	5
Number of observations	103886
Missing cells	0
Missing cells (0%)	0.00%

Variable types

Text	1
Numeric	3
Categorical	1

MISSING CELLS (%)	0.07%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	4.0 MiB
Average record size in memory	40.0 B

Variables

Select Columns ▾

[71]:

[73]: olist_order_payments_dataset.shape

[73]: (103886, 5)

```
[75]: q1 = """
select * from olist_order_reviews_dataset
"""

pd.read_sql(q1,db)
```

	False	review_id	order_id	review_score	review_comment_title	review_comment_message	review_creation_date
0	0	7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb	4	None	None	2018-01-18 00:00:00
1	1	80e641a11e56f04c1ad469d5645fdfde	a548910a1c6147796b98fdf73dbeba33	5	None	None	2018-03-10 00:00:00
2	2	228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b	5	None	None	2018-02-17 00:00:00
3	3	e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b	5	None	Recebi bem antes do prazo estipulado.	2017-04-21 00:00:00
4	4	f7c4243c7fe1938f181bec41a392bdeb	8e6bf81e283fa7e4f11123a3fb894f1	5	None	Parabéns lojas lannister adorei comprar pela l...	2018-03-01 00:00:00
...
99219	99219	574ed12dd733e5fa530cf4bbf39d7c9	2a8c23fee101d4d5662fa670396eb8da	5	None	None	2018-07-07 00:00:00
99220	99220	f3897127253a9592a73be9bdfdf4ed7a	22ec9f0669f784db00fa86d035cf8602	5	None	None	2017-12-09 00:00:00
99221	99221	b3de70c89b1510c4cd3d0649fd302472	55d4004744368f5571d1f590031933e4	5	None	Excelente mochila, entrega super rápida. Super...	2018-03-22 00:00:00
99222	99222	1adeb9d84d72fe4e337617733eb85149	7725825d039fc1f0ceb7635e3f7d9206	4	None	None	2018-07-01 00:00:00
99223	99223	efe49f1d6f951dd88b51e6cc4cc548f	90531360ecb1eec2a1fbb265a0db0508	1	None	meu produto chegou e ja tenho que devolver, po...	2017-07-03 00:00:00

99224 rows × 8 columns

[77]: olist_order_reviews_dataset.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99224 entries, 0 to 99223
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   review_id        99224 non-null   object 
 1   order_id         99224 non-null   object 
 2   review_score     99224 non-null   int64  
 3   review_comment_title  11568 non-null  object 
 4   review_comment_message  40977 non-null  object 
 5   review_creation_date  99224 non-null   object 
 6   review_answer_timestamp  99224 non-null   object 
dtypes: int64(1), object(6)
memory usage: 5.3+ MB
```

[79]: olist_order_reviews_dataset.isna().sum()

```
review_id          0
order_id          0
review_score      0
review_comment_title  87656
review_comment_message  58247
review_creation_date  0
review_answer_timestamp  0
dtype: int64
```

[81]: olist_order_reviews_dataset['review_comment_title'] = olist_order_reviews_dataset['review_comment_title'].fillna('No Title')

[83]: olist_order_reviews_dataset['review_comment_message'] = olist_order_reviews_dataset['review_comment_message'].fillna('No Message')

```
[85]: olist_order_reviews_dataset.isna().sum()
```

```
[85]: review_id          0
order_id            0
review_score        0
review_comment_title 0
review_comment_message 0
review_creation_date 0
review_answer_timestamp 0
dtype: int64
```

```
[87]: olist_order_reviews_dataset.describe().round(2)
```

```
[87]: review_score
```

count	99224.00
mean	4.09
std	1.35
min	1.00
25%	4.00
50%	5.00
75%	5.00
max	5.00

```
[199]: olist_order_reviews_dataset.nunique()
```

```
[199]: review_id          98410
order_id            98673
review_score        5
review_comment_title 4528
review_comment_message 36160
review_creation_date 636
review_answer_timestamp 98248
dtype: int64
```

