Target Business Case Study

By ***Mahesh Sharma***

Problem Statement:

Assuming you are a data analyst/ scientist at Target, you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

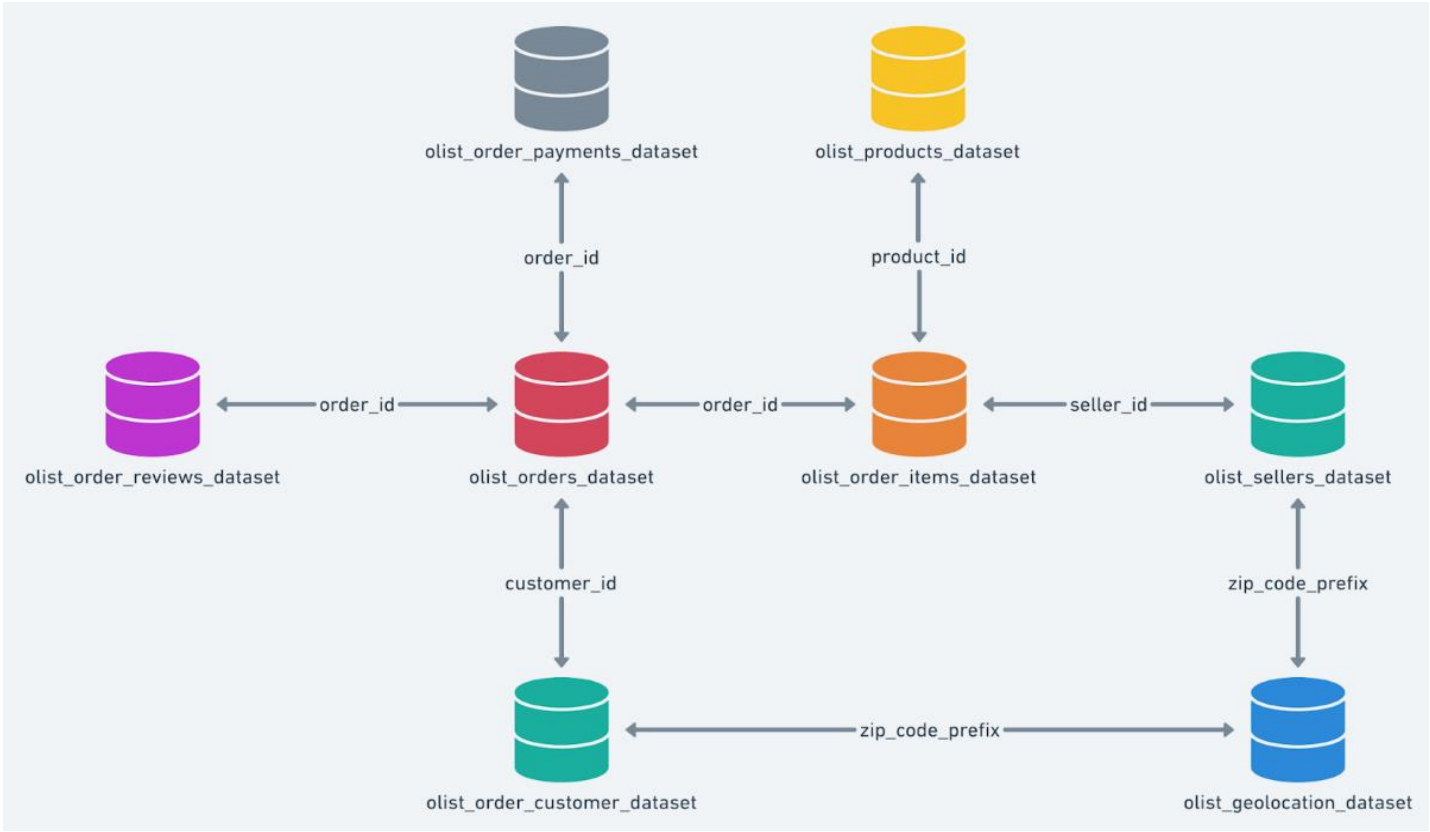
What does 'good' look like?

1. **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**
    1. Data type of all columns in the "customers" table.
    2. Get the time range between which the orders were placed.
    3. Count the Cities & States of customers who ordered during the given period.
2. **In-depth Exploration:**

    1. Is there a growing trend in the no. of orders placed over the past years?
    2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?
    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
3. **Evolution of E-commerce orders in the Brazil region:**
    1. Get the month on month no. of orders placed in each state.
    2. How are the customers distributed across all the states?
4. **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**
    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.
    2. Calculate the Total & Average value of order price for each state.
    3. Calculate the Total & Average value of order freight for each state.

5. **Analysis based on sales, freight and delivery time.**
    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
    2. Find out the top 5 states with the highest & lowest average freight value.
    3. Find out the top 5 states with the highest & lowest average delivery time.
    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.
6. **Analysis based on the payments:**
    1. Find the month on month no. of orders placed using different payment types.
    2. Find the no. of orders placed on the basis of the payment installments that have been paid.



Q 1.1). Querry for fetching the data type of all the column in a table.

*Untitled 3

Untitled 3 RUN SAVE DOWNLOAD

```sql
SELECT COLUMN_NAME, DATA_TYPE
FROM `scaler-420404.target.INFORMATION_SCHEMA.COLUMNS`
WHERE TABLE_NAME = 'customers';
```

Results

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	COLUMN_NAME	DATA_TYPE
1	customer_id	STRING
2	customer_unique_id	STRING
3	customer_zip_code_prefix	INT64
4	customer_city	STRING
5	customer_state	STRING

Q1.2) To get the time range between which the orders were placed.

Untitled 3 RUN SAVE DOWNL

```
1  select
2  min(order_purchase_timestamp) as first_order,
3  max(order_purchase_timestamp) as last_order
4
5  from `target.orders`
```

Results

Query results

JOB INFORMATION | RESULTS | CHART | JSON

Row	first_order	last_order
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

Q1.3) Count the number of Cities and States in our dataset.

Untitled 6 RUN

```
1  select count(distinct geolocation_city) as no_cities,
2  count(distinct geolocation_state) as no_states
3  from `target.geolocation`
4
```

Results

Query results SAVE RESULTS

JOB INFORMATION | RESULTS | CHART

Row	no_cities	no_states
1	8011	27

Q2.1) is there a growing trend in the no. of orders placed over past years ?

Untitled 6 RUN Query compl

```
1  select extract(year from order_purchase_timestamp) as year,
2  count(order_purchase_timestamp) as no_orders
3  from `target.orders`
4  group by year
5  order by year;
6
```

Results.

