Target Business Case:

Target is one of the world's most recognized brands and one of America's leading retailers. 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 business case has information of 100k orders from 2016 to 2018 made at Target in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.

Dataset:
https://drive.google.com/drive/folders/1TGEc66YKbD443nslRi1bWgVd238gJCnb

Data is available in 8 csv files:

1. customers.csv

2. geolocation.csv

3. order_items.csv

4. payments.csv

5. reviews.csv

6. orders.csv

7. products.csv

8. sellers.csv

Each feature or columns of different CSV files are described below:

The customers.csv contain following features:

Features	Description
customer_id	Id of the consumer who made the purchase.
customer_unique_id	Unique Id of the consumer.
customer_zip_code_prefix	Zip Code of the location of the consumer.
customer_city	Name of the City from where order is made.
customer_state	State Code from where order is made(Ex- sao paulo-SP).

The sellers.csv contains following features:

Features	Description
seller_id	Unique Id of the seller registered
seller_zip_code_prefix	Zip Code of the location of the seller.
seller_city	Name of the City of the seller.
seller_state	State Code (Ex- sao paulo-SP)

The order_items.csv contain following features:

Features	Description
order_id	A unique id of order made by the consumers.

order_item_id	A Unique id given to each item ordered in the order.
product_id	A unique id given to each product available on the site.
seller_id	Unique Id of the seller registered in Target.
shipping_limit_date	The date before which shipping of the ordered product must be completed.
price	Actual price of the products ordered .
freight_value	Price rate at which a product is delivered from one point to another.

The geolocations.csv contain following features:

Features	Description
geolocation_zip_code_prefix	first 5 digits of zip code
geolocation_lat	latitude
geolocation_lng	longitude
geolocation_city	city name
geolocation_state	state

The payments.csv contain following features:

Features	Description
order_id	A unique id of order made by the consumers.
payment_sequential	sequences of the payments made in case of EMI.
payment_type	mode of payment used.(Ex-Credit Card)
payment_installments	number of installments in case of EMI purchase.
payment_value	Total amount paid for the purchase order.

The orders.csv contain following features:

Features	Description
order_id	A unique id of order made by the consumers.
customer_id	Id of the consumer who made the purchase.
order_status	status of the order made i.e delivered, shipped etc.
order_purchase_timestamp	Timestamp of the purchase.
order_delivered_carrier_date	delivery date at which carrier made the delivery.
order_delivered_customer_date	date at which customer got the product.
order_estimated_delivery_date	estimated delivery date of the products.

The reviews.csv contain following features:

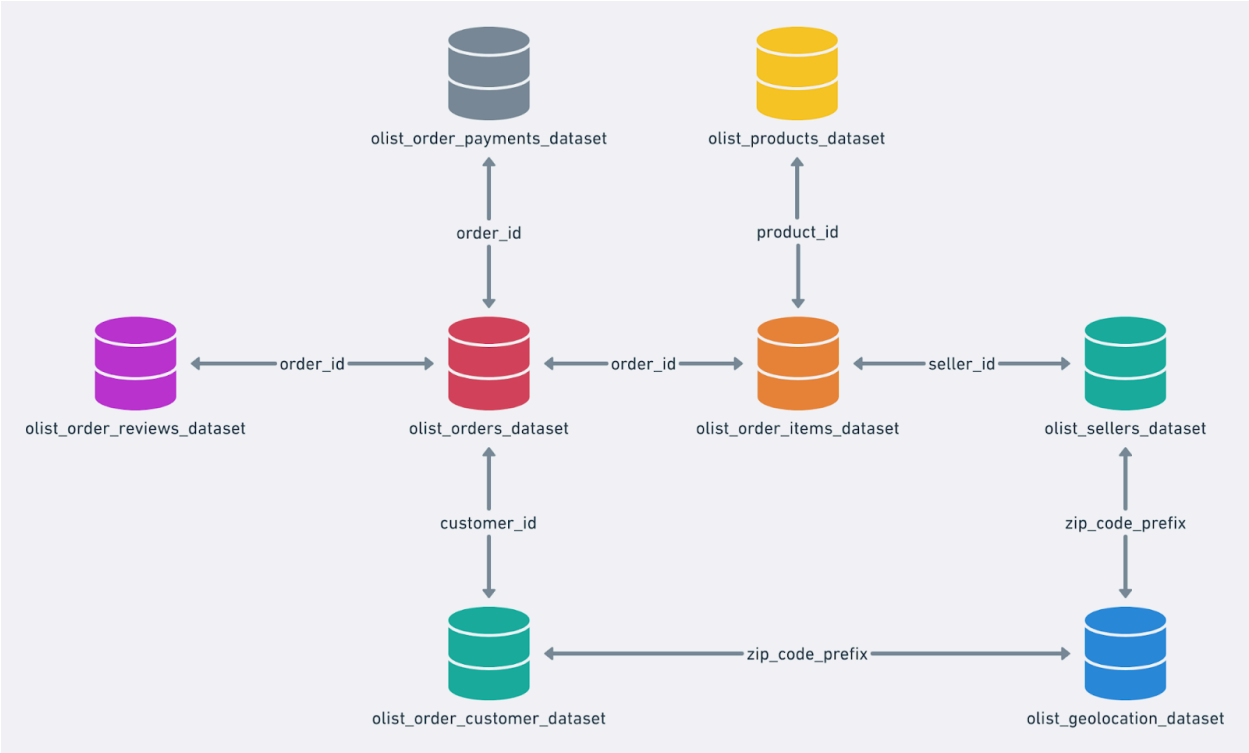
Features	Description
review_id	Id of the review given on the product ordered by the order id.
order_id	A unique id of order made by the consumers.
review_score	review score given by the customer for each order on the scale of 1–5.
review_comment_title	Title of the review
review_comment_message	Review comments posted by the consumer for each order.
review_creation_date	Timestamp of the review when it is created.
review_answer_timestamp	Timestamp of the review answered.

