

**Portfolio Project # 01** **COURSE: DATA ANALYTICS**

SUBMITTED BY

**GROUP E - COHORT 12**EYMAN AMEER   
NAZISH ABDULLAH  
MUHAMMAD JAWAD HASAN

**SUMBITTED TO:** **SIR MISBAH UDDIN**

SUBMISSION DATE: 01-06-2024

Contents

1. Introduction…………………………………………………………………….……………..…02
2. [Data Dictionary 03](#_Toc168092838)
3. [ERD 04](#_Toc168092839)
4. [Objectives 05](#_Toc168092840)
5. [Problems .06](#_Toc168092841)
6. [SQL Queries …………………………………………………………………………….…….07](#_Toc168092844)
7. Visuals…………………………………………………………………………………..….……20
8. [Recommendations………………………………………………………………….….………25](#_Toc168092845)

**Introduction**

In this report dataset for a Brazilian ecommerce store Olist is taken from kaggle website. The data is publicly provided by Olist for the period of Oct2016 to Sep2018.

The dataset has information on 100k orders from 2016 to 2018 made at multiple

marketplaces in Brazil. This is real commercial data, it has been anonymized, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

Olist Store is one of the leading stores in the Brazilian e-commerce market which partners with businesses from all over Brazil to facilitate them to reach a large customer base in all states of Brazil.

The dataset covers all aspects of the revenue streams along with the geographical location of customers and sellers. The store has a significant number of product categories which have multiple products sold under them.

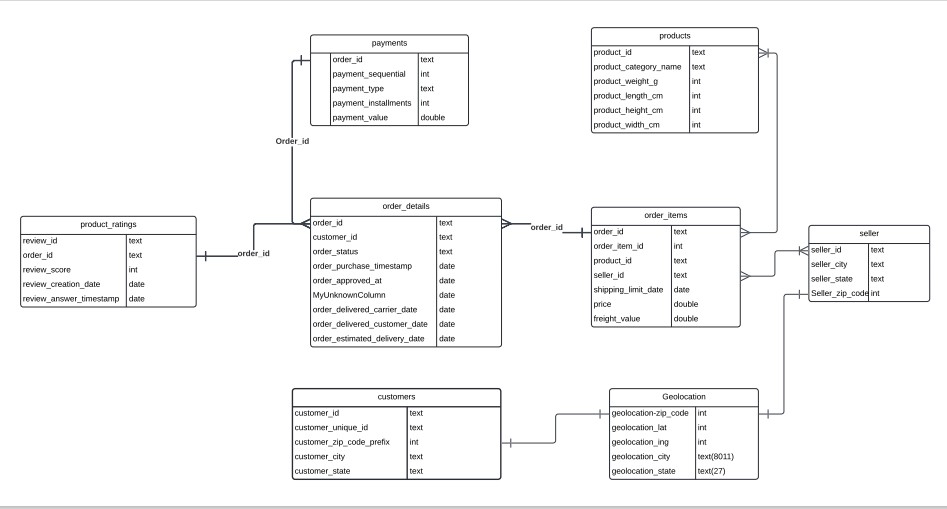
Following are the datasheets provided by the store:

1. Customers
2. Sellers
3. Products
4. Product category
5. Geolocation
6. Order details
7. Order items
8. Order payments
9. Order reviews

**Data Dictionary**



**Entity Relationship Diagram**



The above diagram defines one to one and one-to-many relationship between the fields.

**Objective**

In this report, we have used the given datasets to analyze Olist sales concerning different business variables.

Olist has over 70 product categories, which provides the customer with a single platform to fulfil their shopping needs. In our analysis, we aim to find the most selling products, high sales generating streams, and different variants associated with them.

**Business Problem**

Olist store is one of the leading e-commerce stores in Brazil. Its customer and supplier base is spread across the country with a diversified portfolio of products. To have a clear picture of how and which factors affect the sales of the company, we look into following four dimensions

* How many ***product categories*** and products does the store offer? Which are the most selling product categories and generate high sales revenues?
* What is the pattern in sales generated across ***different states*** in Brazil? How do customer demographics affect the sales?
* Which ***payment methods*** are used while purchasing from the Olist platform? And how does it impact sales?
* What are the ***seasonal trends*** in sales? In which season do customer are more inclined towards purchasing and what are their product preferences?

