

TARGET BUSINESS CASE - SOLUTIONS

-by Prachi Shinde

Q1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.

A. Data type of all columns in the “customers” table.

ANSWER:

```
SELECT column_name, data_type
FROM `business-case-sql-406309.Target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers'
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW
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		

INTERPRETATION:

I have verified the column data types in the “customers” table. Which depicts STRING & INTEGER type data used in the customer table.

B. Get the time range between which the orders were placed.

ANSWER:

```
SELECT
  MIN(order_purchase_timestamp) AS first_order_timestamp,
  MAX(order_purchase_timestamp) AS last_order_timestamp
FROM `business-case-sql-406309.Target.orders`
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	first_order_timestamp	last_order_timestamp			
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC			

INTERPRETATION:

The query allowed us to determine the start and end dates of the data along with the timestamp i.e. from 4th September 2016 at 9:15:19 UTC to 17th October 2018 at 5:30:18 UTC

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

ANSWER:

1	SELECT
2	COUNT(DISTINCT customer_city) AS cities_count,
3	COUNT(DISTINCT customer_state) AS states_count
4	FROM `business-case-sql-406309.Target.customers`
5	WHERE customer_id IN (
6	SELECT DISTINCT customer_id
7	FROM `business-case-sql-406309.Target.orders`)
8	

Query results			
JOB INFORMATION		RESULTS	CHART PREVIEW
Row	cities_count	states_count	
1	4119	27	

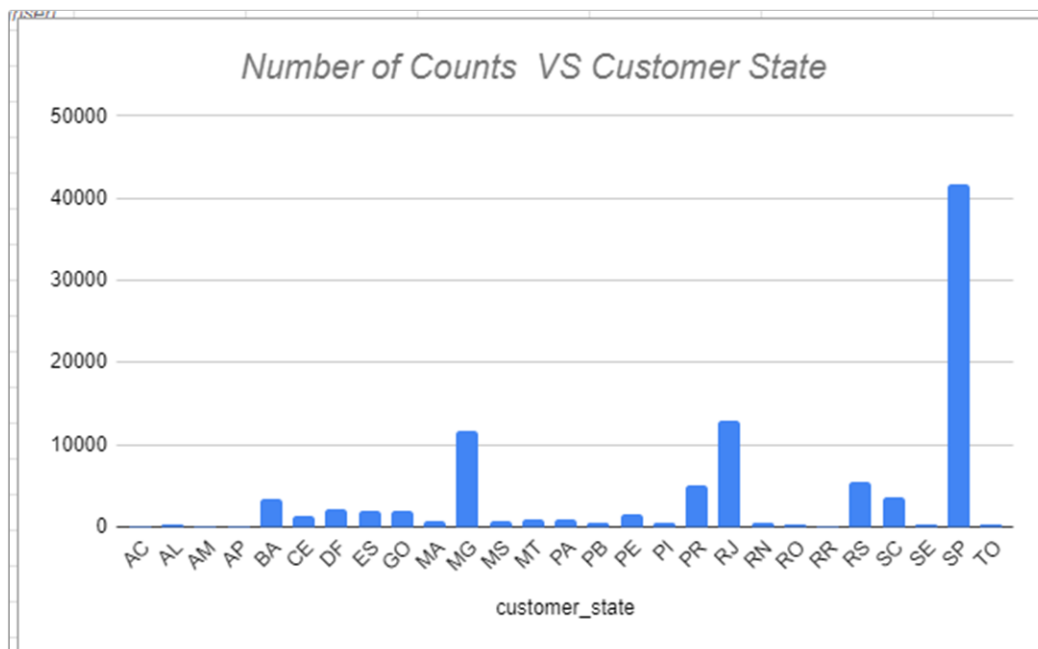
```

158
159 SELECT
160     DISTINCT c.customer_city,
161     c.customer_state,
162     Count(o.customer_id) as `Number_of_orders`
163 FROM `Target.customers` c
164 JOIN `Target.orders` o ON c.customer_id = o.customer_id
165 GROUP BY c.customer_city, c.customer_state
166 ORDER BY `Number_of_orders` DESC

