

Analysis of Brazilian E- Commerce Public Dataset

Mckinley Petalver

General Assumptions:

1. To this analysis, orders whose status is "not available" or "cancelled" are excluded, unless explicitly mentioned to the contrary.
2. Since the payments or other monetary-related metrics recorded in the dataset contained no indication of the currency used, I assumed it was in Brazilian Real, as Olist was specified as a Brazilian e-commerce platform.
3. To compare revenue or sales month-over-month or year-over-year, this report will use the "order_approved_at" dates to reflect finalized transactions. This timestamp can account for any delays in processing or cancellations and might give a more accurate picture of when the sale was confirmed.

Tools used:

1. Excel (for csv checking)
2. PowerBi (charting)
3. SQL (data storage)
4. Python (script for csv to sql)

Technical Exercise

Answers for the following questions with the appropriate visualization.

How many closed deal were won in Q3 of 2018?

Out of 842 data, 93 or 11% were closed during the third quarter of 2018.

Total Deals of Q3 2018

93

2018 closed deals: 839 (-88.92%)

**What is the average
weight of
'electronics' in
kilograms?**

1.28 kg.

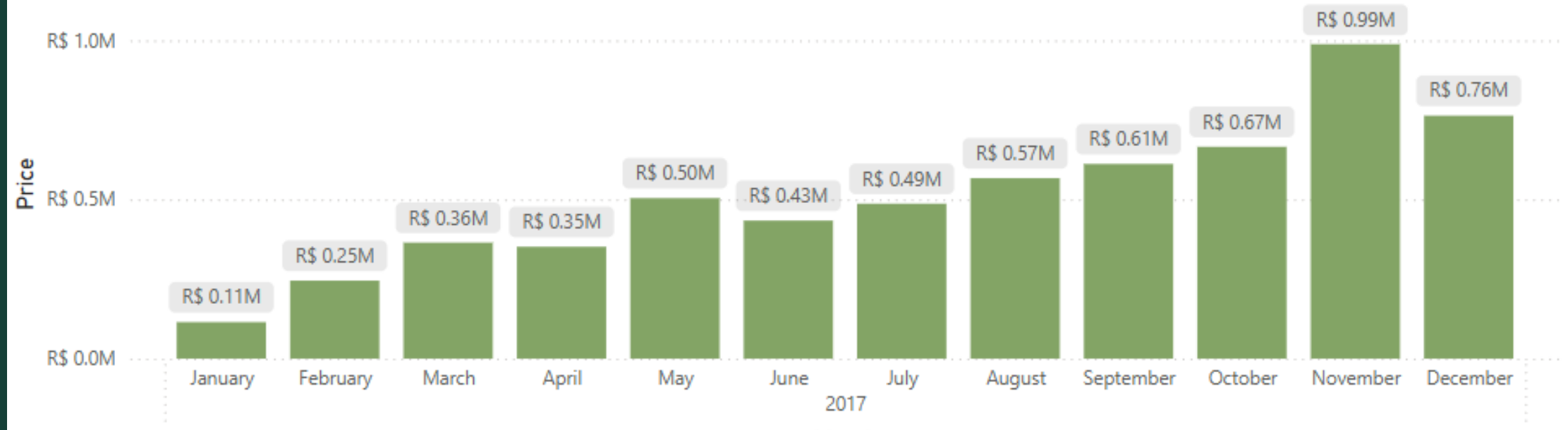
Average of product electronics in kg

1.28

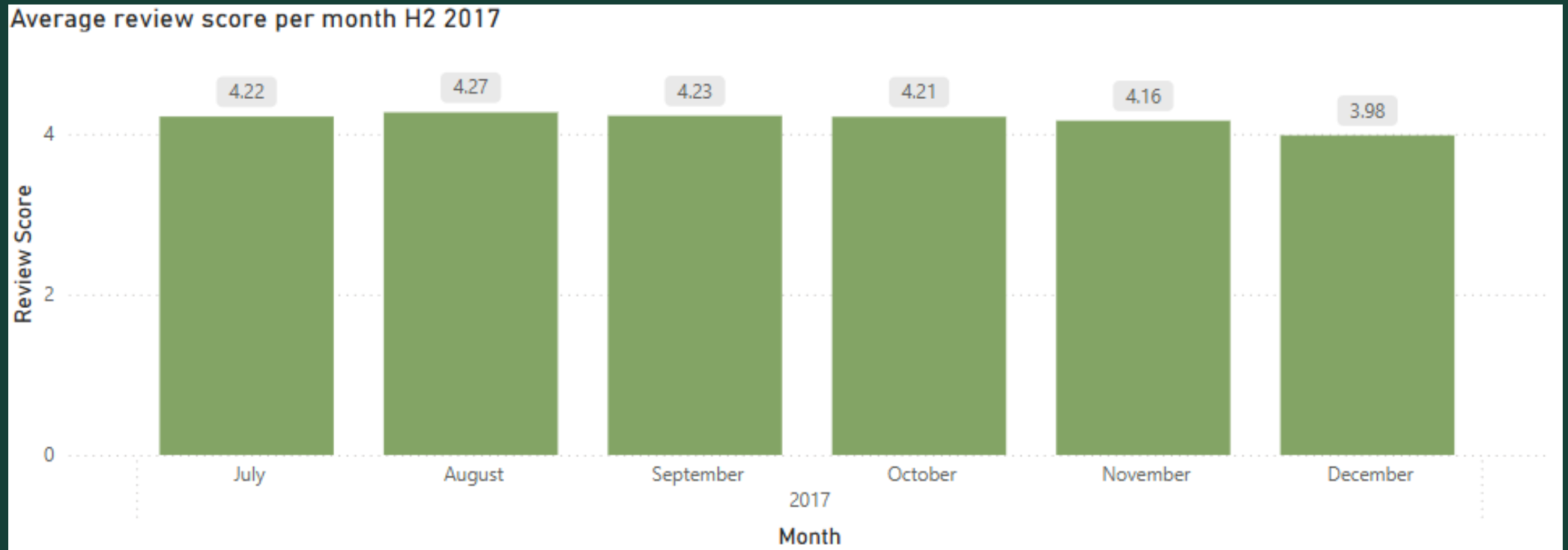
Which month in 2017 has the highest sales?

November has the highest sales based on the product revenue only in Brazilian Real.

Average payment per month 2017

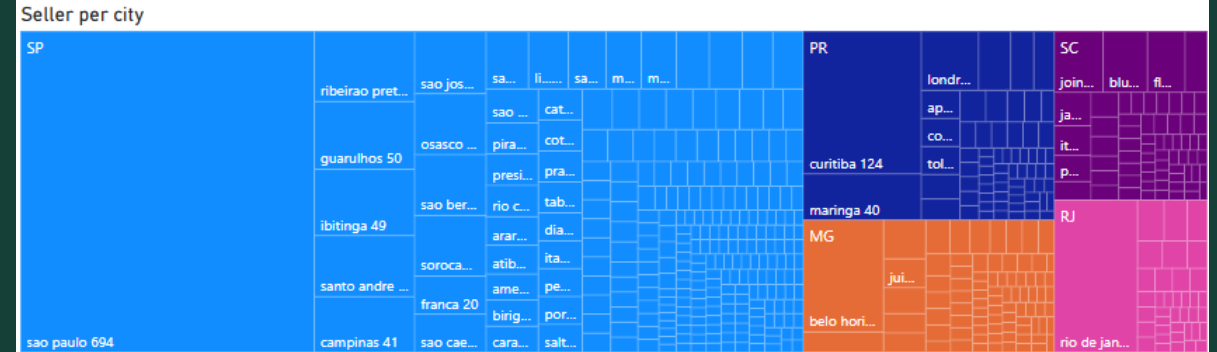


What is the average review score per month H2 2017?



Which state has the highest number of seller per city?

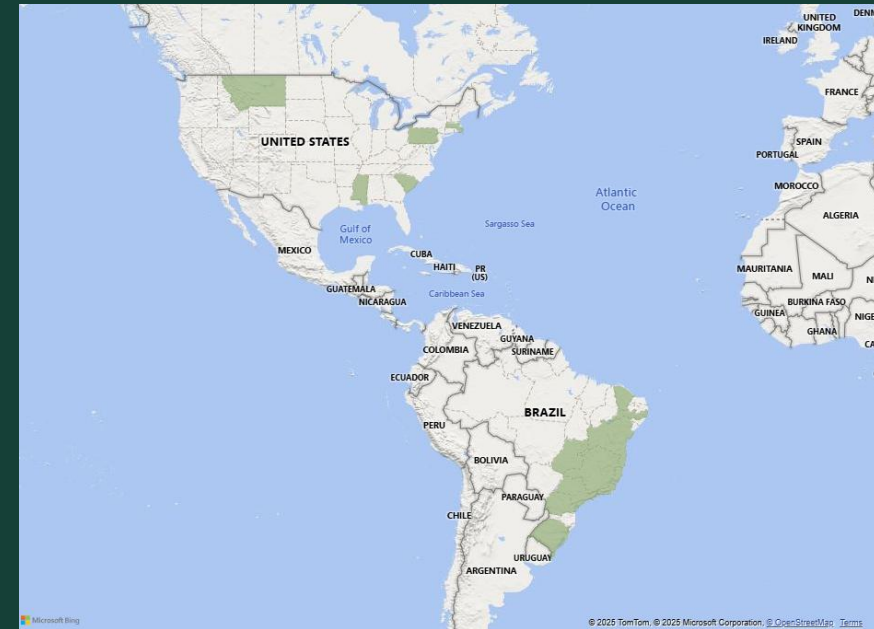
1. SP
2. PR
3. MG
4. SC
5. RJ



seller_state	Count of seller_id
SP	1849
PR	349
MG	244
SC	190
RJ	171
RS	129
GO	40
DF	30
ES	23
BA	19
CE	13
PE	9
PB	6
MS	5
RN	5
MT	4
RO	2
SE	2
AC	1
AM	1
MA	1
PA	1
PI	1
Total	3095

How many states have more than 300 customers in 2018?

