## **Large Retail Store-Case Study**

- 1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset
  - 1.1. Data type of all columns in the "customers" table

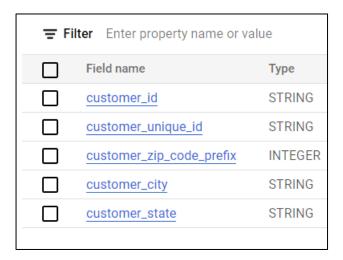
```
customer_id = STRING

customer_unique_id = STRING

customer_zip_code_prefix = INTEGER

customer_city = STRING

customer_state = STRING
```



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

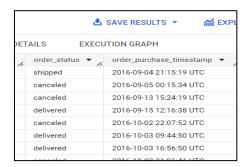
Orders were placed between - 04/09/2016 and 17/10/2018

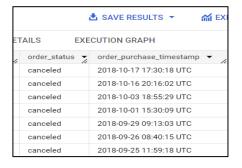
Query-

select\*

from

`large\_retail\_store case-study-390413. large\_retail\_store \_CaseStudy.orders`
order by order\_purchase\_timestamp





1.3. Count the number of Cities and States in our dataset

**Cities**- 4119

States-27

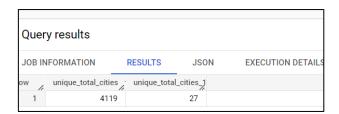
#### Query-

**SELECT** 

COUNT(DISTINCT customer\_city) AS unique\_total\_cities,

COUNT(DISTINCT customer\_state) AS unique\_total\_cities

FROM ` large\_retail\_store -case-study-390413. large\_retail\_store
\_CaseStudy.customers`



## 2. In-depth Exploration:

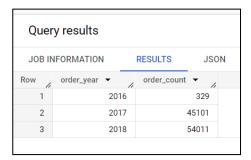
2.1. Is there a growing trend in the no. of orders placed over the past years?

Ans- Yes, the orders have increased over the years.

#### Query-

**SELECT** 

DISTINCT EXTRACT(YEAR FROM (order\_purchase\_timestamp)) as order\_year, count(order\_id)over(partition by EXTRACT(YEAR FROM (order\_purchase\_timestamp))) as order\_count FROM ` large\_retail\_store -case-study-390413. large\_retail\_store \_CaseStudy.orders`
ORDER BY order\_year



## **Insights**:

**a.** Increasing Customer Demand: The rising number of orders over the past years indicates a growing demand for the company's products or services. This can be a

positive sign, as it suggests that customers are increasingly interested in purchasing from the company.

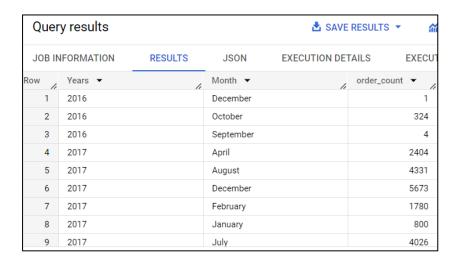
- **b.** Market Expansion Opportunities: The upward trend in order volume signifies potential opportunities for the company to expand its market reach. It indicates that the company's products or services are resonating with customers and can potentially attract a larger customer base.
- **c.** Customer Satisfaction: A consistent increase in orders suggests that customers are satisfied with their previous purchases and are likely to become repeat customers. It indicates that the company is meeting or exceeding customer expectations, leading to increased loyalty and trust

## **Recommendations:**

- **a.** Capacity Planning: With the growing trend in orders, it is crucial for the company to assess its capacity and ensure it can meet the increasing demand. This may involve evaluating production capabilities, inventory management, and logistics to avoid any potential bottlenecks or delays in order fulfilment.
- **b.** Scaling Operations: To accommodate the rising number of orders, the company may need to consider expanding its operations. This could involve increasing staffing levels, investing in automation or technology solutions, or optimizing processes to improve efficiency and scalability.
- c. Enhanced Customer Support: As order volume increases, providing excellent customer support becomes even more critical. The company should focus on strengthening its customer service team, providing timely and helpful responses to inquiries or concerns, and exploring self-service options to empower customers and reduce support ticket volumes.
- **d.** Marketing and Promotions: To capitalize on the growing trend, the company can leverage marketing strategies to further boost sales. This may involve targeted promotions, loyalty programs, referral incentives, or personalized marketing campaigns to engage existing customers and attract new ones.
- **e. Data-Driven Decision Making**: Utilize the data generated from the increased order volume to gain actionable insights. Analyse customer behaviour, purchase patterns, and preferences to identify potential areas for improvement, product expansion, or targeted marketing efforts.

