



SCALER



DATA SCIENCE & MACHINE LEARNING (DSML)

Topic :- BRAZILIAN BUSINESS TARGET CASE SOLUTIONS

BUSINESS CASE STUDIES :- 1

NAME :- SUMIT UMESH AWASTHI

BATCH :- NOV-2023

SUBJECT :- SQL-MODULE

DATE :- 11-JAN-2024

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

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

QUERY :-

```
SELECT  
COLUMN_NAME,  
DATA_TYPE  
from `brazilian-411410.Brazilian_Target_Business_Case_Solution.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

RECOMMENDATION:-

We used the following SQL query in Big Query to confirm the data types of the columns in the "customers" table. The customer table uses data of the STRING and INTEGER types.

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

QUERY :-

```
SELECT  
MAX(order_purchase_timestamp) AS `LAST_ORDER`,  
MIN(order_purchase_timestamp) AS `FIRST_ORDER`  
FROM `Brazilian_Target_Business_Case_Solution.orders`;
```

OUTPUT :-

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

RECOMMENDATION:-

This allowed us to determine the start and end dates of the data i.e. from 4th September 2016 to 17th October 2018.

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

QUERY:-

```
SELECT DISTINCT
c.customer_city,
c.customer_state,
COUNT(o.customer_id) order_count
FROM Brazilian_Target_Business_Case_Solution.orders o JOIN
Brazilian_Target_Business_Case_Solution.customers c
ON o.customer_id =c.customer_id
GROUP BY 1,2
ORDER BY 3 DESC;
```

OUTPUT :-

Row	customer_city	customer_state	order_count
1	sao paulo	SP	15540
2	rio de janeiro	RJ	6882
3	belo horizonte	MG	2773
4	brasilgia	DF	2131
5	curitiba	PR	1521
6	campinas	SP	1444
7	porto alegre	RS	1379
8	salvador	BA	1245
9	guarulhos	SP	1189
10	sao bernardo do campo	SP	938

RECOMMENDATION:-

As we can see, the city of Sao Paulo in the state of SP alone has more orders than the combined orders of the next five cities. This is due to the fact that São Paulo is Brazil's most well-known and prosperous state.

2. In-depth Exploration :-

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

QUERY:-

```
SELECT EXTRACT(YEAR FROM o.order_purchase_timestamp) AS `YEAR`,  
EXTRACT(MONTH FROM o.order_purchase_timestamp) AS `MONTH`,  
COUNT(DISTINCT o.order_id) AS `ORDER_COUNT`  
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN  
`Brazilian_Target_Business_Case_Solution.customers` c  
ON o.customer_id =c.customer_id  
GROUP BY year,month  
ORDER BY year,month;
```

OUTPUT:-

Row	YEAR	MONTH	ORDER_COUNT
1	2016	9	4
2	2016	10	324
3	2016	12	1
4	2017	1	800
5	2017	2	1780
6	2017	3	2682
7	2017	4	2404
8	2017	5	3700
9	2017	6	3245
10	2017	7	4026

RECOMMENDATION:-

Brazil is seeing an increasing trend in e-commerce, according to the order count data. The number of purchases has generally increased, though there have been notable variations. It's crucial to remember that the order count by itself does not represent the rate of business expansion.

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(DISTINCT order_id) AS `ORDER_COUNT`  
FROM `Brazilian_Target_Business_Case_Solution.orders`  
GROUP BY 1  
ORDER BY 1;
```

OUTPUT :-

Row	MONTH	ORDER_COUNT
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959

RECOMMENDATION:-

we can observe some seasonality in the e-commerce orders. The count of orders generally increases from March to August with fluctuations in between. Notably, there is an increase in orders during February and March, coinciding with the Carnival season in Brazil. Additionally, the month of August shows a peak in order count, potentially related to the Festival de Cachaça dedicated to the national liquor, cachaça. It is important to note that further analysis with a larger dataset would be required to validate these seasonality trends.

