Brazilia E-Commerce Analysis

Question and Syntax

```
--1.Retrieve the total number of customers percentage per state.Keep in Desc Order
Select
customer_state,
Count (customer_id) AS Total_Customer,
COUNT (customer_id) * 100 / SUM (Count (customer_id)) Over () As Percentage_Customer
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_customers_dataset]
GROUP by customer_state
ORDER by Percentage_Customer  Desc;
```

- As per data, SP are top states where most of the customers placing orders we cover 41 % of total spends.

Output

Results | Messages

	customer_state	Total_Customer	Percentage_Customer
1	SP	41746	41
2	RJ	12852	12
3	MG	11635	11
4	PR	5045	5
5	RS	5466	5
6	SC	3637	3
7	BA	3380	3
8	GO	2020	2
9	ES	2033	2
10	DF	2140	2
11	CE	1336	1
12	PE	1652	1
13	PB	536	0
14	PA	975	0
15	AC	81	0
16	SE	350	0
17	MA	747	0
18	TO	280	0
19	RO	253	0
20	AL	413	0
21	RR	46	0
22	MT	907	0
23	MS	715	0
24	AM	148	0

Question and Syntax

```
--2.List all unique cities in the customer_city column,Specify top 10.
Select
Distinct TOP 10 customer_city
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_customers_dataset]
```

Output

Results | Messages

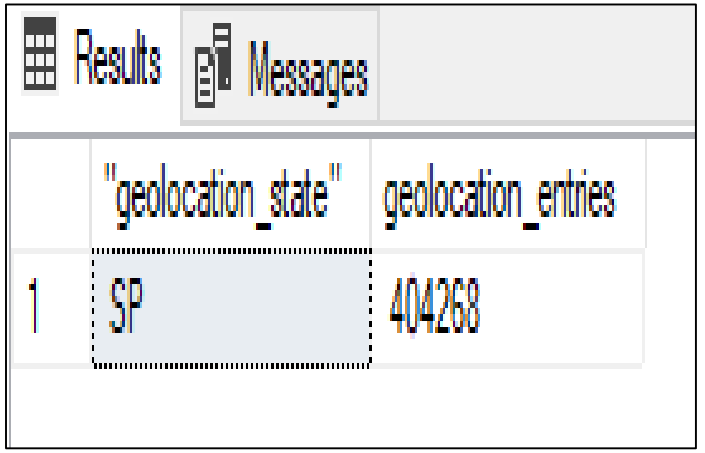
	customer_city
1	franca
2	sao bernardo do campo
3	sao paulo
4	mogi das cruzes
5	campinas
6	jaragua do sul
7	timoteo
8	curitiba
9	belo horizonte
10	montes claros

- As per data, These are Top 10 unique cities where we retrieve from data.

Question and Syntax

```
--3.Identify the states with the most geolocation entries.
SELECT
Top 1
["geolocation_state"],
COUNT (["geolocation_zip_code_prefix"]) As geolocation_entries
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_geolocation_dataset]
GROUP By ["geolocation_state"]
ORDER by geolocation_entries Desc;
```

Output



- As per geolocation data , we Notice SP is state in which most devices zip code fetch and received orders

Question and Syntax

```sql
--4.Find all cities where the geolocation_zip_code_prefix is shared by more than one city.
SELECT
TOP 5
["geolocation_city"],
["geolocation_zip_code_prefix"]
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_geolocation_dataset]
WHERE ["geolocation_zip_code_prefix"]
IN
(SELECT["geolocation_zip_code_prefix"]
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_geolocation_dataset]
GROUP BY ["geolocation_zip_code_prefix"]
HAVING COUNT(DISTINCT ["geolocation_city"]) > 1 );
```

Output

Results | Messages

	"geolocation_city"	"geolocation_zip_code_prefix"
1	aracariguama	"18147"
2	aracoiaba da serra	"18190"
3	ibiuna	"18150"
4	aracariguama	"18147"
5	sao roque	"18136"

- These are geolocation cities in which geolocation zip code prefix is more then one city where we received twice orders.