There are 16 states which had more than 300 customers in 2018.



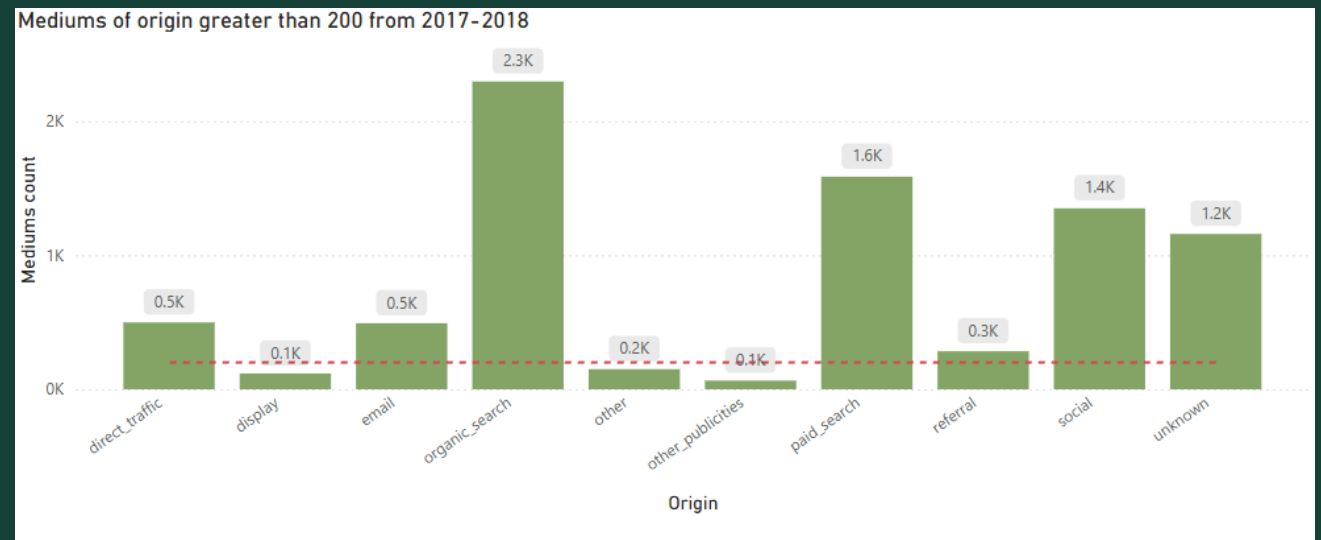
States greater than 300 customers in 2018



How many mediums of origin had more than 200 leads from 2017 to 2018?

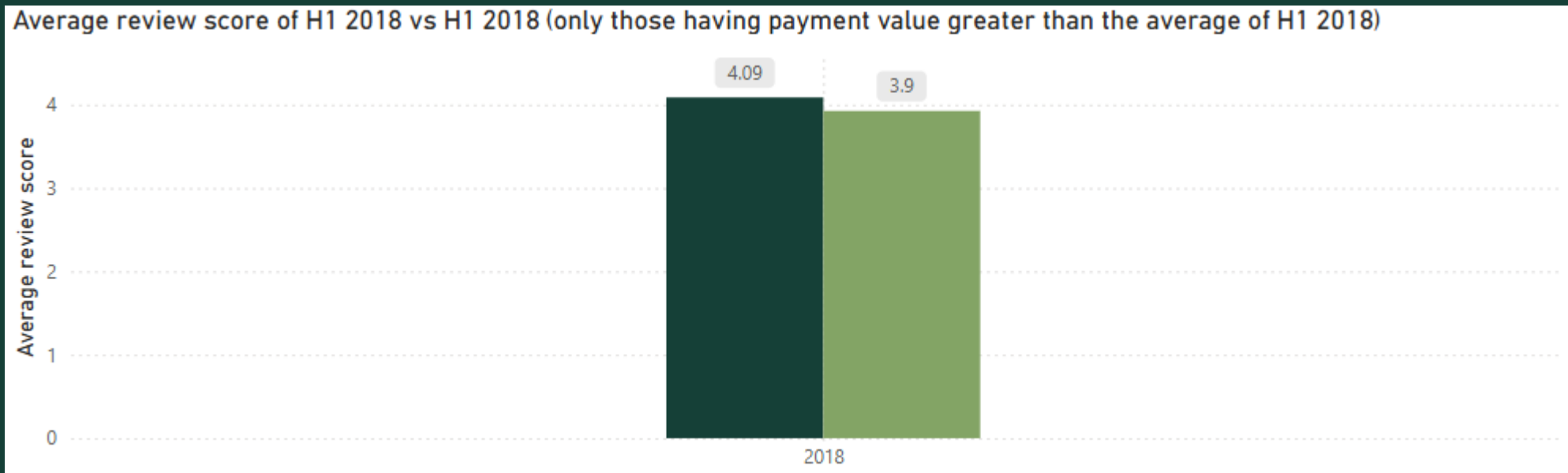
There are 200 mediums of origin.

1. Direct traffic
2. Email
3. Organic search
4. Paid search
5. Referral
6. Social
7. Other mediums



What is the average review score for orders greater than the average value in H1 of 2018?

3.9 average review score

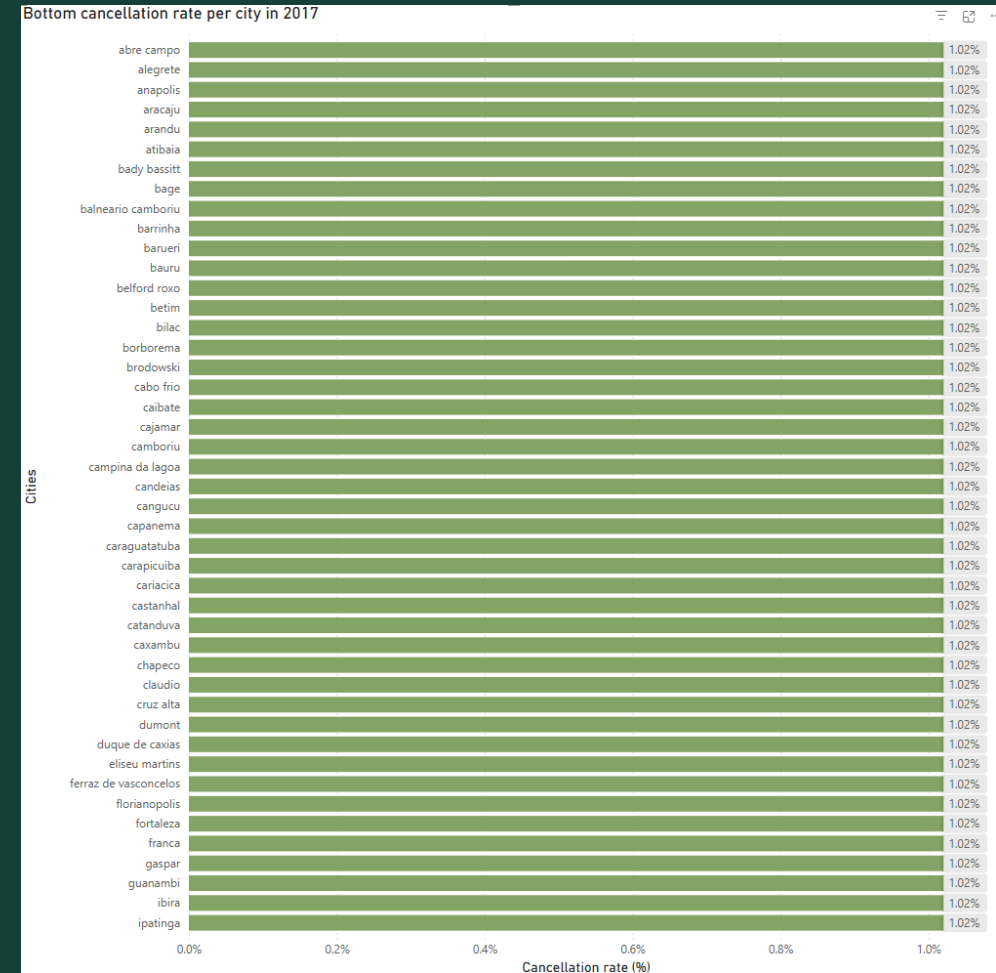


What is the monthly running total payment volume for 'credit_card' in 2017?

