Analyzing Market Trends and Customer Behavior at Target using SQL: A Study of 2016-2018 Data

-Otis Castro S J

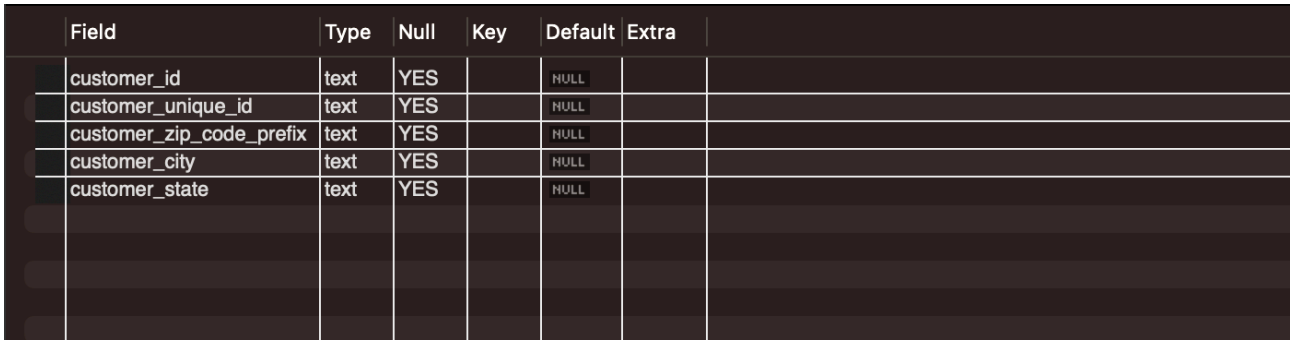
1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

1. Data type of all columns in the "customers" table.

Query:

describe customers;

Screenshot:



Field	Type	Null	Key	Default	Extra
customer_id	text	YES		NULL	
customer_unique_id	text	YES		NULL	
customer_zip_code_prefix	text	YES		NULL	
customer_city	text	YES		NULL	
customer_state	text	YES		NULL	

Insights:

- The customer table consists of five columns.
- Each column in the customer table is set to the ‘text’ data type.

Actionable Items:

- Evaluate the appropriateness of using the ‘text’ data type for each column in the customer table. Consider the expected data size, performance implications, and indexing requirements.
- Implement robust data validation mechanisms to ensure that data entered into the customer table adheres to expected formats and constraints despite using the ‘text’ data type.

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

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 orders;
```

Screenshot:

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

Insights:

- The earliest order was placed on September 4, 2016.
- The latest order was placed on October 17, 2018.

Actionable Items:

- The "orders" table encompasses data spanning 774 days, ranging from September 4, 2016, to October 17, 2018. This extensive timeframe proves invaluable for thorough data analysis.
- It's essential to note that the data for 2016 spans only around four months, while the data for 2018 covers approximately ten months.

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

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

Query:

```
select count(distinct customer_city) as no_of_cities,
count(distinct customer_state) as no_of_states
from orders as o
left join customers as c
on o.customer_id=c.customer_id;
```

Screenshot:

no_of_cities	no_of_states
4119	27

Insights:

- Purchase orders have been generated in a combined total of 27 states and 4,119 cities.

Actionable Items:

- Consider expanding operations geographically in other regions.
- The wide geographic distribution of purchase orders suggests potential opportunities for targeted marketing campaigns or regional sales strategies.
- By analysing purchasing patterns within each state and city, we can tailor approaches to meet the specific needs and preferences of different regions, potentially leading to increased sales and market penetration.

2. In-depth Exploration:

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

Query:

```
select year(order_purchase_timestamp) as year_, count(*) as no_of_orders
from orders
group by year_
order by year_;
```

Screenshot:

year_	no_of_orders
2016	329
2017	45101
2018	54011

Insights:

- There is a noticeable upward trend in the number of orders over the given period, with significant increases from 2016 to 2017 and from 2017 to 2018. This indicates a positive growth trajectory for the business.
- The substantial increase in orders from 2016 to 2018 suggests an expanding market and/or increased customer demand.
- It's essential to note that the data for 2016 spans only around four months, while the data for 2018 covers approximately ten months.

Actionable Items:

- Using historical order data, develop forecasting models to project future demand accurately.
- This enables proactive planning for inventory management, production scheduling, and resource allocation to meet customer needs efficiently.

2. In-depth Exploration:

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

Query:

```
select month(order_purchase_timestamp) as month_no,
monthname(order_purchase_timestamp) as month_,
year(order_purchase_timestamp) as year_, count(*) as no_of_orders
from orders
group by month_no, month_, year_
order by no_of_orders desc;
```

Screenshot:

month_no	month_	year_	no_of_orders
11	November	2017	7544
1	January	2018	7269
3	March	2018	7211
4	April	2018	6939
5	May	2018	6873
2	February	2018	6728
8	August	2018	6512
7	July	2018	6292
6	June	2018	6167
12	December	2017	5673

Insights:

- In 2018, there has been a strong performance in terms of the number of orders, with almost every month showing robust figures.
- The period from November 2017 to August 2018 represents the peak phase of the entire duration, characterized by consistently high order volumes exceeding 5,500 orders each month.
- September and October of both 2016 and 2018 stand out as particularly weak months in terms of order numbers.

Actionable Items:

- Perform an in-depth analysis to identify the primary factors driving the trend of high order volumes from November 2017 to August 2018.
- Leverage the insights gained from the peak phase (November 2017 to August 2018) to optimize inventory management, production schedules, and marketing strategies during high-demand periods.
- Investigate the reasons behind the lower order numbers in certain month. Implement targeted marketing campaigns, promotions, or incentives during these months to stimulate sales and counteract seasonal downturns.

2. In-depth Exploration:

3. During what time of the day, do the Brazilian customers mostly place their orders?

Query:

```
with demo as(
select hour(order_purchase_timestamp) as hrs from orders)
select
case
when hrs between 0 and 6 then "Dawn"
when hrs between 7 and 12 then "Mornings"
when hrs between 13 and 18 then "Afternoon"
when hrs between 19 and 23 then "Night"
end as bins,
count(*) as no_of_orders
from demo
group by bins
order by no_of_orders desc;
```

Screenshot:

bins	no_of_orders
Afternoon	38135
Night	28331
Mornings	27733
Dawn	5242

Insights:

- The Brazilian customers mostly place their orders during afternoons.
- Nights and mornings also witness significant order placements.
- The proportion of orders placed during dawn is notably lower in comparison.

Actionable Items:

- Since the afternoon sees the highest number of orders, consider allocating additional resources during this time to ensure efficient order processing, timely customer support, and optimal inventory management.
- Recognising the significant number of orders during the night, focus on enhancing customer engagement during these hours. Explore opportunities for targeted marketing campaigns, promotions, or special offers.
- Dawn may have the fewest orders, it presents an opportunity for growth and expansion.

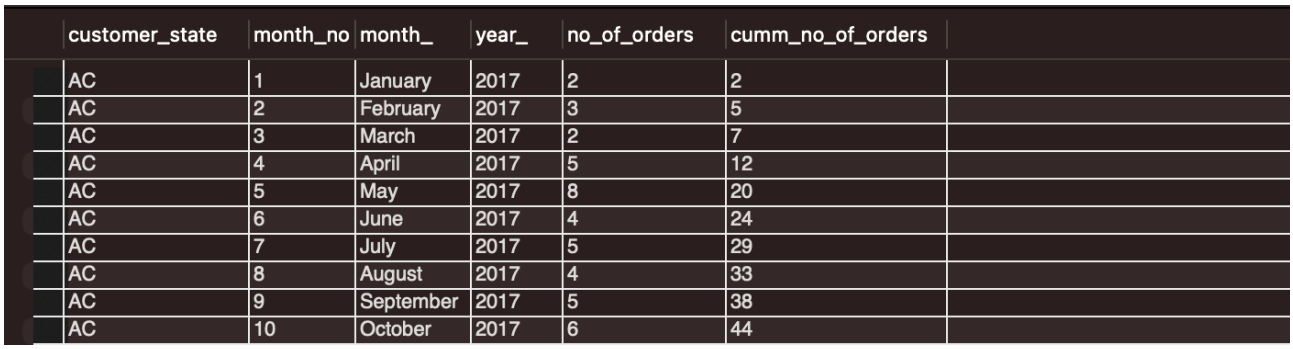
3. Evolution of E-commerce orders in the Brazil region:

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

Query:

```
with demo as(
select c.customer_state, month(o.order_purchase_timestamp) as month_no,
monthname(o.order_purchase_timestamp) as month_,
year(o.order_purchase_timestamp) as year_,
count(*) as no_of_orders
from customers as c
join orders as o
on c.customer_id=o.customer_id
group by customer_state, month_no, month_, year_
order by customer_state, year_, month_no)
select customer_state, month_no, month_, year_, no_of_orders,
sum(no_of_orders) over(partition by customer_state
order by customer_state, year_, month_no
rows between unbounded preceding and current row) as cumm_no_of_orders
from demo;
```

Screenshot:



customer_state	month_no	month_	year_	no_of_orders	cumm_no_of_orders
AC	1	January	2017	2	2
AC	2	February	2017	3	5
AC	3	March	2017	2	7
AC	4	April	2017	5	12
AC	5	May	2017	8	20
AC	6	June	2017	4	24
AC	7	July	2017	5	29
AC	8	August	2017	4	33
AC	9	September	2017	5	38
AC	10	October	2017	6	44

Insights:

- The month on month orders placed in various states exhibit a diverse range, with certain states demonstrating notably stronger performance compared to others.

Actionable Items:

- Analyze the demographics, preferences, and buying behavior of customers in states with stronger order performance. Use this information to segment the market and tailor marketing strategies to target specific customer segments more effectively.
- Form partnerships with local businesses or influencers in states with weaker order performance to increase brand visibility and reach.
- Review the product assortment to ensure it aligns with the preferences and demands of customers in states with stronger order performance.
- Conduct a competitive analysis to understand the strategies and tactics employed by competitors in states with weaker order performance.

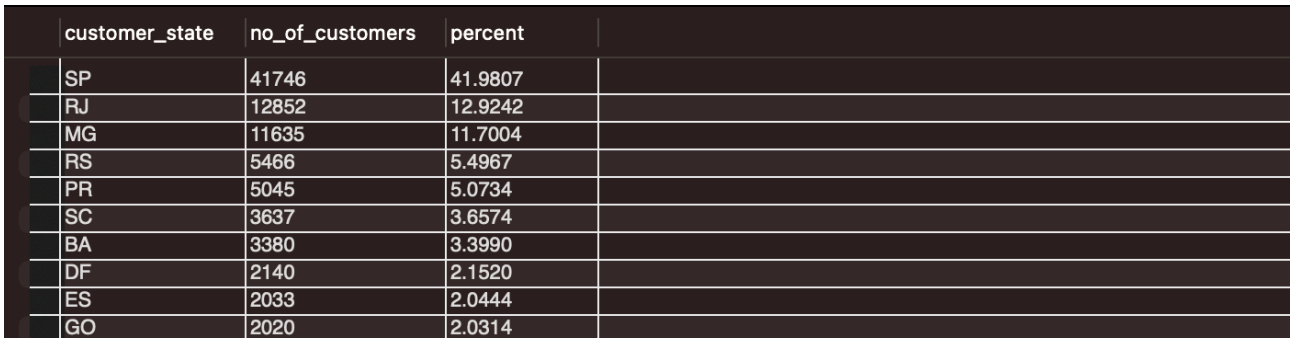
3. Evolution of E-commerce orders in the Brazil region:

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

Query:

```
select customer_state, count(*) as no_of_customers, count(*)/
(
select count(customer_id) from customers
) * 100 as percent
from customers
group by customer_state
order by no_of_customers desc;
```

Screenshot:



customer_state	no_of_customers	percent
SP	41746	41.9807
RJ	12852	12.9242
MG	11635	11.7004
RS	5466	5.4967
PR	5045	5.0734
SC	3637	3.6574
BA	3380	3.3990
DF	2140	2.1520
ES	2033	2.0444
GO	2020	2.0314

Insights:

- SP has the highest proportion of customers, accounting for about 42% of the total customer base.
- SP, RJ, and MG collectively represent roughly two-thirds of the total customer base.
- 15 states individually contribute less than 1% each to the total customer base.

Actionable Items:

- Implement targeted strategies to further engage and retain customers in SP, RJ, and MG, leveraging their significant market share.
- Foster a sense of community and brand loyalty among customers in the top three states.
- The significant gap in customer base between the top three states and the bottom 15 states presents an opportunity for growth in the latter.
- Forge strategic partnerships with local businesses or influencers in the bottom 15 states to increase brand visibility and reach.
- Actively engage with customers in the bottom 15 states to gather feedback and insights on their needs and preferences.

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).

Query:

```
with demo2 as(
with demo as(
select year(o.order_purchase_timestamp) as year_, sum(payment_value) as total
from payments as p
join orders as o
on p.order_id=o.order_id
where year(o.order_purchase_timestamp) between 2017 and 2018
and month(o.order_purchase_timestamp) between 1 and 8
group by year_
order by year_ desc)
select year_, total, lead (total) over() as lead_ from demo)
select round((total - lead_)  /  lead_  *  100,2) as percent_increase from demo2
where lead_ is not null;
```

Screenshot:

Insights:

- The cost of orders from January to August in 2018 increased by approximately 137% compared to the same period in 2017.
- The increased payment value indicates higher consumption levels, which directly impact inventory turnover rates, production schedules, and distribution requirements.

Actionable Items:

- Conduct a thorough analysis of your supply chain processes to ensure they can efficiently support the increased consumption levels.
- Analyze the product mix sold during the January to August periods in 2017 and 2018 to identify which products or categories experienced the highest growth.
- Develop and implement targeted cross-selling and upselling strategies based on customer purchase history and preferences.

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

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

Query:

```
select c.customer_state, round(sum(p.payment_value),2) as total_payment,
round(avg(p.payment_value),2) as avg_payment
from customers as c
join orders as o
on c.customer_id=o.customer_id
join payments as p
on o.order_id=p.order_id
group by c.customer_state
order by avg_payment;
```

Screenshot:

customer_state	total_payment	avg_payment
PB	141545.72	248.33
AC	19680.62	234.29
RO	60866.2	233.2
AP	16262.8	232.33
AL	96962.06	227.08
RR	10064.62	218.8
PA	218295.85	215.92
SE	75246.25	208.44
PI	108523.97	207.11
TO	61485.33	204.27

Insights:

- PB, AC, and RO rank as the top three states based on average payment.
- Ten states exhibit an average payment exceeding 200 units.

Actionable Items:

- Consider introducing high-value or luxury items for customers from the top 5 states with highest average payment, catering to their spending capacity and lifestyle preferences.
- Conduct pricing analyses to identify opportunities for price adjustments or premium pricing models on states with higher average payment
- Identify cross-selling and upselling opportunities to maximise revenue from customers in the top three states.
- Review the product assortment to ensure it meets the needs and preferences of customers in states with low average payment.

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

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

Query:

```
select c.customer_state, round(sum(oi.freight_value),2) as total_freight_value,
round(avg(oi.freight_value),2) as avg_freight_value
from orders as o
join order_items as oi
on o.order_id=oi.order_id
join customers as c
on c.customer_id=o.customer_id
group by c.customer_state
order by avg_freight_value;
```

Screenshot:

customer_state	total_freight_value	avg_freight_value
SP	718723.07	15.15
PR	117851.68	20.53
MG	270853.46	20.63
RJ	305589.31	20.96
DF	50625.5	21.04
SC	89660.26	21.47
RS	135522.74	21.74
ES	49764.6	22.06
GO	53114.98	22.77
MS	19144.03	23.37

Insights:

- SP, PR, and MG offer the lowest average freight value.
- RR, PB, and RO has the highest average freight value.

Actionable Items:

- Initiate negotiations with shipping carriers to secure preferential rates for transporting goods to and from states with low average freight value. Leverage the volume of shipments to negotiate discounted freight rates and reduce overall transportation costs.
- Implement shipment consolidation strategies to reduce the number of individual shipments and maximise freight efficiency.
- Identify inefficiencies in the supply chain and transportation processes that contribute to higher freight costs.
- Conduct route analyses to identify inefficiencies in states with high average freight value.

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 order_id,
datediff(order_delivered_customer_date, order_purchase_timestamp) as time_to_deliver,
datediff(order_delivered_customer_date, order_estimated_delivery_date) as
diff_in_estimated_delivery
from orders;
```

Screenshot:

order_id	time_to_deliver	diff_in_estimated_delivery
e481f51cbdc54678b7cc49136f2d6af7	8	-8
53cdb2fc8bc7dce0b6741e2150273451	14	-6
47770eb9100c2d0c44946d9cf07ec65d	9	-18
949d5b44dbf5de918fe9c16f97b45f8a	14	-13
ad21c59c0840e6cb83a9ceb5573f8159	3	-10
a4591c265e18cb1dcee52889e2d8acc3	17	-6
136cce7faa42fdb2cefd53fdc79a6098	NULL	NULL
6514b8ad8028c9f2cc2374ded245783f	10	-12
76c6e866289321a7c93b82b54852dc33	10	-32
e69bfb5eb88e0ed6a785585b27e16dbf	18	-7

Insights:

- A general trend of delivering products before their estimated date of delivery is observed.

Actionable Items:

- Continuously monitor delivery performance metrics such as on-time delivery rates, order fulfillment times, and customer satisfaction scores.
- Identify areas for improvement and implement corrective actions to maintain a consistent trend of delivering products before their estimated date of delivery.
- Conduct a thorough investigation to identify the root causes of delays in order delivery.
- Keep customers informed about any delays in order delivery through proactive communication channels.

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

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

Query:

```
select c.customer_state, round(avg(freight_value),2) as avg_freight_value
from orders as o
join order_items as oi
on o.order_id=oi.order_id
join customers as c
on c.customer_id=o.customer_id
group by c.customer_state
order by avg_freight_value desc
limit 5;
```

```
select c.customer_state, round(avg(freight_value),2) as avg_freight_value
from orders as o
join order_items as oi
on o.order_id=oi.order_id
join customers as c
on c.customer_id=o.customer_id
group by c.customer_state
order by avg_freight_value
limit 5;
```

Screenshot:

customer_state	avg_freight_value
RR	42.98
PB	42.72
RO	41.07
AC	40.07
PI	39.15

customer_state	avg_freight_value
SP	15.15
PR	20.53
MG	20.63
RJ	20.96
DF	21.04

Insights:

- RR, PB, RO, AC, and PI are the top five states with the highest average freight cost.
- SP, PR, MG, RJ, and DF are the top five states with the least average freight cost.

Actionable Items:

- Establish strategic warehouse locations in or near states with high average freight cost to minimize transportation costs and streamline distribution operations.
- Consider providing value-added delivery services like expedited shipping in states with high average freight cost.
- Explore opportunities for collaborative shipping arrangements with other businesses or suppliers in these states to share transportation costs and reduce shipping expenses.
- Negotiate bulk shipping contracts with carriers and logistics providers to secure discounted rates for shipping goods.
- Investigate alternative transportation modes such as rail, waterway, or intermodal shipping for transporting goods to and from states with high average freight cost.