Query results SAVE RESULTS

JOB INFORMATION | RESULTS | CHART

Row	year	no_orders
1	2016	329
2	2017	45101
3	2018	54011

Q2.2) Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

Untitled 6 RUN Query complet

```
1  select extract(month from order_purchase_timestamp) as month,
2  count(order_purchase_timestamp) as no_orders
3  from `target.orders`
4  group by month
5  order by month;
```

Press Alt+F1 for Accessibility Opt

Results

Query results SAVE R

JOB INFORMATION RESULTS CI

Row	month	no_orders
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959
11	11	7544
12	12	5674

There is a clear seasonal pattern, with the highest number of orders placed in January (10,573) and the lowest number of orders placed in September (4,305)

Q 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

Untitled 6 RUN Query complet

```sql
SELECT

(CASE WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 0 AND 6 THEN 'DAWN'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 7 AND 12 THEN 'MORNING'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 13 AND 18 THEN 'AFTERNOON'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 19 AND 23 THEN 'NIGHT'END) AS
TIME,

COUNT(*) AS NO_ORDERS FROM `target.orders`
 group by TIME
 ORDER BY TIME;

```

Results

Query results SAVE RESU

JOB INFORMATION | RESULTS | CHART | JSO

Row	TIME	NO_ORDERS
1	AFTERNOON	38135
2	DAWN	5242
3	MORNING	27733
4	NIGHT	28331

Conclusion:-

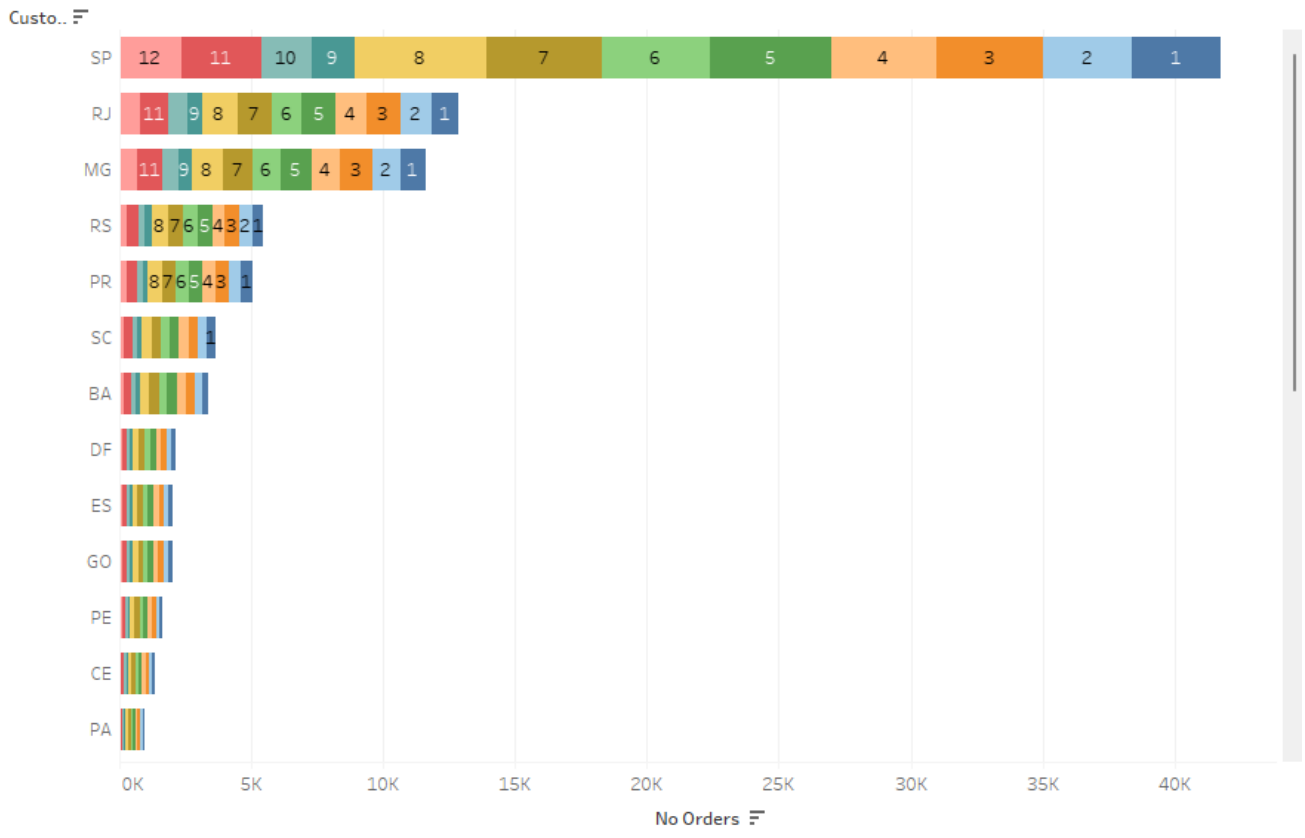
After Studying the aggregated table we come to know that **Brazilian Customer Mostly order in the afternoon** Indicating this the period when people want to shop online. By launching online add campaign and Arranging sufficient Customer Assistant personal during the busiest time can be a good step. **And Customers are buying least during dawn.**

Q3.1) get on month on month no of order placed in each state.

Untitled 6 RUN

```
1  select
2  c.customer_state,
3  extract(month from o.order_purchase_timestamp) as month,
4
5  count(o.order_purchase_timestamp) as no_orders from  `target.customers` as c
6  join `target.orders` as o
7  on c.customer_id = o.customer_id group by c.customer_state, month order by month asc;
```

The same table is extracted for Tableau Visualization for the same data for better comparison