For our analysis, we imported the Olist datasheets in MYSQL Workbench and MS PowerBi.

In the following section, identified problems are explained using queries, run on MYSQL Workbench supported with visuals from the PowerBi dashboard.

**Problems**

On searching through data, following statements were identified for analysis.

**Query no.1**

To calculate the number of orders in each product category placed from Sep 2016 to Aug 2018. This will result in finding out the high-in-demand product categories and the volume of sales they generate.

With Products\_Categorywise AS

(

Select o.order\_id, p.product\_category\_name,

DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s") as

purchase\_date, o.price

from order\_items as o

Join products as p

on o.product\_id= p.product\_id

Join olist\_orders\_dataset as od

on o.order\_id= od.order\_id

where DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s") > '2016-09-31 00:00:00'

and DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s") < '2018-09-00 00:00:00' and

od.order\_status != 'canceled'

)

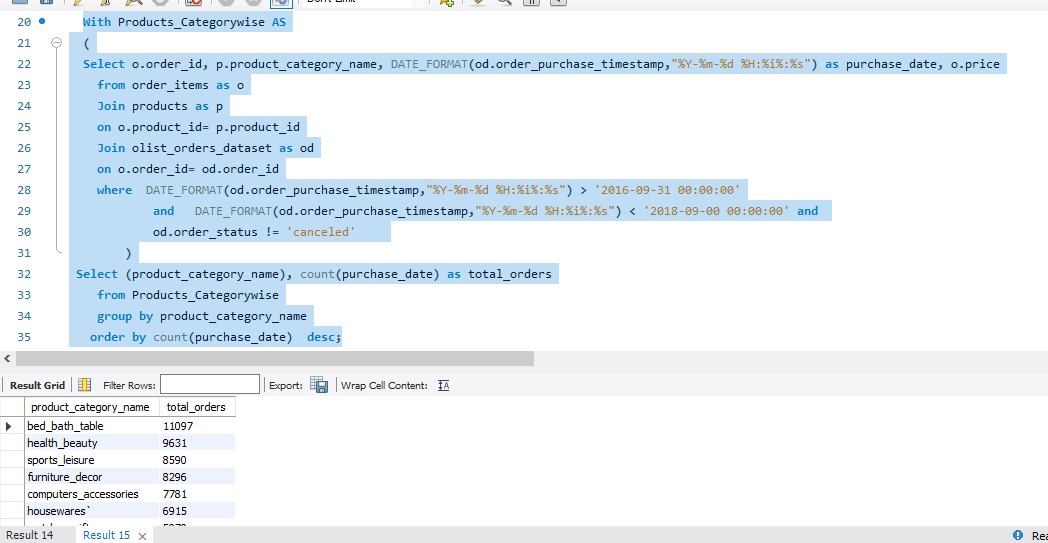
Select (product\_category\_name), count(purchase\_date) as total\_orders

from Products\_Categorywise

group by product\_category\_name

order by count(purchase\_date) desc;

*Initially a CTE named Products\_categorywise is designed to be used in few of the queries. This query converts the text type in date format in order items table and extract selected fields for the orders executed from Oct 2016 to Sep 2018.*



**Insight**

From the above result, it can be concluded that bed and bath had over 11000 orders during the period followed closely by health and beauty products.

**Query no.2**

To calculate the total sales revenue for each product category purchased from Sep 2016 to Aug 2018. This will result in finding out the high sales revenue generating product categories.

With Products\_Categorywise AS

(

Select o.order\_id, p.product\_category\_name, DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s") as purchase\_date, o.price

from order\_items as o

Join products as p

on o.product\_id= p.product\_id

Join olist\_orders\_dataset as od

on o.order\_id= od.order\_id

where DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s") > '2016-09-31 00:00:00'

and DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s") < '2018-09-00 00:00:00' and

od.order\_status != 'canceled'