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

(Dawn, Morning, Afternoon or Night)

- a. 0-6 hrs : Dawn
- b. 7-12 hrs : Mornings
- c. 13-18 hrs : Afternoon

d. 19-23 hrs : Night

QUERY:-

```
SELECT
CASE WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp)
BETWEEN 0 AND 6 THEN 'Dawn'
WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp)
BETWEEN 7 AND 12 THEN 'Mornings'
WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp)
BETWEEN 13 AND 18 THEN 'Afternoon'
WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp)
BETWEEN 19 AND 23 THEN 'Night' END as Hour, COUNT (o.order_id) AS `ORDER_COUNT`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.customers` c
ON o.customer_id =c.customer_id
GROUP BY Hour
ORDER BY order_count;
```

OUTPUT :-

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

RECOMMENDATION:-

Based on the analysis, we found that Brazilian customers tend to place most orders during the daytime, specifically in the afternoon and night. This indicates that customers prefer to shop online when they have leisure time or after completing their daily activities. It's important to note that the assumption here is that the recorded timestamps reflect the correct time zone at the time of purchase

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

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

QUERY:-

```
SELECT
c.customer_state,
EXTRACT(MONTH FROM order_purchase_timestamp) AS `MONTH`,
COUNT(order_purchase_timestamp) AS `ORDER_PLACED`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.customers` c
ON o.customer_id =c.customer_id
GROUP BY 1,2
ORDER BY 1,2;
```

OUTPUT:-

Row	customer_state	MONTH	ORDER_PLACED
1	AC	1	8
2	AC	2	6
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6

RECOMMENDATION:-

The above output illustrates the month-on-month order counts in each state of Brazil, providing valuable insights into the customer purchase trends on a state-by-state basis. It is evident that Sao Paulo (SP) consistently has the highest number of orders in any given month, followed by Riode Janeiro (RJ) and Minas Gerais (MG)

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

QUERY:-

```
SELECT
customer_state,
count(customer_id) AS `No_of_customers`
FROM `Brazilian_Target_Business_Case_Solution.customers`
GROUP BY 1
ORDER BY 2 DESC;
```

OUTPUT :-

Row	customer_state	No_of_customers
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020

RECOMMENDATION:-

Due to its position as the most populated state in Brazil, the statistics indicates that Sao Paulo (SP) has the largest number of customers. It is essential for companies such as Target to comprehend the distribution of clients among Brazilian states and the evolution of e-commerce orders in order to customise marketing tactics, streamline logistics, and improve customer experiences.

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

1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only). You can use the "payment_value" column in the payments table to get the cost of orders.

QUERY :-


```

SELECT
EXTRACT(month FROM o.order_purchase_timestamp) AS `MONTH`,
ROUND((( SUM (CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018
AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
THEN p.payment_value END) - SUM (CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) =
2017
AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
THEN p.payment_value END) ) / SUM (CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) =
2017
AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
THEN p.payment_value END) * 100) ,2) AS `PERCENTAGE_INCREASES`
from `Brazilian_Target_Business_Case_Solution.orders` o join
`Brazilian_Target_Business_Case_Solution.payments` p
ON o.order_id = p.order_id
WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017,2018)
AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
GROUP BY 1
ORDER BY 1;

```

OUTPUT :-

Row	MONTH	PERCENTAGE_INCREASES
1	1	705.13
2	2	239.99
3	3	157.78
4	4	177.84
5	5	94.63
6	6	100.26
7	7	80.04
8	8	51.61

RECOMMENDATION:-

Including only the months of January through August, the total percentage rise in order costs from 2017 to 2018 is 138.53 percent. The month with the largest percentage increase is January, followed by February and April when looking at the increase month by month.

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

QUERY:-

```
SELECT
c.customer_state,
ROUND(SUM(oi.price),2) AS `TOTAL_PRICE`,
ROUND(AVG(oi.price),2) AS `AVG_PRICE`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.order_items` oi
ON o.order_id =oi.order_id JOIN `Brazilian_Target_Business_Case_Solution.customers` c
ON o.customer_id =c.customer_id
GROUP BY 1
ORDER BY 1;
```

OUTPUT:-

Row	customer_state	TOTAL_PRICE	AVG_PRICE
1	AC	15982.95	173.73
2	AL	80314.81	180.89
3	AM	22356.84	135.5
4	AP	13474.3	164.32
5	BA	511349.99	134.6
6	CE	227254.71	153.76
7	DF	302603.94	125.77
8	ES	275037.31	121.91
9	GO	294591.95	126.27
10	MA	119648.22	145.2

RECOMMENDATION:-

The analysis reveals interesting findings. While Sao Paulo (SP) has the highest total price value and total freight value, it surprisingly has the lowest average price value and average freight value among all states. On the other hand, the state of Paraíba (PB) has the highest average price value and average freight value.