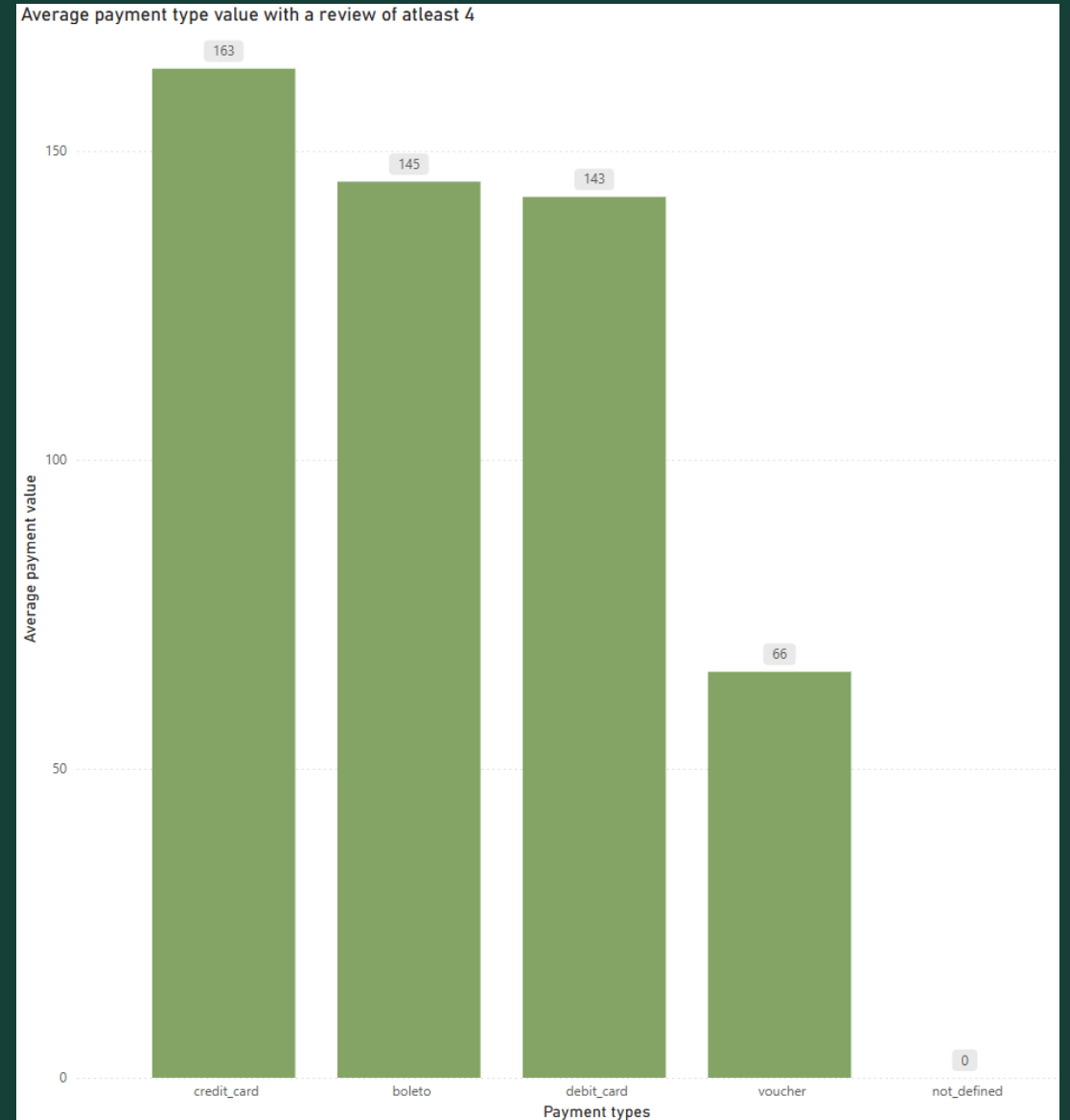


**Which 3 cities have
the lowest
cancellation rate?
What is their
respective total
payment
volume in 2017?**

There are multiple cities who had the same cancellation rate during 2017.
Mostly having 1 volume per city.

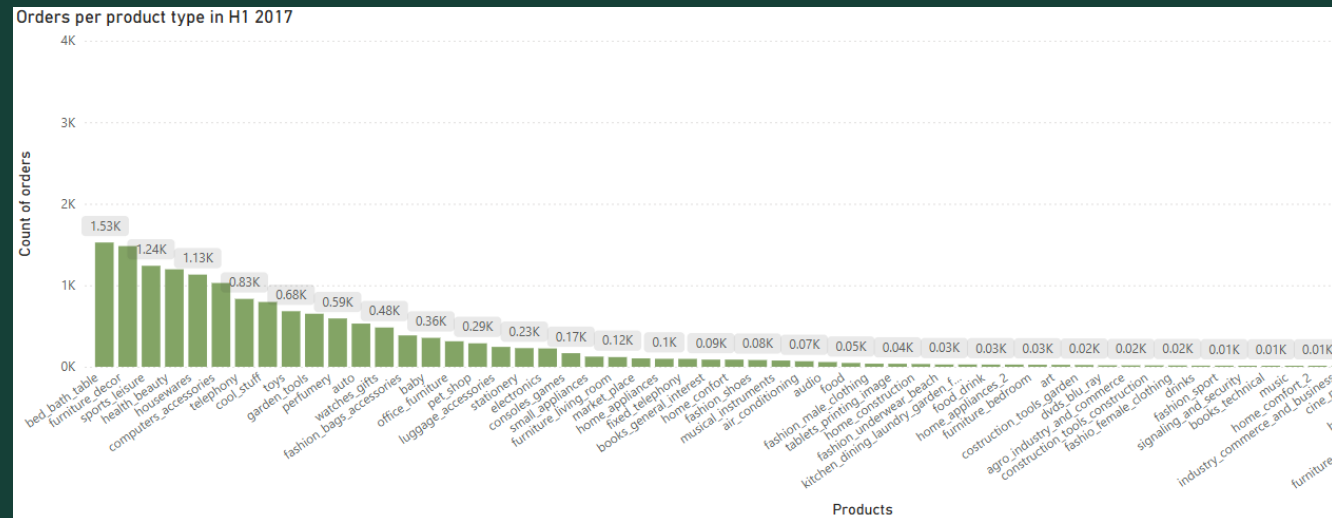


What is the average payment value per payment type for orders from sellers from Rio de Janeiro with a review score of at least 4?



How many orders per product type in H1 2017 vs H2 2017?

Show each product type in English.

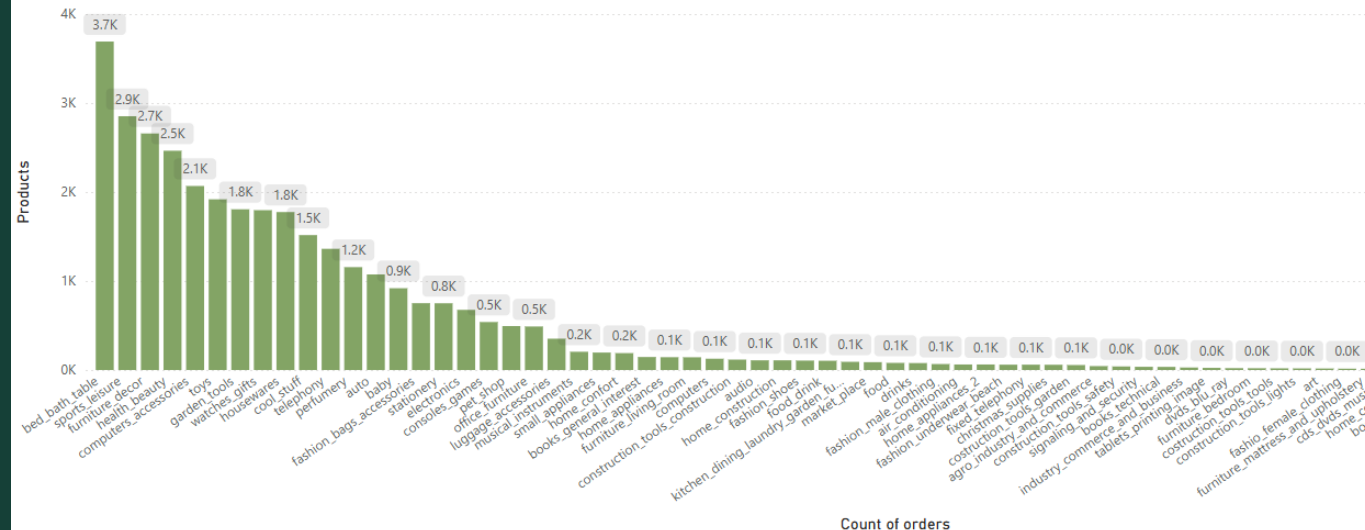


product_category_name_english	Count of product_id
bed_bath_table	1525
furniture_decor	1481
sports_leisure	1239
health_beauty	1196
housewares	1131
computers_accessories	1028
telephony	833
cool_stuff	795
toys	682
garden_tools	651
perfumery	593
auto	530
watches_gifts	482
fashion_bags_accessories	385
baby	355
office_furniture	313
pet_shop	288
luggage_accessories	244
stationery	229
electronics	225
consoles_games	167
small_appliances	125
Total	15885

How many orders per product type in H1 2017 vs H2 2017?

Show each product type in English.

Orders per product type in H2 2017



product_category_name_english	Count of product_id
bed_bath_table	3691
sports_leisure	2852
furniture_decor	2658
health_beauty	2464
computers_accessories	2068
toys	1918
garden_tools	1806
watches_gifts	1796
housewares	1775
cool_stuff	1516
telephony	1362
perfumery	1156
auto	1074
baby	919
fashion_bags_accessories	751
stationery	750
electronics	676
consoles_games	539
pet_shop	494
office_furniture	489
luggage_accessories	352
musical_instruments	205
Total	33989

Brazilian E-Commerce Public Dataset

The Brazilian E-Commerce Public Dataset, provided by [Olist](#), offers a rich source of information on real-world e-commerce transactions in Brazil.

