**SQL Business Case**

**What does 'good' look like?**

**1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**

1.1 **Data type of all columns in the "customers" table**.

**QUERY:**

SELECT

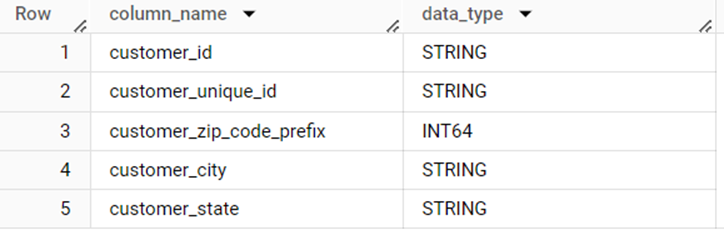
column\_name,

data\_type

FROM `target-411410.target.INFORMATION\_SCHEMA.COLUMNS`

WHERE table\_name = "customers"

**OUTPUT:**

****

**INSIGHTS:**

**From the output, we can see that the customer customer table has 5 columns which has**

**· String datatype for 4 columns**

**· INT64(Integer) datatype for 1 column**

**1.2 Get the time range between which the orders were placed.**

**QUERY:**

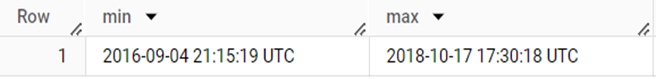
**SELECT**

**MIN(order\_purchase\_timestamp) AS `min`,**

**MAX(order\_purchase\_timestamp) AS `max`**

**FROM `target.orders`**

**OUTPUT:**

****

**INSIGHTS:**

* **Minimum time** at which the order has been placed is **2016-09-04 21:15:19 UTC**
* **Maximum time** at which the order has been placed is **2018-10-17 17:30:18 UTC**

**1.3 Count the Cities & States of customers who ordered during the given period.**

**QUERY:**

**SELECT**

**COUNT(DISTINCT c.customer\_city) AS `city`,**

**COUNT(DISTINCT c.customer\_state) AS `state`**

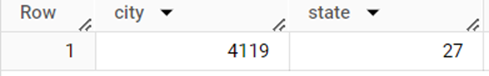
**FROM**

**`target.orders` AS `o` INNER JOIN**

**`target.customers` AS `c`**

**ON c.customer\_id = o.customer\_id**

**OUTPUT:**

****

**INSIGHTS:**

In the given period of time, customers in **4119 cities** have ordered from **27 states**.

2. **In-depth Exploration:**

**2.1** **Is there a growing trend in the no. of orders placed over the past years?**

**QUERY:**

SELECT

EXTRACT(YEAR FROM order\_purchase\_timestamp) AS `year`,

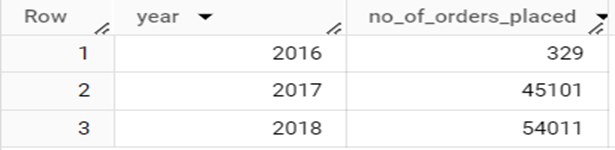
COUNT(order\_id) AS `no\_of\_orders\_placed`

FROM `target.orders`

GROUP BY 1

ORDER BY 1

**OUTPUT:**



**INSIGHTS:**

**Yes**, the number of orders placed has increased, which means there is a **growing trend** in the orders placed over the past years.

**RECOMMENDATIONS:**

To increase the number of orders, few things can be implemented. Some of them are:

· Email Marketing

· Frequent Updates and New Arrivals

· Social Media Engagement

· Promote products by personalized recommendations

· Having a user-friendly website

**2.2 Can we see some kind of monthly seasonality in terms of the no. of orders being placed?**

**QUERY:**

SELECT

EXTRACT(YEAR FROM order\_purchase\_timestamp) AS `year`,

EXTRACT(MONTH FROM order\_purchase\_timestamp) AS `month`,

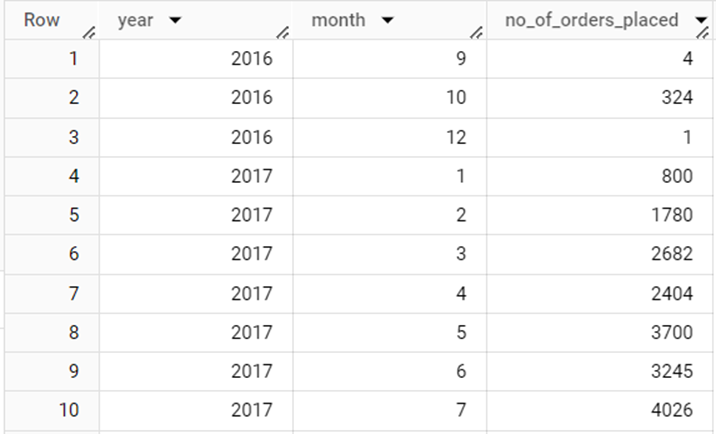
COUNT(order\_id) AS `no\_of\_orders\_placed`

FROM `target.orders`

GROUP BY 1,2

ORDER BY 1,2

**OUTPUT:**



**INSIGHTS:**

* The number of orders placed on each month varies according to year.
* Since the dataset doesn’t have all the months in the year 2016 and 2018, it is hard to guess the seasonality month.
* In the year 2017, the highest number of orders were placed in November and December.
* However, for 2018, the highest number of orders was placed in January and March.

**2.3** **During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)**

**0-6 hrs: Dawn**

**7-12 hrs: Mornings**

**13-18 hrs: Afternoon**

**19-23 hrs: Night**

**QUERY:**

SELECT

(IF(`hour` BETWEEN 0 AND 6,"Dawn",IF( (

`hour` BETWEEN 7 AND 12),"Mornings",IF( (

`hour` BETWEEN 13 AND 18),"Afternoon","Night"

)))) AS `time\_of\_the\_day`,

COUNT(order\_id) AS `no\_of\_orders\_placed`

FROM (SELECT

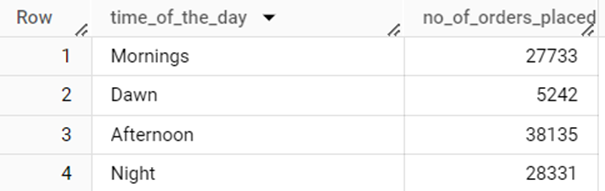
EXTRACT(HOUR FROM order\_purchase\_timestamp) AS `hour`,

order\_id

FROM `target.orders`) AS `t`

GROUP BY 1

**OUTPUT:**

****

**INSIGHTS:**

* Brazilian customers place their orders in the afternoon mostly, followed by night, mornings and dawn.
* This shows that most of the orders are placed on Afternoon.

**3. Evolution of E-commerce orders in the Brazil region:**

**3.1** **Get the month-on-month no. of orders placed in each state.**

**QUERY:**

SELECT

c.customer\_state AS `state`,

EXTRACT(MONTH FROM order\_purchase\_timestamp) AS `month`,

EXTRACT(YEAR FROM order\_purchase\_timestamp) AS `year`,

COUNT(order\_id) AS `no\_of\_orders\_placed`

FROM

`target.orders` AS `o` INNER JOIN

`target.customers` AS `c`

ON o.customer\_id = c.customer\_id

GROUP BY 1,2,3

**OUTPUT:**



**INSIGHTS:**

* The above output shows how many orders have been placed by **each state** in **each month** for **each year**.
* The first row shows that **1048** orders were placed from the state **RJ** in **November 2017.**

