

Target: Comprehensive Data Analysis for Strategic Growth

Business Problem:

Target aims to optimize its operations in Brazil by improving order processing, pricing strategies, and customer satisfaction. Analyzing order trends, payment behaviors, and shipping efficiency can help identify areas for enhancement.

Understanding **customer demographics and product preferences** is crucial for **tailored marketing and inventory management**. Additionally, evaluating **customer reviews and feedback** will provide insights into **service improvements and overall brand perception**.

About Target:

Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

Problem Statement:

- ❖ **Order Processing & Efficiency** – Identify bottlenecks in order fulfillment, delays, and cancellations to improve overall processing time and efficiency.
- ❖ **Pricing & Revenue Optimization** – Analyze pricing trends and customer spending behavior to refine pricing strategies and maximize revenue.
- ❖ **Payment & Shipping Performance** – Assess payment methods, transaction success rates, and freight efficiency to enhance customer convenience and delivery speed.
- ❖ **Customer Satisfaction & Retention** – Evaluate customer reviews and feedback to address service gaps, improve product offerings, and boost customer loyalty.

Dataset Overview:

The dataset consists of **8 CSV files** containing detailed information on **customers, sellers, orders, payments, products, reviews, and geolocation**, providing a **comprehensive view of Target's operations in Brazil**.

- **customers.csv** – Contains **customer details** like **unique ID, zip code, city, and state** for analyzing **consumer demographics**.
- **sellers.csv** – Provides **seller information**, including **unique ID, location zip code, city, and state**.
- **order_items.csv** – Captures details of each **order item**, including **product ID, seller ID, price, shipping deadline, and freight cost**.
- **geolocation.csv** – Stores **location data** with **zip code, latitude, longitude, city, and state** for **mapping purposes**.
- **payments.csv** – Tracks **transaction details** such as **payment type, total amount, and installment breakdown**.
- **orders.csv** – Includes **order processing details** such as **status, timestamps, estimated delivery, and actual delivery dates**.
- **reviews.csv** – Contains **customer feedback** with **review scores, comments, and timestamps of creation and response**.
- **products.csv** – Provides **product-related attributes** like **category, dimensions, weight, and description details**.

This **structured data** will allow for **comprehensive analysis** of **Target's operations**, helping to **optimize processes** and **improve customer experience**.

1. Initial exploration like checking the structure & characteristics of the data.

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

QUERY:

```
SELECT
    column_name,
    data_type
FROM
    `rishabh-sql-project.target.INFORMATION_SCHEMA.COLUMNS`
WHERE
```

```
table_name = 'customers'
```

OUTPUT:

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

INSIGHTS:

- ❖ **Customer Identification:**
 - The **customer_id** and **customer_unique_id** help differentiate each customer.
 - This is crucial for **tracking individual customer activity** and **analyzing their behavior** to improve **services** or **marketing strategies**.
- ❖ **Location-Based Insights:**
 - The **customer_zip_code_prefix**, **customer_city**, and **customer_state** enable **geographical analysis**.
 - Businesses can use this data to **identify popular areas**, **tailor marketing campaigns** to specific regions, and **optimize delivery services**.
- ❖ **Customer Segmentation:**
 - By combining **ID with location data** (**customer_city** and **customer_state**), businesses can create **detailed customer profiles**.
 - This segmentation can lead to **more personalized customer experiences**, **targeted offers**, and **strategic planning** for new store locations or localized services.

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

QUERY:

```
SELECT
    MIN(order_purchase_timestamp) AS first_order,
    MAX(order_purchase_timestamp) AS last_order
FROM
    `rishabh-sql-project.target.orders`
```

OUTPUT:

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

INSIGHTS:

- ❖ **Customer Loyalty:**
 - The **time span of over two years** between the **first and last orders** indicates that the customer might be **returning**, which is a good sign of **loyalty**.
 - This duration suggests the business successfully **retained this customer** over a **significant period**, which is crucial for **long-term success**.
- ❖ **Business Analysis:**
 - **Analyzing order timing** can help pinpoint **when customers are more likely to make purchases**, which can be **vital for strategic planning**.
 - Recognizing patterns like **seasonal buying trends** can guide **inventory management** and **promotional campaigns**.
- ❖ **Customer Engagement:**
 - **Short intervals** between orders suggest **effective engagement strategies**, possibly through **promotions** or by maintaining **high customer satisfaction**.
 - **Longer intervals** might highlight **opportunities for improvement** in **marketing** or **customer service** to encourage **more frequent purchases**.

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

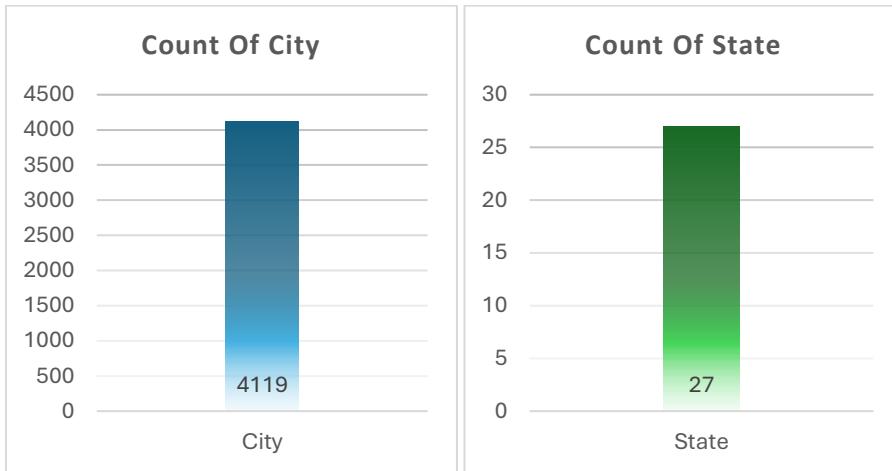
QUERY:

```
SELECT
    COUNT(DISTINCT c.customer_city) AS city_count,
    COUNT(DISTINCT c.customer_state) AS state_count
FROM
    `rishabh-sql-project.target.orders` o
INNER JOIN
    `rishabh-sql-project.target.customers` c
ON
    o.customer_id = c.customer_id
```

OUTPUT:

Row	city_count	state_count
1	4119	27

VISUALIZATION:



INSIGHTS:

- ❖ **Geographic Diversity:**
 - The **high count of cities (4119)** reflects **substantial geographic diversity**, indicating the company serves a **wide and varied customer base**.
 - This **diversity presents opportunities for localized marketing strategies and promotions tailored to specific urban characteristics or needs**.
- ❖ **Market Concentration:**
 - The **count of states (27)** compared to cities suggests a **concentrated presence in fewer states**, highlighting **where the core of the customer base resides**.
 - Understanding **where most customers are located** can help in **planning expansions or enhancing supply chain efficiencies in key states**.
- ❖ **Business Strategy Implications:**
 - The **discrepancy in city and state counts** can guide **targeted customer segmentation efforts**, optimizing **resource allocation** and **marketing campaigns** in **high-engagement areas**.
 - Insights into **geographic distribution** assist in **optimizing logistics, reducing delivery times and costs**, particularly in **states with high order volumes**.

2. In-depth Exploration.

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

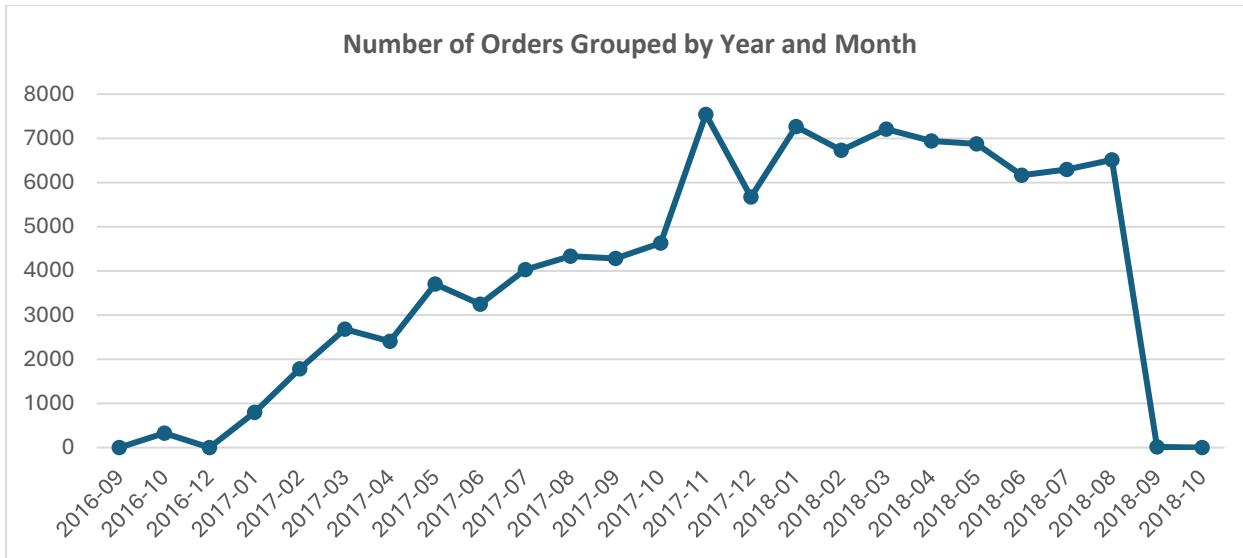
QUERY:

```
SELECT
    EXTRACT(year
    FROM
        order_purchase_timestamp) AS year,
    EXTRACT(month
    FROM
        order_purchase_timestamp) AS month,
    COUNT(*) AS num_of_orders
FROM
    `rishabh-sql-project.target.orders`
GROUP BY
    year,
    month
ORDER BY
    year,
    month
```

OUTPUT: (First 10 Rows)

Row	year	month	num_of_orders
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

VISUALIZATION:



INSIGHTS:

- ❖ **Growth Trends:**
 - Orders steadily increased from 2016 through early 2018, suggesting effective marketing or growing market reach.
 - High order volumes in November 2017 and early 2018 highlight successful periods, possibly due to seasonal sales or promotions.
- ❖ **Potential Issues:**
 - The sharp drops in September and October 2018 need investigation, as they might point to issues like operational disruptions or data recording errors.
 - Investigating these anomalies could provide insights into potential improvements or necessary adjustments in business strategies.
- ❖ **Seasonal Variations:**
 - Seasonal patterns are clear, with spikes in November likely driven by holiday shopping, emphasizing the impact of seasonal promotions.
 - The subsequent drops in December indicate a return to normal buying patterns after the holiday season, showcasing typical consumer behavior trends.

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

QUERY:

```
SELECT
    EXTRACT(month
    FROM
        order_purchase_timestamp) AS month,
```

```
COUNT(*) AS num_of_orders
FROM
`rishabh-sql-project.target.orders`
GROUP BY
month
ORDER BY
month
```

OUTPUT:

Row	month	num_of_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

VISUALIZATION:



INSIGHTS:

- ❖ **Peak Season:**
 - Order volumes peak during May through August, suggesting strong sales in the summer months, possibly due to seasonal promotions or holiday buying behavior.
 - The peak in August with over 10,800 orders could indicate a key marketing period or seasonal product popularity.
- ❖ **Sharp Declines:**
 - There is a significant drop in orders starting in September, with the lowest in October at just under 5,000 orders, indicating a possible seasonal downturn or end of a promotional period.
 - The sharp decline post-August could reflect a return to normal buying patterns after summer promotions.
- ❖ **Year-End Variations:**
 - Order numbers start to recover in November and December, which could be attributed to holiday season shopping, with orders rising to 7,544 in November.
 - Despite a slight increase during the holiday season, the numbers in December (5,674) don't return to the mid-year highs, which could be influenced by market saturation or consumer spending fatigue post-holidays.

