TARGET

Business Case Study

Problem Statement:

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

Here are my interpretations of insights derived from the available data.

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.

Field name	Type
customer_id	STRING
customer_unique_id	STRING
customer_zip_code_prefix	INTEGER
customer_city	STRING
customer_state	STRING

Observation :
The "customers" table is designed with a total of five columns, featuring two identification-related columns, namely *"customer_id" and "customer_unique_id" both having a string* data type. Complementing these, the remaining three columns, designated as "customer_zip_code_prefix," "customer_city," and "customer_state," are dedicated to capturing the diverse aspects of each customer's location where *customer_zip_code_prefix is of integer* data type and the other two showcasing string as its data type.

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

Query -

```
SELECT
  MIN(order_purchase_timestamp) AS oldest_order_record,
  MAX(order_purchase_timestamp) AS latest_order_record
FROM
  `TargetAnalyticsDB.orders`;
```

Output :

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

Observation :
Analyzing the output of the query, it indicates the time range for order placements. The earliest order was on September 4, 2016, at 21:15:19 UTC, while the latest was on October 17, 2018, at 17:30:18 UTC. This implies that the *dataset covers a period of precisely 2 years and 1 month*, capturing the timeline within which these orders were made.

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

Query :

```
SELECT
  COUNT(DISTINCT c.customer_city) as Orders_City_Count,
```

```
  COUNT(DISTINCT c.customer_state) as Orders_State_Count
FROM
  `TargetAnalyticsDB.customers` c
```

Output :

Row	Orders_City_Count	Orders_State_Count
1	4119	27

Observation:

The results of the query reveal valuable insights into the geographical distribution of orders. There are *4,119 unique cities and 27 unique states* from which orders originate. This information underscores the diversity in both city and state representations within the dataset.

2. In-depth Exploration:

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(*) AS ORDER_COUNT
FROM
  `TargetAnalyticsDB.orders`
GROUP BY
  YEAR
ORDER BY
  YEAR;
```

Output :

Row	YEAR	ORDER_COUNT
1	2016	329
2	2017	45101
3	2018	54011

Observation :

1. 2016:
- The dataset commences with a partial-year record from September 4, 2016.
- Logging 329 orders during this period, it might be indicative of a nascent stage or a new beginning, suggesting a starting point for data collection.

2. 2017:
- As the first full year in the dataset, 2017 showcases a remarkable leap in both the duration and order volume.
- Registering 45,101 orders, this substantial growth indicates a significant development from the initial stages of 2016.
- The possibility exists that 2016 served as a foundational year, laying the groundwork for the exponential growth observed in subsequent years.

3. 2018:
- Building upon the momentum of 2017, 2018 sustains the exponential growth trend.
- With 54,011 orders, it further cements the idea that the preceding years were not mere anomalies but rather reflective of a continuing and maturing trend.

Overall Observations:

- The partial data in 2016 could indeed mark a new beginning, with subsequent years showcasing exponential growth.

- The substantial increase in orders over the years indicates a platform's evolving presence, possibly fueled by improved services or expanding market reach.
- This cumulative growth pattern suggests a sustained ***upward trajectory***, emphasizing the significance of the initial data from 2016 in shaping the subsequent narrative of growth and development.

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(*) AS ORDER_COUNT
FROM
  `TargetAnalyticsDB.orders`
GROUP BY
  YEAR,MONTH
ORDER BY
  YEAR,MONTH;
```

Output :

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

Observations :

Year-End Patterns:
- Both 2016 and 2018 exhibit a decline in orders during the closing months of the year (December 2016 and October 2018).
- This recurring trend suggests a potential seasonal influence or external factors common to year-end periods.(In case of 2016 it can be a fresh start which is a potential factor for 2016 number of orders)

Second Quarter Dips (Months 4, 5, 6):
- A slight decrease in orders is observed during the second quarter (April, May, June) in both 2017 and 2018.
- This pattern indicates a possible seasonal or cyclical behavior during these months.