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

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

Query:

```
select c.customer_state,
round(avg(datediff(order_delivered_customer_date, order_purchase_timestamp)),2) as
avg_delivery_time
from orders as o
join order_items as oi
on o.order_id=oi.order_id
join customers as c
on c.customer_id=o.customer_id
group by c.customer_state
order by avg_delivery_time desc
limit 5;
```

```
select c.customer_state,
round(avg(datediff(order_delivered_customer_date, order_purchase_timestamp)),1) as
avg_delivery_time
from orders as o
join order_items as oi
on o.order_id=oi.order_id
join customers as c
on c.customer_id=o.customer_id
group by c.customer_state
order by avg_delivery_time
limit 5;
```

Screenshot:

customer_state	avg_delivery_time
AP	28.22
RR	28.17
AM	26.34
AL	24.45
PA	23.70

customer_state	avg_delivery_time
SP	8.7
PR	11.9
MG	11.9
DF	12.9
SC	15.0

Insights:

- AP, RR, AM, AL, and PA are the top five states with the highest average delivery time.
- SP, PR, MG, DF, and SC are the top five states with the least average delivery time.

Actionable Items:

- Invest in last-mile delivery capabilities, including fleet optimization, driver training, and vehicle maintenance, to expedite delivery times in states with high average delivery time.
- Explore opportunities for shared warehousing facilities, cross-docking arrangements, or joint delivery routes to streamline operations and improve delivery efficiency.
- Invest in transportation infrastructure, logistics facilities, and technology upgrades to support faster and more efficient delivery operations in these states.
- Implement automation technologies and digital solutions to streamline order processing, routing, and delivery scheduling.

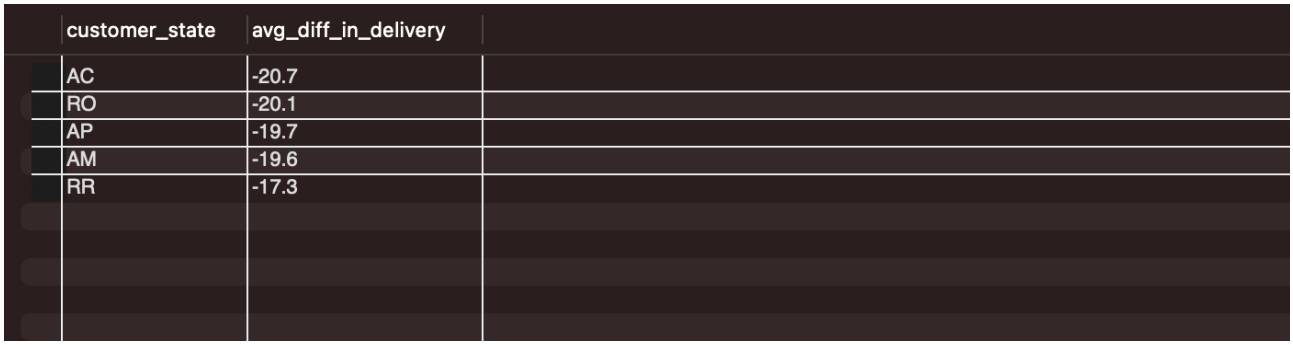
5. Analysis based on sales, freight and delivery time.

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

Query:

```
select c.customer_state,
round(avg(datediff(o.order_delivered_customer_date,o.order_estimated_delivery_date)),
1) as avg_diff_in_delivery
from orders as o
join customers as c
on o.customer_id=c.customer_id
group by c.customer_state
order by avg_diff_in_delivery
limit 5;
```

Screenshot:



customer_state	avg_diff_in_delivery
AC	-20.7
RO	-20.1
AP	-19.7
AM	-19.6
RR	-17.3

Insights:

- AC, RO, AP, AM, and RR the top five states where the order delivery is really fast as compared to the estimated date of delivery.