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:

```

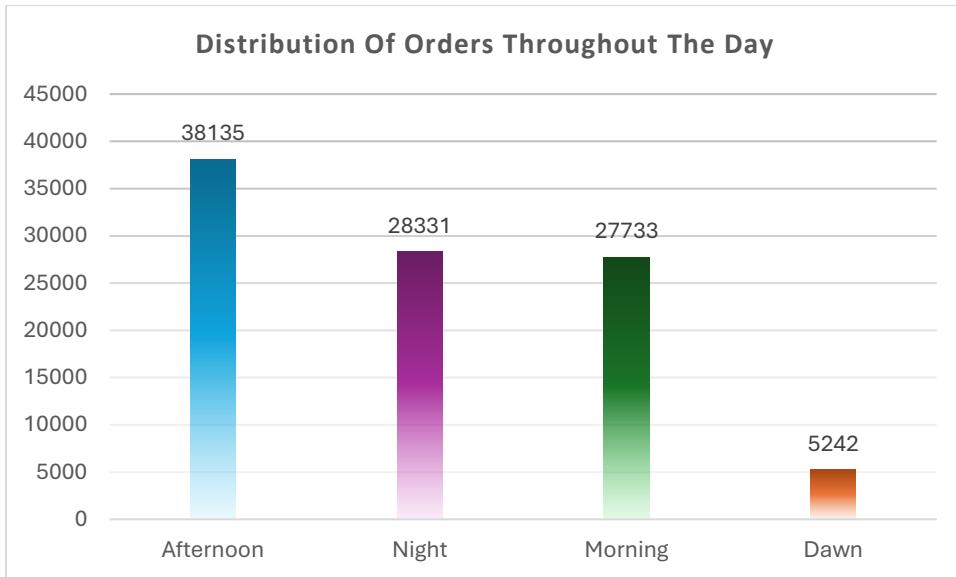
SELECT
  CASE
    WHEN EXTRACT (hour FROM order_purchase_timestamp) BETWEEN 0 AND 6 THEN
      'dawn'
    WHEN EXTRACT (hour FROM order_purchase_timestamp) BETWEEN 7 AND 12 THEN
      'morning'
    WHEN EXTRACT (hour FROM order_purchase_timestamp) BETWEEN 13 AND 18 THEN
      'afternoon'
    WHEN EXTRACT (hour FROM order_purchase_timestamp) BETWEEN 19 AND 23 THEN
      'night'
  END AS time_of_day,
  COUNT(*) AS num_of_orders
FROM
  `rishabh-sql-project.target.orders`
GROUP BY
  time_of_day
ORDER BY
  num_of_orders DESC

```

OUTPUT:

Row	time_of_day	num_of_orders
1	afternoon	38135
2	night	28331
3	morning	27733
4	dawn	5242

VISUALIZATION:



INSIGHTS:

❖ Peak Ordering Times:

- The **afternoon sees the highest number of orders, totaling 38,135**, which suggests that consumers might be using **lunch breaks or afternoon leisure time** to place their orders.
- This period likely aligns with **more active online engagement** and could be an **optimal time for businesses to push notifications or special deals**.

❖ Comparatively Lower Night Activity:

- **Night orders, while lower than afternoon, are still substantial at 28,331**, indicating a **second peak of consumer activity** that could be related to **after-work shopping habits**.
- Businesses could consider **targeting evening promotions or customer service availability enhancements** to capture this **significant segment of the market**.

❖ Morning and Dawn Order Volumes:

- **Mornings see a good amount of activity with 27,733 orders**, potentially capturing the **early birds** or individuals **planning their purchases** after starting their day.
- **Dawn, with only 5,242 orders, is the least active**. This time could be leveraged for **inventory updates and website maintenance** when **user traffic is minimal**.

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

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

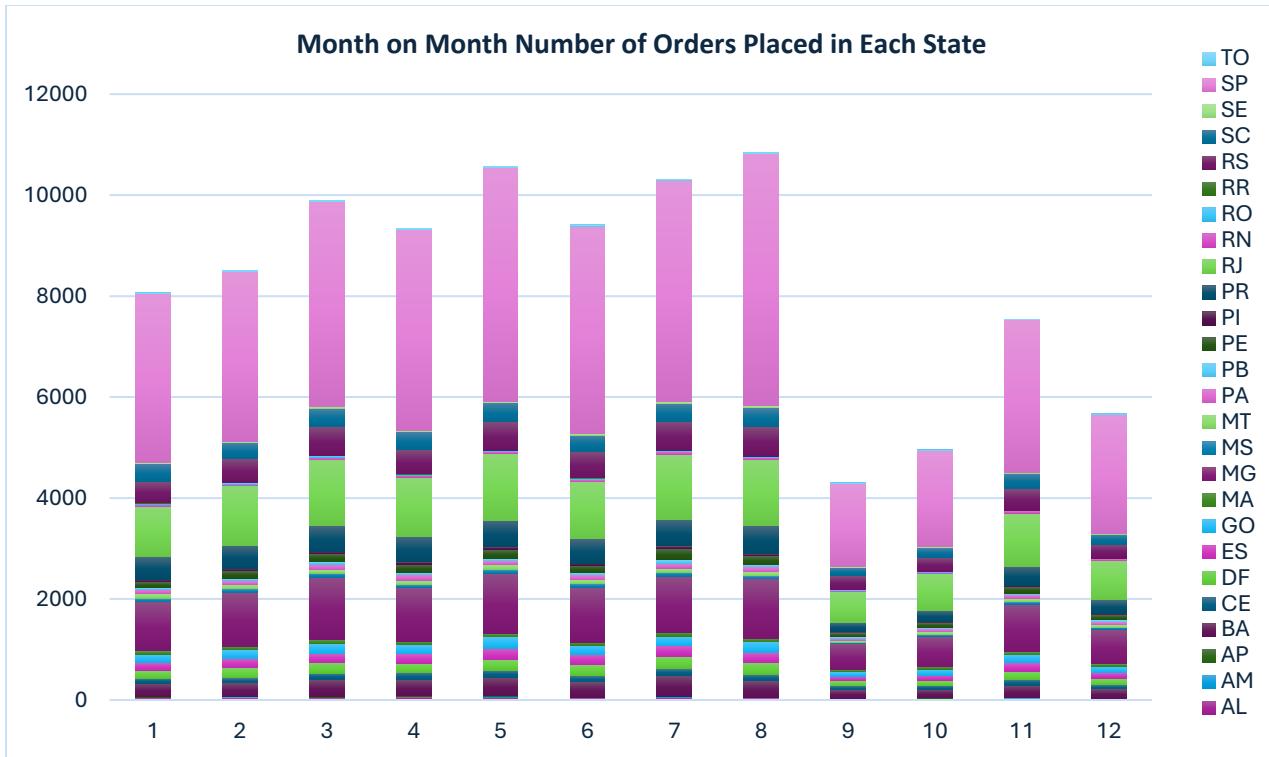
QUERY:

```
SELECT
    EXTRACT(month FROM order_purchase_timestamp) AS month,
    c.customer_state,
    COUNT(*) AS num_of_orders
FROM
    `rishabh-sql-project.target.orders` o
INNER JOIN
    `rishabh-sql-project.target.customers` c
ON
    o.customer_id = c.customer_id
GROUP BY
    c.customer_state,
    month
ORDER BY
    num_of_orders DESC
```

OUTPUT: (First 10 Rows)

Row	month	customer_state	num_of_orders
1	8	SP	4982
2	5	SP	4632
3	7	SP	4381
4	6	SP	4104
5	3	SP	4047
6	4	SP	3967
7	2	SP	3357
8	1	SP	3351
9	11	SP	3012
10	12	SP	2357

VISUALIZATION:



INSIGHTS:

- ❖ **High Seasonal Variation:**
 - The chart shows significant peaks during certain months, indicating high seasonal demand variations, which could be driven by local events or holidays.
 - For instance, the substantial increase in orders during the middle months may suggest a link to seasonal sales or festivals.
- ❖ **State Performance Differences:**
 - SP dominates in order volume across most months, highlighting its significance as a major market.
 - Other states like RJ also show consistent contributions, though much less than SP, indicating a potential focus area for regional growth strategies.
- ❖ **Year-End Sales Impact:**
 - A noticeable decline in orders for most states towards the year-end, with a dip in December, potentially due to the end of holiday promotions.
 - This trend might help in planning better inventory and marketing efforts to sustain or increase sales during this period.

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

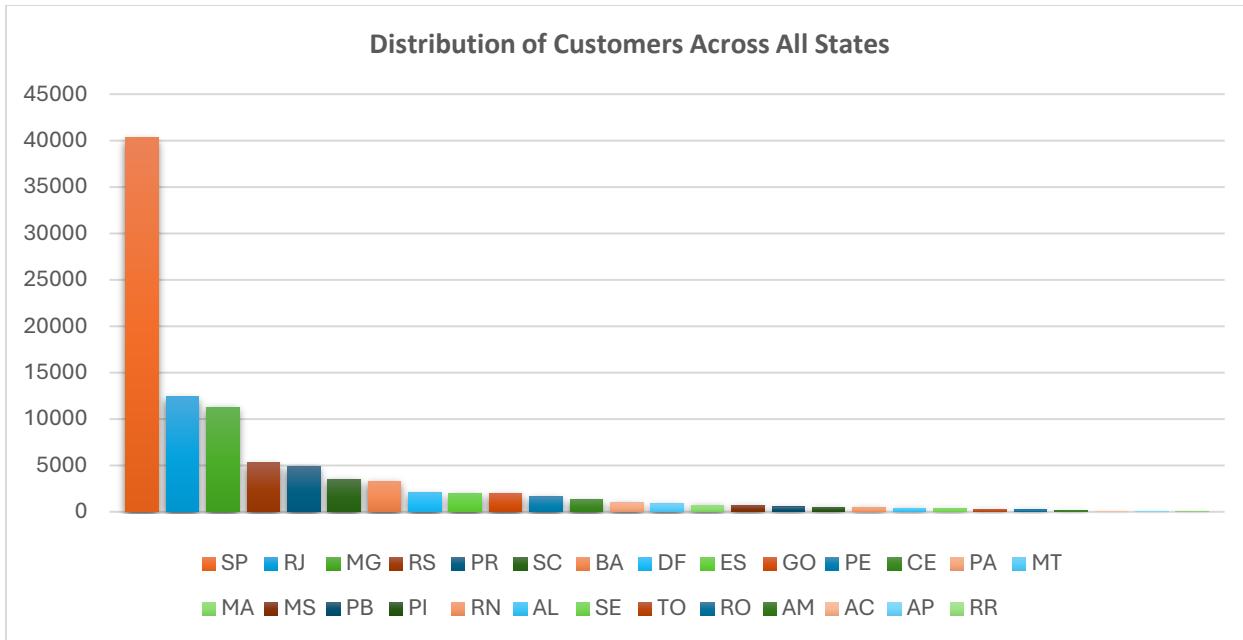
QUERY:

```
SELECT
    customer_state,
    COUNT(DISTINCT(customer_unique_id)) AS num_of_customers
FROM
    `rishabh-sql-project.target.customers`
GROUP BY
    customer_state
ORDER BY
    num_of_customers DESC
```

OUTPUT: (First 10 Rows)

Row	customer_state	num_of_customers
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882
6	SC	3534
7	BA	3277
8	DF	2075
9	ES	1964
10	GO	1952

VISUALIZATION:



INSIGHTS:

- ❖ Major Customer Bases:
 - **SP remains the largest market with over 40,000 customers**, showcasing its importance in overall business operations.
 - **RJ and MG also have substantial customer bases, with over 12,000 and 11,000 customers respectively**, reinforcing their role as critical markets.
- ❖ Significant Secondary Markets:
 - **RS and PR have sizable customer populations of over 5,000 each**, indicating a strong market presence and potential for targeted engagement.
 - **States like SC, BA, and DF, with customers ranging from 2,000 to 3,500**, present opportunities for expanding market penetration.
- ❖ Emerging and Smaller Markets:
 - **States such as GO, PE, and CE have emerging customer bases**, suggesting potential areas for growth and market development.
 - **Smaller states like AC, AP, and RR, though they have the smallest customer counts**, could benefit from specific strategic initiatives to boost the market presence.