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
 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 'Mornings'
   WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 13 AND 18 THEN 'Afternoon'
   ELSE 'Night'
 END AS TIME_ZONE,
 COUNT(*) AS ORDER_COUNT
FROM
 `TargetAnalyticsDB.orders`
GROUP BY
 TIME_ZONE
ORDER BY
 ORDER_COUNT DESC;
```

Output :

Row	TIME_ZONE	ORDER_COUNT
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Observation :
-The analysis reveals a clear *preference for Afternoon* order placements, totaling 38,135, signaling significant customer engagement during those hours.
-Nighttime closely follows with 28,331 orders, reflecting substantial nocturnal online shopping activity.
-Mornings also exhibit strong engagement, recording 27,733 orders, likely aligned with customers making early-day purchases.
-*Dawn sees the lowest* count at 5,242 orders, indicating smaller but noteworthy participation in the early morning.

Overall, businesses can strategically time promotions and advertisements during Afternoon, the peak period, and leverage diverse customer engagement from dawn to late night for targeted initiatives.

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

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

Query :

```
SELECT
   c.customer_state AS STATE,
   EXTRACT(month FROM o.order_purchase_timestamp) AS MONTH,
   COUNT(*) AS ORDER_COUNT
FROM
   `TargetAnalyticsDB.orders` o
LEFT JOIN
   `TargetAnalyticsDB.customers` c ON o.customer_id = c.customer_id
GROUP BY
   STATE, MONTH
ORDER BY
   STATE, MONTH;
```

Output :

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

Observations :
Strategic Insights: Unveiling Regional Dynamics in Year-Long Order Patterns.
This analysis aims to uncover nuanced patterns in regional order dynamics throughout the year, utilizing the median as a robust measure to capture central tendency and mitigate the impact of outliers.
On calculating the median(month wise) order count over the year is 76.

States with Greater than median orders consistently throughout the year include:
- RJ (Rio de Janeiro)
- RS (Rio Grande do Sul)
- SP (São Paulo)
- DF (Distrito Federal)
- PR (Paraná)
- MT (Mato Grosso)
- MA (Maranhão)
- MG (Minas Gerais)
- PE (Pernambuco)
- PA (Pará)
- BA (Bahia)
- CE (Ceará)
- GO (Goiás)
- ES (Espírito Santo)
- SC (Santa Catarina)
- PB (Paraíba)
- MS (Mato Grosso do Sul)

These states demonstrate sustained high market demand and active customer engagement across all months.

States with Lesser than median orders consistently throughout the year include:
- AL (Alabama)
- SE (Sergipe)
- PI (Piauí)
- RN (Rio Grande do Norte)
- AM (Amazonas)
- RR (Roraima)
- TO (Tocantins)
- AC (Acre)
- RO (Rondônia)
- AP (Amapá)

These states exhibit a lower level of market demand or customer activity throughout the year.

Strategic Insights:

- For strategic business decisions, focusing on states with consistently high order counts can optimize resource allocation, marketing efforts, and inventory planning. Understanding regional variations in customer behavior helps tailor strategies to specific market dynamics. This strategic analysis aids companies in refining their approach to different regions, ensuring targeted and effective business operations.

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

Query :