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

```
Query-
SELECT
FORMAT_DATE("%Y",order_purchase_timestamp) as Years,
FORMAT_DATE("%B",order_purchase_timestamp) as Month,
COUNT(*) AS order_count
FROM
`large_retail_store-case-study-390413.large_retail_store_CaseStudy.orders`
GROUP BY 1,2
ORDER BY 1,2
```





## **Insights**

a. Seasonal Peaks in Order Volume-As evident from the line chart above, during the months of March, July, and August in the years 2017 and 2018, a noticeable

increase in the number of orders placed was observed. Also, a sudden peak in the month of November for year 2017. This suggests that there may be external factors, such as holidays, events, or specific trends, influencing customer demand during these months.

## **Recommendations:**

- **a.** Planning and Resource Allocation: The business can utilize this information to effectively allocate resources and plan operations accordingly. By anticipating higher order volumes during March, July, and August, the company can ensure sufficient inventory levels, optimize logistics and delivery processes, and allocate staff and resources adequately to meet the increased demand.
- **b.** Marketing and Promotions: The identified seasonal peaks can be leveraged for targeted marketing campaigns and promotions. By tailoring marketing efforts to align with customer preferences during these months, the company can maximize its reach, engagement, and potential sales. This may include offering special discounts, launching seasonal product lines.
- c. Forecasting and Financial Planning: The observed seasonal patterns can be incorporated into future sales forecasting and financial planning. By considering the historical trends during March, July, and August, the company can develop more accurate revenue projections and budget allocation strategies. This information can be valuable for setting sales targets, evaluating performance, and making informed business decisions
  - 2.3 During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

Ans- Afternoon

Query-

SELECT

```
CASE
 WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 0 AND EXTRACT(HOUR FROM
 order_purchase_timestamp) <7 THEN 'Dawn'</pre>
 WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 7 AND EXTRACT(HOUR FROM
 order_purchase_timestamp) <13 THEN 'Morning'
 WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 13 AND EXTRACT(HOUR
 FROM order_purchase_timestamp) < 19 THEN 'Afternoon'
 WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 19 AND EXTRACT(HOUR
 FROM order_purchase_timestamp) < 24 THEN 'Night'
 ELSE 'Unknown'
 END AS time_of_day,
COUNT(*) AS order_count
FROM ` large_retail_store-case-study-390413. large_retail_store
_CaseStudy.orders`
WHERE order_purchase_timestamp >= '2016-09-01' AND order_purchase_timestamp
<= '2018-10-31'
GROUP BY time_of_day
```

#### ORDER BY order\_count DESC;

JOB IN	FORMATION	RESULTS	JSON	EXEC
Row	time_of_day ▼	li.	order_count ▼	1
1	Afternoon		381	35
2	Night		28331	
3	Morning		277	33
4	Dawn		52	42

## Insights and Recommendations-

Peak Business Hours is between 13:00hrs to 19:00hrs. Some recommendations:

- a. Optimal Staffing-The company can allocate additional resources, such as staff and inventory, cater to the higher demand during these hours. By having an adequate number of employees on the shop floor or customer service representatives available, the company can provide better assistance to customers, reduce wait times, and enhance the overall customer experience.
- b. Marketing and Promotions -Additionally, the company can consider adding new products (promotional products) that customers may be interested in trying. If customers enjoy these products, they may consider purchasing them during their next visit or recommend them to their family members or friends. This strategy will likely increase sales and allow for the introduction of new brands to the shelves based on demand.
- c. Maximize Sales and Minimize Waste They can sell the stock that is closer to its shelf life at a reduced price, which is also known as a clearance sale, so that they don't have to discard the products once they expire.
- d. **Peak Hour Optimization** Understanding the peak hours can assist the company in optimizing its order fulfilment and delivery processes. By aligning their operations to prioritize the processing and dispatching of orders received during this time, the company can aim for faster delivery times and improve overall customer satisfaction
- 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(YEAR FROM o.order_purchase_timestamp) AS year,
EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
c.customer_state,
COUNT(*) AS order_count
FROM ` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders` o
JOIN ` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.customers` c ON o.customer_id = c.customer_id
WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) BETWEEN 2016 AND 2018
GROUP BY year, month, c.customer_state
ORDER BY year, month, c.customer_state
limit 10
```

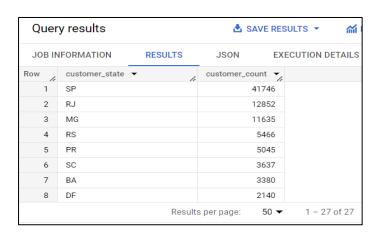


#### 3.2 How are the customers distributed across all the states?

#### Query-

#### **SELECT**