**3.2** **How are the customers distributed across all the states?**

**QUERY:**

SELECT

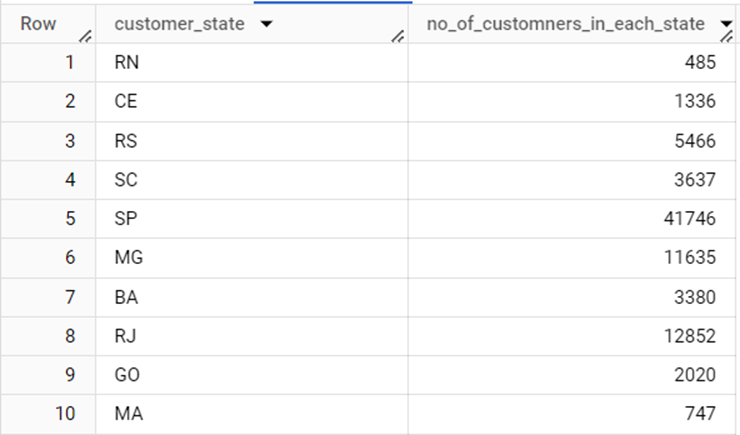
customer\_state,

COUNT(customer\_id) AS `no\_of\_customners\_in\_each\_state`

FROM `target.customers`

GROUP BY 1

**OUTPUT:**



**INSIGHTS:**

* The above output shows how many **customers** are there in **each state**.
* The state RN has 485 customers.

**4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**

4.1 **Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only). You can use the "payment\_value" column in the payments table to get the cost of orders**.

**QUERY:**

WITH `prev\_year` AS

(SELECT

sum(payment\_value) AS `sum1`,

EXTRACT(YEAR FROM order\_purchase\_timestamp) AS `year`

FROM

`target.payments` AS `p` INNER JOIN

`target.orders` AS `o`

ON p.order\_id = o.order\_id

WHERE

(EXTRACT(MONTH FROM order\_purchase\_timestamp) BETWEEN 1 AND 8) AND

EXTRACT(YEAR FROM order\_purchase\_timestamp) = 2017

GROUP BY 2),

`current\_year` AS

(SELECT

sum(payment\_value) AS `sum2`,

EXTRACT(YEAR FROM order\_purchase\_timestamp) AS `year`

FROM

`target.payments` AS `p` INNER JOIN

`target.orders` AS `o`

ON p.order\_id = o.order\_id

WHERE

(EXTRACT(MONTH FROM order\_purchase\_timestamp) BETWEEN 1 AND 8) AND

EXTRACT(YEAR FROM order\_purchase\_timestamp) = 2018

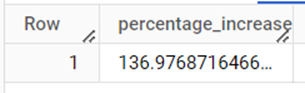
GROUP BY 2)

SELECT

(sum2 - sum1)/sum1 \* 100 AS `percentage\_increase`

FROM prev\_year, current\_year

**OUTPUT:**



**INSIGHTS:**

The percentage of cost of orders placed by customers has been increased by **137%** in 2018 from 2017 based on the months between Jan to Aug.

**4.2** **Calculate the Total & Average value of order price for each state.**

**QUERY:**

SELECT

customer\_state,

ROUND(AVG(payment\_value),0) AS `avg\_price`,

ROUND(SUM(payment\_value),0) AS `total\_price`

FROM

`target.payments` as `p` INNER JOIN

`target.orders` AS `o`

ON p.order\_id = o.order\_id

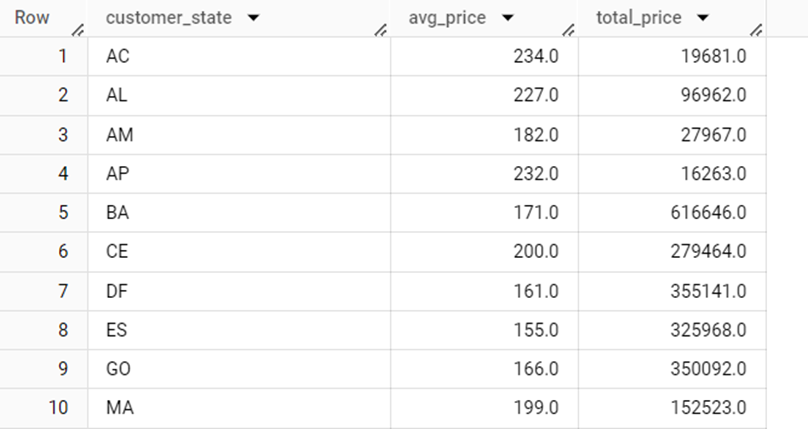
INNER JOIN `target.customers` AS `c`

ON o.customer\_id = c.customer\_id

GROUP BY 1

ORDER BY 1

**OUTPUT:**



**INSIGHTS:**

* The above output shows the average price and total price for each state.
* The state **AC** has an average order value of **234** and total order value of **19681**.