This case study aims to leverage the dataset to gain actionable insights into **customer behavior, sales performance, and logistical efficiency** within the Brazilian online retail landscape.

The logo for Olist Store is displayed on a solid blue background. The word "olist" is in a large, white, lowercase sans-serif font, and the word "store" is in a smaller, white, lowercase sans-serif font directly below it. There are some light blue geometric shapes in the background, including a rectangle in the top left and a vertical bar on the right side.

olist
store

Report approach

Data Acquisition



Data Processing



Data Analysis



- After storing the cleaned and structured data in SQL using Python scripts, Power BI connected to this database through CSV files to build robust data models, establishing relationships, guided with the given ERD and creating calculated measures relevant to business questions.
- Through Power BI's visualization capabilities, various charts and graphs were constructed to explore trends, patterns, and anomalies within the data, such as sales performance over time, customer segmentation, or review score performance.
- By interacting with these dynamic visualizations and analyzing key performance indicators presented in Power BI, actionable insights were identified, enabling data-driven decisions regarding inventory management, marketing strategies, or customer engagement initiatives.

Exploratory data analysis

How did the e-commerce in Brazil performed?

Customer location globally

- Globally, it is not surprising that most customers involved in Brazilian commerce are in Brazil.



Customer location analysis in Brazil

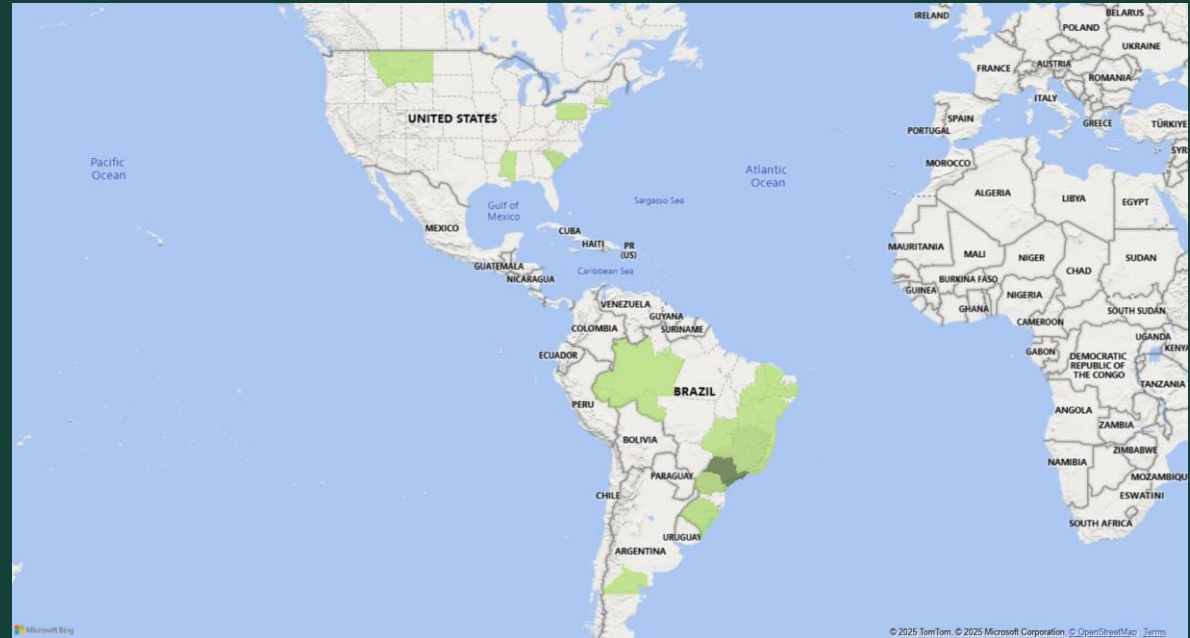
Deeper analysis of customer locations in Brazil highlighted Sao Paulo (SP) state as having the most customers, with neighboring Rio de Janeiro (RJ) and Minas Gerais (MG) also showing significant numbers, possibly due to SP's regional influence.

One assumption for this customer concentration is the potential proximity of sellers within these states, although other factors may also play a role.



Seller location analysis across America

The analysis confirmed the initial assumption about seller distribution. Despite the presence of some sellers internationally, the majority are located across the main states in Brazil. This is primarily due to the high population density of the three states previously discussed within Brazil.

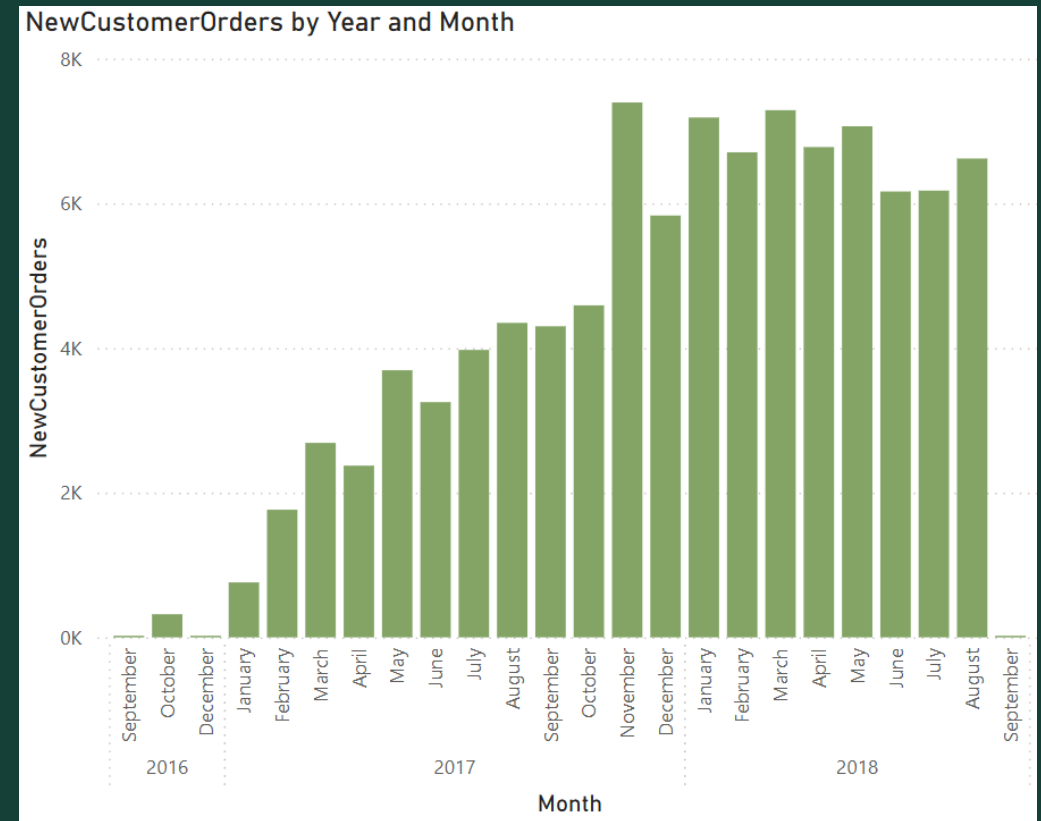


There is significant new customer orders increase from 2017–2018

Given this increase, it might be beneficial to implement a targeted strategy to encourage these new customers to place repeat orders and become long-term customers.

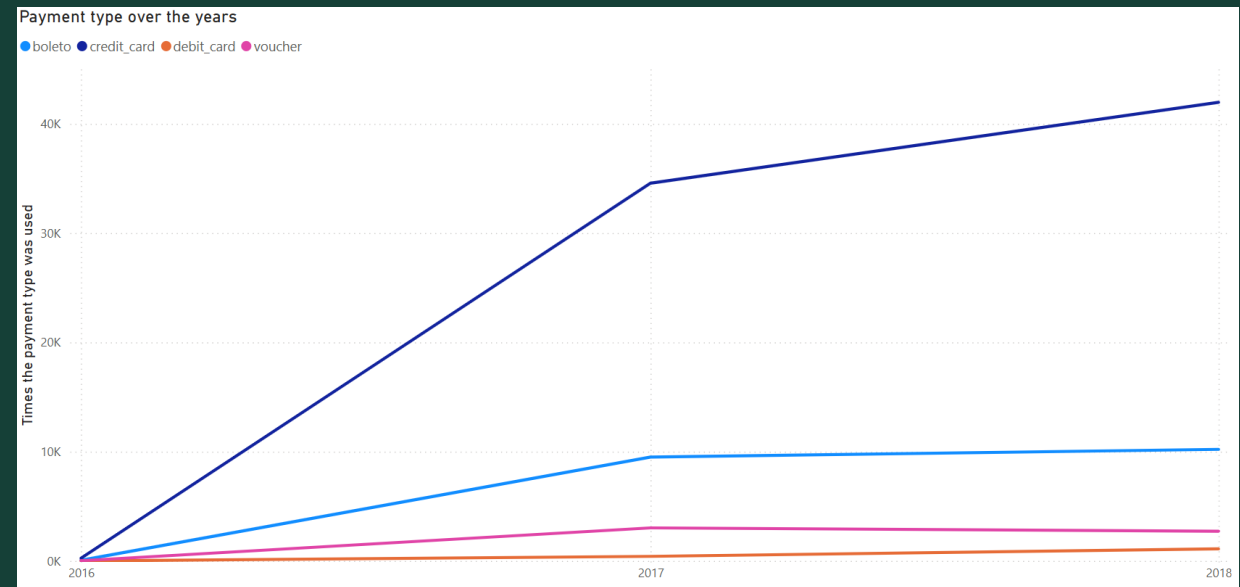
Therefore, a dedicated effort might be required to focus on these new customers to foster repeat business and ensure their retention.