```
customer_state,
COUNT(*) AS customer_count
FROM ` large_retail_store t-case-study-390413.large_retail_store
_CaseStudy.customers`
GROUP BY customer_state
ORDER BY customer_count DESC
```



## Insights:

- a. **Identified High-Demand States**: Over the past 3 years, three states-SP, RJ and MGs in Brazil consistently had the highest number of orders placed. This indicates specific regions with significant customer interest and potential market opportunities.
- b. **Regional Market Understanding**: The company can gain valuable insights into consumer preferences and behaviours by analysing the ordering patterns in these states. This understanding can help tailor marketing strategies and product offerings to effectively target these specific regions.

#### **Actionable Recommendations:**

- a. Regional Targeting: Allocate marketing resources and campaigns to specifically target these high-demand states. This can include localized advertising, promotions, and partnerships with regional influencers or events to maximize brand visibility and customer engagement.
- b. **Inventory Management**: Optimize inventory management to ensure sufficient stock availability in the high-demand states. By monitoring and analysing the specific product preferences and purchasing patterns, the company can adjust its inventory levels accordingly, reducing stockouts and improving customer satisfaction.
- c. Localized Customer Support: Establish dedicated customer support teams or localized contact centers to cater to the high-demand states. This ensures prompt and efficient assistance, addressing customer queries or concerns in a timely manner and enhancing overall customer experience.
- d. Partnerships with Local Suppliers: Explore collaborations with local suppliers or producers in the high-demand states. This can provide access to unique or specialized products, strengthen regional supply chains, and foster a sense of community support, appealing to local customers.
- 4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others

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

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

#### Query-

```
SELECT
```

```
c.customer_state,
ROUND(SUM(oi.price),2) AS total_order_price,
ROUND(AVG(oi.price),2) AS average_order_price
FROM
` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders` o
JOIN
` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.customers` c ON o.customer_id = c.customer_id
JOIN
` large_retail_store -case-study-390413. large_retail_store
CaseStudy.order_items` oi ON o.order_id = oi.order_id
GROUP BY
c.customer_state
ORDER BY total_order_price desc
```

Query results			<b>≛</b> SAVE R	ESULTS ▼
JOB INFORMATION RE		RESULTS	JSON EXECUTION DETAILS	
Row	customer_state ▼	. //	total_order_price 🔻	average_order_price
1	SP		5202955.05	109.65
2	RJ		1824092.67	125.12
3	MG		1585308.03	120.75
4	RS		750304.02	120.34
5	PR		683083.76	119.0
6	90		E20EE2 24	124.65
Results per page: 50 ▼ 1 - 27 of 27				

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

#### Query-

#### SELECT

```
c.customer_state,
ROUND(SUM(oi.freight_value),2) AS total_order_freight,
ROUND(AVG(oi.freight_value),2) AS average_order_freight
FROM
` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders` o
JOIN ` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.customers` c ON o.customer_id = c.customer_id
```

```
JOIN ` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.order_items` oi ON o.order_id = oi.order_id
GROUP BY
c.customer_state
ORDER BY total_order_freight DESC
```

Quer	Query results   ♣ SAVE RESULTS ▼				
JOB IN	IFORMATION	RESULTS	JSON EXE	ECUTION DETAILS	EXECUTION
Row	customer_state	<b>~</b>	total_order_freight	average_order_freigh	
1	SP	, ,	718723.07	15.15	
2	RJ		305589.31	20.96	
3	MG		270853.46	20.63	
4	RS		135522.74	21.74	
5	PR		117851.68	20.53	
6	DΛ		100156 60	26.26	
			Result	s per page: 50 ▼	1 - 27 of 27

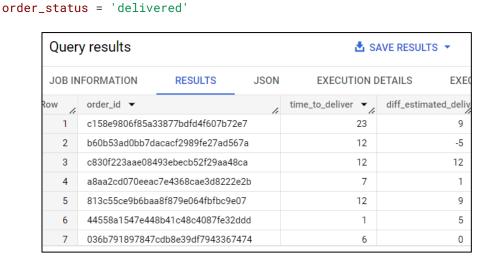
## 5. Analysis based on sales, freight and delivery time.

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. Do this in a single query.

#### Query-

WHERE