**4.3** **Calculate the Total & Average value of order freight for each state.**

**QUERY:**

SELECT

customer\_state,

ROUND(AVG(freight\_value),0) AS `avg\_freight\_value`,

ROUND(SUM(freight\_value),0) AS `total\_freight\_value`

FROM `target.order\_items` as `oi` INNER JOIN

`target.orders` as `o`

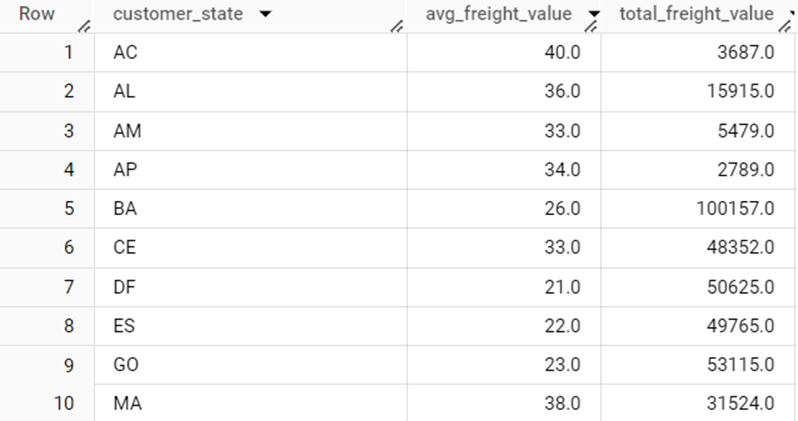
ON oi.order\_id = o.order\_id

INNER JOIN `target.customers` AS `c`

ON o.customer\_id = c.customer\_id

GROUP BY 1

ORDER BY 1

**OUTPUT:** 

**INSIGHTS:**

The above output shows the average freight value and total freight value for each state. The state **AC** has the average freight value of **40** and total freight value of **3687**.

**5. Analysis based on sales, freight and delivery time.**

**5.1 Find the no. of days taken to deliver each order from the order’s purchase date as delivery time.  
Also, calculate the difference (in days) between the estimated & actual delivery date of an order.  
Do this in a single query.** **You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:**

**time\_to\_deliver = order\_delivered\_customer\_date - order\_purchase\_timestamp**

**diff\_estimated\_delivery = order\_delivered\_customer\_date - order\_estimated\_delivery\_date**

**QUERY:**

SELECT

order\_id,

DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY) AS `time\_to\_deliver`,

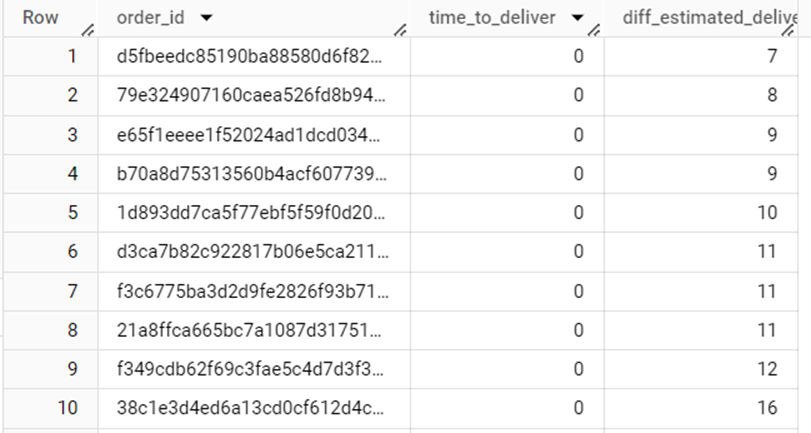
DATE\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,DAY) AS `diff\_estimated\_delivery`,

FROM `target.orders`

WHERE order\_delivered\_customer\_date IS NOT NULL AND order\_status = "delivered"