To encourage these new customers to order again and build loyalty, a focused approach may be necessary.



Which payment type does the customers prefer?

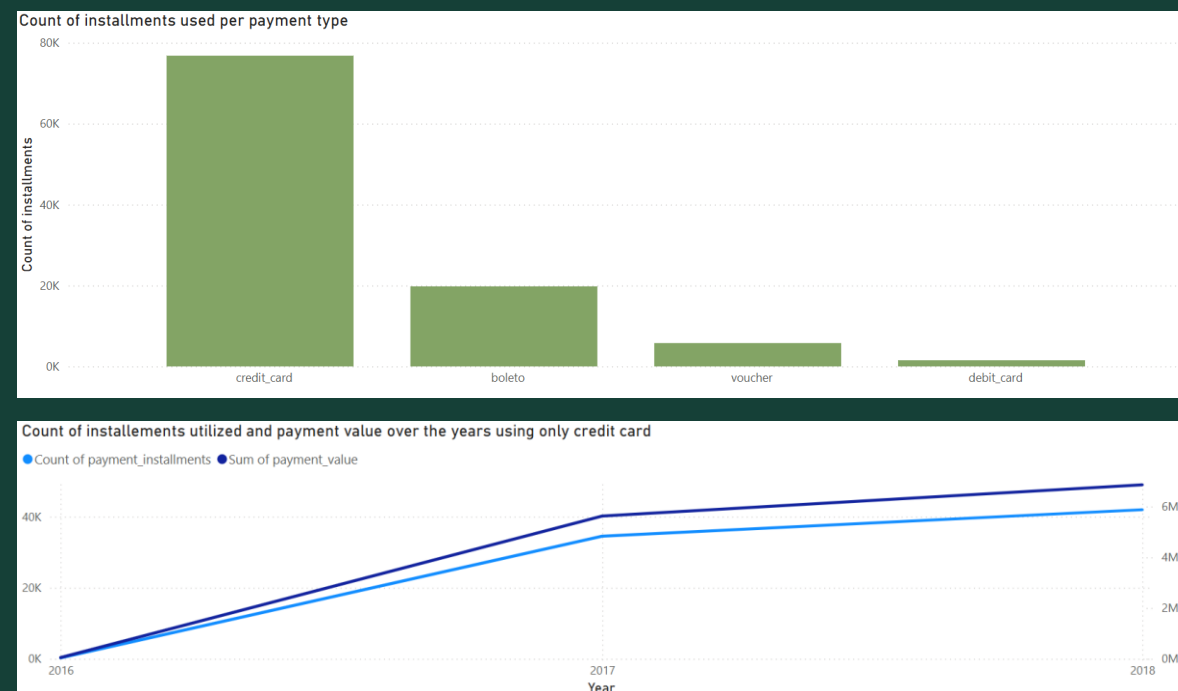
This visualization illustrates the different payment methods customers used over time when placing their orders. Notably, there was a significant shift towards credit card usage throughout the years, and this method maintained a strong lead, creating a considerable difference compared to other payment options.



Why?

The chart emphasizes the prevalence of installment usage across payment types, with the graph showcasing the particular strength of credit cards for installment plans. Notably, there is a strong proportional link between the number of installments and the total payment value.

Another key benefit of credit cards is the option for delayed payment.

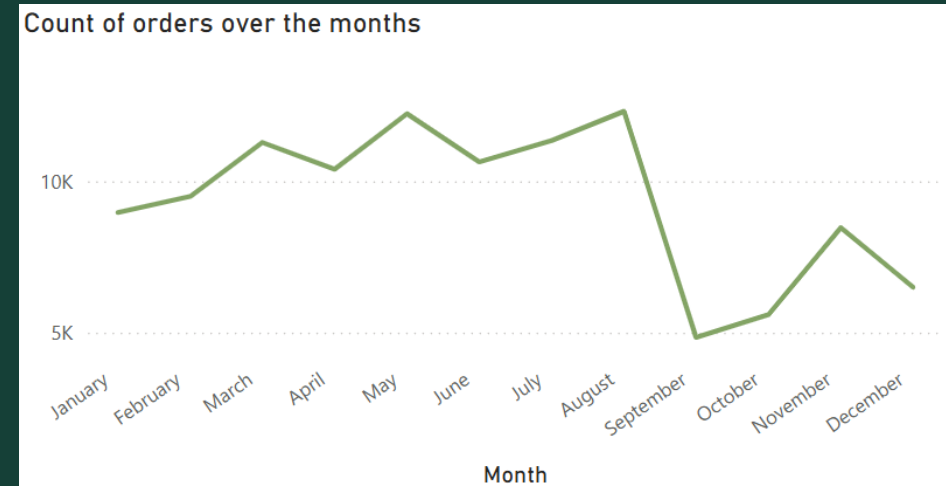
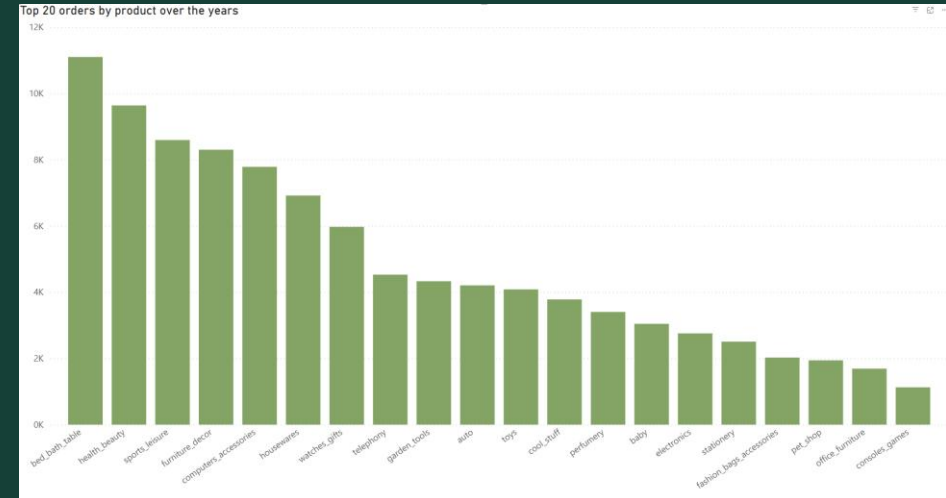


Looking at product categories, the majority of orders, representing 11% of the total, are for items intended for bedrooms, bathrooms, and dining tables.

When examining the purchase patterns in 2017, considering this is the only year where months have the complete months, the first half of the year saw the highest volume of orders.

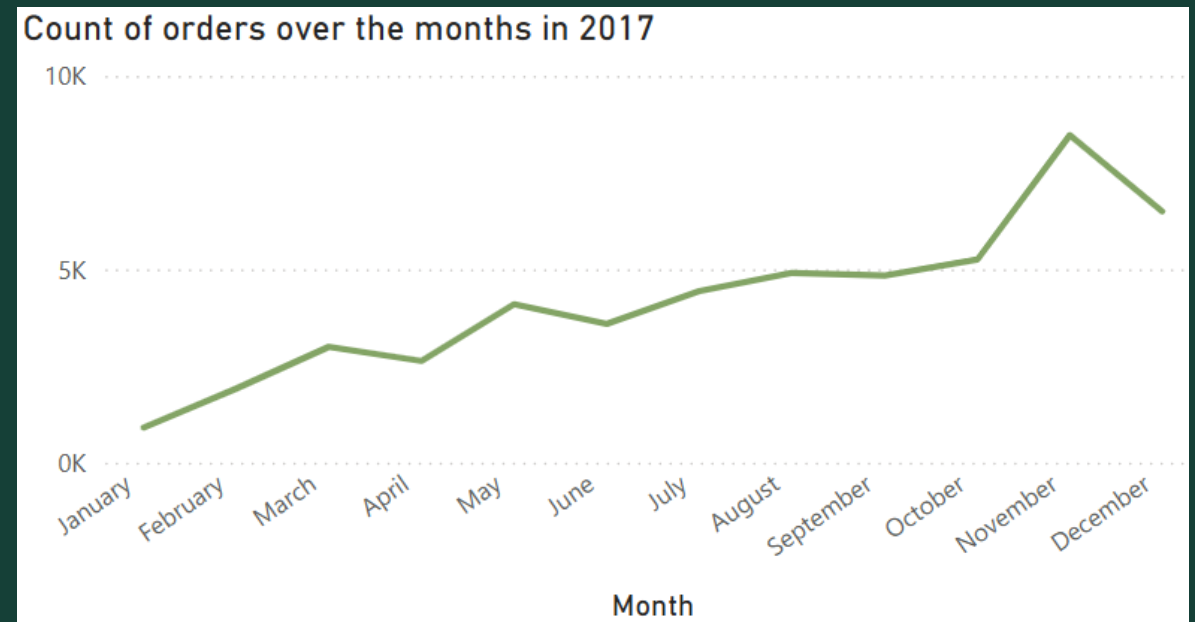
In contrast, the second half experienced a dip, with September showing the lowest performance.