The products.csv contain following features:

Features	Description
product_id	A unique identifier for the proposed project.
product_category_name	Name of the product category
product_name_lenght	length of the string which specifies the name given to the products ordered.
product_description_lenght	length of the description written for each product ordered on the site.

product_photos_qty	Number of photos of each product ordered available on the shopping portal.
product_weight_g	Weight of the products ordered in grams.
product_length_cm	Length of the products ordered in centimeters.
product_height_cm	Height of the products ordered in centimeters.
product_width_cm	width of the product ordered in centimeters.

High level overview of relationship between datasets:



Assume you are a data scientist at Target, and are given this data to analyze and provide some insights and recommendations from it.

What ‘good’ looks like?

1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset
    1. Data type of columns in a table
    2. Time period for which the data is given
    3. Cities and States covered in the dataset
2. In-depth Exploration:
    1. Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?
    2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?
3. Evolution of E-commerce orders in the Brazil region:
    1. Get month on month orders by region, states
    2. How are customers distributed in Brazil
4. Impact on Economy: Analyze the money movemented by e-commerce by looking at order prices, freight and others.
    1. Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only)
    2. Mean & Sum of price and freight value by customer state
5. Analysis on sales, freight and delivery time
    1. Calculate days between purchasing, delivering and estimated delivery
    2. Create columns:
        1. time_to_delivery = order_purchase_timestamp-order_delivered_customer_date
        2. diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date
    3. Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery
    4. Sort the data to get the following:
        1. Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5
        2. Top 5 states with highest/lowest average time to delivery
        3. Top 5 states where delivery is really fast/ not so fast compared to estimated date

6. Payment type analysis:
    1. Month over Month count of orders for different payment types
    2. Distribution of payment installments and count of orders

Evaluation Criteria (80 points)

1. Initial exploration of dataset like checking the characteristics of data (10 points)
2. In-depth Exploration (10 points)
3. Evolution of E-commerce orders in the Brazil region (10 points)
4. Impact on Economy (10 points)
5. Analysis on sales, freight and delivery time (10 points)
6. Payment type analysis (10 points)
7. Actionable Insights (10 points)
8. Recommendations (10 points)

Submission Process:

- Type your insights and recommendations in the text editor
- Convert your solutions notebook into PDF, upload it on the dashboard
- Optionally, you may add images/graphs in the text editor by taking screenshots
- After submitting, you will not be allowed to edit your submission

Solution :

Q1) Import the dataset and do usual exploratory analysis steps like checking the structure &
characteristics of the dataset

1) Data type of columns in a table
2) Time period for which the data is given
3) Cities and States covered in the dataset