```
[89]: olist_order_reviews_dataset['review_score'].value_counts()
```

```
[89]: review_score
5    57328
4    19142
1    11424
3     8179
2     3151
Name: count, dtype: int64
```

```
[91]: # Generate a profile report for the DataFrame df
profile = ProfileReport(olist_order_reviews_dataset)
```

```
# Display the report in the notebook
profile
```

Error displaying widget
Error displaying widget
Error displaying widget

Pandas Profiling Report

Overview Variables Missing values Sample

Overview

Brought to you by YData

Overview

Reproduction

Dataset statistics

Number of variables	7
Number of observations	99224
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	5.3 MiB

Variable types

Text	4
Categorical	1
DateTime	2

Variables

Select Columns ▾

[91]:

[93]: `olist_order_reviews_dataset.shape`

[93]: (99224, 7)

```
[95]: q1 = """
    select * from olist_orders_dataset
"""
pd.read_sql(q1, db)
```

	False	order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date
0	0	e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-04 19:51
1	1	53cdb2fc8bc7dce0b6741e2150273451	b0830fb4747a6c6d20dea0b8c802d7ef	delivered	2018-07-24 20:41:37	2018-07-26 03:24:27	2018-07-26 14:31
2	2	47770eb9100c2d0c44946d9cf07ec65d	41ce2a54c0b03bf3443c3d931a367089	delivered	2018-08-08 08:38:49	2018-08-08 08:55:23	2018-08-08 13:51
3	3	949d5b44dbf5de918fe9c16f97b45f8a	f88197465ea7920adcdbe7375364d82	delivered	2017-11-18 19:28:06	2017-11-18 19:45:59	2017-11-22 13:31
4	4	ad21c59c0840e6cb83a9ceb5573f8159	8ab97904e6daea8866dbdbc4fb7aad2c	delivered	2018-02-13 21:18:39	2018-02-13 22:20:29	2018-02-14 19:41
...
99436	99436	9c5dedf39a927c1b2549525ed64a053c	39bd1228ee8140590ac3aca26f2dfe00	delivered	2017-03-09 09:54:05	2017-03-09 09:54:05	2017-03-10 11:11
99437	99437	63943bddc261676b46f01ca7ac2f7bd8	1fca14ff2861355f6e5f14306ff977a7	delivered	2018-02-06 12:58:58	2018-02-06 13:10:37	2018-02-07 23:21
99438	99438	83c1379a015df1e13d02aae0204711ab	1aa71eb042121263aafe80c1b562c9c	delivered	2017-08-27 14:46:43	2017-08-27 15:04:16	2017-08-28 20:51
99439	99439	11c177c8e97725db2631073c19f07b62	b331b74b18dc79bcd6532d51e1637c1	delivered	2018-01-08 21:28:27	2018-01-08 21:36:21	2018-01-12 15:31
99440	99440	66dea50a8b16d9b4dee7af250b4be1a5	edb027a75a1449115f6b43211ae02a24	delivered	2018-03-08 20:57:30	2018-03-09 11:20:28	2018-03-09 22:11

99441 rows × 9 columns

	order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date
count	99441	99441	99441	99441	99281	97658
unique	99441	99441	8	98875	90733	81018
top	e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2018-04-11 10:48:14	2018-02-27 04:31:10	2018-05-09 15:48:00
freq	1	1	96478	3	9	47

[101]: `olist_orders_dataset.isna().sum()`

```

[101]: order_id          0
customer_id        0
order_status        0
order_purchase_timestamp 0
order_approved_at    160
order_delivered_carrier_date 1783
order_delivered_customer_date 2965
order_estimated_delivery_date 0
dtype: int64

[103]: olist_orders_dataset['order_approved_at'] = olist_orders_dataset['order_approved_at'].fillna('2025-01-01')

[105]: olist_orders_dataset['order_delivered_carrier_date'] = olist_orders_dataset['order_delivered_carrier_date'].fillna('2025-01-01')

[107]: olist_orders_dataset['order_delivered_customer_date'] = olist_orders_dataset['order_delivered_customer_date'].fillna('2025-01-01')