```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON	EXECUTION DET
Row	customer_city	customer_state	Number_of_orders			
1	sao paulo	SP	15540			
2	rio de janeiro	RJ	6882			
3	belo horizonte	MG	2773			
4	brasilia	DF	2131			
5	curitiba	PR	1521			
6	campinas	SP	1444			
7	porto alegre	RS	1379			
8	salvador	BA	1245			
9	guarulhos	SP	1189			
10	sao bernardo do campo	SP	938			



INTERPRETATION:

Initially I counted distinct number of cities along with the states which have ordered items in the given time period which was 4117 cities and 27 states of Brazil. Highest number of orders have been placed by Sao Paulo which is 15540 followed by Rio de Janerio and Belo Horizonte on second and third highest position. This could be because Sao Paulo is largest city in the Brazil which is highly populous and economic centre. Implying richness of the people in Sao Paulo hence large number of orders. I have used excel to clearly visualize the data.

2. In-depth Exploration:

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

ANSWER:

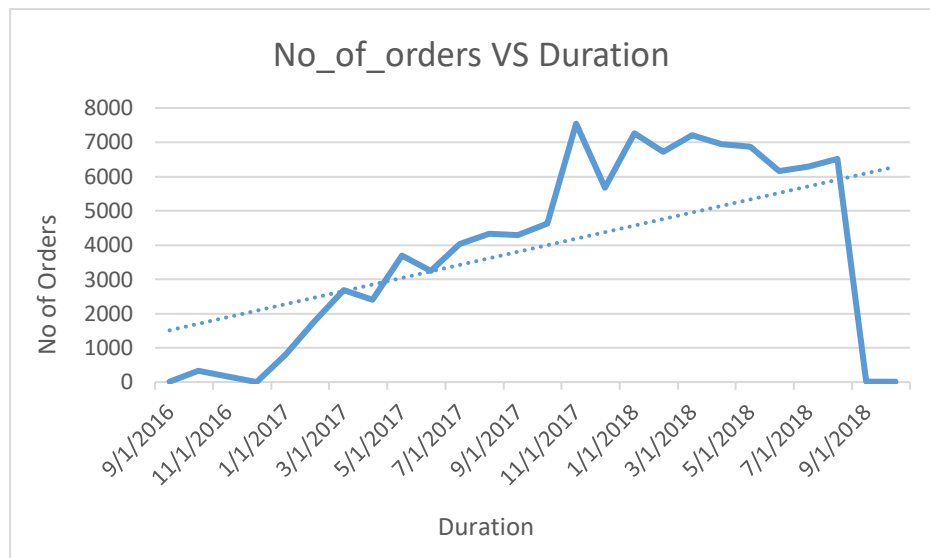
```

1 SELECT
2   EXTRACT(YEAR FROM order_purchase_timestamp) as 'Year',
3   EXTRACT(MONTH FROM order_purchase_timestamp) as 'Month',
4   COUNT(order_id) AS 'order_count'
5 FROM
6   | 'Target.orders' |
7 GROUP by 'Year', 'Month'
8 ORDER by 'Year' asc, 'Month' asc

```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW
Row	Year	Month	order_count	
1	2016	9	4	
2	2016	10	324	
3	2016	12	1	
4	2017	1	800	
5	2017	2	1780	
6	2017	3	2682	
7	2017	4	2404	
8	2017	5	3700	
9	2017	6	3245	
10	2017	7	4026	



INTERPRETATION:

It can be clearly observed from the chart that there is a growing trend in the number of orders over the years in Brazil. The count of orders has shown an overall upward trend, with some fluctuations and has plateaued a bit.

