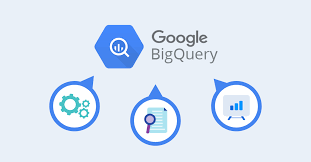
***Business Case: Target SQL***

******

***About Target SQL:-***

*● 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:***

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

***Dataset:***

1. ***customers.csv***
2. ***sellers.csv***
3. ***order\_items.csv***
4. ***geolocation.csv***
5. ***payments.csv***
6. ***reviews.csv***
7. ***orders.csv***
8. ***products.csv***

***What does ‘good’ look like?***

**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.**

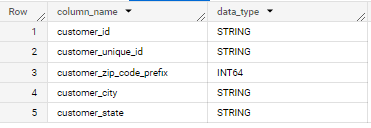
**SELECT**

**column\_name,**

**data\_type,**

**FROM `target\_business\_case`.INFORMATION\_SCHEMA.COLUMNS**

**WHERE table\_name = "customers"**

****

**Insides:- In this Question I’m checking the data types to related customer then i took my project name with customer column then i did apply where condition because wanted take the customer details. So we can check the details in the table.**

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

**SELECT MIN(order\_purchase\_timestamp) AS first\_order\_time, MAX(order\_purchase\_timestamp) AS last\_order\_time**

**FROM `target\_business\_case.orders`**

****

**Insides :- In this query im checking the in what time order were placed. I wrote after the select t mentioned the min and max and calling from the with my project name and with that column name i wanted to get the details. So we can check the details in the table.**

**C. Count the number of Cities and States in our dataset.**

**SELECT COUNT(DISTINCT geolocation\_city) AS city\_count, COUNT(DISTINCT geolocation\_state) AS state\_count**

**FROM `target\_business\_case.geolocation`;**

****

**Insides :- In the query i’m trying the get the details. How many city and states are means total numbers of city and states i took count function then in the From statement i mentioned table name.so we can in the table.**

**II. In-depth Exploration:**

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

**SELECT**

**EXTRACT(YEAR FROM order\_delivered\_customer\_date) AS year,**

**EXTRACT(MONTH FROM order\_delivered\_customer\_date) AS month,**

**COUNT(\*) AS order\_count**

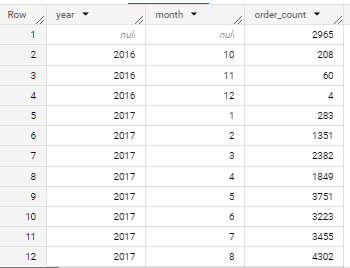
**FROM `target\_business\_case.orders`**

**GROUP BY**

**year, month**

**ORDER BY**

**year, month;**

****

**Insides :- In this query want to get the trend is growing in which year and how many years and in how many months and how many times. So we can check the details as given in table.**

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

**SELECT**

**EXTRACT(MONTH FROM order\_delivered\_customer\_date) AS month,**

**COUNT(\*) AS order\_count**

**FROM**

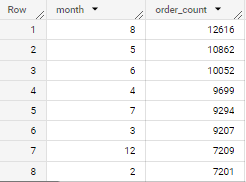
**`target\_business\_case.orders`**

**GROUP BY**

**month**

**ORDER BY**

**order\_count DESC;**

****

**Insides : - In this query wanted to check the seasonality in terms of the orders being placed so here we can check how many months is there and order count in the table.**

**C. 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**

**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'**

**ELSE 'Other'**

**END AS order\_time\_of\_day,**

**COUNT(\*) AS order\_count**

**FROM**

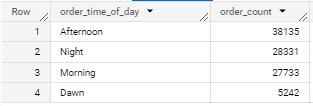
**`target\_business\_case.orders`**

**GROUP BY**

**order\_time\_of\_day**

**ORDER BY**

**order\_count DESC;**

****

**Insides :- i want to categorize the hours of a day into the given time brackets/intervals and find out during which intervals the Brazilian customers usually**