[109]: olist_orders_dataset.isna().sum()

[109]: order_id          0
customer_id        0
order_status        0
order_purchase_timestamp 0
order_approved_at    0
order_delivered_carrier_date 0
order_delivered_customer_date 0
order_estimated_delivery_date 0
dtype: int64

[197]: olist_orders_dataset.nunique()

[197]: order_id          99441
customer_id        99441
order_status         8
order_purchase_timestamp 98875
order_approved_at    90734
order_delivered_carrier_date 81019
order_delivered_customer_date 95665
order_estimated_delivery_date 459
dtype: int64

[111]: olist_orders_dataset['order_status'].unique()

[111]: array(['delivered', 'invoiced', 'shipped', 'processing', 'unavailable',
       'canceled', 'created', 'approved'], dtype=object)

[113]: olist_orders_dataset['order_status'].nunique()

[113]: 8

[115]: olist_orders_dataset['order_status'].value_counts()

[115]: order_status
delivered      96478
shipped        1107
canceled       625
unavailable     609
invoiced        314
processing       301
created          5
approved          2
Name: count, dtype: int64

[117]: # Generate a profile report for the DataFrame df
profile = ProfileReport(olist_orders_dataset)

# Display the report in the notebook
profile

```

Error displaying widget
Error displaying widget
Error displaying widget

Pandas Profiling Report

Overview Variables Missing values Sample

Overview

Brought to you by YData

Overview

Alerts 3

Reproduction

Dataset statistics

Number of variables

8

Variable types

Text

2

Number of observations	99441
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	6.1 MiB
Average record size in memory	64.0 B

Categorical	1
DateTime	5

Variables

Select Columns ▾

[117]:

[119]: olist_orders_dataset.shape

[119]: (99441, 8)

```
[121]: q1 = """
select * from olist_products_dataset
"""
pd.read_sql(q1,db)
```

	False	product_id	product_category_name	product_name_length	product_description_length	product_photos_qty	product_weight_g
0	0	1e9e8ef04dbcff4541ed26657ea517e5	perfumaria	40.0	287.0	1.0	225
1	1	3aa071139cb16b67ca9e5dea641aaa2f	artes	44.0	276.0	1.0	1000
2	2	96bd76ec8810374ed1b65e291975717f	esporte_lazer	46.0	250.0	1.0	154
3	3	cef67bcfe19066a932b7673e239eb23d	bebés	27.0	261.0	1.0	371
4	4	9dc1a7de274444849c219cff195d0b71	utilidades_domesticas	37.0	402.0	4.0	625
...
32946	32946	a0b7d5a992ccda646f2d34e418fff5a0	moveis_decoracao	45.0	67.0	2.0	12300
32947	32947	bf4538d88321d0fd4412a93c974510e6	construcao_ferramentas_iluminacao	41.0	971.0	1.0	1700
32948	32948	9a7c6041fa9592d9d9ef6fce62a71f8c	cama_mesa_banho	50.0	799.0	1.0	1400
32949	32949	83808703fc0706a22e264b9d75f04a2e	informatica_acessorios	60.0	156.0	2.0	700
32950	32950	106392145fca363410d287a815be6de4	cama_mesa_banho	58.0	309.0	1.0	2083

32951 rows × 10 columns

[123]: olist_products_dataset.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32951 entries, 0 to 32950
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   product_id       32951 non-null   object 
 1   product_category_name  32341 non-null   object 
 2   product_name_length  32341 non-null   float64
 3   product_description_length  32341 non-null   float64
 4   product_photos_qty    32341 non-null   float64
 5   product_weight_g     32949 non-null   float64
 6   product_length_cm    32949 non-null   float64
 7   product_height_cm    32949 non-null   float64
 8   product_width_cm     32949 non-null   float64
dtypes: float64(7), object(2)
memory usage: 2.3+ MB
```