```
order_id,
DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,DAY) AS
time_to_deliver,
DATE_DIFF(order_estimated_delivery_date, order_delivered_customer_date,DAY)
AS diff_estimated_delivery
FROM
    large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders`
```



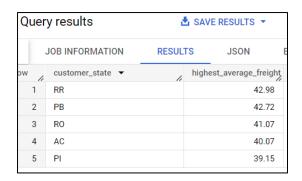
## **Insights**:

- 1. Efficient Delivery Performance: In the past 3 years, orders placed in these 3 states-SP,RJ and MJ of Brazil were consistently delivered before the estimated delivery dates. This indicates a high level of efficiency in the company's delivery processes and logistics operations.
- 2. Customer Satisfaction: Early delivery of orders can lead to increased customer satisfaction and a positive shopping experience. Meeting or surpassing delivery expectations enhances customer loyalty and encourages repeat business.

#### **Actionable Recommendations:**

- 1. Maintain Delivery Excellence: Continue prioritizing efficient delivery operations to uphold the company's reputation for timely order fulfilment. This involves monitoring delivery performance, optimizing logistics routes, and investing in technology and systems that enable accurate tracking and on-time deliveries.
- 2. **Communicate Estimated Delivery Dates**: Ensure clear and transparent communication of estimated delivery dates to customers during the ordering process. This helps manage customer expectations and further enhances satisfaction when orders are delivered ahead of schedule.
- 3. **Collaborate with Delivery Partners**: Maintain strong relationships with delivery partners and carriers to ensure a smooth and efficient delivery process. Regularly evaluate their performance and establish service level agreements to maintain consistent delivery standards.
  - 5.2 Find out the top 5 states with the highest & lowest average freight value

## Query-



Query results						
	JOB INFORMATION	RESUL	ΓS	JSON		E)
w /	customer_state ▼	6	lowe	st_average_f	reight	7
1	SP				15.15	
2	PR				20.53	
3	MG				20.63	
4	RJ				20.96	
5	DF				21.04	

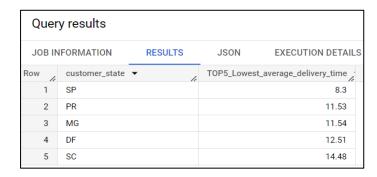
5.3 Find out the top 5 states with the highest & lowest average delivery time

Query for Highest Average Delivery Time-



Query for Lowest Average Delivery Time-

```
SELECT
c.customer_state,
ROUND(AVG(DATE_DIFF(order_delivered_customer_date,
order_purchase_timestamp,DAY)),2) AS TOP5_Lowest_average_delivery_time
FROM
` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders` o
JOIN ` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.customers` c
ON o.customer_id = c.customer_id
GROUP BY
c.customer_state
ORDER BY
TOP5_Lowest_average_delivery_time
LIMIT 5
```



5.4 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(DATE_DIFF(order_delivered_customer_date,order_estimated_delivery_d
ate,DAY))) AS Average_Actual_and_Est_delivery

FROM `large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders` o
JOIN `large_retail_store -case-study-390413. large_retail_store
_CaseStudy.customers` c
ON o.customer_id = c.customer_id

WHERE order_status = "delivered"
GROUP BY 1
ORDER BY 2 ASC
LIMIT 5
```

JOB IN	NFORMATION	RESULTS	JSON EXE
Row	customer_state	<b>~</b>	Average_Actual_and_
1	AC		-20.0
2	AM		-19.0
3	RO		-19.0
4	AP		-19.0
5	RR		-16.0

Note- Negative indicates the orders were delivered way before they were estimated to deliver.

## **Insights**:

a. **Efficient Delivery Performance**: The company demonstrates a consistent ability to deliver orders ahead of the estimated delivery time in the five states of Brazil. This indicates a strong logistics and delivery system in place.