**order the most.so in the table we can check the all details.**

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

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

**SELECT**

**g.geolocation\_state,**

**EXTRACT(YEAR FROM o.order\_purchase\_timestamp) AS year,**

**EXTRACT(MONTH FROM o.order\_purchase\_timestamp) AS month,**

**COUNT(o.order\_id) AS order\_count**

**FROM**

**`target\_business\_case.geolocation` AS g**

**INNER JOIN**

**`target\_business\_case.customers` AS c**

**ON**

**g.geolocation\_zip\_code\_prefix = c.customer\_zip\_code\_prefix**

**INNER JOIN**

**`target\_business\_case.orders` AS o**

**ON**

**o.customer\_id = c.customer\_id**

**GROUP BY**

**g.geolocation\_state,**

**year,**

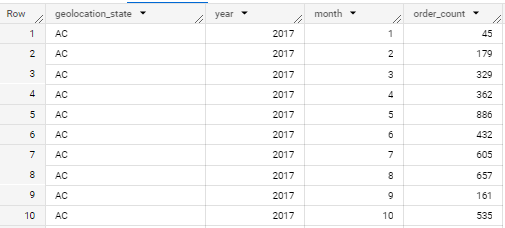
**month**

**ORDER BY**

**g.geolocation\_state,**

**year,**

**Month;**

****

**Insides :- in want to get the no. of orders placed in each state, in each month**

**by customers with oder count and all the details given in the table.**

**B. How are the customers distributed across all the states?**

**SELECT**

**g.geolocation\_state,**

**COUNT(DISTINCT c.customer\_id) AS unique\_customers\_count**

**FROM**

**`target\_business\_case.geolocation` g**

**INNER JOIN `target\_business\_case.customers` c**

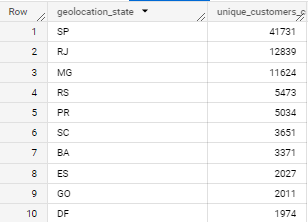
**ON g.geolocation\_zip\_code\_prefix=c.customer\_zip\_code\_prefix**

**GROUP BY**

**g.geolocation\_state**

**ORDER BY**

**unique\_customers\_count DESC**

****

**Insides :- I want to get the no. of unique customers present in each state. And how many uniques customers is there so we can check the as in given table.**

**IV. Impact on Economy: Analyze the money movement by e-commerce by looking at**

**order prices, freight and others.**

**A. Get the % increase in the cost of orders from year 2017 to 2018 (include**

**months between Jan to Aug only).**

**WITH total\_payment AS**

**(**

**SELECT EXTRACT(YEAR FROM o.order\_purchase\_timestamp) AS payment\_year,**

**SUM(p.payment\_value) AS summed\_payment**

**FROM `target\_business\_case.orders` o**

**INNER JOIN `target\_business\_case.payments` p**

**ON o.order\_id=p.order\_id**

**WHERE EXTRACT(MONTH FROM order\_purchase\_timestamp) BETWEEN 1 AND 8 AND**

**EXTRACT(YEAR FROM order\_purchase\_timestamp) IN (2017, 2018)**

**GROUP BY payment\_year**

**),**

**payment\_of\_2018 AS**

**(**

**SELECT summed\_payment FROM total\_payment WHERE payment\_year = 2018**

**),**

**payment\_of\_2017 AS**

**(**

**SELECT summed\_payment FROM total\_payment WHERE payment\_year = 2017**

**)**

**SELECT DISTINCT(ROUND((((SELECT summed\_payment FROM payment\_of\_2018) -**

**(SELECT summed\_payment from payment\_of\_2017))/(SELECT summed\_payment FROM**

**payment\_of\_2017)) \* 100, 2)) AS percentage\_increase FROM total\_payment**

****

**Insides :- payment\_value column in the payments table to get the cost of orders. I was trying the check the increasing percents of the cost of orders.**