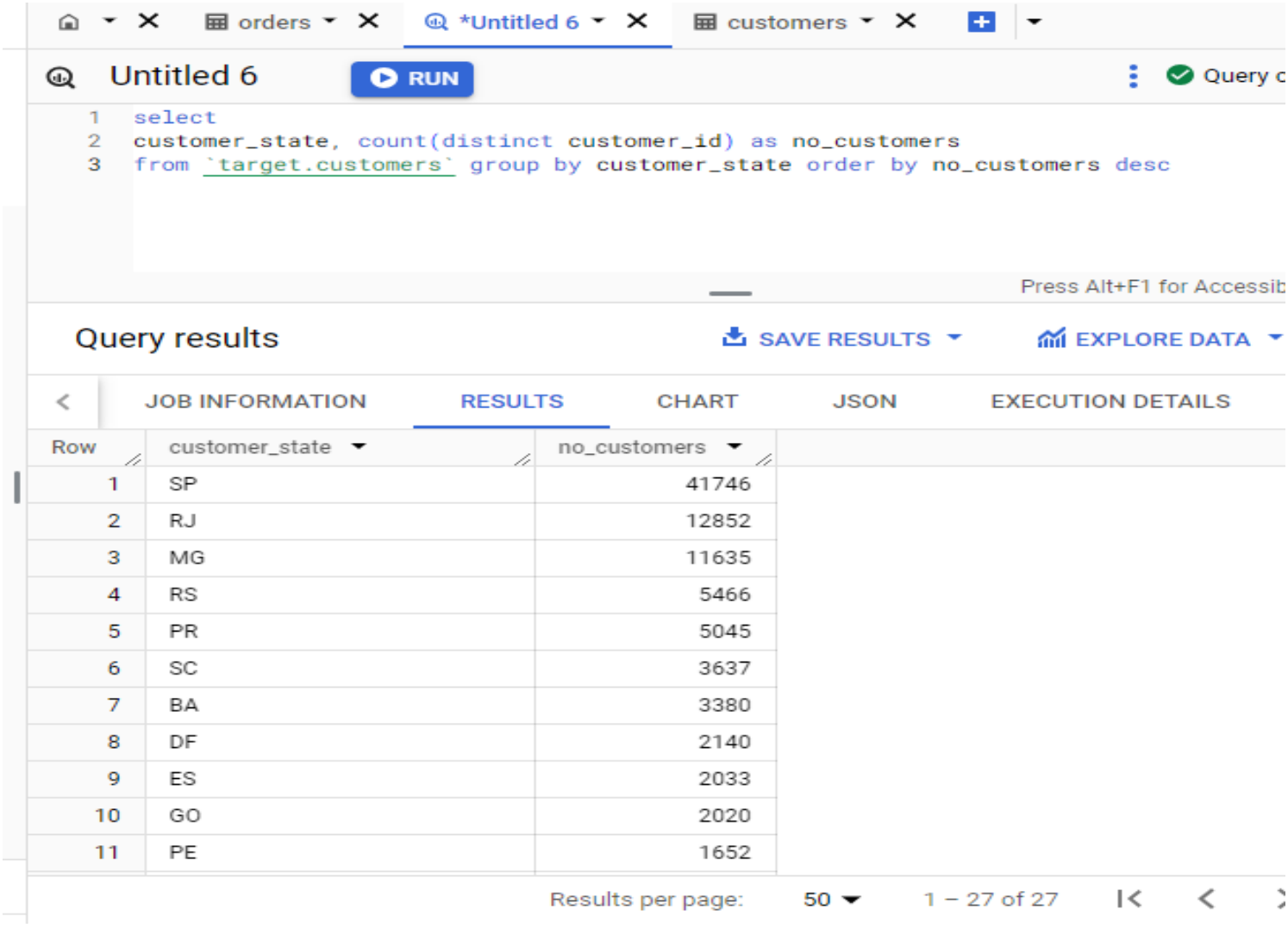


Insights:

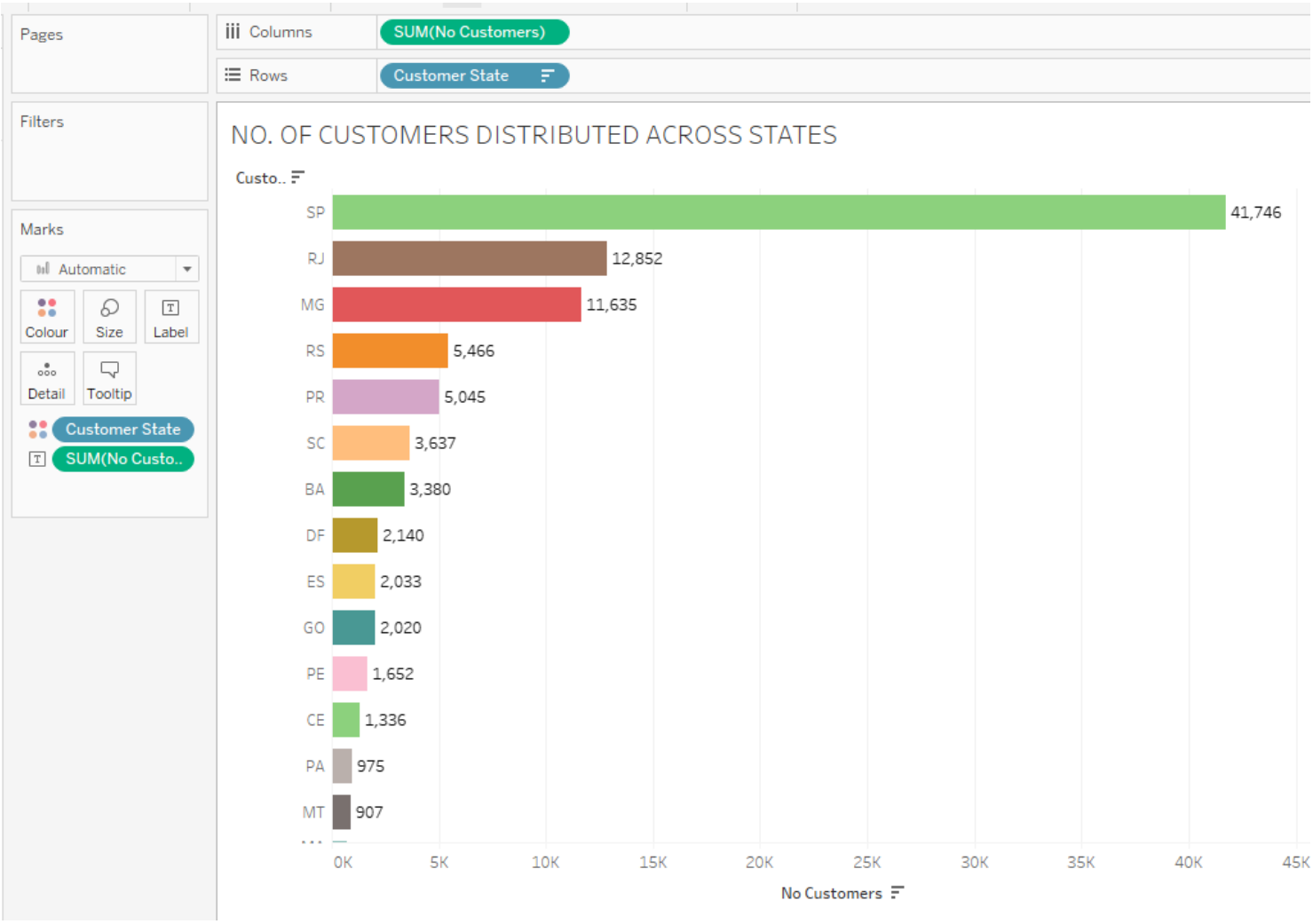
1. We can learn more about the monthly order count for each state by examining the query's results. Over time, we can spot trends, patterns, or seasonality in the order volume for various states. We can use it to determine which states have consistently high order volumes and to pinpoint any months or states where order counts have significantly changed. Here in our data, we can find that for every month the state called SP has the highest number of orders.