4. Impact on Economy.

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

QUERY:

```
WITH YearlyCost AS (
    SELECT
        EXTRACT(year FROM o.order_purchase_timestamp) AS year,
        SUM(p.payment_value) AS total_payment
    FROM
        `rishabh-sql-project.target.orders` o
    INNER JOIN
        `rishabh-sql-project.target.payments` p
    ON
        o.order_id = p.order_id
    WHERE
        EXTRACT(month
            FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
        AND EXTRACT(year
            FROM o.order_purchase_timestamp) BETWEEN 2017 AND 2018
    GROUP BY
        year
)
SELECT
    2017 AS year_2017,
    2018 AS year_2018,
    a.total_payment AS total_payment_2017,
    b.total_payment AS total_payment_2018,
    ROUND(((b.total_payment - a.total_payment) / a.total_payment) * 100, 2) AS
percent_increase
FROM
    YearlyCost a,
    YearlyCost b
WHERE
    a.year = 2017 AND b.year = 2018;
```

OUTPUT:

Row	year_2017	year_2018	total_payment_2017	total_payment_2018	percent_increase
1	2017	2018	3669022.120000...	8694733.840000...	136.98

INSIGHTS:

- ❖ **Significant Growth:**
 - There was a **substantial increase of 136.98% in total payment from 2017 to 2018**, reflecting a **major growth in sales or order value**.

- This suggests **effective marketing strategies** or an **expansion in the customer base** during this period.
- ❖ **Market Expansion:**
 - The **large increase in payments** could indicate **successful entry into new markets** or the **introduction of new products** that resonated well with consumers.
 - It might also reflect an **overall increase in the average order value**, suggesting **customers are purchasing more expensive items or more products per order**.
- ❖ **Economic Implications:**
 - Such **significant year-on-year growth** could have **positive implications for the company's revenue and profitability**, assuming **costs have not increased at the same rate**.
 - This **growth rate could attract potential investors** or be used in **marketing materials** to highlight the company's **robust growth in a competitive market**.

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

QUERY:

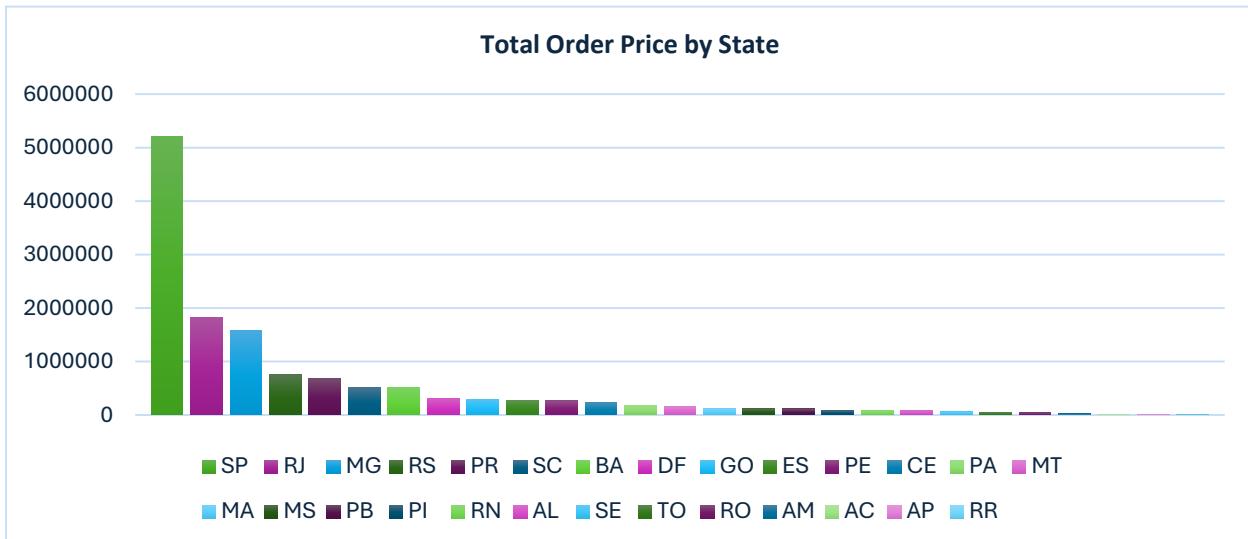
```
WITH StateOrderTotals AS (
  SELECT
    c.customer_state AS state,
    SUM(price) AS total_price,
    COUNT(DISTINCT(o.order_id)) AS num_of_orders
  FROM
    `rishabh-sql-project.target.orders` o
  INNER JOIN
    `rishabh-sql-project.target.order_items` i
  ON
    o.order_id= i.order_id
  INNER JOIN
    `rishabh-sql-project.target.customers` c
  ON
    o.customer_id=c.customer_id
  GROUP BY
    state )
  SELECT
    state,
    total_price,
    num_of_orders,
    (total_price/num_of_orders) AS avg_price
  FROM
    StateOrderTotals
  ORDER BY
```

```
total_price DESC
```

OUTPUT: (First 10 Rows)

Row	state ▾	total_price ▾	num_of_orders ▾	avg_price ▾
1	SP	5202955.0500013418	41375	125.75117945622578
2	RJ	1824092.6699998307	12762	142.93156793604692
3	MG	1585308.0299999055	11544	137.32744542618724
4	RS	750304.02000001865	5432	138.1266605301949
5	PR	683083.76000001514	4998	136.67142056823033
6	SC	520553.34000000561	3612	144.1177574750846
7	BA	511349.99000000581	3358	152.27813877308094
8	DF	302603.93999999797	2125	142.40185411764611
9	GO	294591.949999997	2007	146.78223716990382
10	ES	275037.30999999674	2025	135.82089382715887

VISUALIZATION:





INSIGHTS:

- ❖ **Market Leadership:**
 - ❖ SP continues to dominate both in terms of total sales volume and the number of orders, reflecting its strong market position with over 5 million in sales and around 41,375 orders.
 - ❖ This dominance suggests targeted marketing and sales strategies in SP could yield significant returns due to the large customer base.
- ❖ **Premium Market Indicators:**
 - ❖ DF shows a high average order value at approximately 142.40, despite having fewer orders (2,125), indicating that while the volume is lower, the spending per transaction is high.
 - ❖ This may imply a market with higher disposable incomes or a preference for premium or luxury goods, presenting opportunities for high-margin products.
- ❖ **Emerging Markets with Potential:**
 - ❖ States like GO and ES, with average prices of 146.78 and 135.82 respectively, and a substantial number of orders, show potential as emerging markets that could be ripe for expansion.
 - ❖ The substantial order volumes coupled with reasonable average spending suggest these regions could benefit from increased marketing efforts or expanded distribution networks to boost sales.

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

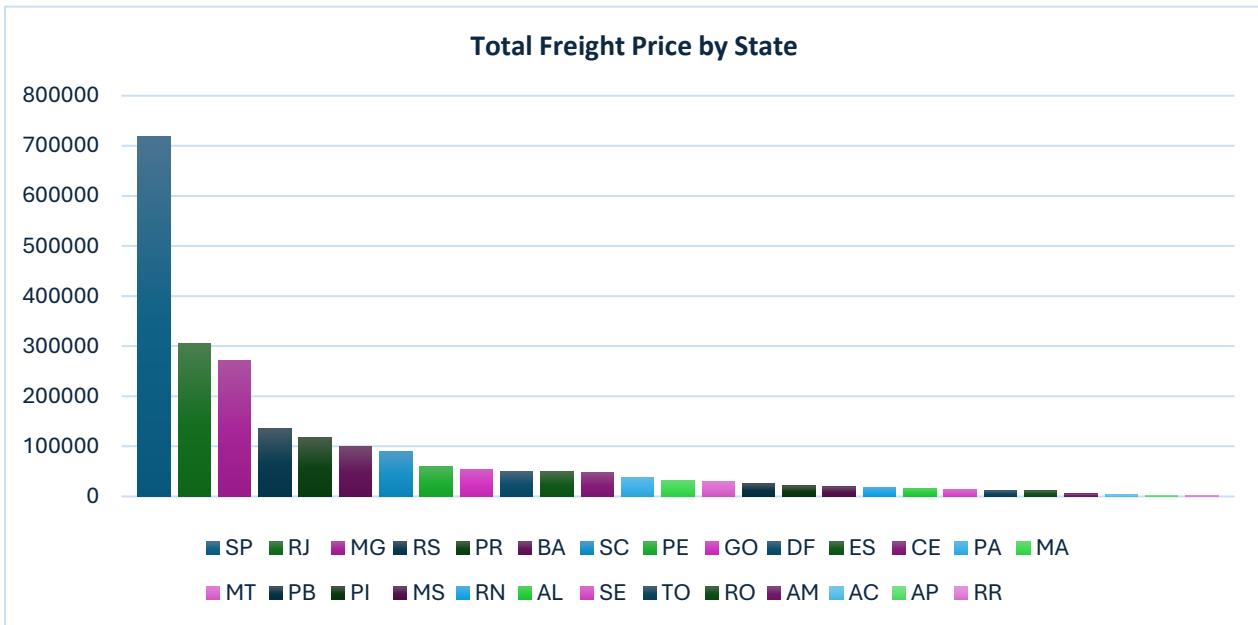
QUERY:

```
WITH StateFreightTotals AS (
  SELECT
    c.customer_state AS state,
    SUM(i.freight_value) AS total_freight,
    COUNT(DISTINCT o.order_id) AS num_of_orders
  FROM
    `rishabh-sql-project.target.orders` o
  INNER JOIN
    `rishabh-sql-project.target.order_items` i ON o.order_id = i.order_id
  INNER JOIN
    `rishabh-sql-project.target.customers` c ON o.customer_id = c.customer_id
  GROUP BY
    state
)
SELECT
  state,
  total_freight,
  num_of_orders,
  (total_freight / num_of_orders) AS avg_freight
FROM
  StateFreightTotals
ORDER BY
  total_freight DESC;
```

OUTPUT: (First 10 Rows)

Row	state ▾	total_freight ▾	num_of_orders ▾	avg_freight ▾
1	SP	718723.0700000088	41375	17.370950332326498
2	RJ	305589.3100000018	12762	23.945252311550053
3	MG	270853.46000000072	11544	23.462704435204497
4	RS	135522.73999999967	5432	24.948958026509512
5	PR	117851.6799999996	4998	23.579767907162786
6	BA	100156.68000000009	3358	29.826289458010748
7	SC	89660.2599999998	3612	24.822884828349892
8	PE	59449.660000000018	1648	36.073822815533994
9	GO	53114.980000000025	2007	26.464862979571514
10	DF	50625.5000000000116	2125	23.823764705882407

VISUALIZATION:



INSIGHTS:

- ❖ **Freight Cost Distribution:**
 - **SP dominates in freight volume**, with a **total freight cost of over 718,000** and the **highest number of orders at 41,375**, reflecting its **central role in logistics and commerce within the region**.