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

**SELECT**

**g.geolocation\_state AS state,**

**SUM(o.price) AS total\_price,**

**AVG(o.price) AS average\_price**

**FROM**

**`target\_business\_case.sellers` AS s**

**JOIN**

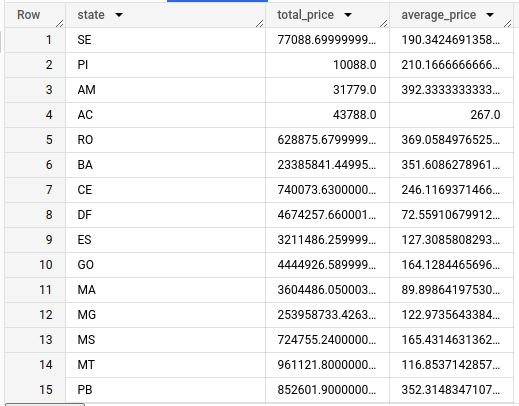
**`target\_business\_case.geolocation` AS g ON s.seller\_zip\_code\_prefix = g.geolocation\_zip\_code\_prefix**

**join**

**`target\_business\_case.order\_items` AS o on o.seller\_id = s.seller\_id**

**GROUP BY**

**state**

****

**Insides :- I want to fetch the total price and the average price of orders for**

**each state. I did the calculate the total & average value of order price for each state and we can check the tables.**

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

**SELECT**

**g.geolocation\_state AS state,**

**SUM(o.freight\_value) AS total\_freight,**

**AVG(o.freight\_value) AS average\_freight**

**FROM**

**`target\_business\_case.sellers` AS s**

**JOIN**

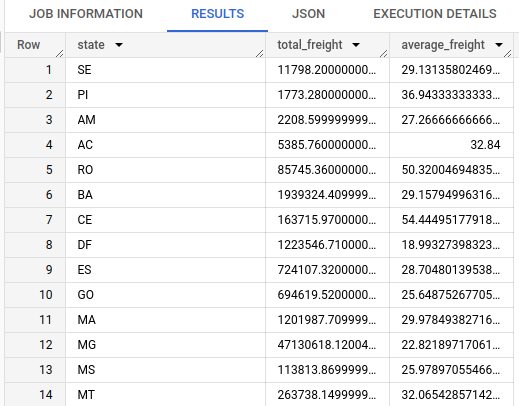
**`target\_business\_case.geolocation` AS g ON s.seller\_zip\_code\_prefix = g.geolocation\_zip\_code\_prefix**

**join**

**`target\_business\_case.order\_items` AS o on o.seller\_id = s.seller\_id**

**GROUP BY**

**State**

****

**Insides :- I want to fetch the total freight value and the average freight value**

**of orders for each state. I did the calculate the total &aAverage value of order freight for each state and we can check the tables.**

**V. Analysis based on sales, freight and delivery time.**

**A. 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.**

**SELECT**

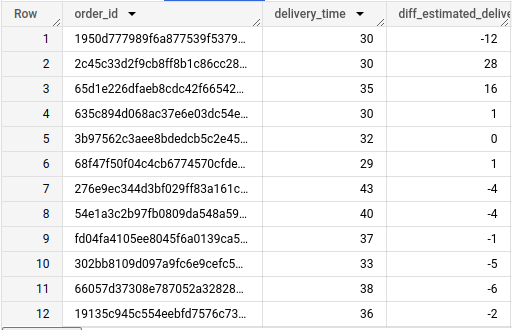
**o.order\_id,**

**DATE\_DIFF(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, DAY) AS delivery\_time,**

**DATE\_DIFF(o.order\_estimated\_delivery\_date, o.order\_delivered\_customer\_date, DAY) AS diff\_estimated\_delivery**

**FROM**

**`target\_business\_case.orders` AS o**

****

**Insides :- The delivery time and the difference between the estimated & actual delivery date checking the given by formula. Finding the no. of days taken to deliver each order from the order’s purchase date as delivery time.**