2. We can target marketing efforts in states with rising order volumes, spot potential operational issues in states with falling order volumes or optimize inventory management based on order trends across different states by analysing these insights.

Q3.2) How are the customers distributed across all the states?



The same table is extracted for Tableau Visualization for the same data for better comparison



Insights:

1. The distribution of clients across states will be shown by analysing the query's results. Which states have the most customers and which states have comparatively fewer consumers can be determined. Here the state called SP has the highest clients and the state called RR has the fewest clients. There are several uses for this information, including: Market targeting, Expansion opportunities and Customer service.

2. We can learn more about the geographic distribution of our client base, spot prospective growth areas, and make wise decisions to optimize our company strategy by looking at the customer distribution between states.

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.

Untitled 6 RUN This query will process 8.14 MB when ru

```
1  with cte as (select *,
2  extract(month from o.order_purchase_timestamp)as month,
3  extract(year from o.order_purchase_timestamp) as year
4
5
6  from `target.payments` as p join `target.orders` as o on p.order_id = o.
   order_id)
7  ,
8  cte1 as (select (case when year = 2017 then sum(payment_value) else 0 end) as
   `2017`,
9  (case when year = 2018 then sum(payment_value) else 0 end) as `2018` from  cte
   where month between 1 and 8 and year between 2017 and 2018 group by year)
10 select round(((sum(`2018`)-sum(`2017`))/sum(`2017`))*100,2) as `% growth`
   from cte1
```

Press Alt+F1 for Accessibility Optic

Query results SAVE RESULTS EXPLORE DATA

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAIL

Row	% growth
1	136.98

Insights:

The findings tell us a growth rate of approximately 137% from 2017 to 2018.

Q 4. 2) Calculate the Total & Average value of order price for each state.

*Untitled 3 | payments

Untitled 3 RUN SAVE DOWNLOAD SHARE SCHEDULE

```
1  select
2  c.customer_state,
3  round(sum(payment_value),2)as Total_value,
4  round(avg(payment_value),2) as average_value from
5  `target.payments` as p join
6  `target.orders` as o on
7  p.order_id = o.order_id
8  join `target.customers` as c on o.customer_id = c.customer_id
9  group by c.customer_state order by Total_value desc;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION

Row	customer_state	Total_value	average_value
1	SP	5998226.96	137.5
2	RJ	2144379.69	158.53
3	MG	1872257.26	154.71
4	RS	890898.54	157.18
5	PR	811156.38	154.15
6	SC	623086.43	165.98
7	BA	616645.82	170.82
8	DF	355141.08	161.13
9	GO	350092.31	165.76
10	ES	325967.55	154.71

Results per page:

Insights:

1. The sum of all order prices for each state is displayed in the "total_order_price" column, which represents the total amount of orders placed.

2. The "average_order_price" column shows the normal order value for each state together with the average order price for that state.

3. We can find states with large total order values, which point to potentially profitable marketplaces, by analysing the results.

4. To develop focused marketing or pricing strategies, it can be helpful to compare the average order prices across states to find areas with higher or lower average spending. 5. To obtain more understanding and make wise judgements based on the data, it's critical to consider the context of each state, such as population, economic variables, or customer behaviour.

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

*Untitled 3 | payments

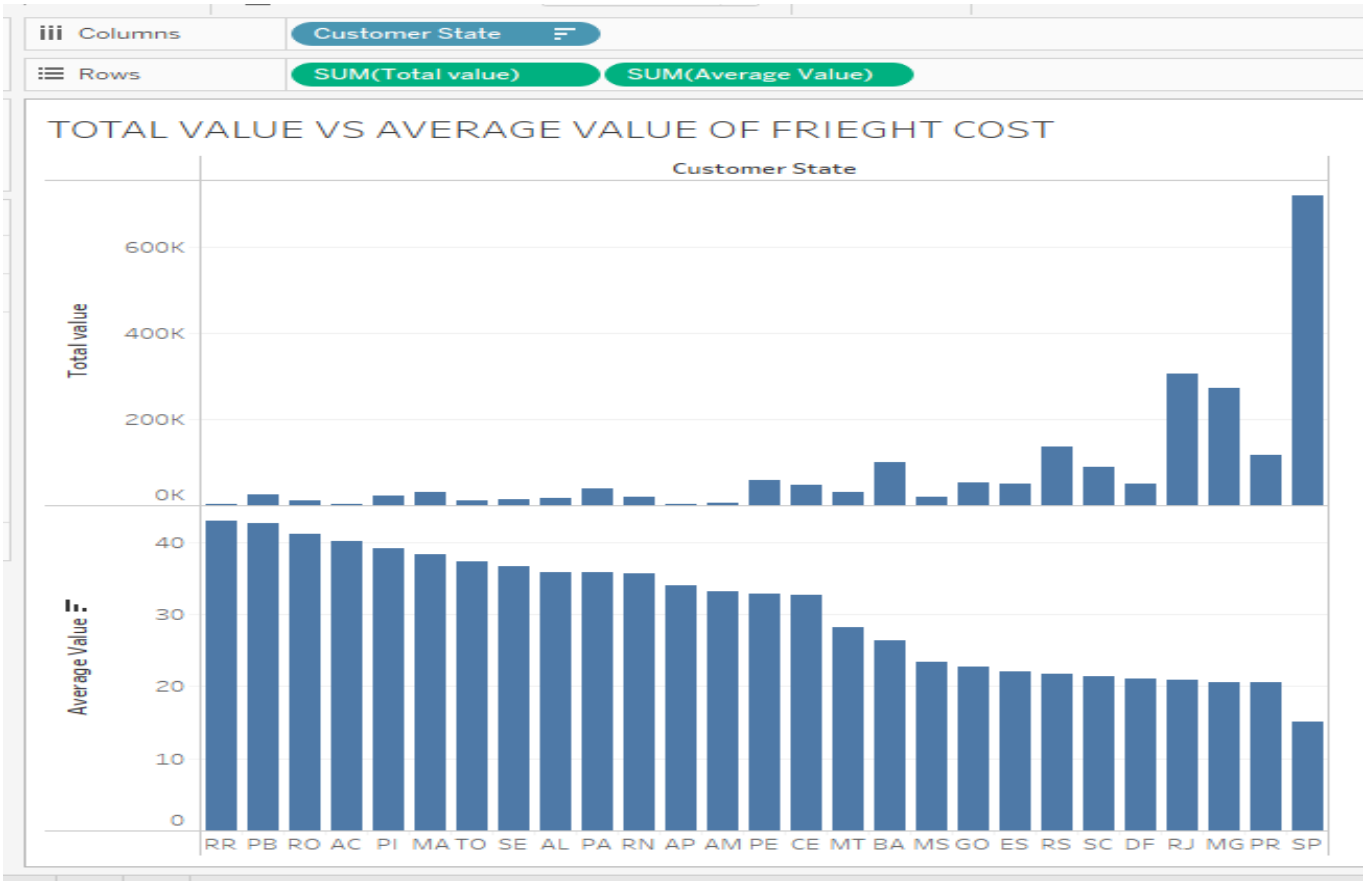
Untitled 3 | RUN | SAVE | DOWNLOAD | SHARE