- **RJ and MG follow as major contributors** with significant freight costs of 305,589 and 270,853 respectively, but with **higher average freight costs compared to SP**, suggesting **longer or more costly delivery routes** or a **higher proportion of premium freight services**.
 - ❖ **High Average Freight Costs:**
 - PI and PB show the highest average freight costs at approximately 48.35 and 48.45 respectively. These elevated averages could be due to **geographical challenges** or **lesser availability of optimized freight routes**.
 - This highlights **potential areas where logistics efficiency can be improved** or where **alternative shipping methods** might be necessary to **reduce costs**.
 - ❖ **Strategic Opportunities for Optimization:**
 - DF and ES, with average freight costs of 23.83 and 24.57 respectively, showcase **more efficient freight pricing strategies** despite their **total and per order freight costs**. This could be indicative of **successful negotiations with shipping providers** or **more effective distribution systems**.
 - **States with lower average freight costs but significant total costs** like SP, RJ, and MG might still benefit from **bulk shipping discounts** or **more strategic logistics planning** to further **reduce per-order costs**.
-

5. Analysis 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.

QUERY:

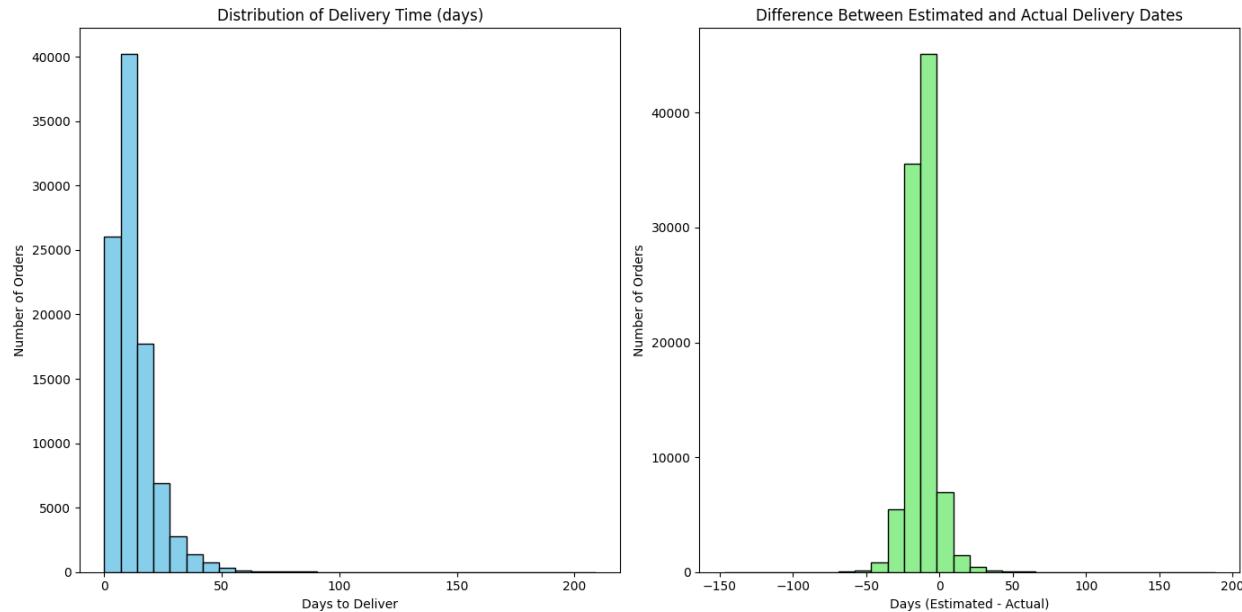
```

SELECT
    order_id,
    TIMESTAMP_DIFF(order_delivered_customer_date, order_purchase_timestamp,
    DAY) AS time_to_deliver,
    TIMESTAMP_DIFF(order_delivered_customer_date,
    order_estimated_delivery_date, DAY) AS diff_estimated_delivery
FROM
    `rishabh-sql-project.target.orders`
WHERE
    order_status = 'delivered';
  
```

OUTPUT: (First 10 Rows)

Row	order_id	time_to_deliver	diff_estimated_delivery
1	bfb0f9bdef84302105ad712db648a6c	54	36
2	98974b076b01553d49ee6467905675a7	43	-6
3	c4b41c36dd589e901f6879f25a74ec1d	36	-14
4	d2292ff2201e74c5db154d1b7ae68cbb	29	-20
5	95e01270fcbae9863423400103359279	30	-19
6	ed8c7b1b3eb256c70ce0c74231e1da88	44	-5
7	5cc475c7c03290048eb2e742cd64cb5e	68	18
8	6b3ee7697a02619a0ace2b3f0aa46bde	47	-2
9	3b2ca3293a7ce539ea2379d704fa37ce	43	-7
10	b2f92b2f7047cd8b35580d629d7b3bfb	43	-7

VISUALIZATION:



INSIGHTS:

- ❖ **Delivery Time Efficiency:**
 - Most orders are delivered within a relatively short timeframe, as indicated by the significant peak in the first chart for delivery days.
 - This suggests a well-optimized logistics operation for most orders.

- ❖ **Estimation Accuracy:**
 - The **second chart shows a large concentration around the zero mark**, indicating that **many orders are delivered on or very close to the estimated delivery dates**.
 - This highlights the **accuracy of the delivery estimations** provided to customers, enhancing **customer satisfaction**.
- ❖ **Outliers and Improvement Areas:**
 - **Some orders are delivered much earlier or later than estimated**, as seen by the **tails extending in both directions from the center of the second chart**.
 - **Analyzing these outliers** could provide insights into **inconsistencies in the delivery process or estimation methods**, offering **opportunities for further optimization**.

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

QUERY:

```
WITH StateFreight AS (
    SELECT
        c.customer_state AS state,
        AVG(i.freight_value) AS avg_freight
    FROM
        `rishabh-sql-project.target.orders` o
    INNER JOIN
        `rishabh-sql-project.target.order_items` i ON o.order_id = i.order_id
    INNER JOIN
        `rishabh-sql-project.target.customers` c ON o.customer_id =
    c.customer_id
    WHERE
        o.order_status = 'delivered'
    GROUP BY
        c.customer_state
),
TopFreight AS (
    SELECT
        state AS top_5_state,
        avg_freight AS top_5_avg_freight,
        ROW_NUMBER() OVER (ORDER BY avg_freight DESC) AS rank
    FROM
        StateFreight
    LIMIT 5
),
BottomFreight AS (
    SELECT
        state AS bottom_5_state,
        avg_freight AS bottom_5_avg_freight,
```

```

        ROW_NUMBER() OVER (ORDER BY avg_freight ASC) AS rank
    FROM
        StateFreight
    LIMIT 5
)

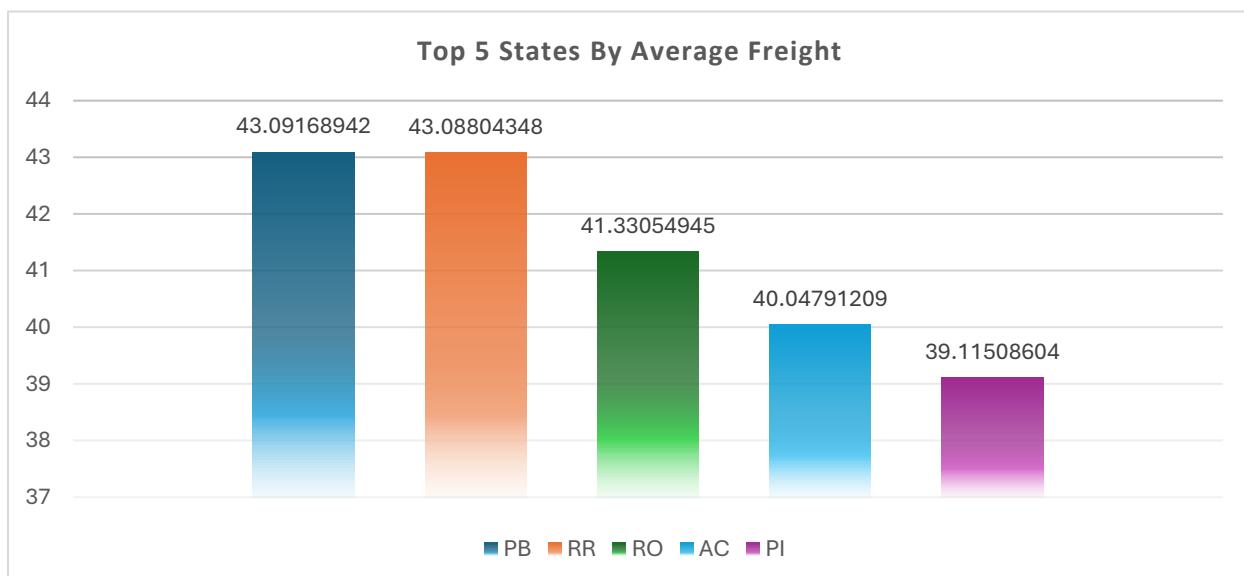
SELECT
    t.top_5_state,
    t.top_5_avg_freight,
    b.bottom_5_state,
    b.bottom_5_avg_freight
FROM
    TopFreight t
JOIN
    BottomFreight b ON t.rank = b.rank;

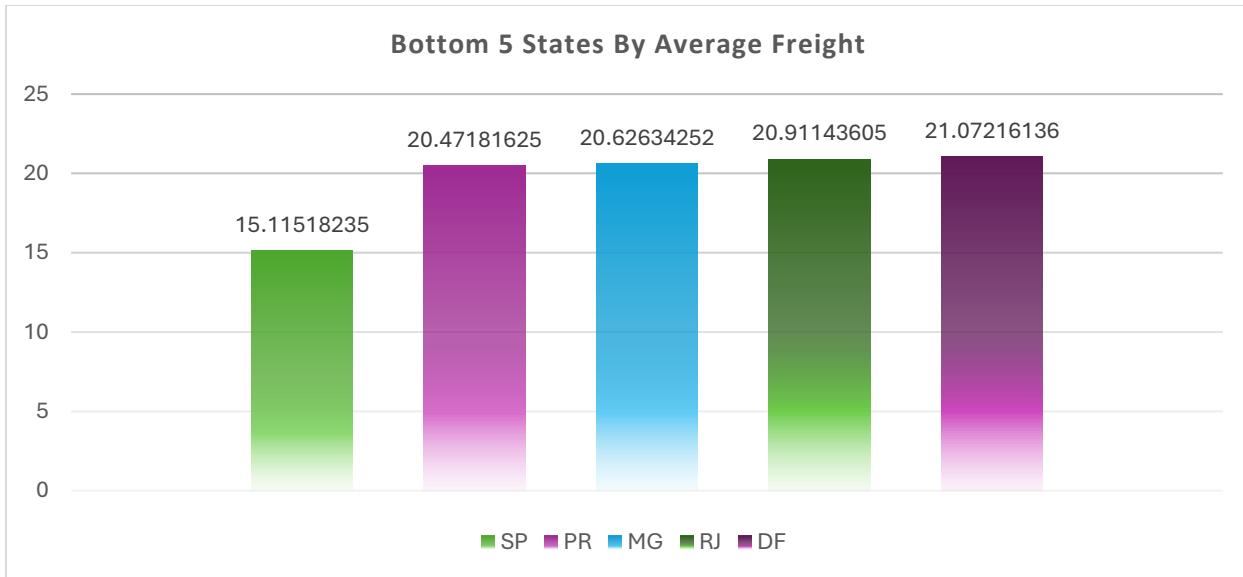
```