**Also, calculating the difference (in days) between the estimated & actual delivery**

**date of an order. All details given in the table.**

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

**WITH state\_freight AS (**

**SELECT**

**g.geolocation\_state AS state,**

**AVG(o.freight\_value) AS average\_freight**

**FROM**

**`target\_business\_case.geolocation` AS g**

**JOIN**

**`target\_business\_case.sellers` AS s ON g.geolocation\_zip\_code\_prefix=s.seller\_zip\_code\_prefix**

**JOIN `target\_business\_case.order\_items` as o ON o.seller\_id=s.seller\_id**

**GROUP BY**

**state**

**)**

**SELECT**

**state,**

**average\_freight**

**FROM (**

**SELECT**

**state,**

**average\_freight**

**FROM**

**State\_freight**

**ORDER BY**

**average\_freight DESC**

**LIMIT**

**5**

**) top\_states**

**UNION ALL**

**SELECT**

**state,**

**average\_freight**

**FROM (**

**SELECT**

**state,**

**average\_freight**

**FROM**

**state\_freight**

**ORDER BY**

**average\_freight ASC**

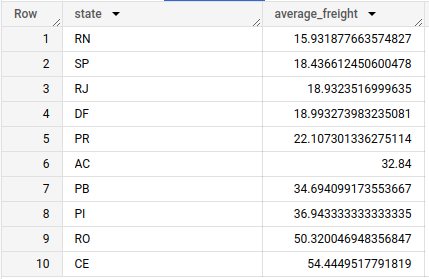
**LIMIT**

**5**

**) bottom\_states**

**ORDER BY**

**average\_freight ASC**

****

**Insides :- I want to find the top 5 & the bottom 5 states arranged in**

**increasing order of the average freight value. So Finding out the top 5 states with the highest & lowest average freight value. In the given table we can check the results.**

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

**WITH state\_delivery\_time AS (**

**SELECT**

**g.geolocation\_state AS state,**

**AVG(DATE\_DIFF(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, DAY)) AS average\_delivery\_time**

**FROM**

**`target\_business\_case.customers` AS c**

**JOIN**

**`target\_business\_case.geolocation` AS g ON c.customer\_zip\_code\_prefix = g.geolocation\_zip\_code\_prefix**

**JOIN**

**`target\_business\_case.orders` as o on o.customer\_id = c.customer\_id**

**GROUP BY**

**state**

**)**

**SELECT**

**state,**

**average\_delivery\_time**

**FROM (**

**SELECT**

**state,**

**average\_delivery\_time**

**FROM**

**state\_delivery\_time**

**ORDER BY**

**average\_delivery\_time DESC**

**LIMIT**

**5**

**) top\_states**

**UNION ALL**

**SELECT**

**state,**

**average\_delivery\_time**

**FROM (**

**SELECT**

**state,**

**average\_delivery\_time**

**FROM**

**State\_delivery\_time**

**ORDER BY**

**average\_delivery\_time ASC**

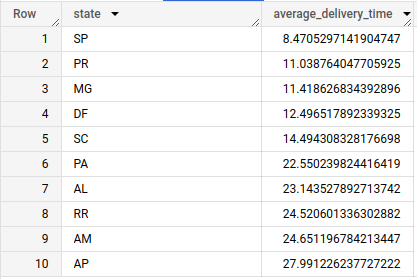
**LIMIT**

**5**

**) bottom\_states**

**ORDER BY**

**average\_delivery\_time ASC**

****

**Insides :- I want to find the top 5 & the bottom 5 states arranged in**

**increasing order of the average delivery time. So Finding out the top 5 states with the highest & lowest average delivery value. In the given table we can check the results.**

**D. 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.**

**WITH state\_delivery\_speed AS (**

**SELECT**

**g.geolocation\_state AS state,**

**AVG(o.order\_estimated\_delivery\_date - o.order\_purchase\_timestamp) - AVG(o.order\_delivered\_customer\_date - o.order\_purchase\_timestamp) AS delivery\_speed**