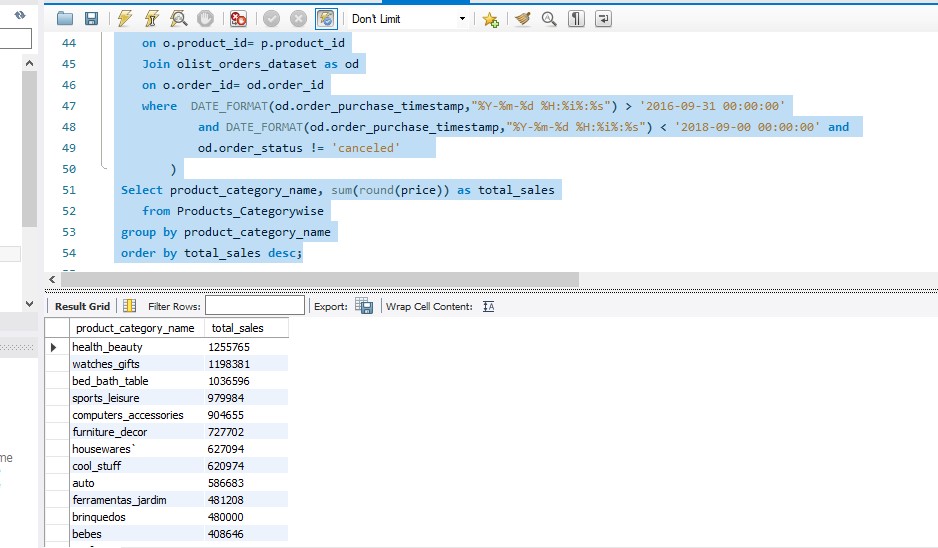
)

Select product\_category\_name, sum(round(price)) as total\_sales

from Products\_Categorywise

group by product\_category\_name

order by total\_sales desc;



**Insight**

Health & beauty products have brought in the highest revenue of 12.5 MN BRL, whereas watches and gifts have relatively fewer orders but have bought a high revenue of almost 1.2 MN BRL.

**Query no.3**

To find average review ratings for top-selling products.

Select p.product\_category\_name, (sum (pr. review\_score)/count(pr.review\_score)) as average\_rating

from order\_items as o

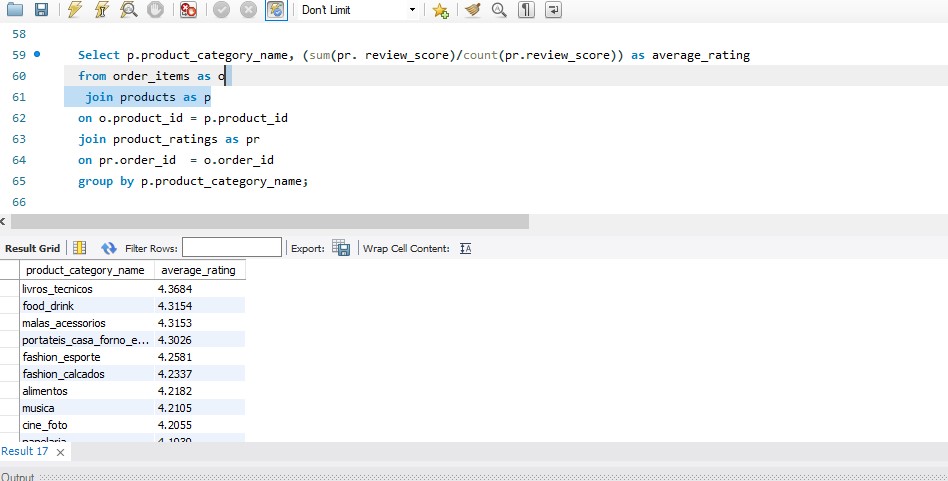
join products as p

on o.product\_id = p.product\_id

join product\_ratings as pr

on pr.order\_id = o.order\_id

group by p.product\_category\_name;



**Insight**

The maximum average review rate is 4.3, which means there are areas where customers' expectations are not met. The underlying reason may be delayed delivery time or high freight charges for low-priced products. The following table shows the average review rating for top-selling product categories

|  |  |
| --- | --- |
| **Product\_Category\_Name with no.orders >5000** | **Average\_Rating** |
| bed\_bath\_table | 3.8957 |
| health\_beauty | 4.1428 |
| sports\_leisure | 4.108 |
| furniture\_decor | 3.9035 |
| computers\_accessories | 3.9308 |
| housewares` | 4.055 |
| watches\_gifts | 4.0192 |

**Query no.4**

To find sales pattern over the quarters in two years.