[125]: olist_products_dataset.describe().round(2)

	product_name_length	product_description_length	product_photos_qty	product_weight_g	product_length_cm	product_height_cm	product_width_cm
count	32341.00	32341.00	32341.00	32949.00	32949.00	32949.00	32949.00
mean	48.48	771.50	2.19	2276.47	30.82	16.94	23.20
std	10.25	635.12	1.74	4282.04	16.91	13.64	12.08
min	5.00	4.00	1.00	0.00	7.00	2.00	6.00
25%	42.00	339.00	1.00	300.00	18.00	8.00	15.00

50%	51.00	595.00	1.00	700.00	25.00	13.00	20.00
75%	57.00	972.00	3.00	1900.00	38.00	21.00	30.00
max	76.00	3992.00	20.00	40425.00	105.00	105.00	118.00

```
[127]: olist_products_dataset.isna().sum()
```

```
[127]: product_id          0
product_category_name    610
product_name_lenght      610
product_description_lenght 610
product_photos_qty        610
product_weight_g           2
product_length_cm          2
product_height_cm          2
product_width_cm          2
dtype: int64
```

```
[129]: olist_products_dataset['product_category_name'] = olist_products_dataset['product_category_name'].fillna('No category name')
```

```
[131]: olist_products_dataset['product_name_lenght'] = olist_products_dataset['product_name_lenght'].fillna('0.0')
```

```
[133]: olist_products_dataset['product_description_lenght'] = olist_products_dataset['product_description_lenght'].fillna('0.0')
```

```
[135]: olist_products_dataset['product_photos_qty'] = olist_products_dataset['product_photos_qty'].fillna('0.0')
```

```
[139]: olist_products_dataset['product_weight_g'] = olist_products_dataset['product_weight_g'].fillna('0.0')
```

```
[141]: olist_products_dataset['product_length_cm'] = olist_products_dataset['product_length_cm'].fillna('0.0')
```

```
[143]: olist_products_dataset['product_height_cm'] = olist_products_dataset['product_height_cm'].fillna('0.0')
```

```
[145]: olist_products_dataset['product_width_cm'] = olist_products_dataset['product_width_cm'].fillna('0.0')
```

```
[147]: olist_products_dataset.isna().sum()
```

```
[147]: product_id          0
product_category_name    0
product_name_lenght      0
product_description_lenght 0
product_photos_qty        0
product_weight_g           0
product_length_cm          0
product_height_cm          0
product_width_cm          0
dtype: int64
```

```
[149]: olist_products_dataset.unique()
```

```
[149]: product_id          32951
product_category_name    74
product_name_lenght      67
product_description_lenght 2961
product_photos_qty        20
product_weight_g          2285
product_length_cm         100
product_height_cm         103
product_width_cm          96
dtype: int64
```

```
[153]: olist_products_dataset['product_category_name'].unique()
```

```
[153]: array(['perfumaria', 'artes', 'esporte_lazer', 'bebes',
       'utilidades_domesticas', 'instrumentos_musicais', 'cool_stuff',
       'moveis_decoracao', 'eletrodomesticos', 'brinquedos',
       'cama_mesa_banho', 'construcao_ferramentas_seguranca',
       'informatica_acessorios', 'beleza_saude', 'malas_acessorios',
       'ferramentas_jardim', 'moveis_escritorio', 'automotivo',
       'eletronicos', 'fashion_calcados', 'telefonia', 'papelaria',
       'fashion_bolsas_e_acessorios', 'pcs', 'casa_construcao',
       'relogios_presentes', 'construcao_ferramentas_construcao',
       'pet_shop', 'eletroportateis', 'agro_industria_e_comercio',
       'No category name', 'moveis_sala', 'sinalizacao_e_seguranca',
       'climatizacao', 'consoles_games', 'livros_interesse_geral',
       'construcao_ferramentas_ferramentas',
       'fashion_underwear_e_moda_praia', 'fashion_roupa_masculina',
       'moveis_cozinha_area_de_servico_jantar_e_jardim',
       'industria_comercio_e_negocios', 'telefonia_fixa',
       'construcao_ferramentas_iluminacao', 'livros_tecnicos',
       'eletrodomesticos_2', 'artigos_de_festas', 'bebidas',
       'market_place', 'la_cuisine', 'construcao_ferramentas_jardim',
       'fashion_roupa_feminina', 'casa_conforto', 'audio',
       'alimentos_bebidas', 'musica', 'alimentos',
       'tablets_impressao_imagem', 'livros_importados',
       'portateis_casa_forno_e_cafe', 'fashion_esporte',
       'artigos_de_natal', 'fashion_roupa_infanto_juvenil',
       'dvds_blu_ray', 'artes_e_artesanato', 'pc_gamer', 'moveis_quarto',
       'cine_foto', 'fraldas_higiene', 'flores', 'casa_conforto_2',
       'portateis_cozinha_e_preparadores_de_alimentos',
       'seguros_e_servicos', 'moveis_colchao_e_estofado',
```

```

`cas_avas_musicais`, atype=object)

[151]: olist_products_dataset['product_category_name'].value_counts()

[151]:
product_category_name
cama_mesa_banho          3029
esporte_lazer             2867
moveis_decoracao          2657
beleza_saude              2444
utilidades_domesticas     2335
...
fashion_roupa_infanto_juvenil    5
casa_comforto_2            5
pc_gamer                   3
seguros_e_servicos          2
cds_dvds_musicais           1
Name: count, Length: 74, dtype: int64