OUTPUT:

Row	top_5_state	top_5_avg_freight	bottom_5_state	bottom_5_avg_freight
1	PB	43.09168941979523	SP	15.115182354461009
2	RR	43.088043478260865	PR	20.471816250663853
3	RO	41.330549450549427	MG	20.626342520904252
4	AC	40.0479120879121	RJ	20.911436046100459
5	PI	39.115086042064966	DF	21.072161358811034

VISUALIZATION:





INSIGHTS:

- ❖ **Freight Cost Variation:**
 - States like PB and RR have the highest average freight charges, indicating potentially longer distances or less efficient logistics routes.
 - This could suggest a need for logistical optimizations or revised shipping strategies in these areas.
- ❖ **Efficiency in Freight Management:**
 - States such as SP and PR have some of the lowest average freight charges, possibly due to better logistics infrastructure or higher order volumes, reducing per-unit costs.
 - This demonstrates efficient supply chain management and possibly higher consumer market density, which can lower transportation costs.
- ❖ **Strategic Implications:**
 - Differences in freight costs can influence business strategies, such as pricing policies or marketing strategies, tailored to specific regions.
 - Analyzing these variations can help businesses plan for better cost management and customer satisfaction by adjusting delivery methods or focusing on local warehouse solutions.

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

QUERY:

```

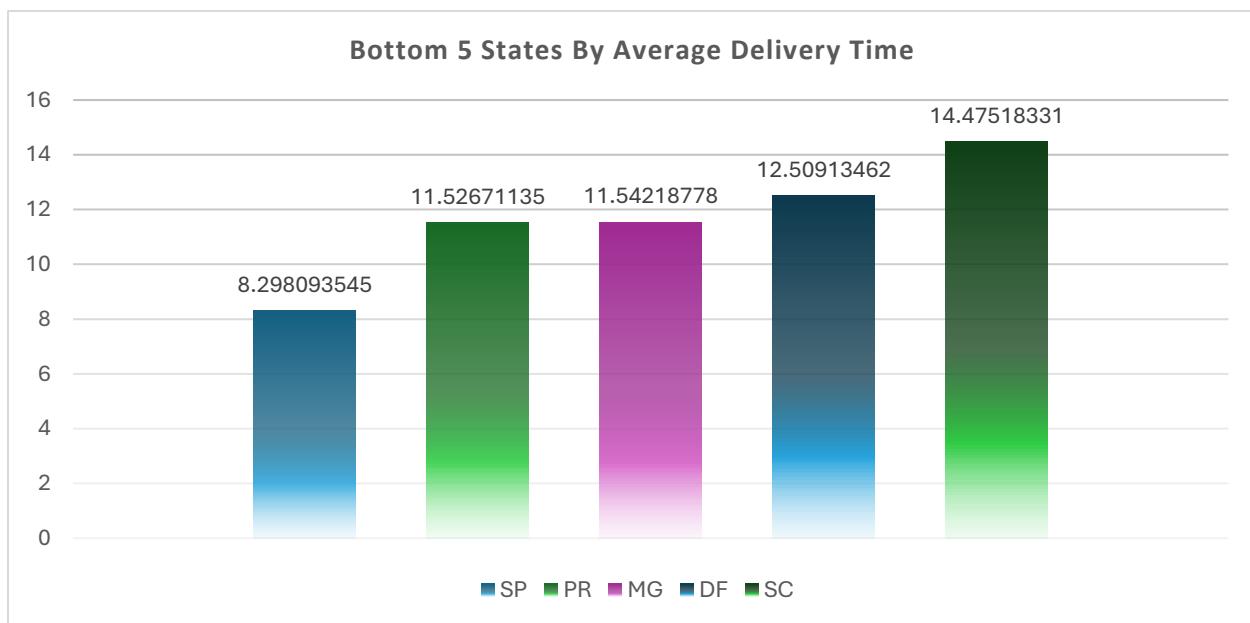
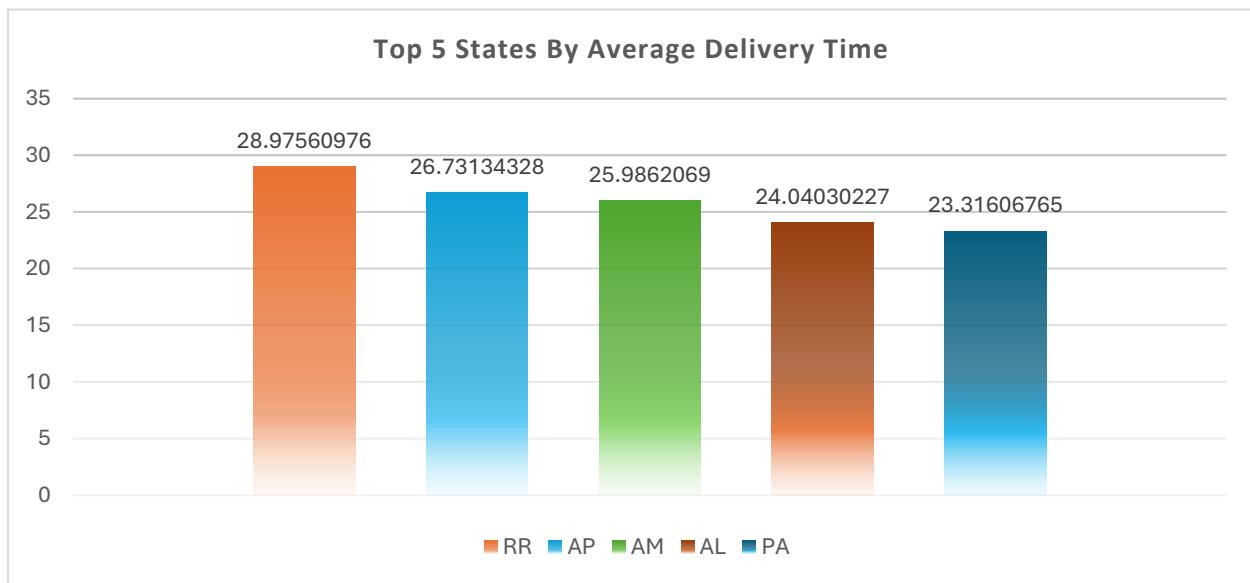
WITH StateDelivery AS (
    SELECT
        c.customer_state AS state,
        AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)) AS avg_delivery_time
    FROM
        `rishabh-sql-project.target.orders` o
    INNER JOIN
        `rishabh-sql-project.target.customers` c ON o.customer_id =
c.customer_id
    WHERE
        o.order_status = 'delivered'
    GROUP BY
        c.customer_state
),
TopDelivery AS (
    SELECT
        state AS top_5_state,
        avg_delivery_time AS top_5_avg_delivery_time,
        ROW_NUMBER() OVER (ORDER BY avg_delivery_time DESC) AS rank
    FROM
        StateDelivery
    LIMIT 5
),
BottomDelivery AS (
    SELECT
        state AS bottom_5_state,
        avg_delivery_time AS bottom_5_avg_delivery_time,
        ROW_NUMBER() OVER (ORDER BY avg_delivery_time) AS rank
    FROM
        StateDelivery
    LIMIT 5
)
SELECT
    t.top_5_state,
    t.top_5_avg_delivery_time,
    b.bottom_5_state,
    b.bottom_5_avg_delivery_time
FROM
    TopDelivery t
JOIN
    BottomDelivery b ON t.rank = b.rank;

```

OUTPUT:

Row	top_5_state	top_5_avg_delivery_time	bottom_5_state	bottom_5_avg_delivery_time
1	RR	28.975609756097565	SP	8.29809354472258
2	AP	26.731343283582088	PR	11.526711354864918
3	AM	25.986206896551714	MG	11.542187775233367
4	AL	24.040302267002534	DF	12.509134615384614
5	PA	23.316067653276956	SC	14.475183305132552

VISUALIZATION:



INSIGHTS:

- ❖ **Extended Delivery Times:**
 - States like RR and AP have the longest average delivery times, suggesting geographical challenges or less optimized logistics networks.
 - These longer times may affect customer satisfaction and could be targeted for logistical improvements.
- ❖ **Efficient Delivery Performance:**
 - States such as SP and PR show remarkably shorter average delivery times, indicating highly efficient delivery systems.
 - This efficiency is likely due to better infrastructure and possibly more concentrated logistics operations.
- ❖ **Strategic Improvement Opportunities:**
 - Comparing the top and bottom states provides valuable insights into where logistic investments could drastically improve service levels.
 - Such analyses can help businesses prioritize which areas might benefit most from increased logistics support or innovative delivery solutions.

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

QUERY:

```
WITH StateDeliverySpeed AS (
  SELECT
    c.customer_state AS state,
    AVG(DATE_DIFF(o.order_delivered_customer_date,
      o.order_estimated_delivery_date, DAY)) AS avg_delivery_diff
  FROM
    `rishabh-sql-project.target.orders` o
  INNER JOIN
    `rishabh-sql-project.target.customers` c ON o.customer_id =
      c.customer_id
  WHERE
    o.order_status = 'delivered' AND
    o.order_delivered_customer_date IS NOT NULL AND
    o.order_estimated_delivery_date IS NOT NULL
  GROUP BY
    c.customer_state
)
SELECT
  state,
  avg_delivery_diff
FROM
  StateDeliverySpeed
```

```

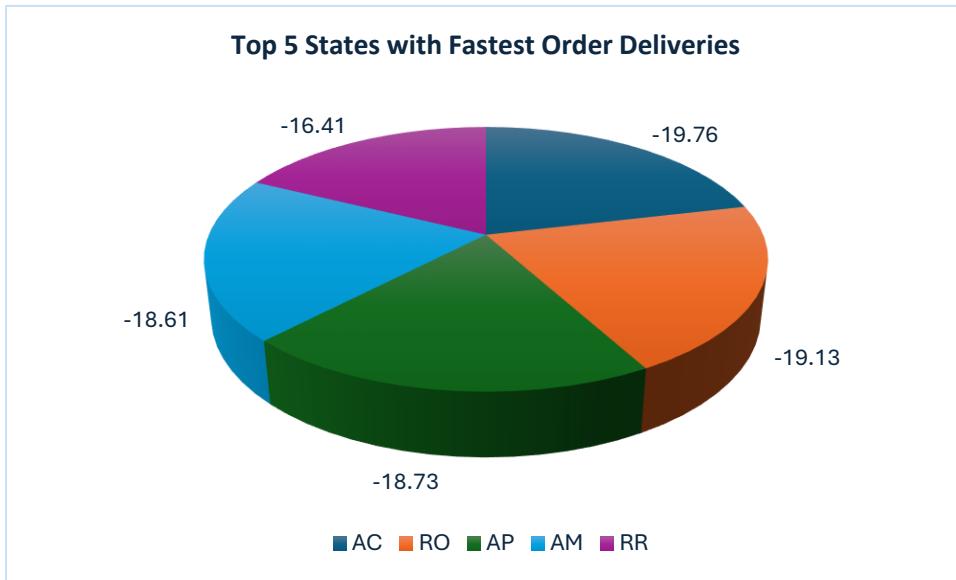
WHERE
    avg_delivery_diff < 0
ORDER BY
    avg_delivery_diff
LIMIT 5;

```

OUTPUT:

Row	state	avg_delivery_diff
1	AC	-19.76249999999996
2	RO	-19.131687242798357
3	AP	-18.731343283582092
4	AM	-18.606896551724137
5	RR	-16.414634146341459

VISUALIZATION:



INSIGHTS:

- ❖ **Exceptional Delivery Efficiency:**
 - The states AC, RO, AP, AM, and RR significantly outperformed their estimated delivery times, showing exceptional logistical efficiency.
 - This suggests a well-coordinated delivery operation in these regions that exceeds expectations.