ORDER BY 2,3

**OUTPUT:**

****

**5.2 Find out the top 5 states with the highest & lowest average freight value.**

**QUERY:**

WITH `MAX` AS

(SELECT

customer\_state,

MAX(avg) AS `highest`

FROM (SELECT

DISTINCT customer\_state,

ROUND(AVG(freight\_value),2) AS `avg`

FROM

`target.customers` AS `c` INNER JOIN

`target.orders` AS `o`

ON c.customer\_id = o.customer\_id

INNER JOIN `target.order\_items` AS `oi`

on o.order\_id = oi.order\_id

GROUP BY 1

ORDER BY 2 DESC

) AS `t`

GROUP BY 1

ORDER BY 2 DESC

LIMIT 5),

`MIN` AS

(SELECT

customer\_state,

MIN(avg) AS `lowest`

FROM (SELECT

DISTINCT customer\_state,

ROUND(AVG(freight\_value),2) AS `avg`

FROM

`target.customers` AS `c` INNER JOIN

`target.orders` AS `o`

ON c.customer\_id = o.customer\_id

INNER JOIN `target.order\_items` AS `oi`

on o.order\_id = oi.order\_id

GROUP BY 1

ORDER BY 2

LIMIT 5

) AS `t`

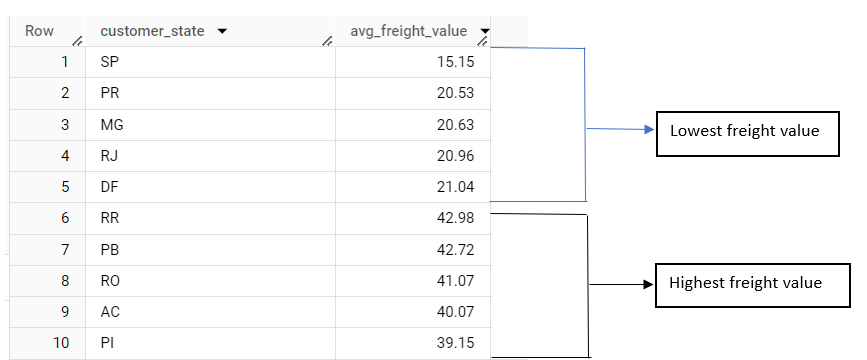
GROUP BY 1

ORDER BY 2 DESC)

(SELECT \* FROM `MAX`)

UNION ALL

(SELECT \* FROM `MIN`)

**OUTPUT:** 

**INSIGHTS:**

* The states **SP, PR, MG,RJ and DF** are the top 5 states with **lowest freight value**.
* The states **RR, PB, RO, AC, PI** are the top 5 states with **highest freight value.**

**5.3 Find out the top 5 states with the highest & lowest average delivery time.**

**QUERY:**

WITH Highest AS

(SELECT

customer\_state,

ROUND(AVG(DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY)),2) AS `average\_time\_to\_deliver`

FROM

`target.orders` AS `o` INNER JOIN

`target.customers` AS `c`

ON o.customer\_id = c.customer\_id

GROUP BY 1

ORDER BY 2 desc

LIMIT 5),

Lowest AS

(SELECT

customer\_state,

ROUND(AVG(DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY)),2) AS `average\_time\_to\_deliver`

FROM

`target.orders` AS `o` INNER JOIN

`target.customers` AS `c`

ON o.customer\_id = c.customer\_id

GROUP BY 1

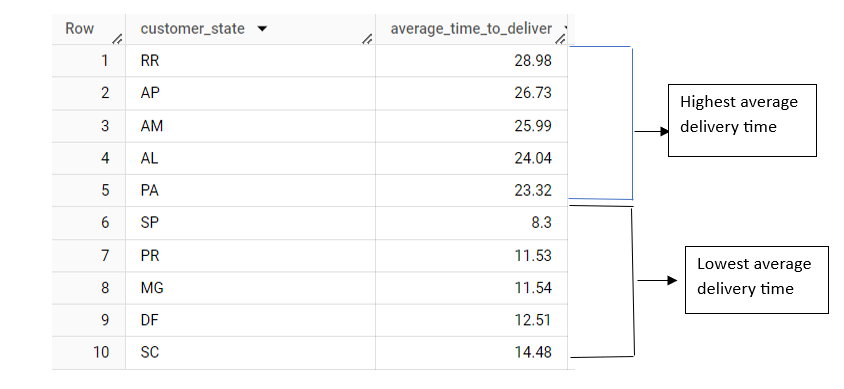
ORDER BY 2 ASC

LIMIT 5)