```

```

[157]: # Generate a profile report for the DataFrame df
profile = ProfileReport(olist_products_dataset)

# Display the report in the notebook
profile

```

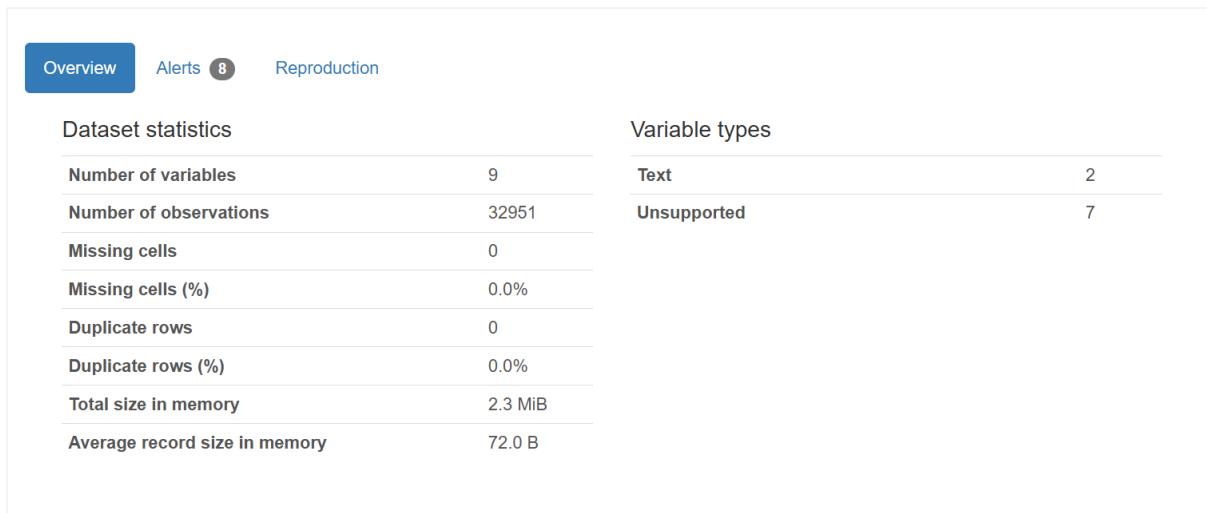
Error displaying widget
Error displaying widget
Error displaying widget

Pandas Profiling Report

Overview Variables Missing values Sample

Overview

Brought to you by YData



Variables

Select Columns ▾

[157]:

```
[159]: olist_products_dataset.shape
```

```
[159]: (32951, 9)
```

```
[161]: q1 = """
select * from olist_sellers_dataset
"""
pd.read_sql(q1,db)
```