- b. **Customer Satisfaction**: The average delivery time being around 20 days earlier than the estimated delivery time suggests that customers receive their orders sooner than expected. This can lead to increased customer satisfaction, as customers appreciate prompt deliveries.
- c. **Competitive Advantage**: The company gains a competitive edge by providing faster delivery compared to the estimated time. This can help attract and retain customers, as timely delivery is often a crucial factor influencing customer buying decisions.

## **Recommendations:**

- a. **Maintain Delivery Efficiency**: The company should continue focusing on maintaining its efficient delivery performance. Timely deliveries not only enhance customer satisfaction but also contribute to positive word-of-mouth and repeat business.
- b. **Set Realistic Delivery Estimates**: While delivering ahead of the estimated time is commendable, it is essential to ensure that the estimated delivery time is realistic and achievable. Setting accurate estimates helps manage customer expectations and prevents potential dissatisfaction caused by delayed deliveries.
- c. Monitor and Optimize Logistics: Regularly monitor and optimize the logistics and delivery processes to identify any areas for improvement. Streamlining operations, utilizing technology for route optimization, and collaborating with reliable carriers can further enhance the efficiency of the delivery system.

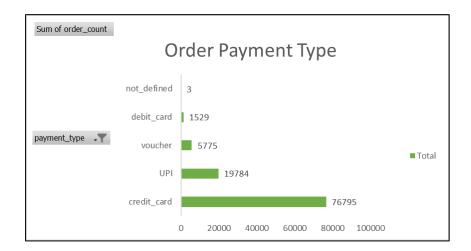
#### 6. Analysis based on the payments

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

#### Query-

```
FORMAT_DATE("%Y-%B",o.order_purchase_timestamp) as Years_Month, p.payment_type,
COUNT(*) AS order_count
FROM
`large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders` o
JOIN `large_retail_store -case-study-390413. large_retail_store
_CaseStudy.payments` p
ON o.order_id = p.order_id
GROUP BY
Years_Month, p.payment_type
```

Query results   ♣ SAVE RESULTS ▼				
JOB IN	FORMATION RESULTS	JSON EXECUTION DE	TAILS EXECUTI	
Row	Years_Month ▼	payment_type ▼	order_count ▼	
1	2016-December	credit_card	1	
2	2016-October	debit_card	2	
3	2016-October	UPI	63	
4	2016-October	credit_card	254	
5	2016-October	voucher	23	
6	2016-September	credit_card	3	
7	2017-April	voucher	202	
8	2017-April	UPI	496	
9	2017-April	credit_card	1846	
10	2017-April	debit_card	27	



## **Insights**:

- a. **Credit Card Dominance**: The analysis reveals that a significant portion of customers prefer paying by credit card when making purchases. This indicates that credit cards are the most widely adopted payment method among the customer base.
- b. **UPI Adoption**: Following credit cards, the second most popular payment method is UPI (Unified Payments Interface). This suggests an increasing acceptance and adoption of digital payment solutions, reflecting the evolving payment landscape and customers' inclination towards convenient and secure transactions.

c. **Payment Diversity**: The presence of multiple payment methods indicates that customers have diverse preferences when it comes to making payments. This highlights the importance of providing a range of payment options to cater to different customer preferences and increase convenience.

## Recommendations:

- a. Strengthen Credit Card Support: Given the popularity of credit cards as the preferred payment method, the e-commerce company should ensure seamless credit card processing, robust security measures, and partnerships with reliable payment gateways. This will enhance the overall payment experience and instill trust among customers.
- b. **Expand Digital Payment Options**: As UPI emerges as a significant payment method, the company should prioritize expanding support for UPI payments. This can be achieved by integrating with UPI-enabled platforms, mobile wallets, or payment aggregators to offer customers a wider range of digital payment options.
- c. Evaluate Additional Payment Methods: Continuously monitor emerging payment trends and customer preferences to identify other popular payment methods in the market. Assess the feasibility and demand for integrating these methods into the ecommerce platform, allowing customers to choose from a broader array of payment options.
  - 6.2 Find the no. of orders placed on the basis of the payment installments that have been paid.

#### Query-

```
SELECT
```

```
p.payment_installments,
COUNT(*) AS order_count
FROM
     large_retail_store -case-study-390413. large_retail_store
_CaseStudy.orders` o
JOIN ` large_retail_store -case-study-390413. large_retail_store
_CaseStudy.payments` p ON o.order_id = p.order_id
GROUP BY
p.payment_installments
```