SELECT \* FROM Highest

UNION ALL

SELECT \* FROM Lowest

**OUTPUT:**

**INSIGHTS:**

* The states **RR, AP, AM, AL, PA** have the **highest average delivery time.**
* The states **SP, PR, MG, DF, SC** have the lowest **average delivery time.**

**5.4 Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.  
 You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.**

**QUERY:**

SELECT

customer\_state,

ROUND(AVG(DATE\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,DAY)),2) AS `average\_difference\_between\_expected\_and\_actual\_delivery\_dates`

FROM

`target.orders` AS `o` INNER JOIN

`target.customers` AS `c`

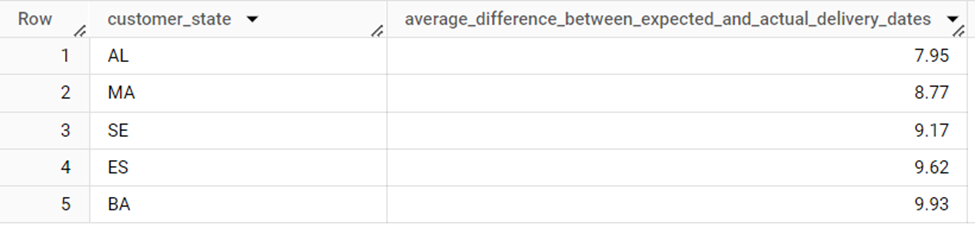
ON o.customer\_id = c.customer\_id

WHERE order\_delivered\_customer\_date IS NOT NULL

GROUP BY 1

ORDER BY 2 ASC

LIMIT 5

**OUTPUT:**

6. **Analysis based on the payments:**

**6.1** **Find the month on month no. of orders placed using different payment types.**

**QUERY:**

SELECT

DISTINCT format\_date('%Y-%m', order\_purchase\_timestamp) as year\_month,

payment\_type,

COUNT(o.order\_id) AS `no\_of\_orders\_placed`

FROM

`target.orders` AS `o` INNER JOIN

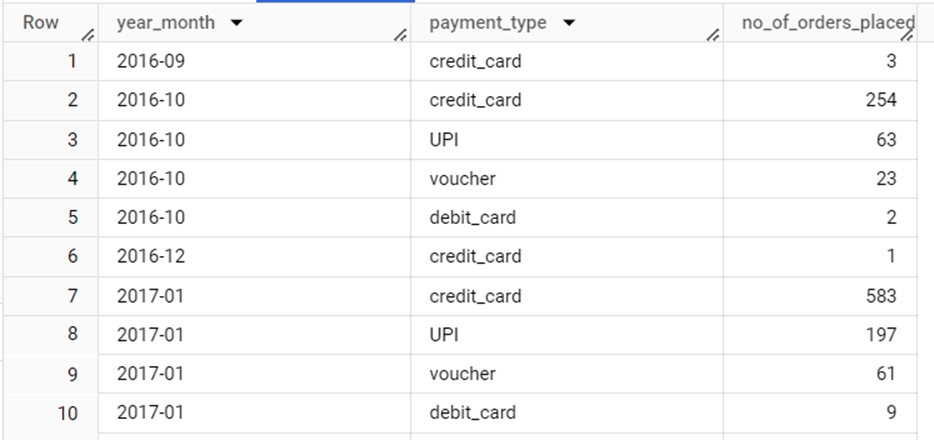
`target.payments` AS `p`

ON o.order\_id = p.order\_id

GROUP BY 1,2

ORDER BY 1

**OUTPUT:**

****

**6.2 Find the no. of orders placed on the basis of the payment installments that have been paid**

**QUERY:**

SELECT

payment\_installments,

COUNT(order\_id) AS `count\_of\_payment\_installment`

FROM

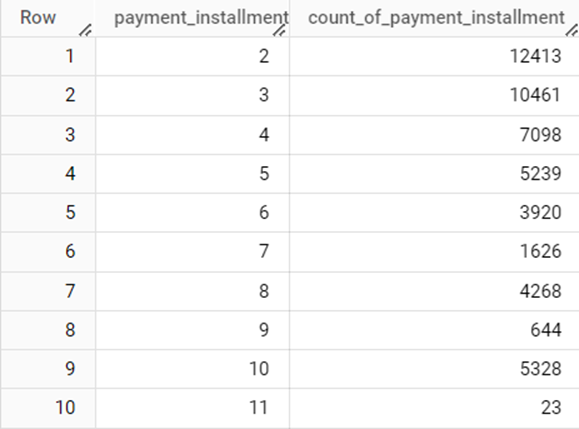
`target.payments`

where payment\_installments >= 2

GROUP BY 1

ORDER BY 1

**OUTPUT:**



**Recommendations:**

### **1. Order Growth & Marketing Strategies**

**Observation**: There's a growing trend in the number of orders placed over the years.

**Recommendation:** To continue driving growth, Target Brazil can focus on:

1. **Email Marketing**: Send personalized promotions and reminders to customers.
2. **Social Media Engagement:** Strengthen presence on platforms popular in Brazil to engage customers with new offers and products.
3. **User-friendly Website**: Simplify navigation and checkout process for a seamless shopping experience.
4. **Product Recommendations:** Use AI to recommend products based on customer browsing and purchase history.

### **2. Monthly Seasonality**

**Observation:** The highest number of orders were placed in November and December of 2017, but for 2018, it shifted to January and March.

**Recommendation:**

1. **Holiday Promotions:** Focus on aggressive promotions in November and December when demand surges due to the holiday season.