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 `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.order_items` oi
ON o.order_id =oi.order_id JOIN `Brazilian_Target_Business_Case_Solution.customers` c
ON o.customer_id = c.customer_id
GROUP BY 1
ORDER BY 1;
```

OUTPUT :-

Row	customer_state	TOTAL_FREIGHT_VALUE	AVG_FREIGHT_VALUE
1	AC	3686.75	40.07
2	AL	15914.59	35.84
3	AM	5478.89	33.21
4	AP	2788.5	34.01
5	BA	100156.68	26.36
6	CE	48351.59	32.71
7	DF	50625.5	21.04
8	ES	49764.6	22.06
9	GO	53114.98	22.77
10	MA	31523.77	38.26

RECOMMENDATION:-

The analysis reveals interesting findings. While Sao Paulo (SP) has the highest total price value and total freight value, it surprisingly has the lowest average price value and average freight value among all states. On the other hand, the state of Paraíba (PB) has the highest average price value and 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,
DATE_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY) AS `DELIVERED_IN_DAYS`,
DATE_DIFF(order_estimated_delivery_date,
order_purchase_timestamp, DAY) AS `ESTIMATED_DELIVERY_IN_DAYS`,
DATE_DIFF(order_estimated_delivery_date,
order_delivered_customer_date, DAY) AS `ESTIMATED_MINUS_ACTUAL_DELIVERY_IN_DAYS`
FROM `Brazilian_Target_Business_Case_Solution.orders`
WHERE DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) IS NOT NULL
ORDER BY DELIVERED_IN_DAYS;
```

OUTPUT :-

Row	order_id	DELIVERED_IN_DAYS	ESTIMATED_DELIVERY_IN_DAYS	ESTIMATED_MINUS_ACTUAL_DELIVERY_IN_DAYS
1	e65f1eeee1f52024ad1dcd03447f7482	0	10	9
2	bb5a519e352b45b714192a02ffe25681	0	26	25
3	434cecee7d1a65fc65358a632b6f725f	0	20	19
4	d3ca7b82c922817b06e5ca21165c5ea2	0	12	11
5	1d893dd7ca5f77ebf5f59f0d2017eee0	0	10	10
6	d5fbeedc85190ba88580d6f82d1d5ed3	0	8	7
7	79e324907160caea526fd8b94389dbbc	0	9	8
8	38c1e3d4ed6a13cd0cf612d4c09766e9	0	17	16
9	8339b608be0d84fca9d8da68b58332c3	0	28	27
10	f349cdb62f69c3fae5c4d7d3f3a4a185	0	13	12

RECOMMENDATION:-

From the result, we found the time duration between purchasing an order, its delivery, and the estimated delivery.

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

QUERY :-

```

SELECT
c.customer_state,
ROUND(AVG(i.freight_value),2) AS `AVG_FREIGHT_VALUE`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.order_items` i
ON o.order_id =i.order_id JOIN `Brazilian_Target_Business_Case_Solution.customers` c
ON o.customer_id =c.customer_id
GROUP BY c.customer_state
ORDER BY AVG_FREIGHT_VALUE;

```

OUTPUT :- FIRST TOP 5 HIGHEST AVG

Row	customer_state	AVG_FREIGHT_VALUE
1	RR	42.98
2	PB	42.72
3	RO	41.07
4	AC	40.07
5	PI	39.15

OUTPUT :- FIRST TOP 5 LOWEST AVG

Row	customer_state	AVG_FREIGHT_VALUE
1	SP	15.15
2	PR	20.53
3	MG	20.63
4	RJ	20.96
5	DF	21.04

RECOMMENDATION:-

From the results, we can observe that Sao Paulo (SP) has the lowest average time to delivery, while Roraima (RR) has the highest average freight value.