Actionable Items:

- Conduct a detailed analysis of the order delivery process in the top five states to identify key factors contributing to their fast delivery speeds.
- Strategically position warehouses and distribution centers in key locations to reduce transit times and improve delivery speed.
- Conduct a thorough review of order processing workflows and logistics operations to identify bottlenecks and inefficiencies in other states.
- Implement advanced logistics management systems and order tracking solutions to enhance visibility and control over the delivery process.

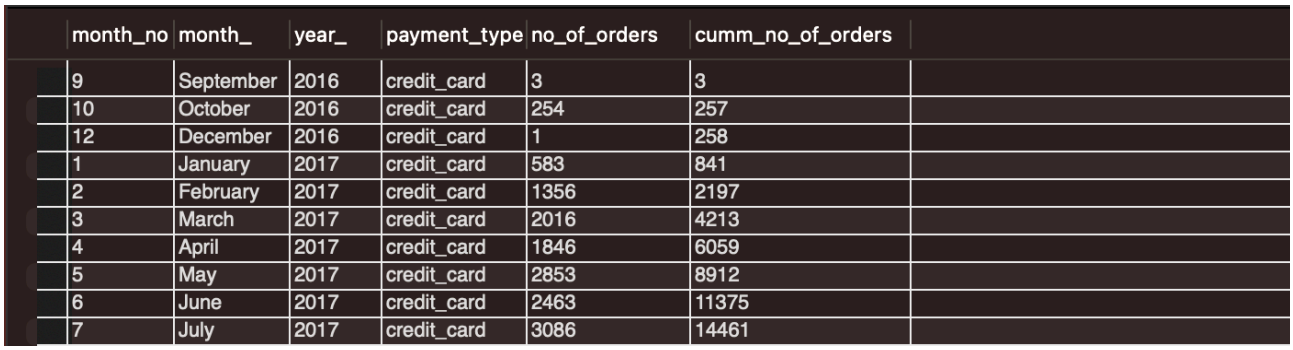
6. Analysis based on the payments:

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

Query:

```
with demo as(
select month(o.order_purchase_timestamp) as month_no,
monthname(o.order_purchase_timestamp) as month_,
year(o.order_purchase_timestamp) as year_, p.payment_type,
count(*) as no_of_orders
from orders as o
join payments as p
on o.order_id=p.order_id
group by year_, month_no, month_, payment_type
order by payment_type, year_, month_no)
select month_no, month_, year_,p ayment_type, no_of_orders,
sum(no_of_orders) over (partition by payment_type
order by payment_type, year_, month_no
rows between unbounded preceding and current row) as cumm_no_of_orders
from demo;
```

Screenshot:



month_no	month_	year_	payment_type	no_of_orders	cumm_no_of_orders
9	September	2016	credit_card	3	3
10	October	2016	credit_card	254	257
12	December	2016	credit_card	1	258
1	January	2017	credit_card	583	841
2	February	2017	credit_card	1356	2197
3	March	2017	credit_card	2016	4213
4	April	2017	credit_card	1846	6059
5	May	2017	credit_card	2853	8912
6	June	2017	credit_card	2463	11375
7	July	2017	credit_card	3086	14461

Insights:

- The number of payments made using credit cards notably exceeds those made through all other payment methods.

Actionable Items:

- Encourage customers to use alternative payment methods by offering incentives or discounts for non-credit card payments to diversify payment channels and reduce reliance on credit cards.
- Offer incentives or rewards for customers who choose to pay with cash or alternative non-credit card methods.
- Provide clear and informative guidance to customers on the various payment methods available and how to use them effectively.
- Streamline credit card processing systems to ensure efficiency and reliability.

6. Analysis based on the payments:

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

Query:

```
select count(*) as no_of_orders from payments
where payment_installments>0;
```

Screenshot:

Insights:

- The total number of orders with at least one installment is 103,884.

Actionable Items:

- Explore additional payment methods or financing options to cater to diverse customer needs and preferences. Consider partnering with third-party payment providers or financial institutions to offer alternative financing solutions, such as buy now, pay later programs or installment loans.
- Provide clear and transparent information about installment payment options, including terms and conditions, fees, and repayment schedules.