- ❖ **Customer Satisfaction Potential:**
 - Faster than expected deliveries in these states likely contribute to higher customer satisfaction and positive customer experience.
 - This could also lead to increased customer loyalty and positive word-of-mouth for businesses operating in these regions.
 - ❖ **Benchmark for Improvement:**
 - These states set a benchmark for other regions. Analyzing the delivery logistics in these states may provide insights that could be applied elsewhere to enhance delivery performance.
 - The substantial negative average differences indicate potential areas of overestimation in delivery times, which could be adjusted to set more realistic expectations without impacting on customer satisfaction.
-

6. Analysis based on the payments.

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

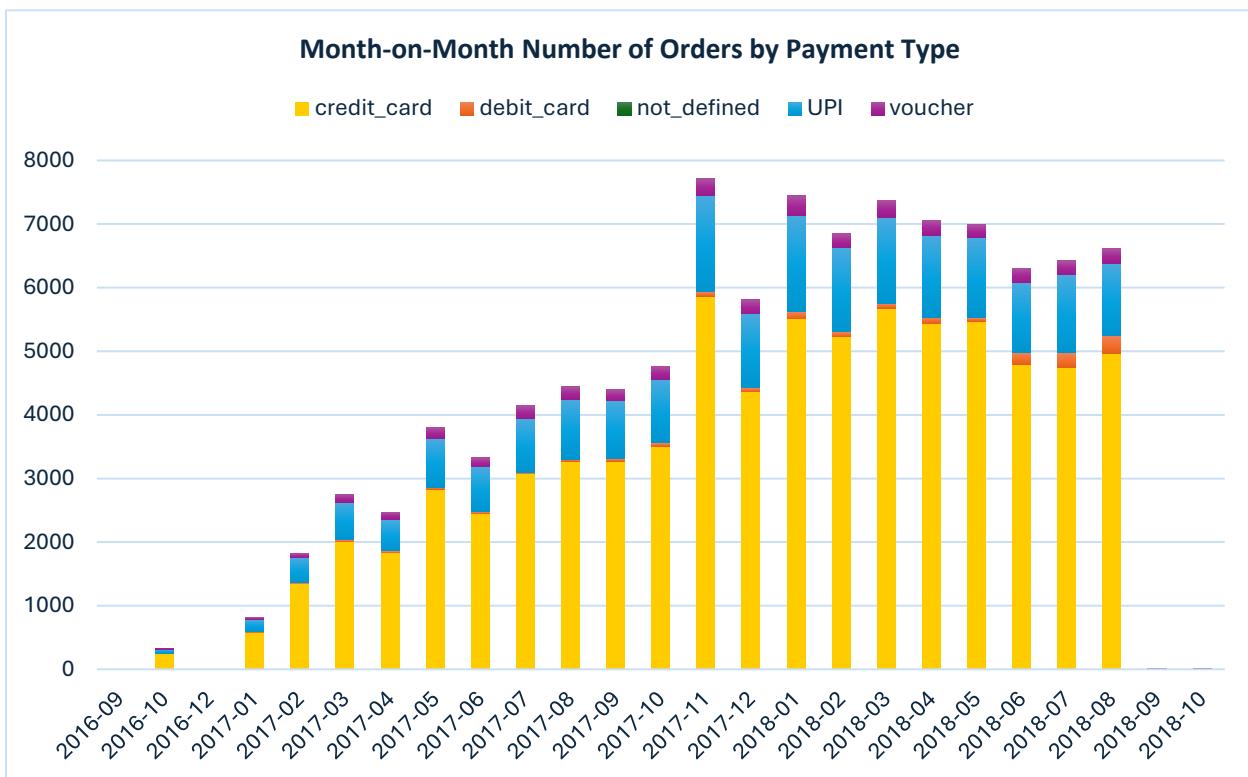
QUERY:

```
WITH MonthlyPayments AS (
  SELECT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS year,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
    p.payment_type,
    COUNT(DISTINCT o.order_id) AS num_of_orders
  FROM
    `rishabh-sql-project.target.orders` o
  INNER JOIN
    `rishabh-sql-project.target.payments` p ON o.order_id = p.order_id
  GROUP BY
    year, month, p.payment_type
)
SELECT
  year,
  month,
  payment_type,
  num_of_orders
FROM
  MonthlyPayments
ORDER BY
  year, month, payment_type;
```

OUTPUT: (First 10 Rows)

Row	year	month	payment_type	num_of_orders
1	2016	9	credit_card	3
2	2016	10	UPI	63
3	2016	10	credit_card	253
4	2016	10	debit_card	2
5	2016	10	voucher	11
6	2016	12	credit_card	1
7	2017	1	UPI	197
8	2017	1	credit_card	582
9	2017	1	debit_card	9
10	2017	1	voucher	33

VISUALIZATION:



INSIGHTS:

- ❖ Credit Card Usage:
 - Credit cards are the most used payment type, highlighting their trust and wide acceptance among consumers.
 - This trend indicates the convenience and benefits (like reward points) that credit cards offer, encouraging frequent use.
- ❖ Adoption of UPI:
 - UPI's growth is visible, reflecting its increasing popularity due to its simplicity and speed of transactions.
 - The rise in UPI transactions could be tied to broader digitalization and the ease of linking bank accounts for direct transfers.
- ❖ Payment Method Variety:
 - Multiple payment options are actively utilized, indicating a robust system that accommodates various consumer preferences.
 - Vouchers and debit cards also show consistent use, suggesting promotions and direct access to funds are valued by different customer segments.

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

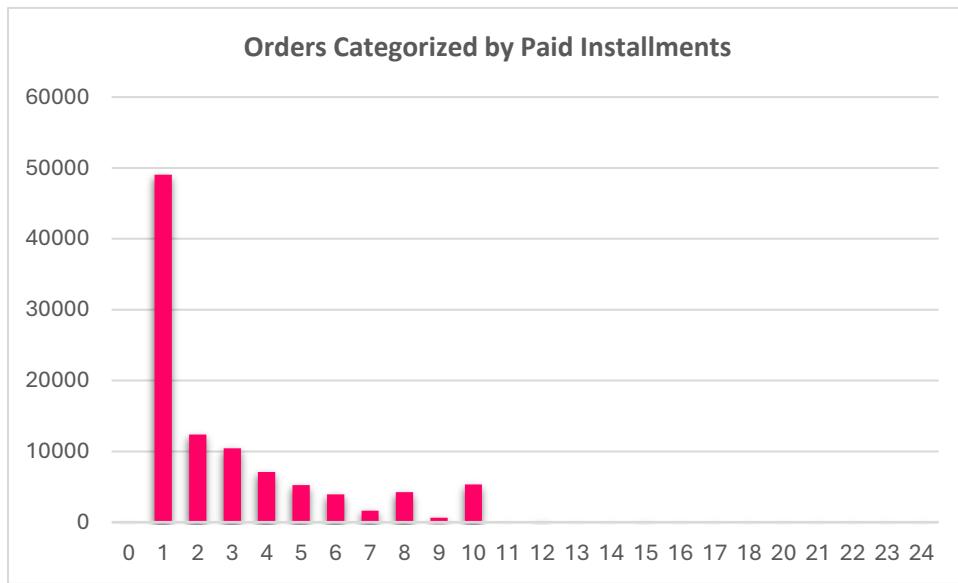
QUERY:

```
SELECT
  payment_installments,
  COUNT(DISTINCT order_id) AS num_of_orders
FROM
  `rishabh-sql-project.target.payments`
GROUP BY
  payment_installments
ORDER BY
  payment_installments;
```

OUTPUT: (First 10 Rows)

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

VISUALIZATION:



INSIGHTS:

- ❖ **Predominance of Single Payments:**
 - **Most consumers prefer to pay in one installment**, possibly to **avoid interest charges or for simpler transaction management**.
 - **Single installment payments dominate**, accounting for the **majority of transactions**, reflecting a **consumer trend towards straightforward payment solutions**.
- ❖ **Decline in Multiple Installments:**

- As the **number of installments increases, the number of orders significantly decreases**, suggesting a **disinclination toward long-term financial commitments in shopping**.
 - Orders decrease sharply after one installment, indicating that **customers may find multi-installment payments less attractive** due to **increased overall cost or complexity**.
- ❖ **Limited Use of High Installments:**
- Orders with **more than seven installments are rare**, highlighting a **low consumer preference for spreading costs over many months**.
 - **Very high installment options (8+ installments)** are scarcely used, which could reflect less favorable terms or the hassle of **long-term financial commitments in consumer perceptions**.
-

7. Additional Questions.

7.1) What is the average review score per product category?