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

ANSWER:

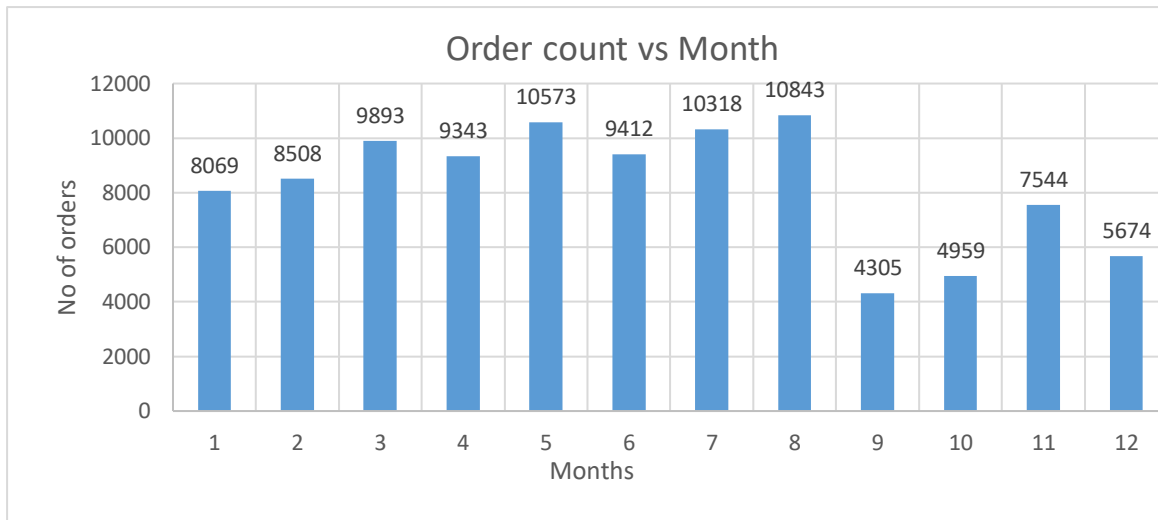
```

1 select
2 extract(month from order_purchase_timestamp) as `Month`,
3 count (distinct order_id) as `Order count`
4 from `Target.orders`
5 group by `Month`
6 order by `Month`
7

```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	J
Row	Month	Order count			
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			



INTERPRETATION:

We can observe some seasonality in the e-commerce orders. The No of orders increases from March to August with slight fluctuations in between. In September orders have decreased drastically till the month of December with some fluctuations.

C. 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

ANSWER:

```

1 select CASE
2   WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 0 AND 6 THEN 'Dawn'
3   WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 7 AND 12 THEN 'Mornings'
4   WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 13 AND 18 THEN 'Afternoon'
5   WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 19 AND 23 THEN 'Night'
6   END as Hour,
7   COUNT (o.order_id) as Order_count
8 FROM `Target.orders` o
9 JOIN `Target.customers` c ON o.customer_id =c.customer_id
10 GROUP BY Hour
11 ORDER BY Order_count;
12

```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	Hour	Order_count			
1	Dawn	5242			
2	Mornings	27733			
3	Night	28331			
4	Afternoon	38135			

INTERPRETATION:

Based on the analysis we can observe clearly that Brazilian customers order maximum during Afternoons and Night which depicts their free time. Whereas, comparatively less orders were placed during Dawn and Mornings which can suggest that they are sleeping and busy doing chores! By studying customer behaviour data we can strategise and give a seamless shopping experience. **In the code we have missed 4 hours while aloting the time slots. This time should also be taken into consideration.**

Q3. Evolution of E-commerce orders in the Brazil region:

A. Get the month on month no. of orders placed in each state.

ANSWER:

<pre>SELECT c.customer_state as `State`, #EXTRACT(YEAR FROM o.order_purchase_timestamp) as `Year`, EXTRACT(MONTH FROM o.order_purchase_timestamp) as `Month`, COUNT(o.order_id) AS `Order_count` FROM `Target.orders` o join `Target.customers` c on o.customer_id = c.customer_id GROUP by `State`,`Month` ORDER by `State`,`Month`</pre>				
Query results				
JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTI				
Row	State ▾	Month ▾	Order_count ▾	
1	AC	1	8	
2	AC	2	6	
3	AC	3	4	
4	AC	4	9	
5	AC	5	10	
6	AC	6	7	
7	AC	7	9	
8	AC	8	7	
9	AC	9	5	
10	AC	10	6	

INTERPRETATION:

Analyzing month on month order count we can conclude Sao Paulo has the highest number of orders in any given month as compared to other States. I could not plot the Graph.