**FROM**

**`target\_business\_case.customers` AS c**

**JOIN**

**`target\_business\_case.geolocation` AS g ON c.customer\_zip\_code\_prefix = g.geolocation\_zip\_code\_prefix**

**INNER JOIN**

**`target\_business\_case.orders` AS o ON o.customer\_id = c.customer\_id**

**WHERE**

**o.order\_status = 'delivered'**

**GROUP BY**

**state**

**)**

**SELECT**

**state,**

**delivery\_speed**

**FROM**

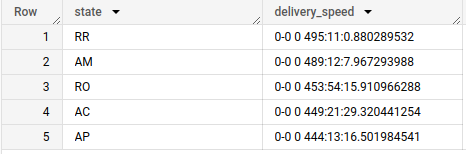
**state\_delivery\_speed**

**ORDER BY**

**delivery\_speed DESC**

**LIMIT**

**5**

****

**Insides :- Including the orders. I’m trying to Finding out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.**

**So i’m differentiatng the averages of actual & estimated delivery**

**date to figure out how fast the delivery was for each state. So we can see the table.**

**VI. Analysis based on the payments:**

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

**SELECT**

**EXTRACT(YEAR FROM order\_purchase\_timestamp) AS year,**

**EXTRACT(MONTH FROM order\_purchase\_timestamp) AS month,**

**p.payment\_type,**

**COUNT(o.order\_id) AS order\_count**

**FROM**

**`target\_business\_case.orders` AS o**

**JOIN**

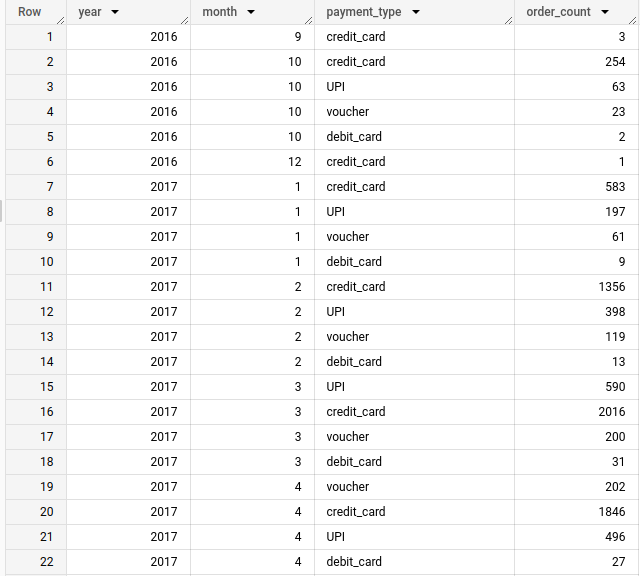
**`target\_business\_case.payments` AS p ON o.order\_id = p.order\_id**

**GROUP BY**

**year, month, p.payment\_type**

**ORDER BY**

**year, month;**

****

**Insides :- I to count the no. of orders placed using different payment**

**methods in each month over the past years. Finding the on month no. of orders placed with different payment types. So with the help of join i did and we can see the results in the table.**

**B. Find the no. of orders placed on the basis of the payment installments that have**

**been paid.**

**SELECT**

**payment\_installments,**

**COUNT(DISTINCT o.order\_id) AS order\_count**

**FROM**

**`target\_business\_case.orders` AS o**

**JOIN**

**`target\_business\_case.payments` AS p ON o.order\_id = p.order\_id**

**WHERE**

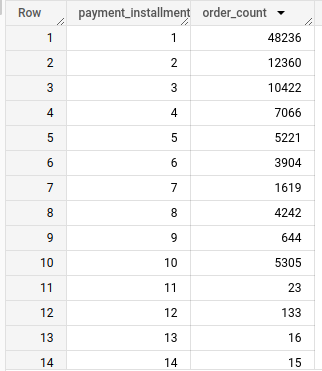
**p.payment\_sequential = 1**

**GROUP BY**

**payment\_installments**

**ORDER BY**

**payment\_installments**

****

**Insides :- I want to count the no. of orders placed based on the no. of**

**payment installments at least one installment has been successfully paid.**

**Finding the no. of orders placed on the basis of the payment installments that have**

**been paid. So i called the tables and mentioned the columns names there and took the obj payment\_sequential so in this table we can get the details of payments and orders count.**