QUERY:

```

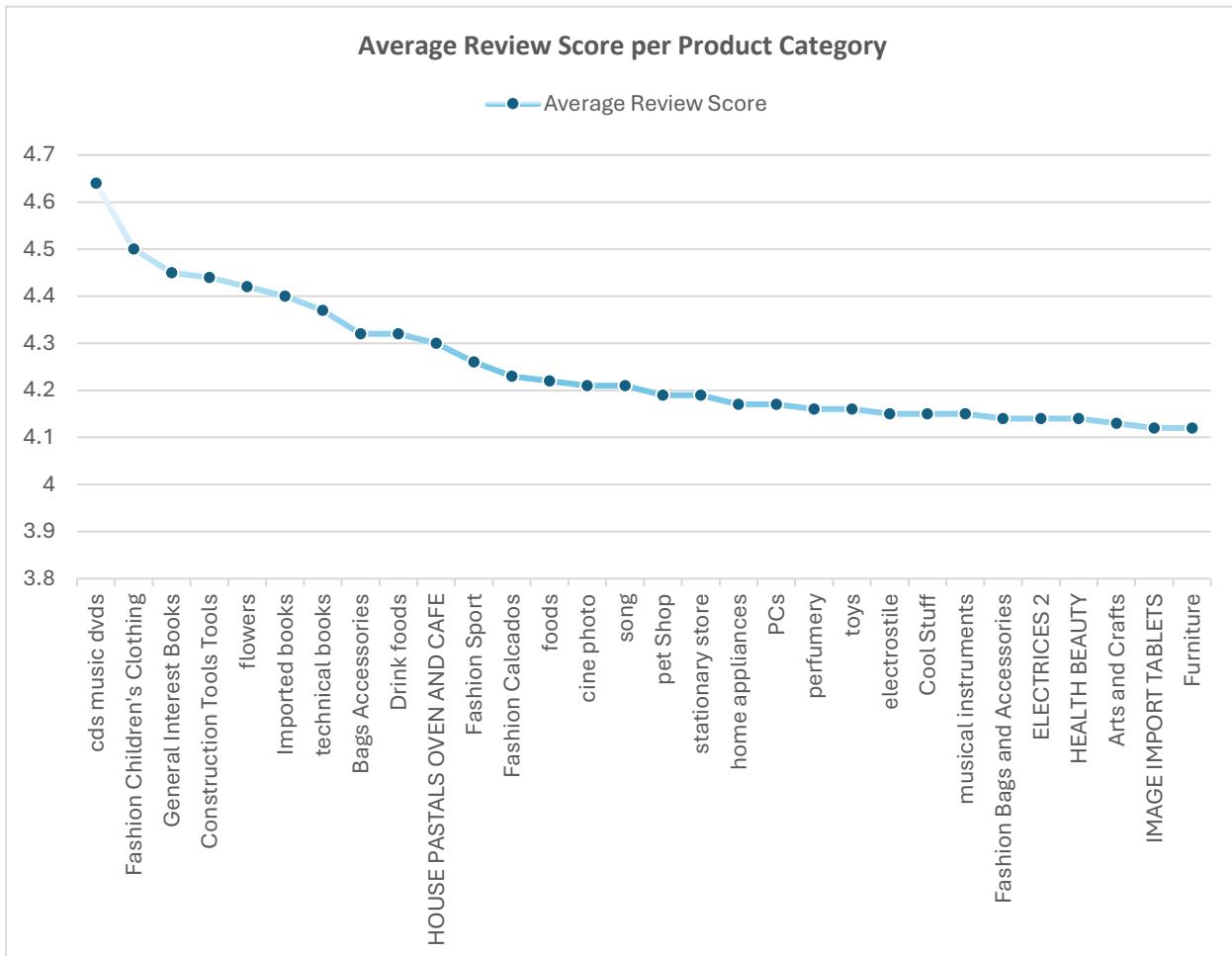
SELECT
    p.`product category` AS product_category,
    AVG(r.review_score) AS average_review_score
FROM
    `rishabh-sql-project.target.order_reviews` r
JOIN
    `rishabh-sql-project.target.orders` o ON r.order_id = o.order_id
JOIN
    `rishabh-sql-project.target.order_items` oi ON o.order_id = oi.order_id
JOIN
    `rishabh-sql-project.target.products` p ON oi.product_id = p.product_id
GROUP BY
    p.`product category`
ORDER BY
    average_review_score DESC;

```

OUTPUT: (First 10 Rows)

Row	product_category ▾	average_review_score ▾
1	cds music dvds	4.64
2	Fashion Children's Clothing	4.5
3	General Interest Books	4.45
4	Construction Tools Tools	4.44
5	flowers	4.42
6	Imported books	4.4
7	technical books	4.37
8	Bags Accessories	4.32
9	Drink foods	4.32
10	HOUSE PASTALS OVEN AND CAFE	4.3

VISUALIZATION: (For the Top 30 Product Category)



INSIGHTS:

- ❖ **High-Performing Categories:**
 - **Categories like "cds music dvds" and "Fashion Children's Clothing" top the list with scores near 4.5, indicating strong customer satisfaction and potential areas for highlighting in marketing efforts.**
 - **"General Interest Books" and "Construction Tools Tools" also perform well, showing both high quality and possibly a strong fit between product expectations and delivery.**
- ❖ **Categories with Room for Enhancement:**
 - **"Hygiene diapers" and "insurance and services" have the lowest scores, which suggests customer expectations are not being met effectively in these categories.**
 - **Categories related to home comfort and personal electronics like "PC Gamer" and "Furniture office" also show lower satisfaction, indicating potential issues with product quality or customer service.**
- ❖ **Opportunities for Business Strategy Adjustments:**
 - **There is a gradual decline in review scores as we move from entertainment and essential categories to more niche segments such as "Fashion Men's Clothing" and specialized equipment.**
 - **Improving customer feedback loops in lower-scoring categories could help pinpoint issues, leading to better product offerings or enhanced customer support.**

7.2) How has the average product price trended over the years?

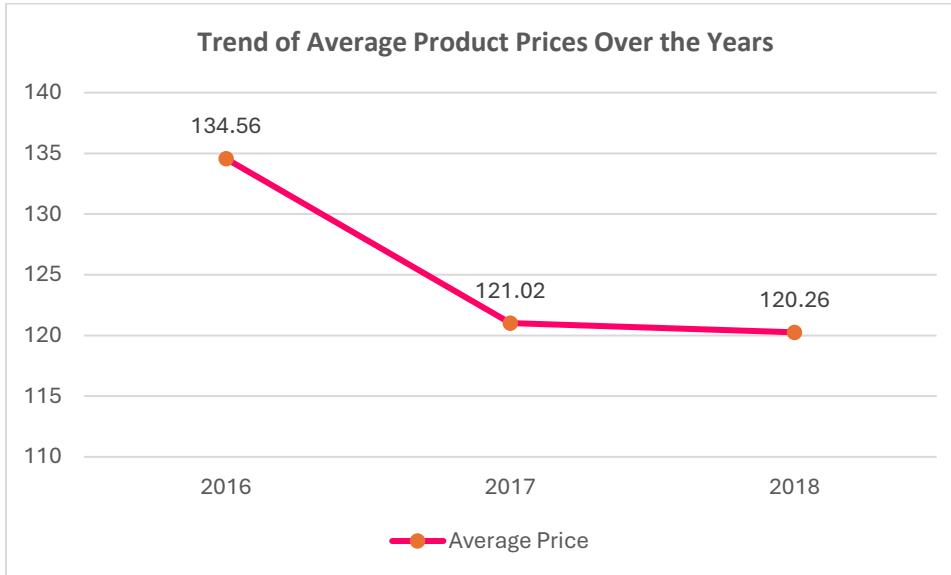
QUERY:

```
SELECT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS year,
    AVG(i.price) AS average_price
FROM
    `rishabh-sql-project.target.orders` o
JOIN
    `rishabh-sql-project.target.order_items` i ON o.order_id = i.order_id
GROUP BY
    year
ORDER BY
    year;
```

OUTPUT:

Row	year	average_price
1	2016	134.55654054054057
2	2017	121.0248305284674
3	2018	120.26264816985862

VISUALIZATION:



INSIGHTS:

- ❖ **Declining Price Trend:**
 - The average product prices showed a downward trend from 2016 through 2018, decreasing each year, which indicates a potential shift towards more affordable product offerings over time.
 - The consistent decline suggests that retailers might be responding to competitive pressures or changes in consumer demand favoring lower-priced items.
- ❖ **Market Adaptation:**
 - The gradual reduction in prices could reflect an adaptation strategy by sellers to attract more customers in a price-sensitive market.
 - This pricing strategy might also be influenced by the introduction of cost-effective manufacturing technologies or bulk purchasing efficiencies gained over the years.
- ❖ **Potential Impact on Sales Volume:**
 - By lowering prices, sellers could be aiming to increase the volume of sales, which is a common tactic in competitive markets to maintain or grow market share.

- Although this could **lead to thinner profit margins per unit**, the **overall revenue might benefit** from the **higher volume of sales, compensating for the lower individual prices**.

7.3) How does product photo quantity relate to sales volume?

QUERY:

```

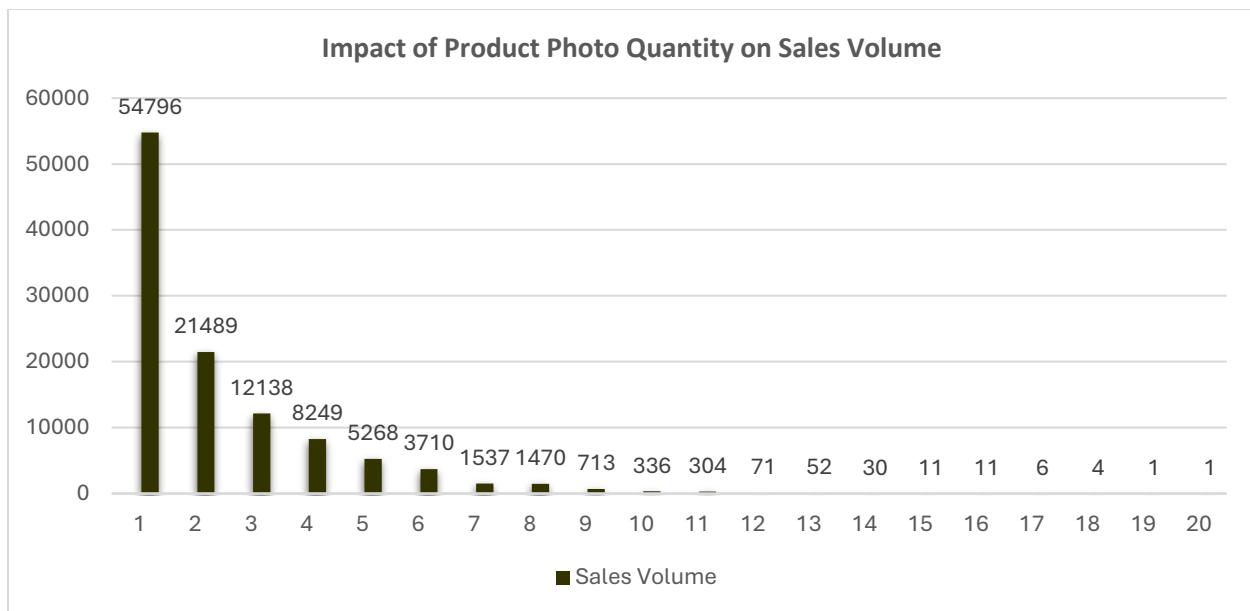
SELECT
    p.product_photos_qty AS photo_quantity,
    COUNT(o.order_id) AS sales_volume
FROM
    `rishabh-sql-project.target.products` p
JOIN
    `rishabh-sql-project.target.order_items` oi ON p.product_id =
oi.product_id
JOIN
    `rishabh-sql-project.target.orders` o ON oi.order_id = o.order_id
WHERE
    o.order_status = 'delivered'
GROUP BY
    p.product_photos_qty
ORDER BY
    sales_volume DESC;

```

OUTPUT: (First 10 Rows)

Row	photo_quantity	sales_volume
1	1	54796
2	2	21489
3	3	12138
4	4	8249
5	5	5268
6	6	3710
7	null	1537
8	7	1470
9	8	713
10	10	336

VISUALIZATION:



INSIGHTS:

- ❖ **Photo Quantity and Sales Correlation:**
 - Products with a single photo represent the highest sales volume, suggesting simplicity or established product recognition may influence buyer decisions.
 - As photo quantity increases, sales volume shows a sharp decline after the first photo, indicating diminishing returns on additional images.
- ❖ **Optimal Photo Quantity:**
 - Sales volumes are considerably high for products with up to 4 photos, supporting the use of multiple images to provide adequate product views and details.
 - Beyond the four photos, sales volume decreases significantly, which may reflect niche products or over-saturation of visual information.
- ❖ **Market Strategy Considerations:**
 - The trend suggests that having a few high-quality images may be more effective than a larger quantity of photos, focusing on quality over quantity.
 - Products with no photos (null) still maintain some sales, potentially indicating purchases based on brand loyalty or product specifications alone.

7.4) Analysis of payment type preferences across different states.