```
1  select
2  c.customer_state,
3  round(sum(freight_value),2)as Total_value,
4  round(avg(freight_value),2) as average_value from
5  `target.order_items` as p join
6  `target.orders` as o on
7  p.order_id = o.order_id
8  join `target.customers` as c on o.customer_id = c.customer_id
9  group by c.customer_state order by Total_value desc;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DET

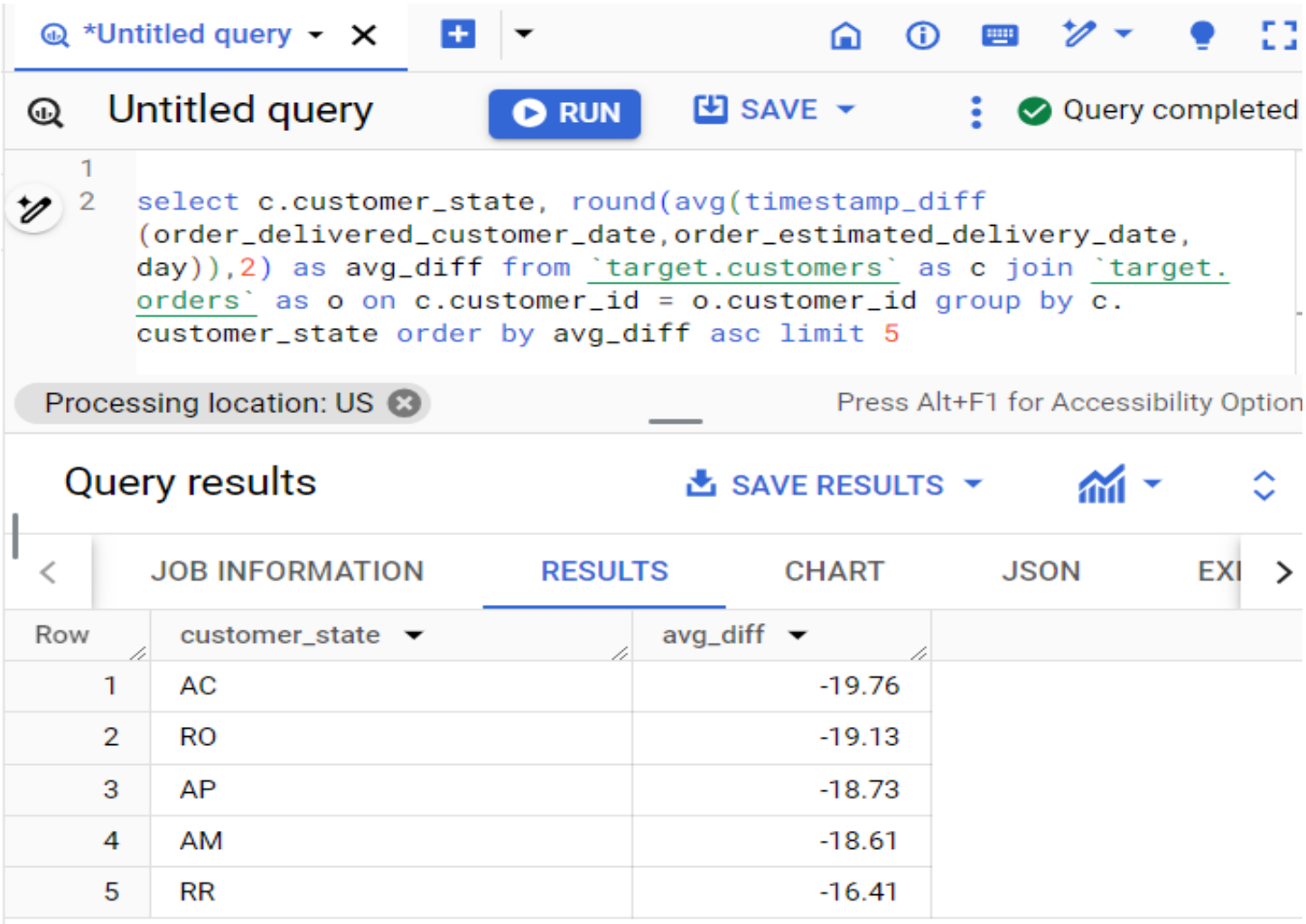
Row	customer_state	Total_value	average_value
1	SP	718723.07	15.15
2	RJ	305589.31	20.96
3	MG	270853.46	20.63
4	RS	135522.74	21.74
5	PR	117851.68	20.53
6	BA	100156.68	26.36
7	SC	89660.26	21.47
8	PE	59449.66	32.92
9	GO	53114.98	22.77
10	DF	50625.5	21.04



Insights:

1. We can find states with high total freight costs, here in our case a state called SP, by analysing the results, which could point to regions with higher shipping prices or logistical difficulties.
2. When optimizing logistics operations or pricing strategies, it might be helpful to discover regions with higher or lower average shipping prices by comparing the average order freight costs across states.
3. Understanding the differences in order freight rates between states can offer information about local shipping habits, supplier locations, or client preferences that can be used to optimize processes and cut costs.

Q5.1) 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.



Row	customer_state	avg_diff
1	AC	-19.76
2	RO	-19.13
3	AP	-18.73
4	AM	-18.61
5	RR	-16.41

Insights:

1. Insights into the effectiveness of the delivery process, including any delays or early deliveries compared to the projected timeframe, can be gained by analysing the delivery_time and diff_estimated_delivery columns.

2. These columns can be further examined to find trends, outliers, or elements that affect delivery times or discrepancies between estimated and actual delivery dates.

3. These insights can be applied to manage customer expectations, enhance customer satisfaction, optimize the delivery process, and improve logistics operations.

Q5.2) Find out the top 5 states with the highest & lowest average freight value.