Select

Case

when extract(month from purchase\_date) between 10 and 12 and extract(year from purchase\_date) = 2016 then 'Qtr12017'

when extract(month from purchase\_date) between 01 and 03 and extract(Year from purchase\_date) = 2017 then 'Qtr22017'

when extract(month from purchase\_date) between 04 and 06 and extract(year from purchase\_date) = 2017 then 'Qtr32017'

when extract(month from purchase\_date) between 07 and 09 and extract(year from purchase\_date) = 2017 then 'Qtr42017'

else

case

when extract(month from purchase\_date) between 10 and 12 and extract(year from purchase\_date) = 2017 then 'Qtr12018'

when extract(month from purchase\_date) between 01 and 03 and extract(Year from purchase\_date) = 2018 then 'Qtr22018'

when extract(month from purchase\_date) between 04 and 06 and extract(year from purchase\_date) = 2018 then 'Qtr32018'

when extract(month from purchase\_date) between 07 and 09 and extract(year from purchase\_date) = 2018 then 'Qtr42018'

else null

End

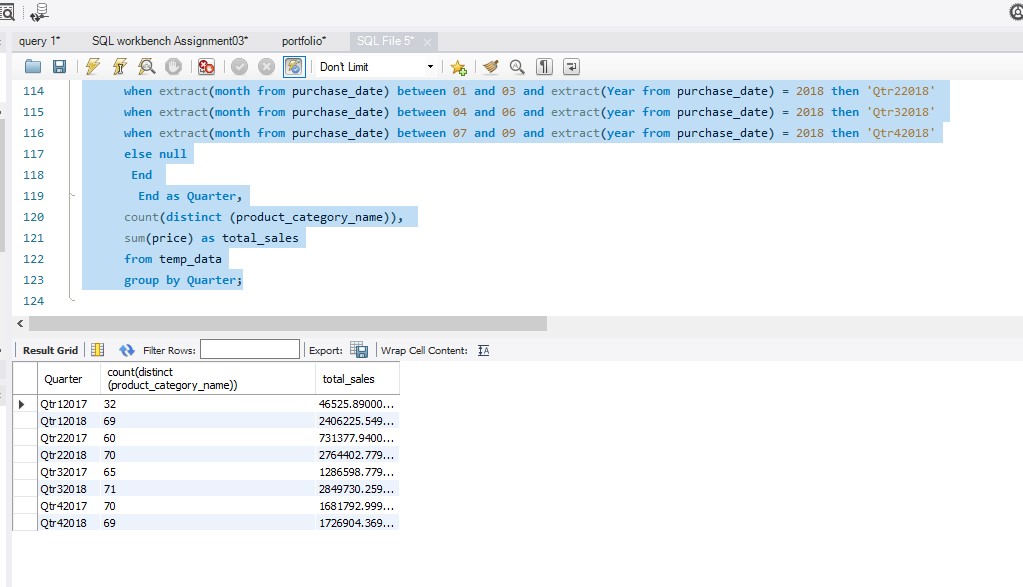
End as Quarter,

count(distinct (product\_category\_name)),

sum(price) as total\_sales

from temp\_data

group by Quarter;



**Insight**

There was a sharp rise in sales from quarter 1 of 2017 to quarter 1 of 2018 because Olist introduced new product categories in Jan 2017. This is also verified by an increase in sales in 2nd quarter of 2017 only. The gap between the revenues in the following quarters shows a decline as we see more product categories subsequently introduced in 2017.

**Query no.5**

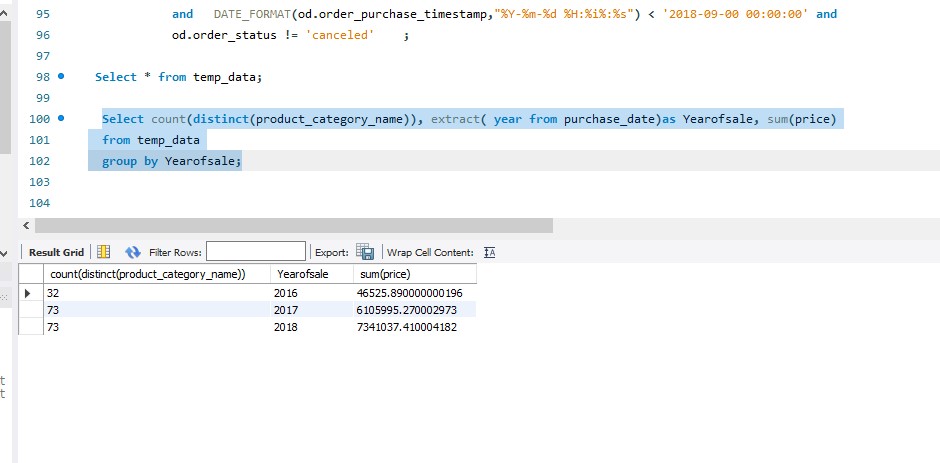
To find number of product categories sold during two years.