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

QUERY :-

```

SELECT
c.customer_state,
ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,o.order_purchase_timestamp,DAY)),2) AS
`AVG_TIME_TO_DILIVERY`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.order_items` i
ON o.order_id =i.order_id JOIN `Brazilian_Target_Business_Case_Solution.customers` c
ON o.customer_id =c.customer_id
GROUP BY c.customer_state
ORDER BY AVG_TIME_TO_DILIVERY;

```

OUTPUT :- FIRST TOP 5 HIGHEST AVG_TIME_TO_DILIVERY

Row	customer_state	AVG_TIME_TO_DILIVERY
1	RR	27.83
2	AP	27.75
3	AM	25.96
4	AL	23.99
5	PA	23.3

OUTPUT :- FIRST TOP 5 LOWEST AVG_TIME_TO_DILIVERY

Row	customer_state	AVG_TIME_TO_DILIVERY
1	SP	8.26
2	PR	11.48
3	MG	11.52
4	DF	12.5
5	SC	14.52

RECOMMENDATION:-

From the results, we can observe that Sao Paulo (SP) has the lowest average time to delivery, while Roraima (RR) has the highest average time to delivery.

- Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery. You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

QUERY :-

```
SELECT
c.customer_state,
ROUND(AVG(DATE_DIFF(o.order_estimated_delivery_date,o.order_delivered_customer_date,DAY)),2)
AS `DIFF_ESTIMATED_DILIVERY`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.order_items` i
ON o.order_id =i.order_id JOIN `Brazilian_Target_Business_Case_Solution.customers` c
ON o.customer_id = c.customer_id
GROUP BY c.customer_state
ORDER BY DIFF_ESTIMATED_DILIVERY DESC;
```

OUTPUT :-

Row	customer_state	DIFF_ESTIMATED_DILIVERY
1	AC	20.01
2	RO	19.08
3	AM	18.98
4	AP	17.44
5	RR	17.43
6	MT	13.64
7	PA	13.37
8	RS	13.2
9	RN	13.06
10	PE	12.55
11	PR	12.53

RECOMMENDATION:-

Understanding the sales, freight, and delivery time dynamics is crucial for businesses to optimize their operations and improve customer satisfaction. By leveraging SQL queries and analyzing state-wise patterns, companies can gain valuable insights into sales trends, optimize their logistics, and make data-driven decisions to enhance their overall efficiency and customer experience.

6. Analysis based on the payments:-

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

QUERY :-

```
SELECT
p.payment_type,
EXTRACT(MONTH FROM o.order_purchase_timestamp) AS `MONTH`,
COUNT(DISTINCT o.order_id) AS `ORDER_COUNT`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.payments` p
ON o.order_id = p.order_id
GROUP BY 1,2
ORDER BY 1,2;
```

OUTPUT :-

Row	payment_type	MONTH	ORDER_COUNT
1	UPI	1	1715
2	UPI	2	1723
3	UPI	3	1942
4	UPI	4	1783
5	UPI	5	2035
6	UPI	6	1807
7	UPI	7	2074
8	UPI	8	2077
9	UPI	9	903
10	UPI	10	1056

RECOMMENDATION:-

The analysis shows an overall uptrend from January to August and another uptrend from September to November. Credit card transactions are the most popular payment method, followed by UPI. Debit card transactions are the least preferred option. Notably, credit card transactions are rapidly increasing compared to other payment methods, possibly due to benefits like “buy now, pay later” options or cashback received using credit cards.

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

QUERY :-


```

SELECT
p.payment_installments,
COUNT(o.order_id) AS `ORDER_COUNT`
FROM `Brazilian_Target_Business_Case_Solution.orders` o JOIN
`Brazilian_Target_Business_Case_Solution.payments` p
ON o.order_id =p.order_id
WHERE o.order_status != 'canceled'
GROUP BY 1
ORDER BY 2 DESC;

```

OUTPUT :-

Row	payment_installments	ORDER_COUNT
1	1	52184
2	2	12353
3	3	10392
4	4	7056
5	10	5292
6	5	5209
7	8	4239
8	6	3898
9	7	1620
10	9	638
11	12	133

RECOMMENDATION:-

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

7. Actionable Insights & Recommendations

Enhance the logistics Since the analysis shows that there are some problems with delivery times in some areas, so the business may want to think about streamlining its shipping procedures or collaborating with reputable shipping companies to speed up delivery times. Prioritize marketing during the busiest months. According to the data, there may be some seasonality characteristics in Brazil's e-commerce sales, with peaks occurring in particular months. To take advantage of the increasing demand, the business may concentrate on marketing during these busy months. Enhance the projected delivery schedules. According to the investigation, there are certain problems with scheduled delivery times not being fulfilled, which can leave clients unhappy. Through increased precision in its projected delivery windows, the business might enhance client contentment and allegiance.