Untitled RUN MORE SAVE DOWNLOAD SHARE SCHEDULE This query will process 14.5

```sql
select high.customer_state, low.customer_state, high.avg_high_freight, low.avg_low_freight from
(select
c.customer_state,
round(avg(oi.freight_value),2) as avg_high_freight,
row_number() over(order by round(avg(oi.freight_value),2) desc) as row_num1

from `target.customers` as c join `target.orders` as o on c.customer_id = o.customer_id join `target.order_items` as oi on
o.order_id = oi.order_id group by c.customer_state
limit 5) as high join

(select
c.customer_state,
round(avg(oi.freight_value),2) as avg_low_freight,
row_number() over(order by round(avg(oi.freight_value),2)) as row_num2

from `target.customers` as c join `target.orders` as o on c.customer_id = o.customer_id join `target.order_items` as oi on
o.order_id = oi.order_id group by c.customer_state
limit 5 ) as low on high.row_num1 = low.row_num2;
```

Press Alt+F1 for Accessibility Op

Query results SAVE RESULTS EXPLORE DATA

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	customer_state	customer_state_1	avg_high_freight	avg_low_freight
1	RR	SP	42.98	15.15
2	PB	PR	42.72	20.53
3	RO	MG	41.07	20.63
4	AC	RJ	40.07	20.96
5	PI	DF	39.15	21.04

Insights:

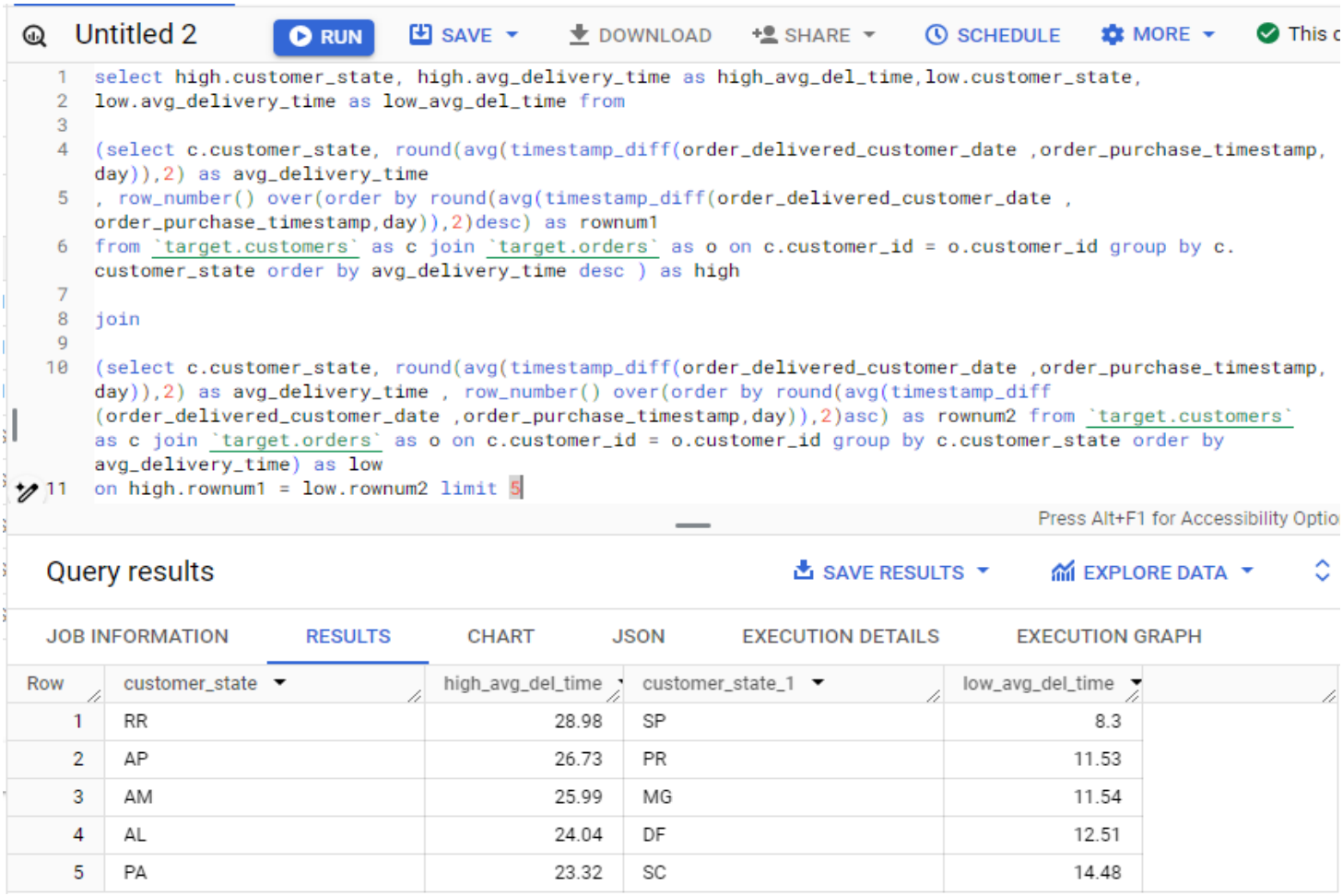
1. The states with the highest average freight values like states called RR and PB may experience greater shipping prices due to reasons like remote locations, higher transportation costs, or supply chain difficulties.

2. It might be useful for our company to try to optimize logistics operations or save costs to locate places with relatively reduced shipping prices by looking at the states with the lowest average freight values like states such as SP and PR.

3. This data can help us develop focused initiatives, bargain freight costs, or spot possible opportunities to reduce costs in our supply chain operations.

4. When assessing the data and drawing conclusions from these insights, it is crucial to consider additional elements like distance, transportation infrastructure, carrier availability, or regional economic variations.

Q5. 3)Find out the top 5 states with the highest & lowest average delivery time.



Row	customer_state	high_avg_del_time	customer_state_1	low_avg_del_time
1	RR	28.98	SP	8.3
2	AP	26.73	PR	11.53
3	AM	25.99	MG	11.54
4	AL	24.04	DF	12.51
5	PA	23.32	SC	14.48

Insights:

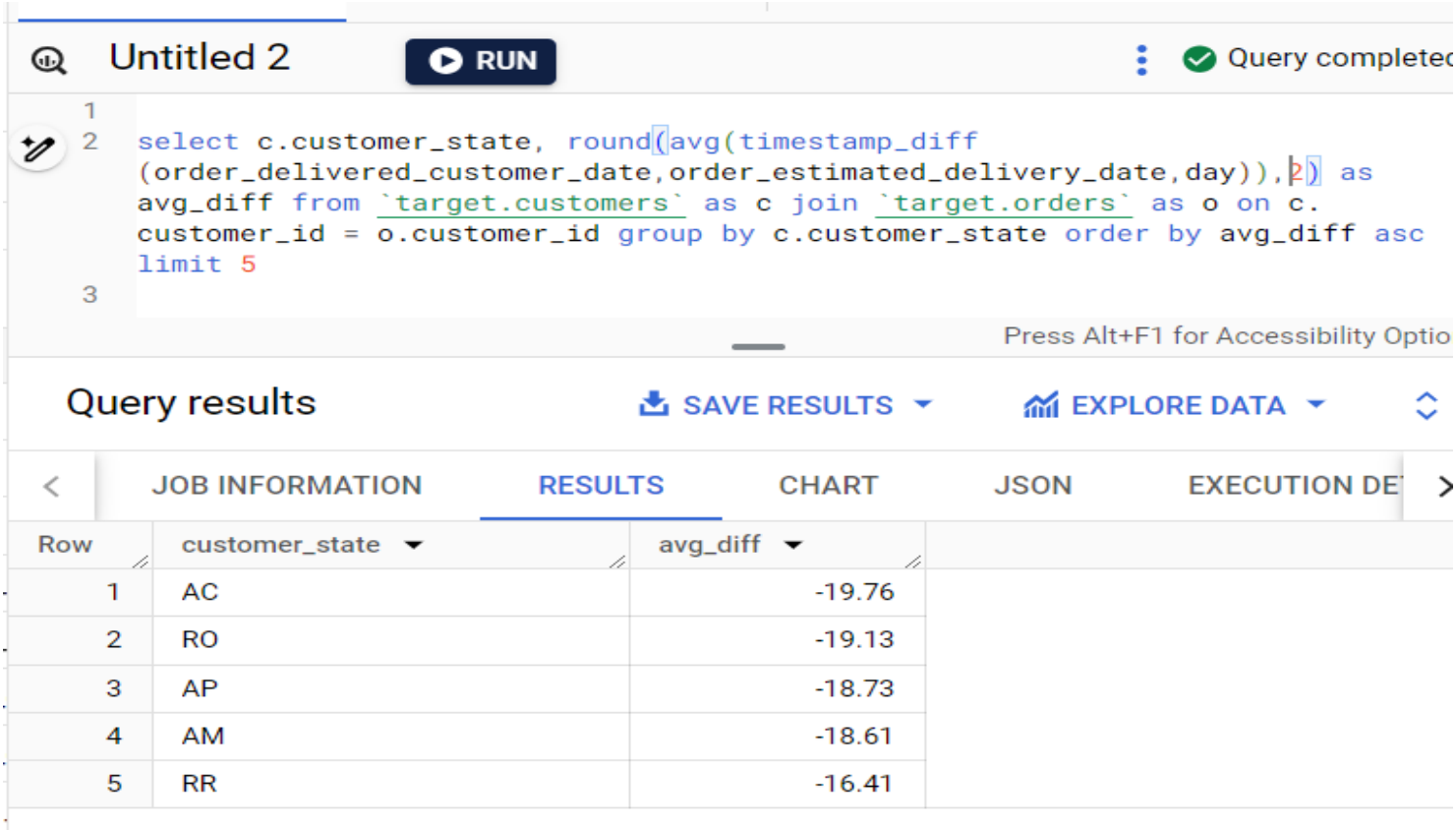
1. Finding areas with effective delivery operations, quicker transit times, or solid logistics networks can be done by looking at the states like SP and PR with the lowest average delivery times and states called RR and AP with highest average delivery times.

2. These insights can be helpful for our company looking to improve customer satisfaction, operational efficiency, delivery process optimization, and setting reasonable expectations for customers based on regional delivery time patterns.

3. When evaluating the data and drawing conclusions from these insights, it's crucial to take additional elements into account, such as population density, the distinction between urban and rural locations, customer expectations, or unique logistical restrictions.

4. Utilizing this information, our company can concentrate on areas where delivery efficiency improvements can be made, thereby improving customer experiences and operational efficiencies.

Q 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.



Insights:

1. Our company operating in these states called AC, RO, AP, and AM where average delivery speed is highest can take advantage of the quicker delivery times by highlighting their rapid and dependable service, thereby drawing more clients, and boosting client satisfaction.

2. These data can help us improve our operations, enhance customer experience, optimize logistics, or look for expansion prospects in areas with a track record of quick order delivery.

Q6.1)Find the month on month no. of orders placed using different payment types.

orders | *Untitled query

Untitled query RUN SAVE DOWNLOAD Query comple

```
1  select format_timestamp('%Y-%m', o.order_purchase_timestamp) AS
2  month,count(o.order_id),
3  payment_type
4  from `target.payments`as p join `target.orders`as o on p.order_id = o.order_id
   group by payment_type,month order by month
```