2. **Early-Year Campaigns:** Given the spike in early 2018, invest in "New Year" campaigns and clearance sales for unsold inventory in January and March.

### **3. Order Placement Times**

**Observation:** Most orders are placed in the afternoon, followed by night.

**Recommendation:**

1. **Timed Discounts**: Implement flash sales or time-limited discounts during afternoon and evening hours to further capitalize on high customer activity.
2. **Push Notifications:** Notify customers of special offers during these peak periods to boost sales.

### **4. State-Wise Distribution & Economic Impact**

**Observation:** Certain states, like AC, have relatively lower order values and higher freight costs, while others like SP have lower freight costs.

**Recommendation:**

1. **Logistics Optimization:** Explore regional distribution centers in high freight cost areas (e.g., AC, RO) to reduce shipping costs and delivery times.
2. **Targeted Discounts:** Offer state-specific discounts to incentivize purchases in states with lower order activity or higher average freight costs.

### **5. Freight Value & Delivery Time**

**Observation:** The states with the lowest average freight values are SP, PR, MG, while RR, PB, and RO have the highest. Delivery times are faster in states like SP, PR, and DF.

**Recommendation:**

1. **Delivery Service Partnership:** Establish partnerships with local delivery providers in regions with slower delivery times (RR, AP) to improve shipping efficiency.
2. **Incentivize Faster Delivery:** Provide express delivery options with incentives like free or discounted rates to improve customer satisfaction.

### **6. Payment Trends**

**Observation:** There's a variety of payment types used, and a significant number of customers use installment payments.

**Recommendation:**

1. **Installment-Based Promotions:** Offer flexible installment plans with lower interest rates to attract more customers, especially for higher-ticket items.
2. **Diverse Payment Options:** Continue expanding payment methods, ensuring popular options in Brazil (like boleto bancário) are well-supported.

### **7. Delivery Time Optimization**

**Observation:** States like RR, AP, AM have the highest delivery times, while SP and PR have the lowest.

**Recommendation:**

1. **Improve Delivery Forecasts:** Enhance logistics forecasting to reduce delivery delays in states with long shipping times, offering more accurate delivery estimates.
2. **Fast Delivery Incentives:** In states where deliveries are quicker than estimated, offer next-day delivery promotions to convert more sales.