1) Data type of columns in a table

```
describe customers;
describe geolocation;
describe order_items;
describe order_reviews;
```

```
describe orders;
describe payments;
describe products;
describe reviews;
describe sellers;
```

2) Time period for which the data is given

```
SELECT
        MIN(order_purchase_timestamp) AS min_date,
        MAX(order_purchase_timestamp) AS max_date,
        datediff(MAX(order_purchase_timestamp), MIN(order_purchase_timestamp)) AS
days_difference
FROM orders;
```

3) Cities and States covered in the dataset

- **Distinct Cities**

```
SELECT
        DISTINCT geolocation_city AS cities
FROM
        geolocation;
```

- **Distinct States**

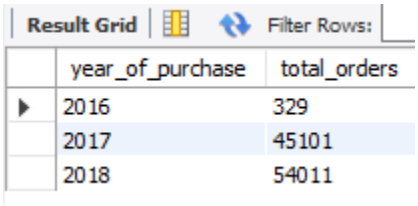
```
SELECT
        DISTINCT geolocation_state AS states
FROM
        geolocation;
```

Q2) In-depth Exploration:

1. **Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?**
2. **What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?**

1) **Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?**
   - **To verify if there is an increasing trend on e-commerce in Brazil, we can check if the number of orders has increased over the years. If yes then we can say that there is an increasing trend. So to get year wise data and their count we can use the following query :**

```
SELECT
        YEAR(order_purchase_timestamp) as year_of_purchase,
        count(*) as total_orders
FROM orders
GROUP BY year_of_purchase
ORDER BY year_of_purchase;
```



year_of_purchase	total_orders
2016	329
2017	45101
2018	54011

Conclusion: So we can see the number of orders placed is increasing continuously and there is an increasing trend on e-commerce in Brazil.

   - **To check seasonality where there are more orders placed. Can we see some seasonality with peaks at specific months**

Link used to refer months and seasons : Seasons and Months

Use comparison year wise.

```
SELECT
	CASE
		WHEN  MONTH(order_purchase_timestamp) in (1, 2, 3) then 'Summer'
		WHEN  MONTH(order_purchase_timestamp) in (4, 5, 6) then 'Autumn'
		WHEN  MONTH(order_purchase_timestamp) in (7, 8, 9) then 'Winter'
		WHEN  MONTH(order_purchase_timestamp) in (10, 11, 12) then 'Spring'
	END AS season,
COUNT(*) AS orders_per_season
FROM orders
GROUP BY season
ORDER BY orders_per_season DESC;
```

season	orders_per_season
Autumn	29328
Summer	26470
Winter	25466
Spring	18177

Considering Months **(Jan, Feb, Mar)** As **'Summer'**, **(Apr, May, Jun)** As **'Autumn', (Jul, Aug, Sep)** As **'Winter'** and **(Oct, Nov, Dec)** As **'Spring'** seasons. So when we count total orders placed per season, we can see that **Autumn season** has more orders placed, following **Summer**, **Winter**, and **Spring** season.

Conclusion: So we can see the number of orders placed in Autumn Season are more as compared to any other Season in Brazil.

2) What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

```
SELECT
CASE

when hour(order_purchase_timestamp) BETWEEN 5 AND 11 then 'Morning'
when hour(order_purchase_timestamp) BETWEEN 12 AND 16 then 'Afternoon'
when hour(order_purchase_timestamp) BETWEEN 17 AND 20 then 'Evening'
else 'Night'

end AS time_purchased,
COUNT(*) as orders_placed
FROM orders
GROUP BY time_purchased
ORDER BY orders_placed DESC;
```

	time_purchased	orders_placed
▶	Afternoon	32211
	Evening	24094
	Morning	22428
	Night	20708

Considering Hours (24 Hours Format Clock) **5 - 11** As **'Morning'**, **12 - 16** As **'Afternoon'**, **17 - 20** As **'Evening'** and **21 - 4** As **'Night'** time. So when we count total orders placed based on time, we can see that most customers place orders in Afternoon following **Evening**, **Morning**, and **Night** time..

Conclusion: So we can see the number of orders placed in the Afternoon time period is more as compared to any other time in a day.