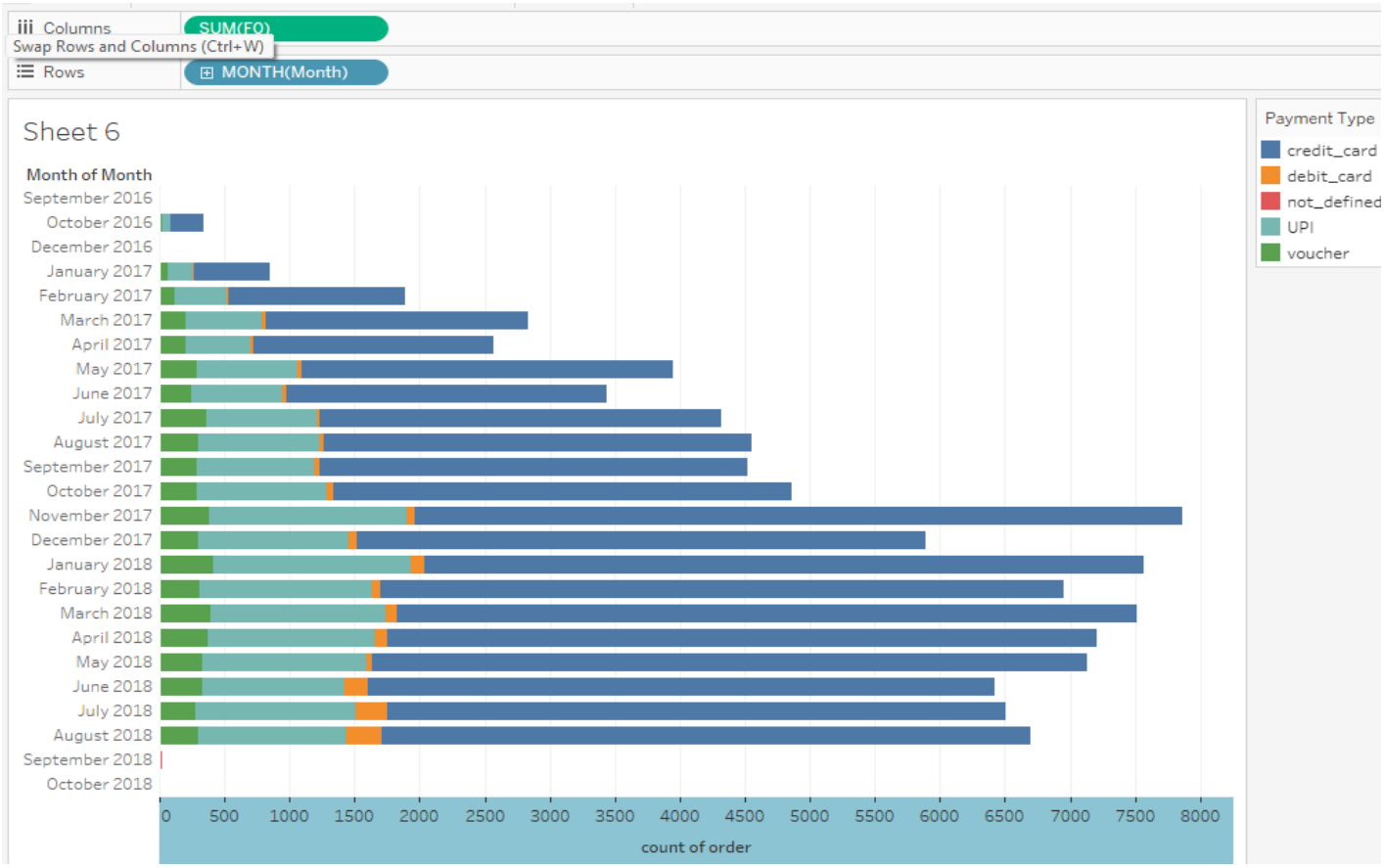
Press Alt+F1 for Accessibility Op

Query results SAVE RESULTS EXPLORE DATA

RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	month	f0_	payment_type
1	2016-09	3	credit_card
2	2016-10	254	credit_card
3	2016-10	23	voucher
4	2016-10	2	debit_card
5	2016-10	63	UPI
6	2016-12	1	credit_card
7	2017-01	61	voucher

The same table is extracted for Tableau Visualization for the same data for better comparison



Insights

1. We identify that credit card as a payment method was most used in November 2017.

2. To analyze seasonality, identify peak months, or evaluate the effects of marketing efforts or outside variables on consumer behavior, tracking the month-to-month trends in order counts can be helpful.

3. Based on the payment preferences noticed during various months, these insights might help firms optimize their payment procedures, customize marketing campaigns, or enhance customer experiences.

Q6.2)Find the no. of orders placed on the basis of the payment installments that have been paid.

Untitled query RUN DOWNLOAD SAVE Query complet

```
1  select count(DISTINCT order_id) as count_orders , payment_installments from
   `target.payments` group by payment_installments order by payment_installments
```

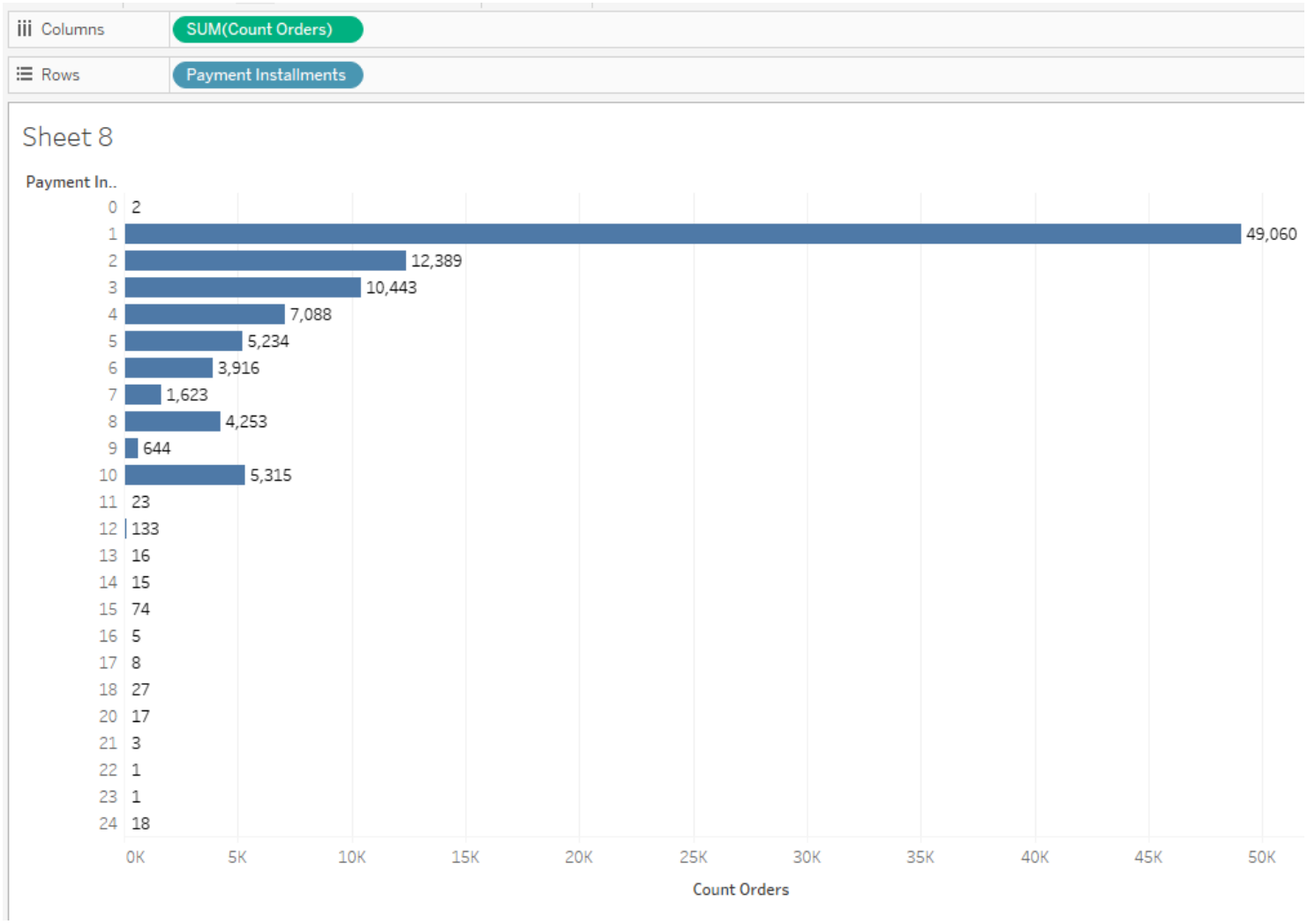
Processing location: US Press Alt+F1 for Accessibility Optio

Query results SAVE RESULTS EXPLORE DATA

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAI

Row	count_orders	payment_installment
1	2	0
2	49060	1
3	12389	2
4	10443	3
5	7088	4
6	5234	5
7	3916	6
8	1623	7
9	4253	8
10	644	9

The same table is extracted for Tableau Visualization for the same data for better comparison



Insights:

1. 49060 orders were placed where payment installment was 1.

2. This analysis can help determine whether payment installment alternatives are popular or preferred by clients.

3. Customers' preferences for budgeting or financing may be discerned by whether they tend to select a particular number of payment installments.

4. Monitoring the distribution of orders according to payment installments might reveal information about the buying habits of clients and their preference for flexible payment methods.