This query will support the result derived from our previous query.

Select count(distinct(product\_category\_name)), extract( year from purchase\_date)as Yearofsale, sum(price)

from temp\_data

group by Yearofsale;



**Insight**

From the above table, it is evident that sale revenue has substantially increased in year 2017 due to the addition of multiple new product categories on the store’s platform.

**Query no.6**

To find number of customers residing in different states of Brazil.

This query will provide insight on customer demographics.

Select

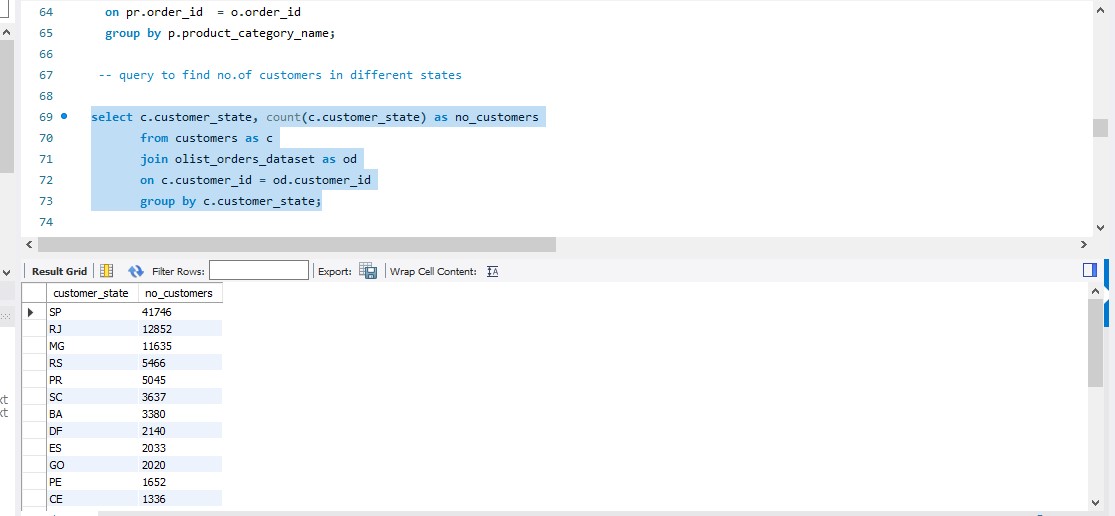
c.customer\_state, count(c.customer\_state) as no\_customers

from customers as c

join olist\_orders\_dataset as od

on c.customer\_id = od.customer\_id

Group by c.customer\_state;



**Insight**

It is observed that the highest number of customers are from SP, and that’s comprehensible as São Paulo is the largest and most populous state in Brazil, located in the Southeast Region. It has more than 600 municipalities, among which few are known for being extremely advanced in technology.

Due to high awareness and accessibility to internet and networks Olist makes high profit from urban areas.

**Query no.7**

To find where are the most sellers located in different states of Brazil.

Select

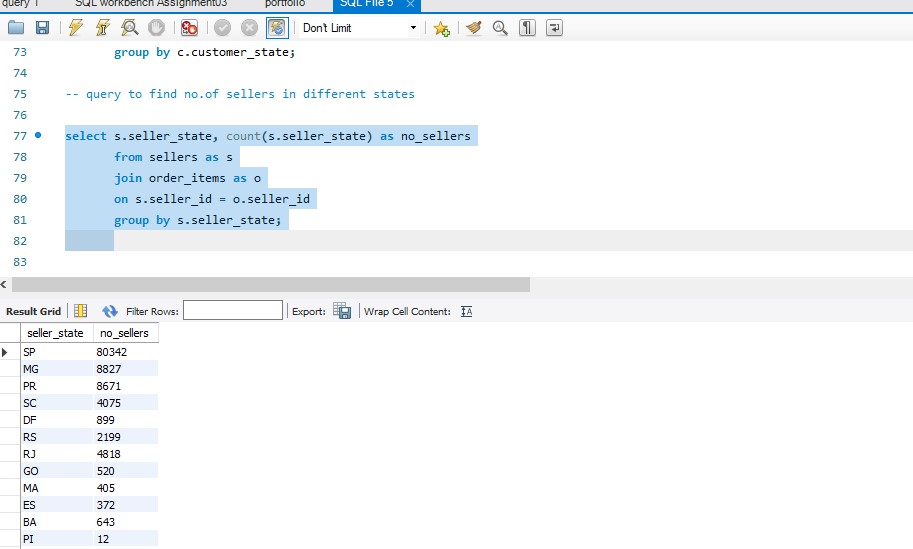
s.seller\_state, count(s.seller\_state) as no\_sellers