```
SELECT
   customer_state AS STATE,
   COUNT(DISTINCT customer_id) AS CUSTOMER_COUNT,
   ROUND((COUNT(DISTINCT customer_id) / SUM(COUNT(DISTINCT customer_id)) OVER ()) * 100,
1) AS RELATIVE_PERCENTAGE
FROM
   `TargetAnalyticsDB.customers`
GROUP BY
   STATE
ORDER BY
   RELATIVE_PERCENTAGE DESC;
```

Output:

Row	STATE	CUSTOMER_COUNT	RELATIVE_PERCENTAGE
1	SP	41746	42.0
2	RJ	12852	12.9
3	MG	11635	11.7
4	RS	5466	5.5
5	PR	5045	5.1
6	SC	3637	3.7
7	BA	3380	3.4
8	DF	2140	2.2
9	ES	2033	2.0
10	GO	2020	2.0

Integrated Insights:

Customer Distribution by State:

Top 5 States:
São Paulo (SP) leads with an impressive customer count of 41,746, constituting 42.0% of the total customer base. Following closely are Rio de Janeiro (RJ) with 12,852 customers (12.9%), and Minas Gerais (MG) with 11,635 customers (11.7%). Rio Grande do Sul (RS) and Paraná (PR) also command notable market shares at 5.5% and 5.1%, respectively. These states represent lucrative opportunities for business expansion and growth, given their substantial customer bases.

Lowest 5 States:
On the other end, Roraima (RR) holds the lowest customer count at 46 (0.0%), followed by Amapá (AP) with 68 customers (0.1%). Acre (AC), Amazonas (AM), and Alagoas (AL) contribute 0.1% to 0.4% of the total customer base. While these states present smaller markets, they offer potential for strategic initiatives to establish brand presence and capture untapped customer segments.

Strategic Considerations:
For strategic considerations, focusing on *high customer base states* like São Paulo, Rio de Janeiro, and Minas Gerais is paramount. These regions, with their significant market shares, should be the focal point for *resource allocation, marketing strategies, and business expansion* efforts. Simultaneously, targeted initiatives in *lower customer base states,* such as Roraima, Amapá, Acre, Amazonas, and Alagoas, are essential to penetrate niche markets, *enhance brand awareness, and tap into growth opportunities*. Tailoring approaches based on regional dynamics ensures a balanced and effective business strategy.

Integrated Insights of Customer Count and Orders (state wise):
The synthesis of order count and state-wise customer distribution paints a *comprehensive picture of market dynamics.* States like São Paulo, Rio de Janeiro, and Minas Gerais not only exhibit high customer counts but also demonstrate consistently elevated order counts throughout the year. This alignment signifies robust market engagement, suggesting these states as strategic focal points for businesses. On the contrary, states with lower customer counts, such as Roraima, Amapá, Acre, Amazonas, and Alagoas, also reflect proportionally lower order counts. Crafting tailored strategies for these regions is imperative, focusing on expanding the customer base, increasing order frequency, and enhancing brand visibility. The integrated approach allows for data defined decision-making, ensuring a balanced and effective market penetration strategy.

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.

Query :

```
WITH YearlyOrdersCost AS (
    SELECT
        EXTRACT(YEAR FROM o.order_purchase_timestamp) AS year,
        ROUND(SUM(p.payment_value), 4) AS total_orders_cost
    FROM
        `TargetAnalyticsDB.orders` o
    INNER JOIN
        `TargetAnalyticsDB.payments` p ON p.order_id = o.order_id
    WHERE
        EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
        AND EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017, 2018)
    GROUP BY
        year
    ORDER BY
        year
)

SELECT
    year,
    total_orders_cost,
    ROUND(
        COALESCE(
                        ((total_orders_cost - LAG(total_orders_cost) OVER (ORDER BY year)) / LAG(total_orders_cost) OVER (ORDER BY year)) * 100,
            0
        ),
        2
    ) AS percentage_increase
FROM
    YearlyOrdersCost;
```

Output :

Row	year	total_orders_cost	percentage_increase
1	2017	3669022.12	0.0
2	2018	8694733.84	136.98

Observation :

The data indicates a substantial increase in total orders cost from 2017 to 2018, with the total rising from $3,669,022.12 in 2017 to $8,694,733.84 in 2018. This represents a remarkable percentage *growth of ~137%*, reflecting a significant expansion in business activity or market demand during the transition period. Further analysis and consideration of external factors would be essential to comprehend the driving forces behind this notable upswing and its potential implications.

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

Query :

```
WITH OrdersInfo AS (
    SELECT
        o.order_id,
        o.customer_id,
        SUM(p.payment_value) AS order_amount
    FROM
        `TargetAnalyticsDB.orders` o
        LEFT JOIN `TargetAnalyticsDB.payments` p ON o.order_id = p.order_id
    GROUP BY
        o.order_id, o.customer_id
)