False		seller_id	seller_zip_code_prefix	seller_city	seller_state
0	0	3442f8959a84dea7ee197c632cb2df15	13023	campinas	SP
1	1	d1b65fc7debc3361ea86b5f14c68d2e2	13844	mogi guacu	SP
2	2	ce3ad9de960102d0677a81f5d0bb7b2d	20031	rio de janeiro	RJ
3	3	60f30002a11555b660003dd50a1b1c2	4105	sao paulo	SP

		codigos de barras			
4	4	51a04a8a6bdcb23deccc82b0b80742cf	12914	braganca paulista	SP
...
3090	3090	98dddbc4601dd4443ca174359b237166	87111	sarandi	PR
3091	3091	f8201cab383e484732366d1906e2fdfa	88137	palhoca	SC
3092	3092	74871d19219c7d518d0090283e03c137	4650	sao paulo	SP
3093	3093	e603cf3fec55f8697c9059638d6c8eb5	96080	pelotas	RS
3094	3094	9e25199f6fef7e7c347120ff175652c3b	12051	taubate	SP

3095 rows × 5 columns

```
[163]: olist_sellers_dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3095 entries, 0 to 3094
Data columns (total 4 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   seller_id        3095 non-null   object 
 1   seller_zip_code_prefix 3095 non-null   int64  
 2   seller_city       3095 non-null   object 
 3   seller_state      3095 non-null   object 
dtypes: int64(1), object(3)
memory_usage: 96.8+ KB
```

```
[165]: olist_sellers_dataset.describe().round(2)
```

```
[165]: seller_zip_code_prefix
```

count	3095.00
mean	32291.06
std	32713.45
min	1001.00
25%	7093.50
50%	14940.00
75%	64552.50
max	99730.00

```
[195]: olist_sellers_dataset.nunique()
```

```
[195]: seller_id          3095  
        seller_zip_code_prefix 2246  
        seller_city            611  
        seller_state           23  
        dtype: int64
```

```
[167]: olist_sellers_dataset['seller_city'].nunique()
```

[167]: 611

```
[169]: olist_sellers_dataset['seller_state'].nunique()
```

[169]: 23

```
[171]: olist_sellers_dataset['seller_state'].unique()
```

```
[171]: array(['SP', 'RJ', 'PE', 'PR', 'GO', 'SC', 'BA', 'DF', 'RS', 'MG', 'RN',
       'MT', 'CE', 'PB', 'AC', 'ES', 'RO', 'PI', 'MS', 'SE', 'MA', 'AM',
       'PA'], dtype=object)
```

```
[177]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# لتفعيل عرض الرسوم داخل Jupyter
%matplotlib inline

# قراءة البيانات من ملف CSV
data = pd.read_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\Brazilian E-Commerce.csv")

# تأكيد من الأعمدة المتاحة في البيانات
print(data.columns)

# day one, the age and speed of 13 cars:
x = np.array(data['seller_city']) # استخدم العمود المناسب
y = np.array(data['seller_zip_code_prefix']) # استخدم العمود المناسب
plt.scatter(x, y)

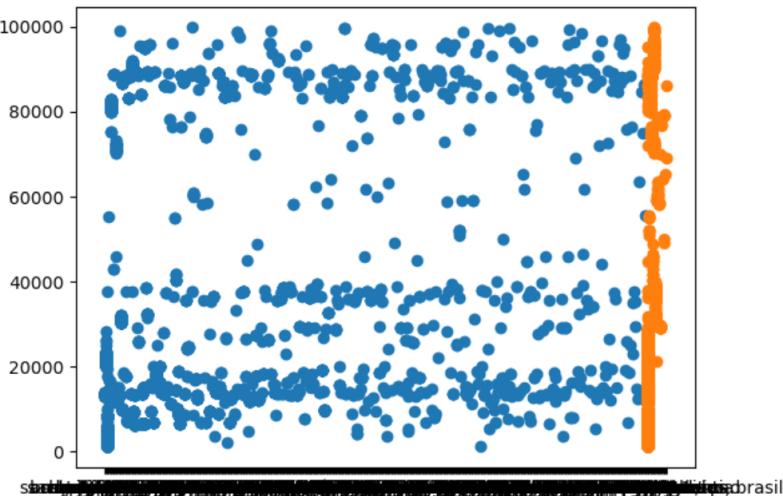
# day two, the age and speed of 15 cars:
x = np.array(data['seller_state']) # استخدم العمود المناسب
```

```
y = np.array(data['seller_zip_code_prefix']) # استخدم العمود المناسب
```

```
plt.scatter(x, y)
```

```
# عرض الرسوم  
plt.show()
```

```
Index(['seller_id', 'seller_zip_code_prefix', 'seller_city', 'seller_state'], dtype='object')
```



```
[179]: # Generate a profile report for the DataFrame df  
profile = ProfileReport(olist_sellers_dataset)  
  
# Display the report in the notebook  
profile
```