from sellers as s

join order\_items as o

on s.seller\_id = o.seller\_id

group by s.seller\_state;



**Insights**

The majority of the suppliers are from SP. As SP is the most populated and developed state in Brazil, local businesses in this area are well aware of the usage of e-commerce platforms and modern modes of executing business transactions.

**Query no.8**

To find delivery days in different states

Select c.customer\_state as state , DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s") as purchase\_date,

DATE\_FORMAT(od.order\_delivered\_customer\_date,"%Y-%m-%d %H:%i%:%s") as delivery\_date, datediff(DATE\_FORMAT(od.order\_delivered\_customer\_date,"%Y-%m-%d %H:%i%:%s"),

DATE\_FORMAT(od.order\_purchase\_timestamp,"%Y-%m-%d %H:%i%:%s")

) as delivery\_days

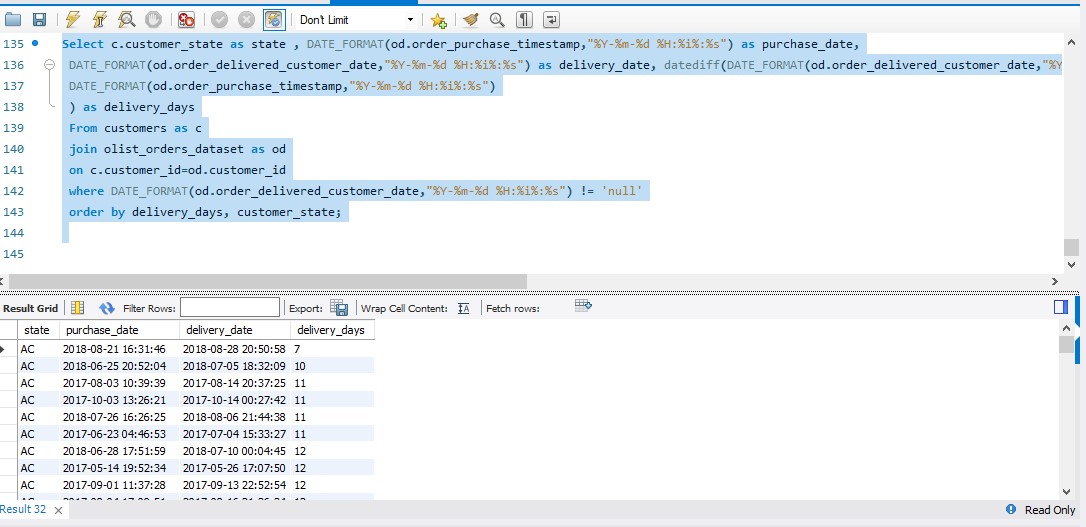
From customers as c

join olist\_orders\_dataset as od

on c.customer\_id=od.customer\_id

where DATE\_FORMAT(od.order\_delivered\_customer\_date,"%Y-%m-%d %H:%i%:%s") != 'null'

order by delivery\_days, customer\_state;



**Insights**



The above table clearly shows that average delivery period offered to customers in states having low sales is above 20 days.

**Query no.9**

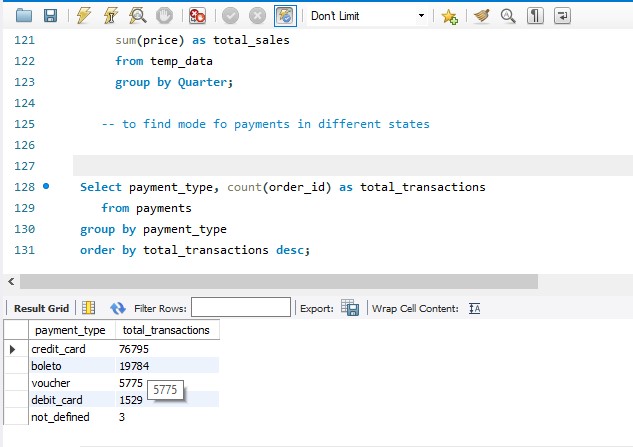
To find customer preferences for using various modes of payment in different states.