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

1. Get month on month orders by region, states

```
SELECT
        DISTINCT c.customer_state,
        MONTHNAME(order_purchase_timestamp) AS 'month',
        COUNT(*)
        OVER
        (
                PARTITION BY c.customer_state,
                YEAR  (order_purchase_timestamp),
                MONTH(order_purchase_timestamp)
                ORDER BY MONTH(order_purchase_timestamp)
        ) AS total_orders
FROM orders o
JOIN customers c
ON o.customer_id = c.customer_id
```

	customer_state	month	year	total_orders
▶	AC	January	2017	2
	AC	February	2017	3
	AC	March	2017	2
	AC	April	2017	5
	AC	May	2017	8
	AC	June	2017	4
	AC	July	2017	5

We can see that from State ‘AC’ there were 2 orders in January, 3 Orders in February and so on.

Conclusion: We can see the number of orders per month placed for all the states in Brazil.

2. **If we want to see increase or decrease in orders comparing each month, eg. If we want to see what is the growth in orders as compared to previous month we can do the following:**

```
SELECT
        customer_state,
        month,
        year,
        total_orders,
        ROUND((total_orders - LAG (total_orders)
        OVER (
                PARTITION BY customer_state, year
   ))/ LAG (total_orders) OVER (
                PARTITION BY customer_state, year
        ) * 100,2) AS difference_perc_growth
FROM (
        SELECT
                DISTINCT c.customer_state,
                MONTH(order_purchase_timestamp) AS 'month',
                YEAR(order_purchase_timestamp) AS 'year',
                COUNT(*)
                OVER(
                        PARTITION BY c.customer_state,
YEAR(order_purchase_timestamp), MONTH(order_purchase_timestamp)
                        ORDER BY MONTH(order_purchase_timestamp)
                ) AS total_orders
        FROM orders o
        JOIN customers c
        ON o.customer_id = c.customer_id
) AS table1
ORDER BY customer_state, year, month;
```

customer_state	month	year	total_orders	difference_perc_growth
AC	1	2017	2	NULL
AC	2	2017	3	50.00
AC	3	2017	2	-33.33
AC	4	2017	5	150.00
AC	5	2017	8	60.00
AC	6	2017	4	-50.00

The above query will give us total orders per month and the percentage of orders increased/decreased with respect to the previous month for each state for each year. With this we can analyze that each state has what percentage of orders has increased/decreased for each year.

3. How are customers distributed in Brazil

```
SELECT
        customer_state,
        total_customers,
        ROUND((total_customers/sum_cust) * 100,3) AS
customers_distributed_percentage
        FROM
        (
                SELECT
                        c.customer_state,
                        COUNT(*) as total_customers,
                        SUM(COUNT(*)) OVER() as sum_cust
                FROM orders o
                JOIN customers c
                ON o.customer_id = c.customer_id
                GROUP BY customer_state
        ) AS table1
ORDER BY customers_distributed_percentage DESC;
```

customer_state	total_customers	customers_distributed_percentage
SP	41746	41.981
RJ	12852	12.924
MG	11635	11.700
RS	5466	5.497
PR	5045	5.073
SC	3637	3.657

In the results above we can see that **most of the customers** are from the State abbreviation **'SP'** which is **41.981%**. **Followed by 'RJ', 'MG', 'RS', 'PR' and 'SC'.**

Conclusion: The following is the distribution of customers who are registered with Target around Brazil.

4) **Analysis on sales, freight and delivery time**
   a) **Calculate days between purchasing, delivering and estimated delivery**
   b) **Create columns:**
      i) **time_to_delivery = order_purchase_timestamp-order_delivered_customer_date**
      ii) **diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date**

```
SELECT
        o.order_id,
    c.customer_id,
        c.customer_state,

DATEDIFF(order_delivered_customer_date,order_purchase_timestamp)
AS time_to_delivery,

DATEDIFF(order_estimated_delivery_date,order_delivered_customer_dat
e) AS diff_estimated_delivery

FROM orders o
JOIN customers c
ON c.customer_id = o.customer_id;
```

order_id	customer_id	customer_state	time_to_delivery	diff_estimated_delivery
00010242fe8c5a6d1ba2dd792cb16214	3ce436f183e68e07877b285a838db11a	RJ	7	9
00018f77f2f0320c557190d7a144bdd3	f6dd3ec061db4e3987629fe6b26e5cce	SP	16	3
000229ec398224ef6ca0657da4fc703e	6489ae5e4333f3693df5ad4372dab6d3	MG	8	14
00024acbcdf0a6daa1e931b038114c75	d4eb9395c8c0431ee92fce09860c5a06	SP	6	6
00042b26cf59d7ce69dfabb4e55b4fd9	58dbd0b2d70206bf40e62cd34e84d795	SP	25	16
00048cc3ae777c65dbb7d2a0634bc1ea	816cbea969fe5b689b39cfc97a506742	MG	7	15