Question and Syntax

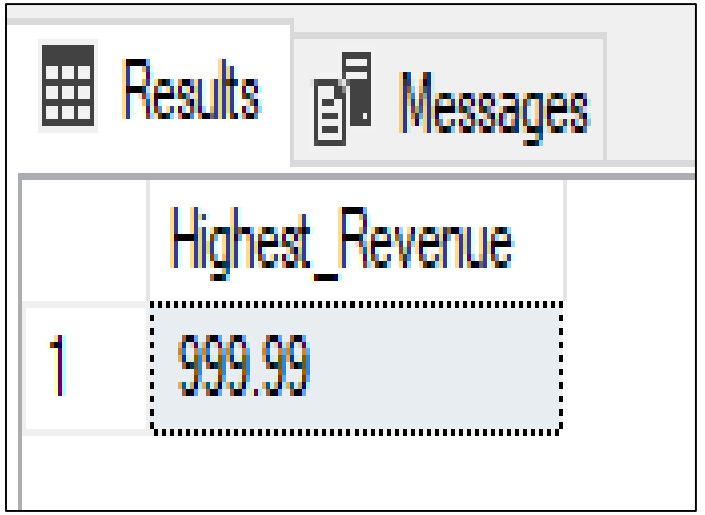
```
--5.Retrieve the highest revenue (price) from all orders.

SELECT

MAX ([price]) As Highest_Revenue

FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_order_items_dataset]
```

Output

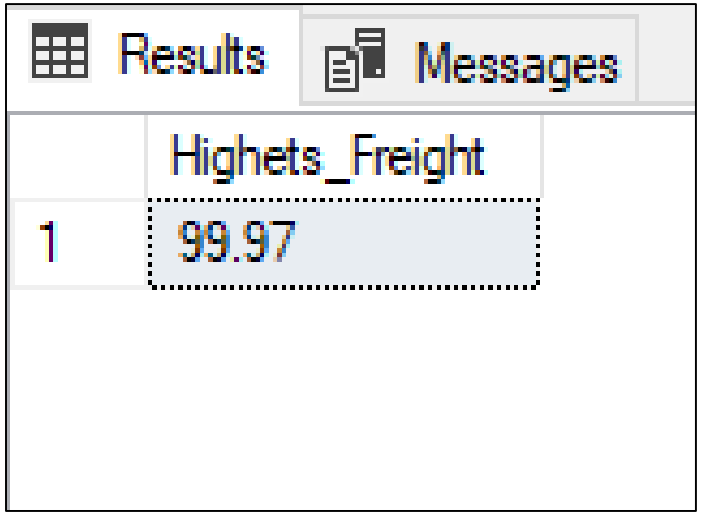


- The Highest revenue recorded from orders are 999.

Question and Syntax

```
--6.Calculate the Max freight_value for delivered orders.
SELECT
MAX (["freight_value"]) As Highets_Freight
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_order_items_dataset]
```

Output



- The Maximum freight value recorded in delivered orders is 99.97

Question and Syntax

```sql
--7.Count the number of orders made using each payment_type.
SELECT
payment_type,
Count (order_id) As No_of_Orders
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_order_payments_dataset]
GROUP by payment_type
ORDER BY No_of_Orders Desc;
```

Output

Results | Messages

	payment_type	No_of_Orders
1	credit_card	76795
2	boleto	19784
3	voucher	5775
4	debit_card	1529
5	not_defined	3

- As per checking Credit Cards contributed max amount in payment type , followed by boleto and voucher

Question and Syntax

```sql
--8.Find the top 5 total payment_value for orders with more than one installment.
SELECT
Top 5
[payment_installments],
SUM(CAST([payment_value] AS FLOAT)) AS Total_Payment_Value
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_order_payments_dataset]
WHERE [payment_installments] > 1
GROUP BY [payment_installments]
ORDER BY Total_Payment_Value DESC;
```

Output

Results | Messages

	payment_installments	Total_Payment_Value
1	10	2211577.34
2	2	1579283.03000001
3	3	1491103.8
4	8	1313423.34
5	4	1163907.61

- As per checking installment data , we retrieve for big amount payment value we have done more one installment

.