B. How are the customers distributed across all the states?

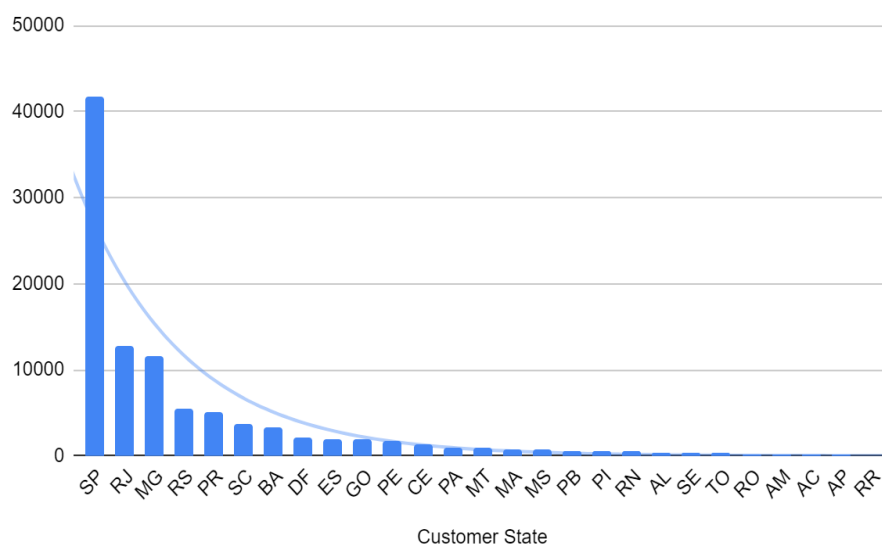
ANSWER:

```
1 Select
2 COUNT( distinct customer_id) AS 'No of Customer',
3 customer_state as 'Customer State'
4 FROM 'Target.customers'
5 GROUP BY customer_state
6 ORDER BY 'No of Customer' desc
7 |
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON	E
Row	No of Customer	Customer State				
1	41746	SP				
2	12852	RJ				
3	11635	MG				
4	5466	RS				
5	5045	PR				
6	3637	SC				
7	3380	BA				
8	2140	DF				
9	2033	ES				
10	2020	GO				

No of Orders VS Customer State



INTERPRETATION:

Once more, the data analysis demonstrates that São Paulo (SP) holds the highest customer count, a phenomenon attributable to its position as Brazil's most populous state. This outcome mirrors the earlier analysis, highlighting a positive correlation between a state's population and its order count.

Q4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

A. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).

ANSWER:

<pre>1 SELECT 2 EXTRACT(month FROM o.order_purchase_timestamp) as `Month`, 3 ROUND((4 (SUM (CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018 AND 5 EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 6 THEN p.payment_value END) 7 - 8 SUM (CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 AND 9 EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 10 THEN p.payment_value END)) 11 / 12 SUM (CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 AND 13 EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 14 THEN p.payment_value END) 15 * 100) ,2) AS `Percentage_increase` 16 FROM `Target.orders` o 17 join `Target.payments` p ON o.order_id =p.order_id 18 WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017,2018) AND 19 EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 20 GROUP BY `Month` 21 ORDER BY `Month`</pre>			
Query results			
JOB INFORMATION		RESULTS	CHART PREVIEW JSON
Row	Month	Percentage_increase	
1	1	705.13	
2	2	239.99	
3	3	157.78	
4	4	177.84	
5	5	94.63	
6	6	100.26	
7	7	80.04	
8	8	51.61	

INTERPRETATION:

January shows the highest percentage increase which is 705%, followed by February (239%) and April (177%).

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

ANSWER:

```

1 select
2   c.customer_state as `State`,
3   round(sum(oi.price),2) as `Total price`,
4   round(avg(oi.price),2) as `Average price`
5 from `Target.order_items` oi
6 join `Target.orders` o on oi.order_id = o.order_id
7 join `Target.customers` c on c.customer_id = o.customer_id
8 group by `State`
9 order by `State`

```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON	EXE
Row	State ▾	Total price ▾	Average price ▾			
1	AC	15982.95	173.73			
2	AL	80314.81	180.89			
3	AM	22356.84	135.5			
4	AP	13474.3	164.32			
5	BA	511349.99	134.6			
6	CE	227254.71	153.76			
7	DF	302603.94	125.77			
8	ES	275037.31	121.91			
9	GO	294591.95	126.27			
10	MA	119648.22	145.2			

Row	State ▾	Total price ▾	Average price ▾			
1	RR	7829.43	150.57			
27	SP	5202955.05	109.65			

INTERPRETATION:

The analysis reveals São Paulo (SP) has the highest total price value and surprisingly has the lowest average price value. On the other hand, the state of Roraima (RR) has the highest average price value and lowest Total Price. This can be justified by the number of orders placed and number of customers ordered the items has always been highest by SP.

Second table was retrieved by calling highest and lowest ranks for the Total Price and Average Price.

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

ANSWER:

1	select
2	c.customer_state as `State`,
3	round(sum(oi.freight_value),2) as `Total freight value`,
4	round(avg(oi.freight_value),2) as `Average freight value`
5	from `Target.order_items` oi
6	join `Target.orders` o on oi.order_id = o.order_id
7	join `Target.customers` c on c.customer_id = o.customer_id
8	group by `State`
9	order by `State` #, `Total freight value`, `Average freight value`

Query results				
JOB INFORMATION		RESULTS	CHART	PREVIEW
Row	State	Total freight value	Average freight value	
1	AC	3686.75	40.07	
2	AL	15914.59	35.84	
3	AM	5478.89	33.21	
4	AP	2788.5	34.01	
5	BA	100156.68	26.36	
6	CE	48351.59	32.71	
7	DF	50625.5	21.04	
8	ES	49764.6	22.06	
9	GO	53114.98	22.77	
10	MA	31523.77	38.26	

Row	State	Total freight value	Average freight value
1	RR	2235.19	42.98
27	SP	718723.07	15.15

INTERPRETATION:


The analysis reveals São Paulo (SP) has the highest total freight value, it surprisingly has the lowest average freight value among all states. On the other hand, the state of Roraima (RR) has the lowest total freight value and highest average freight value. This can be justified by the number of orders placed and number of customers ordered the items has always been highest by SP.

Second table was retrieved by calling highest and lowest ranks for the Total freight value and Average freight value.

Q5. Analysis based on sales, freight and delivery time.

- A. 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.

ANSWER:

<pre>1 SELECT 2 order_id, 3 DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,DAY) AS `Delivery Time`, 4 DATE_DIFF(order_estimated_delivery_date,order_delivered_customer_date,DAY) AS `Diff_estimated_delivery` 5 FROM `Target.orders` 6 WHERE DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,DAY) IS NOT NULL 7 ORDER BY `Delivery Time`</pre>				
Query results 				
JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTION DETAILS				
Row	order_id ▾	Delivery Time ▾	Diff_estimated_delivery ▾	
1	e65f1eeee1f52024ad1dcd03447f7482	0	9	
2	bb5a519e352b45b714192a02ffe25681	0	25	
3	434cecee7d1a65fc65358a632b6f725f	0	19	
4	d3ca7b82c922817b06e5ca21165c5ea2	0	11	
5	1d893dd7ca5f77ebf5f59f0d2017eee0	0	10	
6	d5fbeedc85190ba88580d6f82d1d5ed3	0	7	
7	79e324907160caea526fd8b94389dbbc	0	8	
8	38c1e3d4ed6a13cd0cf612d4c09766e9	0	16	
9	8339b608be0d84fca9d8da68b58332c3	0	27	
10	f349cdb62f69c3fae5c4d7d3f3a4a185	0	12	

INTERPRETATION:

From the analysis we found the number of days taken to deliver each order and difference between the estimated delivery dates. From the 10 rows we can see delivery has been done before estimated time! This doesn't not conclude whether all deliveries have been done before stipulated estimated delivery date.

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

ANSWER: Top 5 Highest -