This low figure for September is likely due to the fact that data collection began and concluded in September of different years (2016 and 2018), leading to a potential bias when observing trends across multiple years.



In contrast to the earlier observation, analyzing only the 2017 data indicates that the final quarter of the year was likely the period with the highest customer purchase activity.

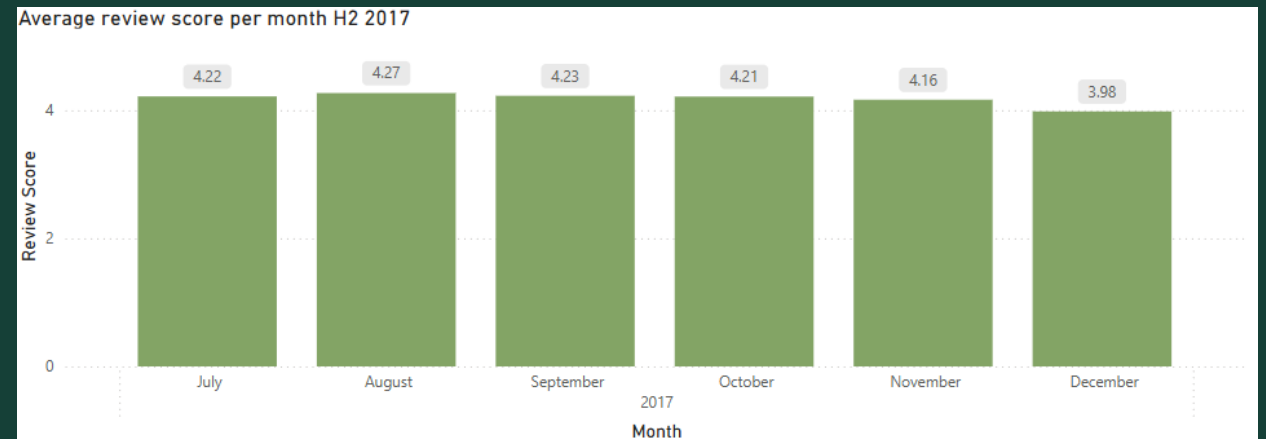
This pattern were likely due to the fact of the seasonal events happening during these months. It also happened to be the shortest quarter because of the holidays.



The fact that Q4 had the most orders highlights the significant impact of review scores for online retailers during this peak shopping period.

Looking at the second half of 2017, the marketplace benefited from a high average review score during its busiest months, contributing to an overall average of 4.11.

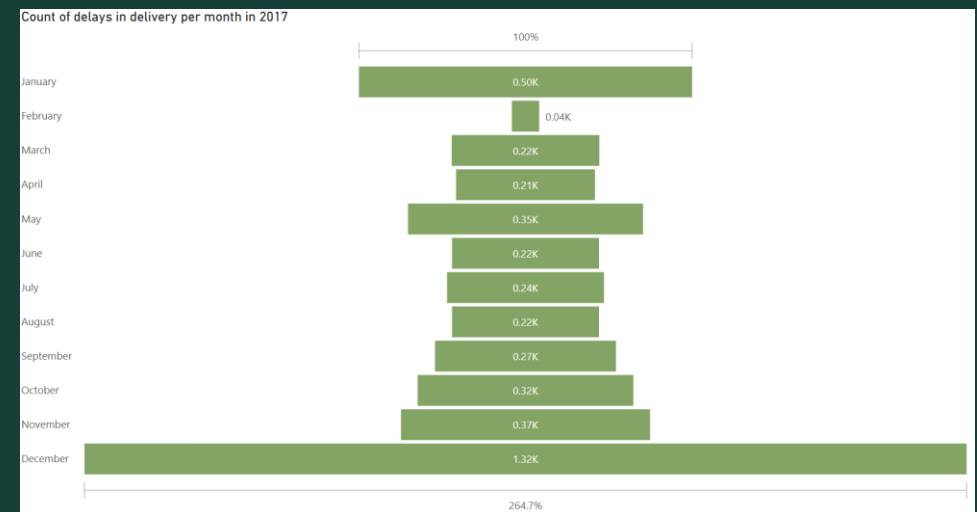
Although, a significant lower average review score was observed in December.



Interestingly, December 2017 also recorded the highest number of reviews received up to that point.

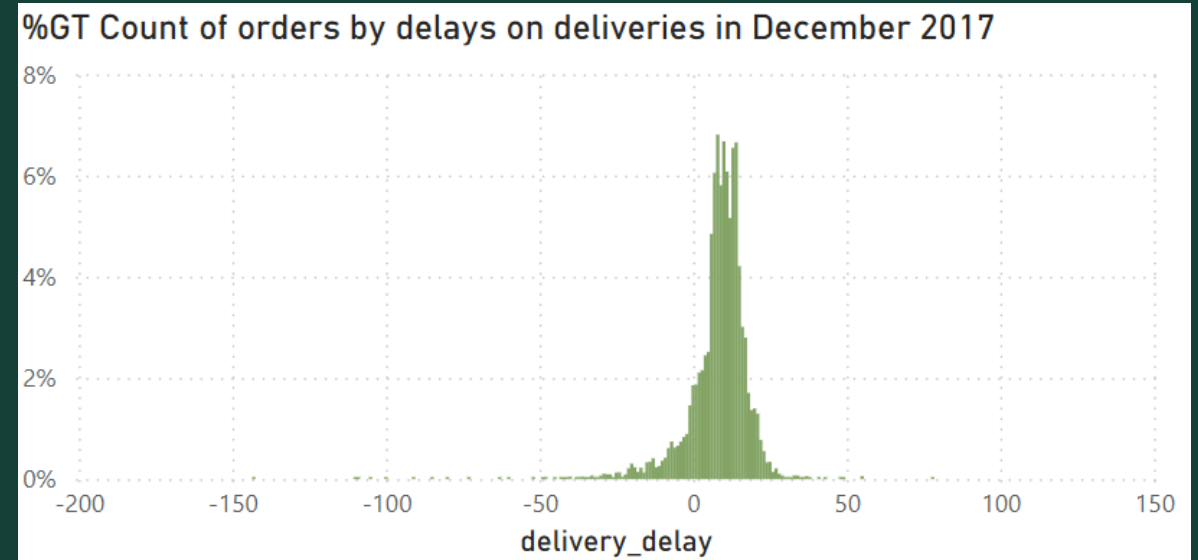
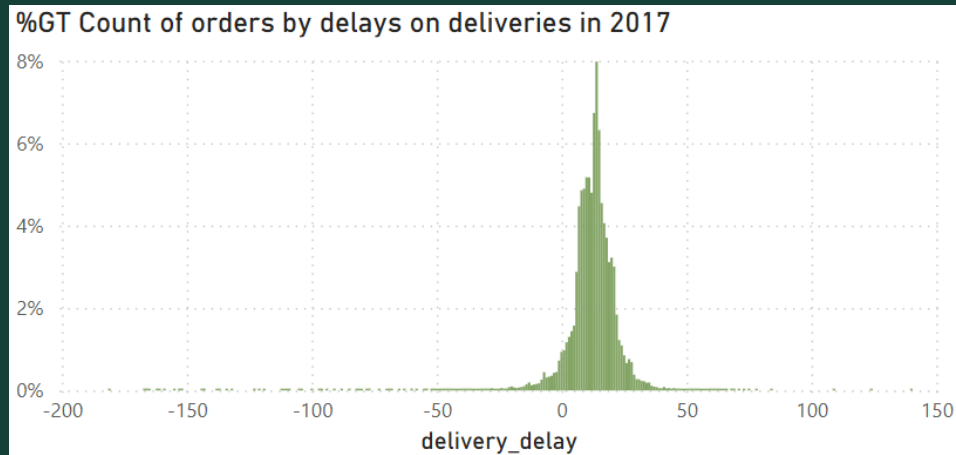
The likely cause for December (and potentially the entire fourth quarter) having the highest volume of reviews coupled with the lowest average score seen so far could be attributed to delivery delays during that period.

This surge in reviews in December is also likely a consequence of the high number of orders placed in November, as illustrated in a previous chart.



An overview of delivery timeliness in 2017

Positive figures represent deliveries that arrived before the estimated date, and negative figures represent deliveries that were delayed.