SELECT
    c.customer_state AS STATE,
    ROUND(SUM(order_amount), 2) AS TOTAL_ORDER_VALUE,
    ROUND(AVG(order_amount), 2) AS AVG_ORDER_VALUE
FROM
    OrdersInfo oi
    LEFT JOIN `TargetAnalyticsDB.customers` c ON oi.customer_id = c.customer_id
GROUP BY
    STATE
ORDER BY
    STATE;
```

Output:

Row	STATE	TOTAL_ORDER_VALUE	AVG_ORDER_VALUE
1	AC	19680.62	242.97
2	AL	96252.7	234.19
3	AM	27846.44	189.43
4	AP	16262.8	239.16
5	BA	607041.45	181.53
6	CE	274549.81	207.52
7	DF	351487.63	165.72
8	ES	323105.75	160.11
9	GO	340626.24	170.48
10	MA	150691.7	204.74

Observation :

Examining the data on total and average order values across states reveals intriguing patterns. *São Paulo (SP) emerges as a commerce hub* with the highest total order value, reaching $5,878,935.99, albeit with a comparatively lower average order value of $142.95. This might signify a high volume of transactions with a focus on affordability. Conversely, states like *Paraíba (PB) demonstrate a* *distinct pattern*, boasting a remarkable average order value of $264.63, implying potentially higher-value transactions despite a lower total order value of $140,519.57. This suggests a market where consumers may be inclined towards more substantial purchases. The data hints at diverse consumer preferences and market dynamics across states, emphasizing the need for nuanced strategies to cater to varied economic landscapes.

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

Query :

```
WITH ProcessedData AS (
  SELECT DISTINCT
    o.order_id,
    o.customer_id,
    oi.freight_value
  FROM
    `TargetAnalyticsDB.orders` o
    LEFT JOIN `TargetAnalyticsDB.order_items` oi ON o.order_id = oi.order_id
  ORDER BY
    o.order_id
)

SELECT
  c.customer_state AS STATE,
  ROUND(SUM(freight_value), 2) AS TOTAL_FREIGHT,
  ROUND(AVG(freight_value), 2) AS AVG_FREIGHT
FROM
  ProcessedData
  LEFT JOIN `TargetAnalyticsDB.customers` c ON ProcessedData.customer_id = c.customer_id
GROUP BY
  STATE
ORDER BY
  STATE;
```

Output :

Row	STATE	TOTAL_FREIGHT	AVG_FREIGHT
1	AC	3386.56	41.3
2	AL	15074.91	36.15
3	AM	4942.31	33.17
4	AP	2413.32	34.98
5	BA	89908.47	26.27
6	CE	44450.64	33.02
7	DF	46395.03	21.32
8	ES	45881.51	22.11
9	GO	47407.43	23.05
10	MA	28682.96	37.94

Observation :

Analyzing the freight data for different states in Brazil yields several insights:

1. Regional Disparities in Freight Costs:

- There is a notable variation in both total and average freight costs among states. States like São Paulo (SP) exhibit the highest total freight cost ($646,582.28) but with a lower average freight cost ($15.27), suggesting a higher volume of shipments with relatively lower individual transportation costs.

2. Efficiency and Affordability in São Paulo (SP):

- São Paulo's lower average freight cost indicates potential efficiencies in logistics, likely due to its developed infrastructure and high shipping volumes. This efficiency may contribute to cost savings for businesses operating in or servicing the São Paulo market.

3. Paraíba's (PB) High Average Freight Cost:
Paraíba stands out with a higher average freight cost ($41.55), implying that individual shipments in this state tend to incur higher transportation expenses. This may indicate unique logistics challenges, longer distances, or a preference for express shipping services.

4. Freight Patterns Reflecting Economic Activities:
States with higher total freight costs, such as Minas Gerais (MG) and Rio de Janeiro (RJ), might indicate robust economic activities and higher demand for transportation services. Conversely, states with lower total freight costs may have less intense economic activity or more localized supply chains.

5. Potential Market Opportunities:
States with lower total and average freight costs may present opportunities for businesses to expand their operations, as reduced transportation expenses could make their products more competitive in those markets.

6. Impact of Geographical Factors:
Geographical factors, such as the vast distances in states like Amazonas (AM) and Roraima (RR), might contribute to higher average freight costs. Understanding these geographic challenges is crucial for optimizing logistics strategies in such regions.

7. Consideration of Industry-Specific Influences:
The disparities in freight costs may also be influenced by the nature of industries predominant in each state. Industries with unique shipping requirements or those dealing with specialized products may contribute to variations in freight costs.

In conclusion, a comprehensive understanding of the freight data involves considering the interplay of economic, logistical, and geographical factors. Businesses can leverage these insights to tailor their logistics strategies, optimize costs, and identify potential growth opportunities in different states.

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.
   Query :