Here we can get all the details like customer state, id, order time_to_delivery, diff_estimated_delivery for each order.

c) **Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery**

```
SELECT
        s.seller_state,
        ROUND(SUM(ot.freight_value) / COUNT(ot.freight_value),2) AS
mean_freight_value,
        ROUND(SUM(DATEDIFF(o.order_delivered_customer_date,o.order_purc
hase_timestamp)) /
COUNT(DATEDIFF(o.order_delivered_customer_date,o.order_purchase_timesta
mp)),2) AS mean_delivery_time,
        ROUND(SUM(DATEDIFF(o.order_estimated_delivery_date,o.order_deliv
ered_customer_date)) /
COUNT(DATEDIFF(o.order_estimated_delivery_date,o.order_delivered_custome
r_date)),2) AS mean_diff_estimated_delivery

FROM order_items ot
JOIN sellers s
ON s.seller_id = ot.seller_id
JOIN orders o
ON ot.order_id = o.order_id
GROUP BY s.seller_state;
```

seller_state	mean_freight_value	mean_delivery_time	mean_diff_estimated_delivery
SP	18.45	12.21	11.30
MG	24.08	12.75	13.47
PR	22.72	13.32	14.23
SC	26.15	13.52	14.20
DF	20.57	12.44	13.18

Here we can see all the mean values of freight, delivery time, difference estimated delivery date.

d) **Sort the data to get the following:**
   i) **Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5**

HIGHEST FREIGHT AVERAGE:

```
SELECT
s.seller_state,
        ROUND(SUM(ot.freight_value) / COUNT(ot.freight_value),2) AS
avg_freight_value,
ROUND(SUM(DATEDIFF(o.order_delivered_customer_date,o.order_purc
hase_timestamp)) /
```

```
COUNT(DATEDIFF(o.order_delivered_customer_date,o.order_purchase_
timestamp)),2) AS avg_delivery_time,
ROUND(SUM(DATEDIFF(o.order_estimated_delivery_date,o.order_deliv
ered_customer_date)) /
COUNT(DATEDIFF(o.order_estimated_delivery_date,o.order_delivered_c
ustomer_date)),2) AS avg_diff_estimated_delivery

FROM order_items ot
JOIN sellers s
ON s.seller_id = ot.seller_id
JOIN orders o
ON ot.order_id = o.order_id
GROUP BY s.seller_state
ORDER BY avg_freight_value DESC
LIMIT 5;
```

seller_state	avg_freight_value	avg_delivery_time	avg_diff_estimated_delivery
RO	50.91	17.43	24.50
CE	46.38	17.90	13.38
PB	39.19	12.38	19.78
PI	36.94	13.73	15.00
AC	32.84	NULL	NULL

Here are the 5 states that have the highest average freight value, so state **'RO'** has highest average freight i.e **50.91** as compared to other states, which is followed by state **'CE'** with average freight of **46.38** and the rest.

LOWEST FREIGHT AVERAGE:

```
SELECT
s.seller_state,
        ROUND(SUM(ot.freight_value) / COUNT(ot.freight_value),2) AS
avg_freight_value,
ROUND(SUM(DATEDIFF(o.order_delivered_customer_date,o.order_purc
hase_timestamp)) /
COUNT(DATEDIFF(o.order_delivered_customer_date,o.order_purchase_
timestamp)),2) AS avg_delivery_time,
ROUND(SUM(DATEDIFF(o.order_estimated_delivery_date,o.order_deliv
ered_customer_date)) /
COUNT(DATEDIFF(o.order_estimated_delivery_date,o.order_delivered_c
ustomer_date)),2) AS avg_diff_estimated_delivery
```

```
FROM order_items ot
JOIN sellers s
ON s.seller_id = ot.seller_id
JOIN orders o
ON ot.order_id = o.order_id
GROUP BY s.seller_state
ORDER BY avg_freight_value
LIMIT 5;
```

seller_state	avg_freight_value	avg_delivery_time	avg_diff_estimated_delivery
SP	18.45	12.21	11.30
PA	19.39	13.25	11.38
RJ	19.47	11.94	12.51
DF	20.57	12.44	13.18
PR	22.72	13.32	14.23