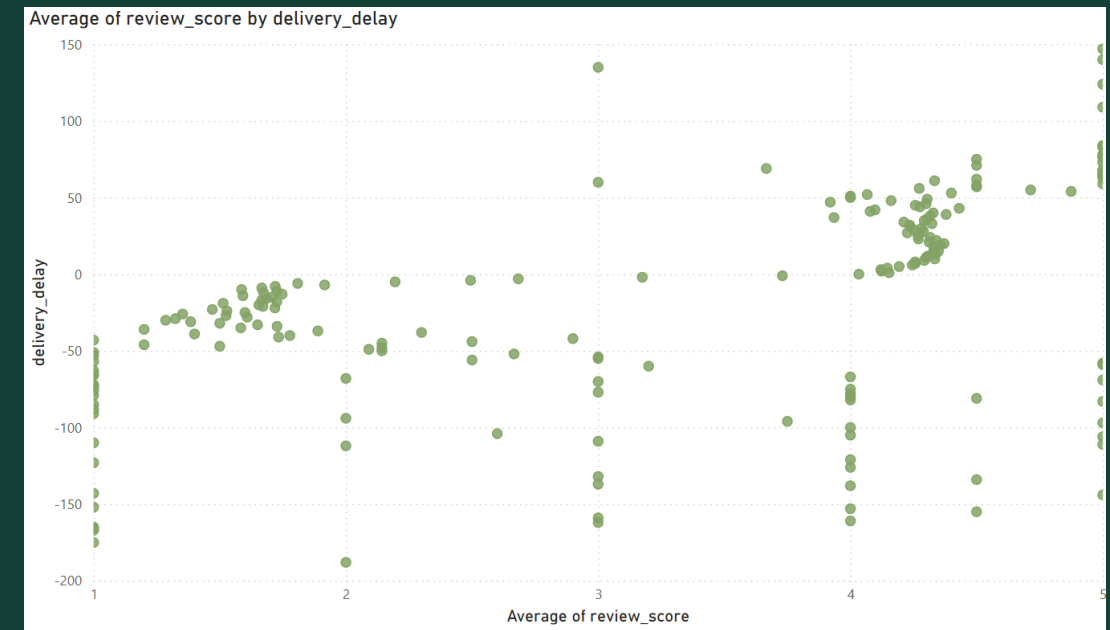
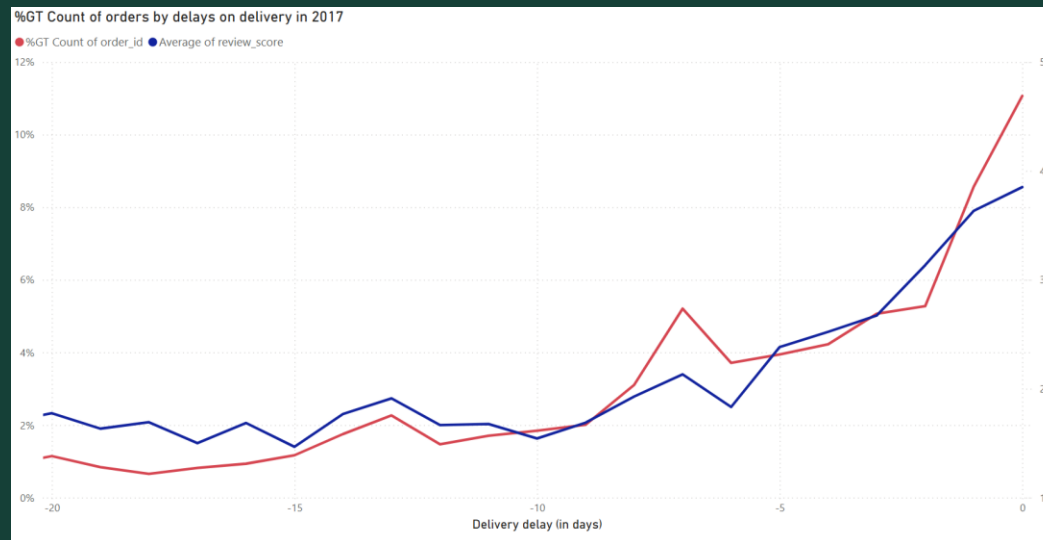


Upon closer examination, it appears that where customers were located was likely not the primary cause of late deliveries.

The purpose of emphasizing this visualization is to propose a connection between the delivery of goods via trucks and the resulting review scores. As seen in previous charts, SP again shows the highest number of delays, which is likely as its high sales volume means more deliveries, and consequently, more potential for delays in that area.

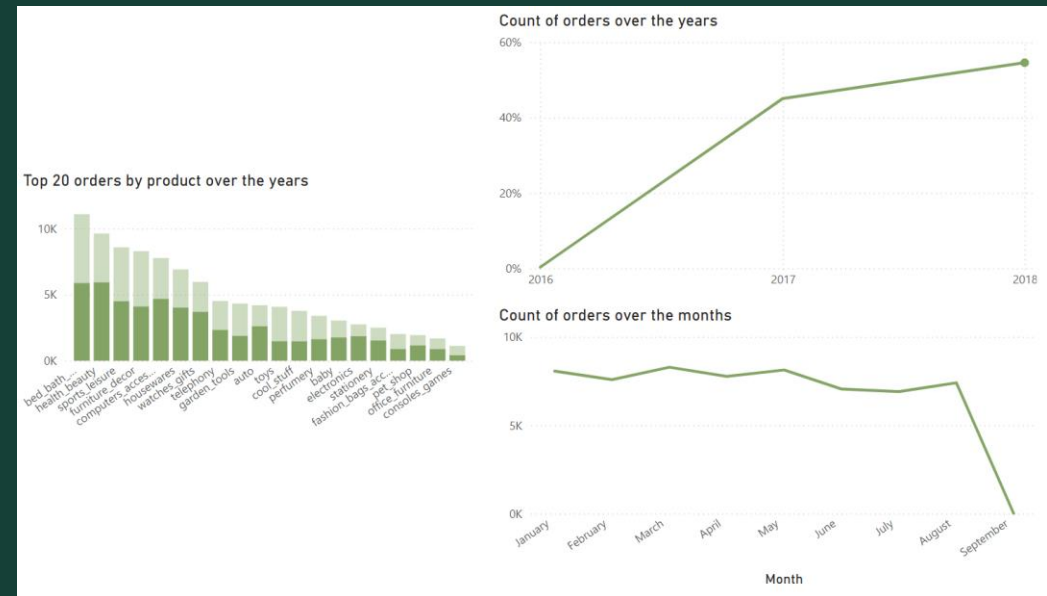


The data suggests a trend where more significant delays in order deliveries lead to poorer customer ratings.



KEY TAKEAWAYS & ACTION POINTS

- Olist performed very well on selling orders. For instance, comparison between 2017 and 2018 showcased the increase on sales by 21%, even though only up to September of 2018 was recorded in the dataset.



Drilled through chart for 2018 only.

KEY TAKEAWAYS & ACTION POINTS

- Based on the analysis, the data indicates that Brazilian e-commerce activity is highly concentrated in Brazil – particularly in states like São Paulo, Rio de Janeiro, and Minas Gerais – driven by both customer density and seller distribution.
- The significant increase in new customer orders from 2017 to 2018 presents a prime opportunity to convert these first-time buyers into repeat customers. To capitalize on this growth, it is recommended to implement targeted retention strategies such as personalized marketing campaigns, loyalty programs, and follow-up incentives to encourage repeat purchases.

KEY TAKEAWAYS & ACTION POINTS

- Additionally, since the analysis reveals a strong customer preference for credit card payments – especially when installment options are available – enhancing these payment features could further drive sales. The observed seasonal trends, particularly the dip in performance during September and the delivery delays in Q4 that resulted in lower review scores, suggest that logistics improvements are critical.
- Optimizing delivery processes (for example, by leveraging more reliable transportation methods and advanced tracking systems) could alleviate these issues, leading to improved customer satisfaction. These strategic initiatives not only aim to boost retention and customer loyalty but also promise to enhance overall operational efficiency and revenue growth, as corroborated by similar findings across industry analyses online.

```

1 import pandas as pd
2 from sqlalchemy import create_engine
3 import os
4
5 # Database connection details
6 db_user = 'root'
7 db_password = '****'
8 db_host = 'localhost'
9 db_name = 'paymongo'
10
11 # Directory containing the CSV files
12 csv_directory = r'C:\Users\Maki\Downloads\2024 PayMongo Data Analyst Assessment Kit\2024\Dataset'
13
14 # Create a connection to the database
15 engine = create_engine(f'mysql+mysqlconnector://{db_user}:{db_password}@{db_host}/{db_name}')
16
17 # Get a list of all files in the directory
18 all_files = os.listdir(csv_directory)
19
20 # Filter for CSV files
21 csv_files = [f for f in all_files if f.endswith('.csv')]
22
23 # Iterate through each CSV file
24 for csv_file in csv_files:
25     try:
26         # Construct the full file path
27         csv_file_path = os.path.join(csv_directory, csv_file)
28
29         # Read the CSV file into a DataFrame
30         df = pd.read_csv(csv_file_path)
31
32         # Extract the table name from the filename (remove the .csv extension)
33         table_name = os.path.splitext(csv_file)[0]
34         table_name = table_name.lower() # Convert to lowercase for consistency (optional)
35
36         # Import the DataFrame into the MySQL table
37         df.to_sql(table_name, con=engine, if_exists='replace', index=False)
38
39         print(f"Data from '{csv_file}' imported successfully into table '{table_name}'")
40
41     except Exception as e:
42         print(f"Error importing data from '{csv_file}': {e}")
43
44 print("\nFinished processing all CSV files.")
45
46

```



Importing csv files to local SQL database through python script

SQLTOOLS

CONNECTIONS

MySQL Local root@localhost:3306/...

paymongo database

Tables

customers

geolocation

olist_closed_deals_dataset

olist_customers_dataset

olist_marketing_qualified_leads_...

olist_order_items_dataset

olist_order_payments_dataset

olist_order_reviews_dataset

olist_orders_dataset

olist_products_dataset

olist_sellers_dataset

product_category_name_translat...

Views

BOOKMARKS

QUERY HISTORY

MySQL Local

select * FROM olist_orders_dataset 3/2...