Error displaying widget

Error displaying widget

Error displaying widget

Pandas Profiling Report

Overview

Variables

Interactions

Correlations

Missing values

Sample

Overview

Brought to you by YData

Overview

Alerts 4

Reproduction

Dataset statistics

Number of variables	4
Number of observations	3095
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	96.8 KiB
Average record size in memory	32.0 B

Variable types

Text	2
Numeric	1
Categorical	1

Variables

Select Columns ▾

```
[179]:
```

```
[181]: olist_sellers_dataset.shape
```

```
[181]: (3095, 4)
```

```
[183]: q1 = """
    select * from product_category_name_translation
"""
pd.read_sql(q1,db)
```

```
[183]:   False  product_category_name  product_category_name_english
0      0        beleza_saude          health_beauty
1      1  informatica_acessorios  computers_accessories
2      2        automotivo            auto
3      3  cama_mesa_banho        bed_bath_table
4      4  moveis_decoracao        furniture_decor
...
66     66           flores            flowers
67     67  artes_e_artesanato  arts_and_craftmanship
68     68        fraldas_higiene  diapers_and_hygiene
69     69  fashion_roupa_infantil  fashion_childrens_clothes
70     70  seguros_e_servicos  security_and_services
```

71 rows × 3 columns

```
[185]: product_category_name_translation.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 71 entries, 0 to 70
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   product_category_name    71 non-null   object 
 1   product_category_name_english 71 non-null   object 
dtypes: object(2)
memory usage: 1.2+ KB
```

```
[187]: product_category_name_translation.describe().round(2)
```

```
[187]:   product_category_name  product_category_name_english
count                71                  71
unique               71                  71
top                 beleza_saude          health_beauty
freq                  1                  1
```

```
[189]: product_category_name_translation.nunique()
```

```
product_category_name      71
product_category_name_english  71
dtype: int64
```

```
[191]: # Generate a profile report for the DataFrame df
profile = ProfileReport(product_category_name_translation)

# Display the report in the notebook
profile
```

Error displaying widget
Error displaying widget
Error displaying widget

Pandas Profiling Report

Overview Variables Missing values Sample

Overview

Brought to you by YData

Overview

Alerts 2

Reproduction

Dataset statistics

Number of variables

2

Variable types

Text

2

Number of observations	71
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.2 KiB
Average record size in memory	17.9 B

Variables

Select Columns ▾

[191]:

```
[193]: product_category_name_translation.shape
```

```
[193]: (71, 2)
```

```
[219]: olist_customers_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_c...
```

```
[221]: olist_geolocation_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_g...
```

```
[223]: olist_order_items_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_o...
```

```
[225]: olist_order_payments_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_o...
```

```
[227]: olist_order_reviews_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_o...
```

```
[229]: olist_orders_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_o...
```

```
[231]: olist_products_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_p...
```

```
[233]: olist_sellers_dataset.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\olist_se...
```

```
[235]: product_category_name_translation.to_csv(r"C:\Users\20109\OneDrive - Egyptian E-Learning University\Desktop\Dashboard\Data Analysis\7- Brazilian Data Analysis\...
```

```
[ ]:
```