Here are the 5 states that have the lowest average freight value, so state **'SP'** has lowest average freight i.e **18.45** as compared to other states, which is followed by state **'PA'** with average freight of **19.39** and the rest.

Conclusion: So we can say that the lowest freight average cost in Brazil is 18.45 which is in the state SP and the highest freight average cost is 50.91 which is in the state 'RO'.

ii) Top 5 states with highest/lowest average time to delivery

LOWEST TIME TO DELIVERY AVERAGE:

```
SELECT
s.seller_state,
ROUND(SUM(DATEDIFF(o.order_delivered_customer_date,o.order_purc
hase_timestamp)) /
COUNT(DATEDIFF(o.order_delivered_customer_date,o.order_purchase_
timestamp)),2) AS avg_delivery_time
FROM order_items ot
JOIN sellers s
ON s.seller_id = ot.seller_id
JOIN orders o
ON ot.order_id = o.order_id
WHERE order_status = 'delivered'
```

```
GROUP BY s.seller_state
ORDER BY avg_delivery_time
LIMIT 5;
```

seller_state	avg_delivery_time
RS	11.49
RJ	11.94
SP	12.21
MS	12.30
PB	12.38

Here are the 5 states that have the lowest average freight value, so state **'RS'** has lowest average freight i.e **11.49 days** as compared to other states, which is followed by state **'RJ'** with average freight of **11.94 days** and the rest.

HIGHEST TIME TO DELIVERY AVERAGE:

```
SELECT
s.seller_state,
ROUND(SUM(DATEDIFF(o.order_delivered_customer_date,o.order_purc
hase_timestamp)) /
COUNT(DATEDIFF(o.order_delivered_customer_date,o.order_purchase_
timestamp)),2) AS avg_delivery_time
FROM order_items ot
JOIN sellers s
ON s.seller_id = ot.seller_id
JOIN orders o
ON ot.order_id = o.order_id
GROUP BY s.seller_state
ORDER BY avg_delivery_time DESC
LIMIT 5;
```

Here are the 5 states that have the highest average time to delivery value, so state **'AM'** has highest average time to delivery i.e **48.00 days** as compared to other states, which is followed by state **'CE'** with average freight of **17.90 days** and the rest.

seller_state	avg_delivery_time
AM	48.00
CE	17.90
MA	17.65
RO	17.43
MT	14.69

iii) Top 5 states where delivery is really fast/ not so fast compared to estimated date

FAST Delivery:

```
SELECT
        seller_state,
   avg_diff_estimated_delivery,
   avg_delivery_time,
        avg_diff_estimated_delivery - avg_delivery_time AS early_by
FROM(
SELECT
        s.seller_state,

ROUND(SUM(DATEDIFF(o.order_delivered_customer_date,o.order_purc
hase_timestamp)) /
COUNT(DATEDIFF(o.order_delivered_customer_date,o.order_purchase_
timestamp)),2) AS avg_delivery_time,

ROUND(SUM(DATEDIFF(o.order_estimated_delivery_date,o.order_deliv
ered_customer_date)) /
COUNT(DATEDIFF(o.order_estimated_delivery_date,o.order_delivered_c
ustomer_date)),2) AS avg_diff_estimated_delivery
FROM order_items ot
JOIN sellers s
ON s.seller_id = ot.seller_id
JOIN orders o
ON ot.order_id = o.order_id
WHERE o.order_status = 'delivered'
GROUP BY s.seller_state
HAVING avg_diff_estimated_delivery > avg_delivery_time) AS t
ORDER BY early_by DESC
LIMIT 5;
```

	seller_state	avg_diff_estimated_delivery	avg_delivery_time	early_by
▶	PB	19.78	12.38	7.40
	RO	24.50	17.43	7.07
	MS	17.38	12.30	5.08
	RS	16.33	11.49	4.84
	SE	17.30	12.60	4.70

Here are the 5 states that deliver products before their estimated delivery average time, so state **'PB'** has the highest early_by value i.e **7.40 days** as compared to other states, which is followed by state **'RO'** with average freight of **7.07 days** and the rest.