Question and Syntax

```sql
--9.Retrieve the list of top 5  orders that were delivered after their estimated delivery date.
SELECT TOP 5
["order_id"],
["order_estimated_delivery_date"],
["order_delivered_customer_date"],
DATEDIFF(DAY, ["order_estimated_delivery_date"], ["order_delivered_customer_date"]) AS Delay_In_Days
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_orders_dataset]
WHERE ["order_delivered_customer_date"] > ["order_estimated_delivery_date"]
ORDER BY Delay_In_Days DESC;
```

Question and Syntax

Results | Messages

	"order_id"	"order_estimated_delivery_date"	"order_delivered_customer_date"	Delay_In_Days
1	"1b3190b2dfa9d789e1f14c05b647a14a"	2018-03-15 00:00:00	2018-09-19 23:24:07	188
2	ca07593549f1816d26a572e06dc1eab6	2017-03-22 00:00:00	2017-09-19 14:36:39	181
3	"47b40429ed8cce3aee9199792275433f"	2018-01-19 00:00:00	2018-07-13 20:51:31	175
4	"2fe324febf907e3ea3f2aa9650869fa5"	2017-04-05 00:00:00	2017-09-19 17:00:07	167
5	"285ab9426d6982034523a855f55a885e"	2017-04-06 00:00:00	2017-09-19 14:00:04	166

- As per checking estimated delivery date highlight significant logistical inefficiencies, requiring focused investigation and targeted interventions to enhance delivery performance

Question and Syntax

```
--10.Calculate the average delivery time for orders

SELECT
AVG(DATEDIFF(DAY, ["order_purchase_timestamp"], ["order_delivered_customer_date"])) AS Average_Delivery_Time
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_orders_dataset]
WHERE ["order_delivered_customer_date"] IS NOT NULL
AND ["order_purchase_timestamp"] IS NOT NULL;
```

Output

Results | Messages

	Average_Delivery_Time
1	-1271

- As per checking delivery timer orders we found average delivery time is (-1,271)

Question and Syntax

```
--11.Find the top 10 total revenue (price) generated by each seller by joining Order_Item and Seller_Dataset
SELECT
Top 10
[Brazilian E-Commerce  Dataset ].[dbo].[olist_sellers_dataset].["seller_id"],
ROUND (SUM(CAST([Brazilian E-Commerce  Dataset ].[dbo].[olist_order_items_dataset].["price"] AS FLOAT)),2) AS "Total_Revenue"
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_sellers_dataset]
LEFT JOIN [Brazilian E-Commerce  Dataset ].[dbo].[olist_order_items_dataset]
ON [Brazilian E-Commerce  Dataset ].[dbo].[olist_sellers_dataset].["seller_id"] = [Brazilian E-Commerce  Dataset ].[dbo].[olist_order_items_dataset].["seller_id"]
GROUP BY [Brazilian E-Commerce  Dataset ].[dbo].[olist_sellers_dataset].["seller_id"]
ORDER BY "Total_Revenue" Desc;
```

Output

	"seller_id"	Total_Revenue
1	"4869f7a5dfa277a7dca6462dcf3b52b2"	229472.63
2	"53243585a1d6dc2643021fd1853d8905"	222776.05
3	"4a3ca9315b744ce9f8e9374361493884"	200472.92
4	fa1c13f2614d7b5c4749cbc52fecda94	194042.03
5	"7c67e1448b00f6e969d365cea6b010ab"	187923.89
6	"7e93a43ef30c4f03f38b393420bc753a"	176431.87
7	da8622b14eb17ae2831f4ac5b9dab84a	160236.57
8	"7a67c85e85bb2ce8582c35f2203ad736"	141745.53
9	"1025f0e2d44d7041d6cf58b6550e0bfa"	138968.55
10	"955fee9216a65b617aa5c0531780ce60"	135171.7

- These are top 10 sellers who generating highest revenue .

Question and Syntax

```
--12. Find top 10 Seller state, keep order in Desc order
SELECT
["seller_state"],
COUNT (["seller_id"]) As No_of_sellers
FROM [Brazilian E-Commerce  Dataset ].[dbo].[olist_sellers_dataset]
Group by ["seller_state"]
ORDER by No_of_sellers Desc;
```

Output

	"seller_state"	No_of_sellers
1	SP	1849
2	PR	349
3	MG	244
4	SC	190
5	RJ	170
6	RS	128
7	GO	40
8	DF	30
9	ES	23
10	BA	19

- These are top 10 state which we received highest No of sellers .