```
SELECT
  o.order_id,
  CONCAT(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY), ' Days') AS Delivery_Time,
  CASE
    WHEN DATE_DIFF(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY) > 0 THEN
      CONCAT(DATE_DIFF(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY), ' Days Late')
    WHEN DATE_DIFF(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY) < 0 THEN
      CONCAT(DATE_DIFF(o.order_estimated_delivery_date, o.order_delivered_customer_date, DAY), ' Days Early')
```

```
        ELSE 'Timely Delivery'
    END AS Expected_Delivery_Difference
FROM
    `TargetAnalyticsDB.orders` o;
```

Output :

Row	order_id	Delivery_Time	Expected_Delivery_Difference
1	1950d777989f6a877539f53795...	30 Days	12 Days Late
2	2c45c33d2f9cb8ff8b1c86cc28c...	30 Days	28 Days Early
3	65d1e226dfaeb8cdc42f665422...	35 Days	16 Days Early
4	635c894d068ac37e6e03dc54ec...	30 Days	1 Days Early
5	3b97562c3aee8bdedcb5c2e45a...	32 Days	Timely Delivery
6	68f47f50f04c4cb6774570cfde3...	29 Days	1 Days Early
7	276e9ec344d3bf029ff83a161c6...	43 Days	4 Days Late
8	54e1a3c2b97fb0809da548a59f...	40 Days	4 Days Late
9	fd04fa4105ee8045f6a0139ca5b...	37 Days	1 Days Late
10	302bb8109d097a9fc6e9cefc59...	33 Days	5 Days Late

Observation :

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

Query :

```
WITH ProcessedData AS (
    SELECT DISTINCT
        o.order_id,
        o.customer_id,
        oi.freight_value
    FROM
        `TargetAnalyticsDB.orders` o
        LEFT JOIN `TargetAnalyticsDB.order_items` oi ON o.order_id = oi.order_id
)

SELECT STATE, AVG_FREIGHT
FROM (
  SELECT
      c.customer_state AS STATE,
      ROUND(AVG(freight_value), 2) AS AVG_FREIGHT,
      ROW_NUMBER() OVER (ORDER BY AVG(freight_value) DESC) AS high_rank,
      ROW_NUMBER() OVER (ORDER BY AVG(freight_value) ASC) AS low_rank
  FROM
      ProcessedData
      LEFT JOIN `TargetAnalyticsDB.customers` c ON ProcessedData.customer_id = c.customer_id
  GROUP BY
      STATE
)
WHERE high_rank <= 5 OR low_rank <= 5
ORDER BY AVG_FREIGHT DESC;
```

Output :

Row	STATE	AVG_FREIGHT
1	RR	42.26
2	RO	42.13
3	PB	41.55
4	AC	41.3
5	PI	39.01
6	DF	21.32
7	RJ	21.1
8	MG	20.78
9	PR	20.46
10	SP	15.27

Observation :

The output reveals a *distinct contrast in average freight values among Brazilian states*, highlighting the geographical variability in shipping costs. *Roraima (RR) tops the list with the highest average freight of $42.26*, closely followed by Rondônia (RO) and Paraíba (PB), each reflecting elevated shipping expenses. On the other end, *São Paulo (SP) emerges as the state with the lowest average freight at $15.27*, showcasing a cost-effective transportation landscape. This analysis offers valuable insights into regional disparities, enabling businesses to strategize logistics and distribution plans based on the varying economic factors influencing shipping expenditures.

The disparity in average freight values underscores the importance of considering state-specific logistics dynamics for effective supply chain management. While states like São Paulo exhibit cost-efficiency, others like Roraima present higher transportation costs, emphasizing the need for tailored shipping strategies. Businesses can leverage this information to optimize resource allocation, streamline operations, and enhance overall efficiency in the complex landscape of Brazilian logistics.

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

Query :