Select payment\_type, count(order\_id) as total\_transactions

from payments

group by payment\_type

order by total\_transactions desc;



**Insight**

The above query result reflects on extensive usage of credit cards for payments by customers. The reason behind this trend may be promotional campaigns introduced by Olist in partnership with credit card companies. Further, it is also observed that payments are received in multiple installments when purchases are made via credit cards.

**Query no. 10**

To find number of products in each category

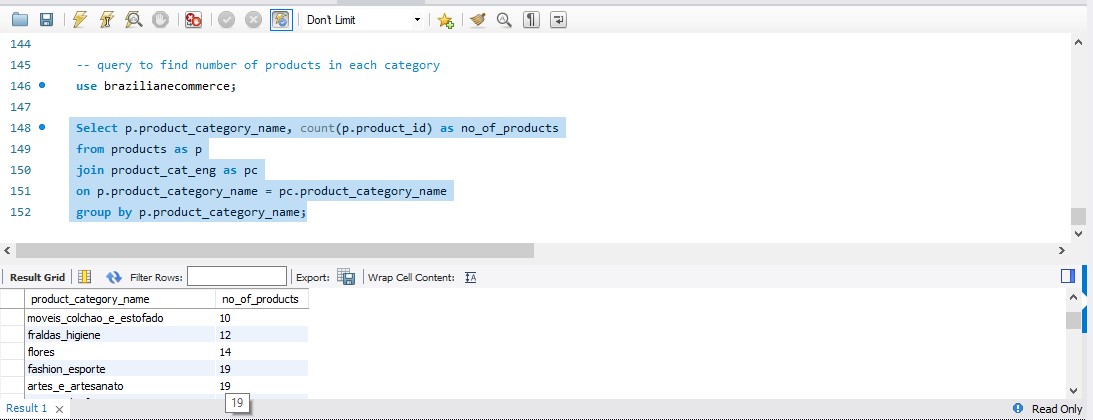
Select p.product\_category\_name, count(p.product\_id) as no\_of\_products

from products as p

join product\_cat\_eng as pc

on p.product\_category\_name = pc.product\_category\_name

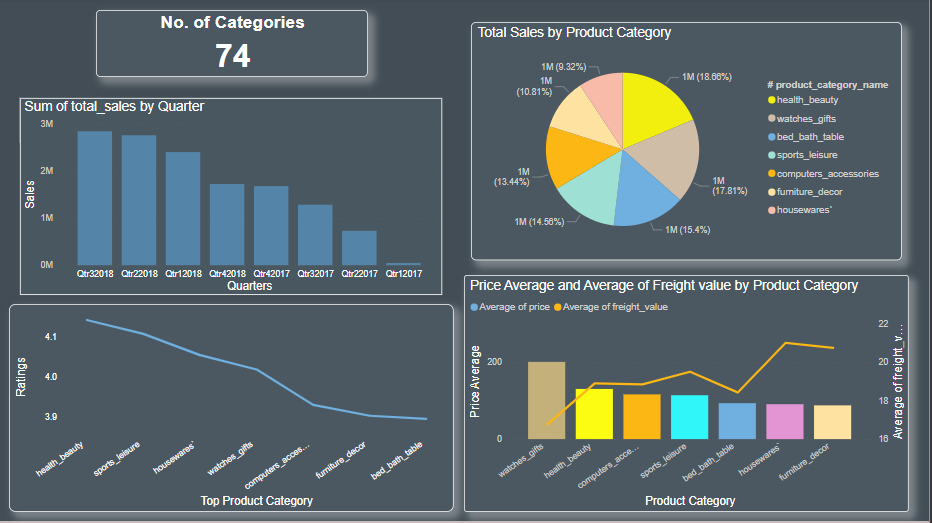
group by p.product\_category\_name;