```
1
2 select
3   c.customer_state as `State`,
4   round(avg(oi.freight_value),2) as `Average freight value`
5 from `Target.order_items` oi
6 join `Target.orders` o on oi.order_id = o.order_id
7 join `Target.customers` c on c.customer_id = o.customer_id
8 group by `State`
9 order by `Average freight value` desc
10 limit 5 |
11
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	State ▾	Average freight value ▾			
1	RR	42.98			
2	PB	42.72			
3	RO	41.07			
4	AC	40.07			
5	PI	39.15			

Top 5 Lowest -

```
2 select
3   c.customer_state as `State`,
4   round(avg(oi.freight_value),2) as `Average freight value`
5 from `Target.order_items` oi
6 join `Target.orders` o on oi.order_id = o.order_id
7 join `Target.customers` c on c.customer_id = o.customer_id
8 group by `State`
9 order by `Average freight value` asc
10 limit 5
11
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	State ▾	Average freight value ▾			
1	SP	15.15			
2	PR	20.53			
3	MG	20.63			
4	RJ	20.96			
5	DF	21.04			

INTERPRETATION:

The analysis reveals São Paulo (SP) has the lowest average freight value among all states. On the other hand, the state of Roraima (RR) has the highest average freight value.

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

ANSWER:

```
1 SELECT
2   c.customer_state as 'State',
3   ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,o.order_purchase_timestamp,DAY)),2) AS 'Average_delivery_time'
4 FROM   'Target.orders' o
5 JOIN   'Target.customers' c ON o.customer_id =c.customer_id
6 GROUP BY 'State'
7 ORDER BY 'Average_delivery_time' DESC
8 LIMIT 5
9
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	State ▼	Average_delivery_time			
1	RR	28.98			
2	AP	26.73			
3	AM	25.99			
4	AL	24.04			
5	PA	23.32			

```
1 SELECT
2   c.customer_state as 'State',
3   ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,o.order_purchase_timestamp,DAY)),2) AS 'Average_delivery_time'
4 FROM   'Target.orders' o
5 JOIN   'Target.customers' c ON o.customer_id =c.customer_id
6 GROUP BY 'State'
7 ORDER BY 'Average_delivery_time' ASC
8 LIMIT 5
9
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	State ▼	Average_delivery_time			
1	SP	8.3			
2	PR	11.53			
3	MG	11.54			
4	DF	12.51			
5	SC	14.48			

INTERPRETATION:

The analysis reveals São Paulo (SP) has the lowest average delivery time among all states. On the other hand, the state of Roraima (RR) has the highest delivery time. Which depicts orders are delivered

within 8 days from purchase time in SP whereas it takes longer in RR that can be upto 28 days from the order of purchases. Delivery time should be reduced in other states.

D. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

ANSWER:

```

1 SELECT
2   c.customer_state as `State`,
3   ROUND(AVG(DATE_DIFF(o.order_estimated_delivery_date,o.order_delivered_customer_date,DAY)),2) AS
4     `diff_estimated_delivery`
5 FROM `Target.orders` o
6 JOIN `Target.customers` c ON o.customer_id =c.customer_id
7 GROUP BY `State`
8 ORDER BY `diff_estimated_delivery` DESC
9 LIMIT 5

```

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	State	diff_estimated_delivery			
1	AC	19.76			
2	RO	19.13			
3	AP	18.73			
4	AM	18.61			
5	RR	16.41			

INTERPRETATION:

Listed top 5 states where delivery has been faster than other states, we can clearly see state AC's average delivery time has been 20 days prior the estimated delivery date.

Q6. Analysis based on the payments:

A. Find the month on month no. of orders placed using different payment types. **ANSWER:**