```
WITH ProcessedData AS (
  SELECT
    c.customer_state AS STATE,
    CAST(ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, SECOND))) AS INT64) AS avg_delivery_time_seconds,
    DENSE_RANK() OVER (ORDER BY
ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp,
SECOND))) ASC) AS rank_asc,
    DENSE_RANK() OVER (ORDER BY
ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp,
SECOND))) DESC) AS rank_desc
  FROM
    `TargetAnalyticsDB.orders` o
    LEFT JOIN `TargetAnalyticsDB.customers` c ON o.customer_id = c.customer_id
  WHERE
    o.order_status = 'delivered'
  GROUP BY
    STATE
)

SELECT
  STATE,
  CONCAT(
    CAST(FLOOR(avg_delivery_time_seconds / (24 * 60 * 60)) AS STRING), ' days ',
    CAST(FLOOR(MOD(avg_delivery_time_seconds, (24 * 60 * 60)) / 3600) AS STRING), ' hours ',
```

```
  CAST(FLOOR(MOD(MOD(avg_delivery_time_seconds, (24 * 60 * 60)), 3600) / 60) AS STRING), ' minutes'
 ) AS Avg_Delivery_Time
FROM
 ProcessedData
WHERE
 rank_asc <= 5 OR rank_desc <= 5
ORDER BY
 rank_asc;
```

Output :

Row	STATE	Avg_Delivery_Time
1	SP	8 days 18 hours 16 minutes
2	PR	11 days 23 hours 47 minutes
3	MG	12 days 0 hours 12 minutes
4	DF	12 days 23 hours 13 minutes
5	SC	14 days 22 hours 54 minutes
6	PA	23 days 18 hours 33 minutes
7	AL	24 days 13 hours 3 minutes
8	AM	26 days 10 hours 13 minutes
9	AP	27 days 4 hours 26 minutes
10	RR	29 days 9 hours 18 minutes

Observation :

1. State with the Shortest Average Delivery Time:
- State: São Paulo (SP)
- Avg Delivery Time: 8 days 18 hours 16 minutes

2. State with the Longest Average Delivery Time:
- State: Roraima (RR)
- Avg Delivery Time: 29 days 9 hours 18 minutes

3. Range of Average Delivery Time:
- The average delivery time across all states ranges from 8 days 18 hours 16 minutes (São Paulo - SP) to 29 days 9 hours 18 minutes (Roraima - RR).

4. Observations:
- States in the southern region (São Paulo - SP, Paraná - PR, Minas Gerais - MG, Distrito Federal - DF, Santa Catarina - SC) generally have shorter average delivery times.
- Northern states (Pará - PA, Alagoas - AL, Amazonas - AM, Amapá - AP, Roraima - RR) have longer average delivery times.
- There is a significant difference in delivery times between the fastest and slowest states.

5. Potential Factors Influencing Delivery Time:
- Geographic distance from distribution centers.
- Transportation infrastructure and logistics efficiency.
- Population density and urbanization.
- Local customs and regulatory processes.

6. Suggestions for Improvement:
- Explore ways to optimize logistics and distribution in states with longer delivery times.
- Invest in improving transportation infrastructure in regions with slower delivery times.
- Consider regional variations in demand and adjust inventory levels accordingly.

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 :

```
WITH ProcessedData AS (
  SELECT
    c.customer_state AS STATE,
    CAST(ROUND(ABS(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, SECOND)) - AVG(TIMESTAMP_DIFF(o.order_estimated_delivery_date,
o.order_purchase_timestamp, SECOND)))) AS INT64) AS Delivery_Diff_Seconds
  FROM
    `TargetAnalyticsDB.orders` o
  LEFT JOIN
    `TargetAnalyticsDB.customers` c ON o.customer_id = c.customer_id
  WHERE
    o.order_status = 'delivered'
  GROUP BY
    1
)