MySQL Local:session.sql X

C:\Users\Waki> AppData\Local\vscode-sqltools>Data>session>MySQL Local:session.sql

1 select * FROM olist_orders_dataset

MySQL Local: select * FROM o...

order_id	customer_id	order_status	order_purchase...	order_approved...	order_delivered...
e481f51cbd54678b7cc49136f2d6af7	9ef432eb6251297304676186b10a928d	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-04 19:55:00
53cdb2fc8bc7dce0b6741e2150273451	b0830fb4747ac6d20dea0b8c802d7ef	delivered	2018-07-24 20:41:37	2018-07-26 03:24:27	2018-07-26 14:31:00
47770eb9100c2d0ca4946d9cfc07ece5d	41ce2a54c0b03bf3443c3d931a367089	delivered	2018-08-08 08:38:49	2018-08-08 08:55:23	2018-08-08 13:50:00
949d5b44dbf5de918f9ec16f97b45f8a	f88197465ea7920adcdcbec7375364d82	delivered	2017-11-18 19:28:06	2017-11-18 19:45:59	2017-11-22 13:39:59
ad21c59c0840e6cb3a9ceb5573f8159	8ab97904e6daea8866dbdbc4fb7aad2c	delivered	2018-02-13 21:18:39	2018-02-13 22:20:29	2018-02-14 19:46:34
a4591c265e18cb1dcee52889e2d8acc3	503740e9ca751ccdda7ba28e9ab8f608	delivered	2017-07-09 21:57:05	2017-07-09 22:10:13	2017-07-11 14:58:04
136cce7faa42fdb2cef53f9c79a6098	ed0271e0b7da060a393796590e7b737a	invoiced	2017-04-11 12:22:08	2017-04-13 13:25:17	NULL
6514b8ad8028c9f2cc2374ded245783f	9bdf00b4b3b52b5526ffa42d37d47f222	delivered	2017-05-16 13:10:30	2017-05-16 13:22:11	2017-05-22 10:07:46
76ce866289321a7c93b82b54852dc33	f54a9f0eb351c431402b8461ea51999	delivered	2017-01-23 18:29:09	2017-01-25 02:50:47	2017-01-26 14:16:31
e69bfb5eb88e0ed6a785585b27e16dbf	31ad1d1b63eb9962463f764d4e6e0c9d	delivered	2017-07-29 11:55:02	2017-07-29 12:05:32	2017-08-10 19:45:24
e6ce16cb79ec1d90b1da9085a6118aeb	494dded5b201313c6d4e7f100595b95c	delivered	2017-05-16 19:41:10	2017-05-16 19:50:18	2017-05-18 11:40:40
34513ce0c4fab462a55830c0989c7edb	7711cf621483d843aaf681855097bc37	delivered	2017-07-13 19:58:11	2017-07-13 20:10:08	2017-07-14 18:43:29
82566a660a982b15fb06e904cd32918	d3eb374c766bc6214e0c830b17ee2341	delivered	2018-06-07 10:06:19	2018-06-09 03:13:12	2018-06-11 13:29:00
5ff96c15d0b717ac6ad1f3d77225a350	19402a48fe860416adf93348aba37740	delivered	2018-07-25 17:44:10	2018-07-25 17:55:14	2018-07-26 13:16:00
432aaf21d85167c2c86ec9448c4e42cc	3df704f53d3f1d4818840b34ec672a9f	delivered	2018-03-01 14:14:28	2018-03-01 15:10:47	2018-03-02 21:09:20
dc3b6b511fac050b97cd5c05de84dc3	3b6828a50ffe546942b7a47d70ac0fc	delivered	2018-06-07 19:03:12	2018-06-12 23:31:02	2018-06-11 14:58:04
403b978360c04a622354cd531062e5f	738b086814c6fcc74b8cc583f8516ee3	delivered	2018-01-02 19:00:43	2018-01-02 19:09:04	2018-01-03 18:19:09
116f0b09343b49556bbad5f35bee0cdf	3187789bec990987628d7a9beb4dd6ac	delivered	2017-12-26 23:41:31	2017-12-26 23:50:22	2017-12-28 18:33:05
85ce859fd6dc634de8d2f1e290444043	059f7fc5719c7da6cbafe370971a8d70	delivered	2017-11-21 00:03:41	2017-11-21 00:14:22	2017-11-23 21:32:26
83018ec114eeeb641c97e08f7b4e926f	7f8c8b9c2ae27bf3300f670c3d478be8	delivered	2017-10-26 15:54:26	2017-10-26 16:08:14	2017-10-26 21:46:53
203096f03d82e0dfbc41ebc2ebdcfb7	d2b091571da224a1b36412c18bc3bbfe	delivered	2017-09-18 14:31:30	2017-09-19 04:04:09	2017-10-06 17:50:03
f848643eec1d69395095eb3840d2051e	4fa1cd166fa598be6de80fa84eade43	delivered	2018-03-15 08:52:40	2018-03-15 09:09:31	2018-03-15 19:52:48
2807d0e504d644894d41672727bc139f	72ae281627a6102d9b3718528b420f8a	delivered	2018-02-03 20:37:35	2018-02-03 20:50:22	2018-02-05 22:37:28
95266dbfb7e20354baba07964dac78d5	a166da34890074091a942054b36e4265	delivered	2018-01-08 07:55:29	2018-01-08 08:07:31	2018-01-24 23:16:37
f3e7c359154d965827355f396b1fdac	62b423aab58096ca514ba6aa06be2f98	delivered	2018-08-09 11:44:40	2018-08-10 03:24:51	2018-08-10 12:29:00
fbf9ac61453ac64ce8ad9783d7d0af6	3a874b4d4c4b6543206ff5d89287f0c3	delivered	2018-02-20 23:46:53	2018-02-22 02:30:46	2018-02-26 22:25:22
acce194856392f074dbf9dada14d8d82	7e20bf5ca92da68200643bda76c504c6	delivered	2018-06-04 00:00:13	2018-06-05 00:35:10	2018-06-05 13:24:00
dd78f56c0270f1906639c11b925620ea	8b212b9525f9e74e85e37ed6df37693e	delivered	2018-03-12 01:50:26	2018-03-12 03:28:34	2018-03-12 21:06:37
91b2a010e1e45e6ba3d133fa997597be	cce89a605105b148387c52e286ac8335	delivered	2018-05-02 11:45:38	2018-05-03 12:55:01	2018-05-10 16:16:00
ecab90c9933c58908d3d6add7c6f5ae3	761df82fed9778854c6dafdaeb567e4	delivered	2018-02-25 13:50:30	2018-02-25 14:47:35	2018-02-26 22:28:50
f70a0af17df5a6cd9a7196128bd354	456dc10730fbdb3a34615447ea195d643	delivered	2017-08-10 11:58:33	2017-08-12 02:45:24	2017-08-17 15:35:07
1790eea0b567cf50911c057cf20f90f9	52142aa69d8d0e1247ab0cada0f76023	delivered	2018-04-16 21:15:39	2018-04-16 22:10:26	2018-04-18 13:05:09
989225ba6d0eb5d78335f7e01de2ae7	816f8653d5361cbf94e58c33f2502a5c	delivered	2017-12-12 13:56:04	2017-12-14 13:54:13	2017-12-16 00:18:57
d887b52c6516eb39e8cd44a5f8b60f7	d9ef95f98d8da3b492bb8c0447910498	delivered	2018-02-03 12:38:58	2018-02-03 12:50:30	2018-02-05 21:26:53
b276e4f8c0fb86bd82fce578f21713e0	cf8ffeddf027932e51e4eae73b384059	delivered	2018-07-29 23:34:51	2018-07-29 23:45:15	2018-07-30 14:43:00
8563039e855156e48fccc4d611a3196	5f16605299d69860e06067eae2d2f9	delivered	2018-02-17 15:59:46	2018-02-17 16:15:34	2018-02-20 23:03:56
60550084e6b4c0cb89a87df1f3e5ebd9	f5458ddc3545711efa883dd7ae7c4497	delivered	2018-02-21 18:15:12	2018-02-23 02:10:52	2018-02-27 18:52:09
5acce57f8d9df55fa48e212a641a69d	295ae9b35379e07727387ff64354b6f	delivered	2017-07-31 21:37:10	2017-08-02 02:56:02	2017-08-03 18:32:48
434d158e96bdd6972ad6e6d73ddcd22	2a1dfb647f32f4390e7b057c67458536	delivered	2018-06-01 12:23:13	2018-06-05 03:35:15	2018-06-08 11:49:00
7206b86ea789983f7a273ea7f0bc2a8	3391c4bc11a817e7973e498b0b023158	delivered	2018-03-26 17:12:18	2018-03-26 17:28:27	2018-03-28 17:22:53
1e7aff52cddb2451ace09d0f848c3699	ddaff536587109b89777e0353215e150	delivered	2017-05-19 18:53:40	2017-05-19 19:05:17	2017-05-22 10:16:07

CONSOLE RE-RUN QUERY EXPORT OPEN

1-50 of 99441

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SQL CONSOLE COMMENTS

PS C:\Users\Waki>

Ln 1, Col 35 Spaces: 4 UTF-8 CRLF () SQL

Viewing and exploring all data through local SQL database

Table: closed deals (842 rows)

Importing data to Power Bi



Configuring table relationship in Power Bi based on the given ERD