```
1 select
2   extract(month from o.order_purchase_timestamp) as `Month`,
3   count (distinct p.order_id) as `Order count`,
4   p.payment_type as `Payment_types`
5 from `Target.orders` o
6 join `Target.payments` p on p.order_id = o.order_id
7 group by `Month`,p.payment_type
8 order by p.payment_type,`Month`
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON
Row	Month	Order count	Payment_types		
1	1	1715	UPI		
2	2	1723	UPI		
3	3	1942	UPI		
4	4	1783	UPI		
5	5	2035	UPI		
6	6	1807	UPI		
7	7	2074	UPI		
8	8	2077	UPI		
9	9	903	UPI		
10	10	1056	UPI		

```
1 select
2   extract(month from o.order_purchase_timestamp) as `Month`,
3   count (distinct p.order_id) as `Order count`,
4   p.payment_type as `Payment_types`
5 from `Target.orders` o
6 join `Target.payments` p on p.order_id = o.order_id
7 group by `Month`,p.payment_type
8 order by `Month`,p.payment_type
```

Row	Month	Order count	Payment_types
1	1	1715	UPI
2	1	6093	credit_card
3	1	118	debit_card
4	1	337	voucher
5	2	1723	UPI
6	2	6582	credit_card
7	2	82	debit_card
8	2	288	voucher
9	3	1942	UPI
10	3	7682	credit_card

Row	Order count	Payment_types
1	76505	credit_card
2	19784	UPI
3	3866	voucher
4	1528	debit_card
5	3	not_defined

INTERPRETATION:

I have mentioned two types of data segregation method for ease of convinence. From the above data we can see people have used credit card as the highest means of payment followed by UPI transcatons.By looking at the month on month analysis we can see there is increase in the transactions made by the customers.

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

ANSWER:

```
1 select
2 p.payment_installments as `Payment_installments`,
3 count(distinct o.order_id) as `Order count`,
4 from `Target.orders` o
5 join `Target.payments` p on p.order_id = o.order_id
6 WHERE o.order_status != 'canceled'
7 group by `Payment_installments`
8 order by `Order count` desc, `Payment_installments`
9
```

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW
Row	Payment_installments	Order count		
1	1	48732		
2	2	12329		
3	3	10374		
4	4	7046		
5	10	5279		
6	5	5204		
7	8	4224		
8	6	3894		
9	7	1617		
10	9	638		

INTERPRETATION:

The analysis reveals that the majority of orders (maximum count) have only one payment installment. The highest number of installments is 24, which is associated with 18 orders.

Actionable Insights:

- São Paulo (SP) leads in orders, urging a focus on other states for growth.
- Leverage seasonal variations for strategic marketing during peak periods.
- Improve delivery times to enhance customer satisfaction and encourage repeat purchases.
- High-order states (SP and RJ) benefit from customer retention strategies.
- Analyze demographics for targeted marketing and increased sales.
- Seize opportunities by offering discounts during off-peak months.

Recommendations:

- Optimize logistics for faster delivery and improved customer satisfaction.
- Implement customer retention strategies, loyalty programs, and personalized offers.
- Evaluate pricing and freight fees for competitiveness and profitability.
- Invest in technology for an enhanced e-commerce experience.
- Collaborate with sellers for expanded product offerings and improved quality.
- Leverage social media and influencers for increased brand awareness.
- Enhance customer service through chat support and timely responses.
- Monitor competitors for strategic adjustments.

Conclusion - Key Takeaways:

- São Paulo (SP) dominates; focus on other states for growth.
- Tailor strategies based on customer demographics.
- Use discounts strategically during off-peak seasons.