QUERY:

```
WITH StatePaymentPreferences AS (
    SELECT
```

```

        c.customer_state AS state,
        p.payment_type,
        COUNT(*) AS num_payments
    FROM
        `rishabh-sql-project.target.payments` p
    JOIN
        `rishabh-sql-project.target.orders` o ON p.order_id = o.order_id
    JOIN
        `rishabh-sql-project.target.customers` c ON o.customer_id = c.customer_id
    WHERE
        o.order_status = 'delivered'
    GROUP BY
        c.customer_state, p.payment_type
    )

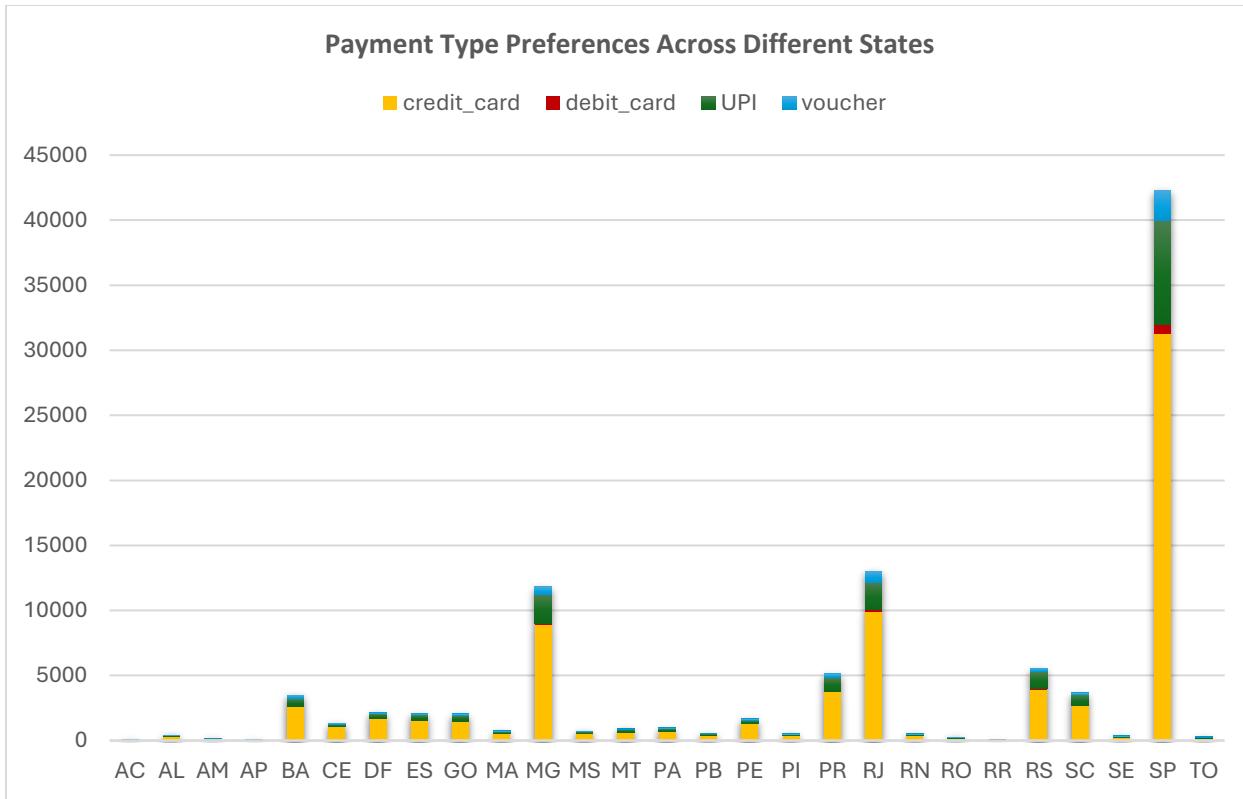
SELECT
    state,
    payment_type,
    num_payments
FROM
    StatePaymentPreferences
ORDER BY
    state, num_payments DESC;

```

OUTPUT: (First 10 Rows)

Row	state ▾	payment_type ▾	num_payments ▾
1	AC	credit_card	60
2	AC	UPI	16
3	AC	voucher	5
4	AC	debit_card	2
5	AL	credit_card	331
6	AL	UPI	64
7	AL	voucher	13
8	AL	debit_card	3
9	AM	credit_card	121
10	AM	UPI	21

VISUALIZATION:



INSIGHTS:

- ❖ **Credit Card Dominance:**
 - Credit cards remain the most popular payment method, with significant usage across all states, emphasizing their dominance in the payment landscape.
 - States like SP, MG, and RJ show the highest numbers, reflecting a strong preference for credit cards in economically active regions.
- ❖ **UPI Adoption:**
 - UPI shows strong growth in states like SP, MG, and RJ, indicating its acceptance in urban and technologically advanced areas.
 - The widespread adoption of UPI can be attributed to the convenience and security it offers, appealing to a tech-savvy consumer base.
- ❖ **Alternative Payment Methods:**
 - Vouchers and debit cards are less common but have notable usage in specific states like SP and BA, suggesting targeted promotions or consumer preferences in specialized segments.
 - The relatively lower numbers for vouchers and debit cards highlight their specific roles in the payment ecosystem, possibly used more for certain types of transactions or in promotional strategies.

7.5) Which seller states have the most delayed shipments?

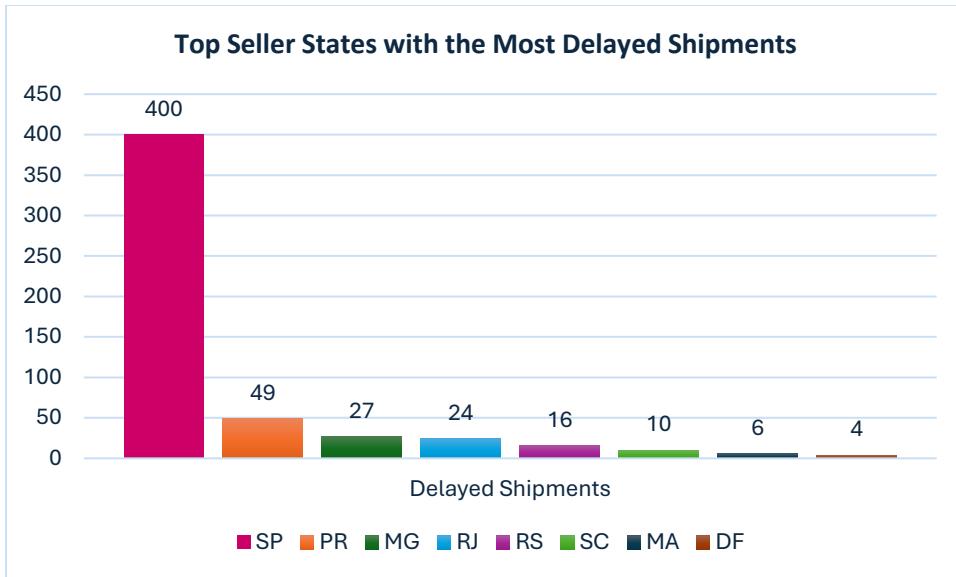
QUERY:

```
SELECT
    s.seller_state AS state,
    COUNT(*) AS delayed_shipments
FROM
    `rishabh-sql-project.target.orders` o
JOIN
    `rishabh-sql-project.target.order_items` oi ON o.order_id = oi.order_id
JOIN
    `rishabh-sql-project.target.sellers` s ON oi.seller_id = s.seller_id
WHERE
    o.order_status = 'delivered' AND
    o.order_delivered_carrier_date > o.order_estimated_delivery_date
GROUP BY
    s.seller_state
ORDER BY
    delayed_shipments DESC;
```

OUTPUT:

Row	state ▾	delayed_shipments ▾
1	SP	400
2	PR	49
3	MG	27
4	RJ	24
5	RS	16
6	SC	10
7	MA	6
8	DF	4

VISUALIZATION:



INSIGHTS:

- ❖ **Concentration in SP:**
 - SP shows a significantly higher number of delayed shipments compared to other states, indicating potential issues with logistics or a higher volume of orders.
 - This concentration suggests that businesses in SP may need to optimize their delivery processes or adjust expectations with customers.
- ❖ **Comparative Analysis:**
 - Other states like PR, MG, and RJ also experience delayed shipments, but their numbers are considerably lower than SP, showing a less severe impact.
 - These states could represent more efficient handling of deliveries or a lower overall sales volume, which needs further exploration.
- ❖ **Broader Implications:**
 - The disparity in delayed shipments among states could influence seller ratings and customer satisfaction across regions.
 - Understanding regional discrepancies can help e-commerce platforms and sellers improve their operational strategies and customer service interventions.

8. Actionable Insights & Recommendations.

8.1) Actionable Insights:

1. Customer Behavior Analysis:

- Analyze customer data to identify **frequent buyers** and develop **loyalty programs** tailored to their **purchase patterns**.

2. Market Expansion Opportunities:

- Use **geographic data** to target **under-served areas** for expansion, focusing on **cities with high customer engagement but low market penetration**.

3. Inventory and Marketing Strategy:

- Adapt **marketing and inventory strategies** based on **seasonal buying trends** to optimize **resource allocation and sales**.

4. Localized Marketing Campaigns:

- Implement **localized marketing strategies** in **densely populated customer zones** to enhance **direct engagement** and **increase sales**.

5. Marketing Effectiveness:

- Leverage **data on high order volumes** during specific periods to plan **effective promotional campaigns** that coincide with these peaks.

6. Strategic Market Focus:

- Concentrate resources and marketing efforts in **states like SP**, where **market penetration is already strong**, to further **solidify the brand's presence**.

7. Promotional Timing:

- Optimize the **timing of marketing promotions** to match **peak ordering times**, enhancing the chances of **customer purchases**.

8. Payment Method Preferences:

- Promote the **use of credit cards through incentives**, considering their **dominance as a preferred payment method** among consumers.

9. Product Presentation Optimization:

- Streamline product photo strategies by focusing on **quality over quantity**, ensuring **photos effectively showcase product benefits**.

10. Logistical Enhancements:

- Address **delivery efficiency disparities across states** by upgrading **logistics infrastructure** to **reduce delivery times** and **improve customer satisfaction**.

8.2) Recommendations:

1. Enhanced Customer Relationship Management:

- Upgrade CRM systems to provide deeper insights into customer behaviors and preferences, aiding in more personalized marketing and services.

2. Targeted Expansion Initiatives:

- Target new regions for business expansion, focusing on high-growth potential areas identified through geographic data analysis.

3. Seasonal Inventory Adjustments:

- Adjust inventory levels dynamically according to seasonal trends to ensure optimal stock availability and minimize excess.

4. City-Specific Campaigns:

- Launch city-specific marketing campaigns to engage with customers on a more personal level, increasing brand loyalty and recognition.

5. Strategic Promotional Activities:

- Schedule promotions during historically high sales periods to maximize revenue and capitalize on consumer buying habits.

6. Reinforce Existing Markets:

- Invest in marketing and community engagement in states with high sales volumes to maintain and grow the customer base.

7. Optimal Timing for Promotions:

- Align promotional activities with identified peak ordering times to capture the maximum audience.

8. Credit Card Usage Incentives:

- Create incentives for using credit cards, such as discounts or reward points, to encourage higher spending.

9. Image Quality Focus in Listings:

- Enhance product listings with high-quality images that highlight key features, improving customer purchase confidence.

10. Delivery Efficiency Improvement:

- Invest in logistical improvements in regions with noted delivery inefficiencies to enhance overall customer satisfaction and trust in the brand.
-

Created By – Rishabh Dev Sahu