5) **Payment type analysis:**
   a) **Month over Month count of orders for different payment types**

```
SELECT
        DISTINCT p.payment_type,
   YEAR(order_purchase_timestamp) as 'year',
        MONTHNAME(order_purchase_timestamp) AS 'month',
        COUNT(*)
        OVER(
                PARTITION BY p.payment_type,
                YEAR(order_purchase_timestamp),
                MONTH(order_purchase_timestamp)
                ORDER BY MONTH(order_purchase_timestamp)
        ) AS total_orders
FROM payments p
JOIN orders o
ON o.order_id= p.order_id
ORDER BY 'year';
```

   b) **Distribution of payment installments and count of orders**

```
SELECT
        payment_installments,
   total_orders,
   ROUND((total_orders/sum_orders) * 100,2)  AS orders_per_installments
FROM (
```

```
        SELECT payment_installments,COUNT(*) as total_orders,
SUM(COUNT(*)) OVER() as sum_orders
        FROM payments
        WHERE payment_installments > 0 AND (payment_type = 'credit_card'
OR payment_type = 'debit_card')
        GROUP BY payment_installments
) AS T
ORDER BY orders_per_installments DESC;
```

payment_installments	total_orders	orders_per_installments
1	26984	34.45
2	12413	15.85
3	10461	13.36
4	7098	9.06
10	5328	6.80
5	5239	6.69

Here we can see that most of the customers prefer **ONE month EMI** to purchase any product. As **34.45%** of orders are placed with one month installments followed by **TWO months EMI** which has around **15.85%** of orders placed. And followed by the rest.

Conclusion: We can see that around 63% of orders have EMIs 3 Months and below. So we can conclude that in Brazil most people prefer Maximum of 3 Months of EMI plans.

- **States that have less than 10% of total users**

```
SELECT
customer_state,
        total_customers,
        ROUND((total_customers/sum_cust) * 100,3) AS
customers_distributed_percentage
FROM
(
                SELECT
                        c.customer_state,
                        COUNT(*) as total_customers,
                        SUM(COUNT(*)) OVER() as sum_cust
                FROM orders o
                JOIN customers c
                ON o.customer_id = c.customer_id
                GROUP BY customer_state
) AS table1
```

```
ORDER BY customers_distributed_percentage;
```

customer_state	total_customers	customers_distributed_percentage
RR	46	0.046
AP	68	0.068
AC	81	0.081
AM	148	0.149
RO	253	0.254
TO	280	0.282

- **Payment Type Details:**

```
SELECT
        payment_type,
        COUNT(*) total_orders
FROM payments
GROUP BY payment_type;
```

payment_type	total_orders
credit_card	76795
UPI	19784
voucher	5775
debit_card	1529
not_defined	3

Conclusion: Most of the customers make a purchase using Credit Cards.

Q7) Actionable Insights:

From the insights above we can say that :

- **For states which have more delay in delivering the orders we can improve our delivery network and solve the problem.**
- **There are around 25 states which have less than 10% of the registered users on the platform. We can try to increase our customers in the following states which will help us increase sales in these regions.**
- **People are adapting to e-commerce as there are 99441 records in the order table, in a span of around 2.1 Years of time. Which we can improve moving forward**

Q8) Recommendations

- **From the insights above we can say that there is an upward trend in E-Commerce in Brazil. And we can increase our e-commerce business in the Brazil region.**
- **People from Brazil mostly purchase during Afternoon period. So we can show our best selling items, or we can show offers to users during the afternoon time period.**
- **The number of orders placed in Autumn Season are more as compared to any other Season in Brazil so we can increase our offers, staff, reduce the time to delivery during this period of time.**
- **Most of our customers use Credit Cards as their payment type, so we can tie up with any banks which will give users some offers and can also issue credit cards to eligible users.**
- **Around 63% of orders have EMIs 3 Months and below. So we can give no cost EMI options to the users in Brazil as most people prefer Maximum of 3 Months of EMI plans. Which can also help increase sales.**