SELECT
  STATE,
  CONCAT(
    CAST(FLOOR(Delivery_Diff_Seconds / (24 * 60 * 60)) AS STRING), ' days ',
    CAST(FLOOR(MOD(Delivery_Diff_Seconds, (24 * 60 * 60)) / 3600) AS STRING), ' hours ',
    CAST(FLOOR(MOD(MOD(Delivery_Diff_Seconds, (24 * 60 * 60)), 3600) / 60) AS STRING), '
minutes'
  ) AS Delivery_Actual_Estimate_Diff
FROM
  ProcessedData
ORDER BY
  Delivery_Diff_Seconds
LIMIT
  5;
```

Output :

Row	STATE	Delivery_Actual_Estimate_Diff
1	AL	8 days 0 hours 46 minutes
2	MA	8 days 21 hours 18 minutes
3	SE	9 days 7 hours 53 minutes
4	ES	9 days 19 hours 7 minutes
5	BA	10 days 2 hours 22 minutes

Observation :
The list of top 5 states with fast delivery times encompasses diverse regions of Brazil. States such as Alagoas (AL), Maranhão (MA), Sergipe (SE), Espírito Santo (ES), and Bahia (BA) represent both *Northeastern and Southeastern regions.* These are the 5 states with the lowest average difference between actual and estimated delivery time.

6. Analysis based on the payments:

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

Query :

```
WITH Calendar AS (
  SELECT DISTINCT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS YEAR,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS MONTH
```

```
  FROM
    `TargetAnalyticsDB.orders` o
),

PaymentChannels AS (
  SELECT DISTINCT
    payment_type
  FROM
    `TargetAnalyticsDB.payments`
)

SELECT
  c.YEAR,
  c.MONTH,
  UPPER(pc.payment_type) AS PAYMENT_CHANNEL,
  COALESCE(COUNT(p.order_id), 0) AS PAY_COUNT
FROM
  Calendar c
CROSS JOIN
  PaymentChannels pc
LEFT JOIN
  `TargetAnalyticsDB.orders` o ON EXTRACT(YEAR FROM o.order_purchase_timestamp) = c.YEAR
AND EXTRACT(MONTH FROM o.order_purchase_timestamp) = c.MONTH
LEFT JOIN
  `TargetAnalyticsDB.payments` p ON p.order_id = o.order_id AND p.payment_type =
pc.payment_type
GROUP BY
  1, 2, 3
ORDER BY
  1, 2, 3;
```

Output :

Row	YEAR	MONTH	PAYMENT_CHANNEL	PAY_COUNT
1	2016	9	CREDIT_CARD	3
2	2016	9	DEBIT_CARD	0
3	2016	9	NOT_DEFINED	0
4	2016	9	UPI	0
5	2016	9	VOUCHER	0
6	2016	10	CREDIT_CARD	254
7	2016	10	DEBIT_CARD	2
8	2016	10	NOT_DEFINED	0
9	2016	10	UPI	63
10	2016	10	VOUCHER	23

Observation :

The dataset provides a comprehensive overview of payment distribution across various channels from September 2016 to October 2018. Notable patterns include a *gradual increase in credit card and UPI transactions*, with October 2016 experiencing a substantial surge. Debit card payments show sporadic peaks, while voucher transactions remain relatively low but exhibit intermittent increases. The data underscores dynamic shifts in payment preferences, with 2017 witnessing *heightened activity in January, March, and November.* Credit card payments consistently dominate the landscape, maintaining a steady presence in 2018, while September 2018 stands out for a significant rise in voucher payments. This analysis illuminates evolving consumer payment behavior and highlights periods of heightened activity in specific channels throughout the observed time frame.

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

Query :

```
SELECT payment_installments, COUNT(DISTINCT order_id) AS num_orders
FROM `TargetAnalyticsDB.payments`
WHERE payment_value > 0
GROUP BY payment_installments;
```

Output :

Row	payment_installments	num_orders
1	0	2
2	1	49057
3	2	12389
4	3	10443
5	4	7088
6	5	5234
7	6	3916
8	7	1623
9	8	4253
10	9	644

Observation :

The output provides a comprehensive overview of the distribution of orders based on the number of payment installments made. Notably, *a significant portion of orders (49,057) is associated with a single payment installment*, indicating a prevalent practice of immediate full payment. As the number of installments increases, the frequency of orders generally decreases, with a gradual decline observed. *The data suggests that a substantial portion of consumers opt for fewer installments or a lump-sum payment*, possibly influenced by factors such as financial preferences or available payment options. There are sporadic occurrences of orders with higher installment counts, but these are less common. Overall, this analysis sheds light on the payment behavior and preferences of consumers.