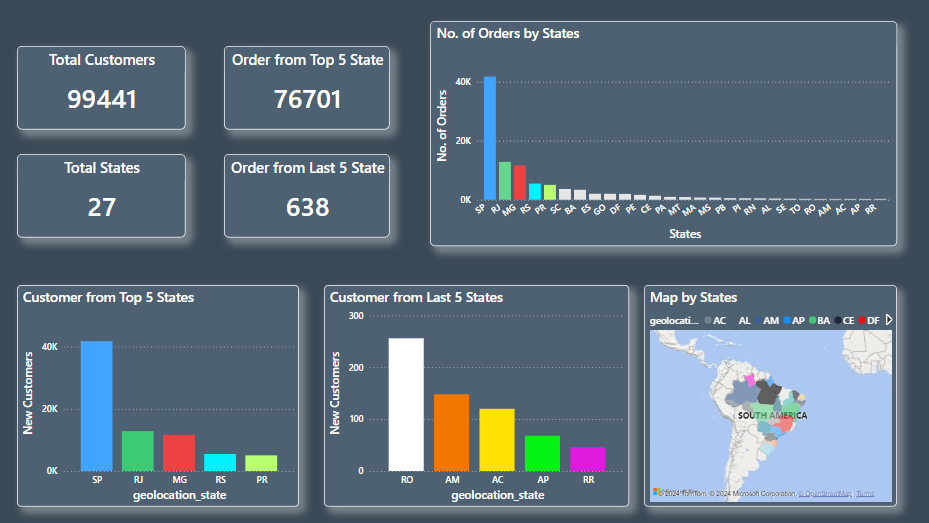
**Dashboards**

The dataset has been studied in detail using MS PowerBi and dashboards have been designed to give a holistic synopsis of our report.

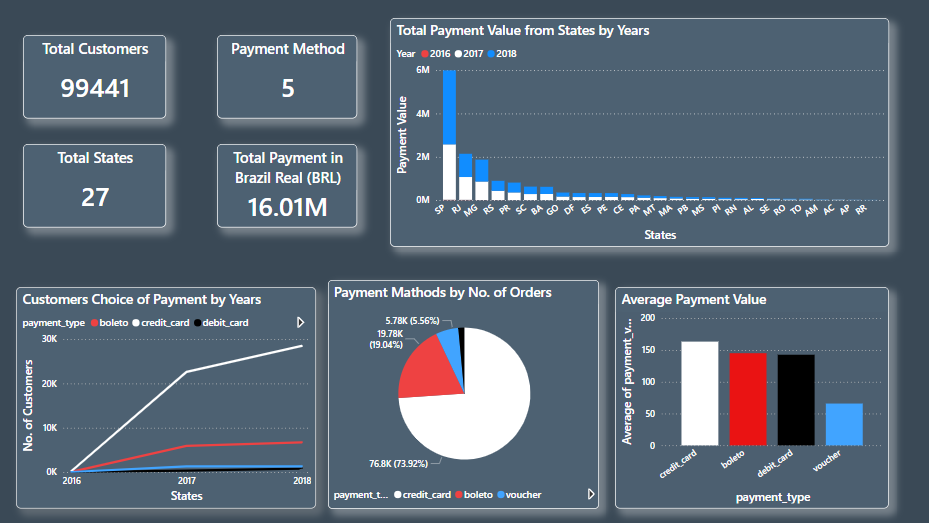
**Analysis by Product Category**

****

**Regional Disparity Analysis**

****

**Payment Method Analysis**

****

**Seasonal Analysis**

**A screenshot of a graph

Description automatically generated**

**Recommendations**

* Increasing trend in sales indicates opportunities for further growth, hence Olist should strengthen its network across the country and adopt strategies to reach more number of customers.
* Olist should focus on partnering with sellers who provide competitive products in the categories of bed & bath, health & beauty, sports leisure, and furniture decoration.
* The Brazilian potential customer base is enormous and with appropriate market analysis, Olist can penetrate untapped regions.
* At the same time by adopting technology advancements and understanding customer demands, the increasing trend in sales cannot only be retained but also taken to the next level.
* In SP and other developed regions, further partnerships should be made to capitalize on the increasing trend of online shopping.
* For low-sales regions, implement targeted marketing campaigns to raise awareness and stimulate demand.
* Ensure the availability of popular payment methods in each region.
* Launch marketing campaigns and promotions around a high-demand season of summer to maximize sales.