Question 1.1
Data type of columns in a table

Query:
```
SELECT *
FROM CustomersSales.INFORMATION_SCHEMA.COLUMNS
```

Output:

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH | PREVIEW

Ro	table_catalog	table_schema	table_name	column_name	ordinal_p	is_nullable	data_type	is_generated
1	sharat-target-business-case	CustomersS...	order_items	order_id	1	YES	STRING	NEVER
2	sharat-target-business-case	CustomersS...	order_items	order_item_id	2	YES	INT64	NEVER
3	sharat-target-business-case	CustomersS...	order_items	product_id	3	YES	STRING	NEVER
4	sharat-target-business-case	CustomersS...	order_items	seller_id	4	YES	STRING	NEVER
5	sharat-target-business-case	CustomersS...	order_items	shipping_limit_date	5	YES	TIMESTAMP	NEVER
6	sharat-target-business-case	CustomersS...	order_items	price	6	YES	FLOAT64	NEVER
7	sharat-target-business-case	CustomersS...	order_items	freight_value	7	YES	FLOAT64	NEVER
8	sharat-target-business-case	CustomersS...	sellers	seller_id	1	YES	STRING	NEVER
9	sharat-target-business-case	CustomersS...	sellers	seller_zip_code_prefix	2	YES	INT64	NEVER
10	sharat-target-business-case	CustomersS...	sellers	seller_city	3	YES	STRING	NEVER
11	sharat-target-business-case	CustomersS...	sellers	seller_state	4	YES	STRING	NEVER
12	sharat-target-business-case	CustomersS...	geolocation	geolocation_zip_code_pre...	1	YES	INT64	NEVER
13	sharat-target-business-case	CustomersS...	geolocation	geolocation_lat	2	YES	FLOAT64	NEVER

Insight's:
- Data Type matches with Data inside each Table and their Columns

Question 1.2

Time period for which the data is given

Query:

```
SELECT min(order_purchase_timestamp) as timeperiod_start , max(order_purchase_timestamp) as timeperiod_end
FROM `sharat-target-business-case.CustomersSales.orders` ;
```

Output:

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

Insight's:

Data given is in the time period between 4th Sep 2016 & 17th Oct 2018

Question 1.3

Cities and States of customers ordered during the given period

Query:

```
select distinct customer_state , customer_city,
from `CustomersSales.orders` A
inner join `CustomersSales.customers` B
on A.customer_id = B.customer_id
order by 1,2;
```

Output:

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GR

Row	customer_state	customer_city
1	AC	brasileia
2	AC	cruzeiro do sul
3	AC	epitaciolandia
4	AC	manoel urbano
5	AC	porto acre
6	AC	rio branco
7	AC	senador guiomard
8	AC	xapuri
9	AL	agua branca
10	AL	anadia

Insight's:

Customer from 4310 Different Cities of 27 States have ordered in the given time period

Recommendations:

Although Target has coverage in all states[26 Districts and 1 Federal District]…it has possibility to expand to furthermore cities in each state as their total 5564 Municipalities in Brazil

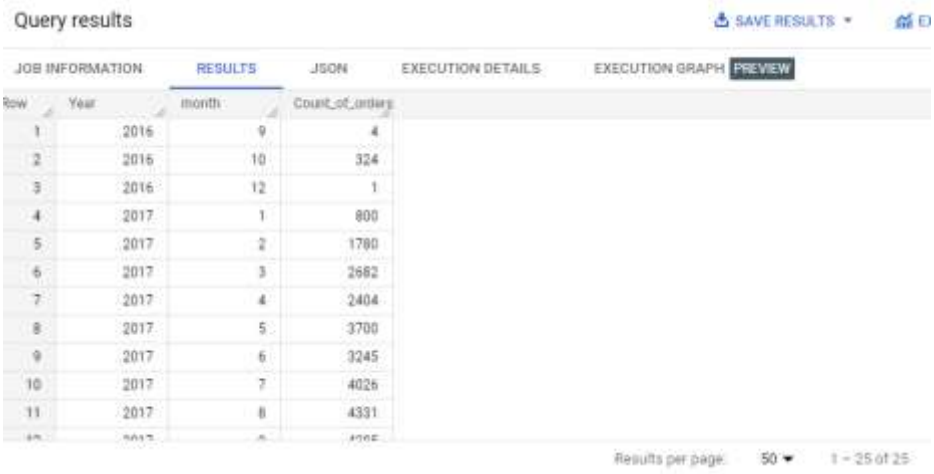
Question 2.1

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?

Query:

```
select extract(year from order_purchase_timestamp) as Year,
       extract(month from order_purchase_timestamp) as month,
       count(order_id) as Count_of_orders
from `CustomersSales.orders`
group by 1,2
order by 1,2;
```

Output:



Insight's:

- Sales Qty has an increasing trending in the 2017 year and particularly highest sales in month of November may be due to "Double Eleven - China's commemoration of Single's Day" which is considered one of the main day for sales
- There is growing trend in Ecommerce in brazil ,as can be seen from order Quantity which have increased from 2017 to 2018

Recommendations:

Order Qty has increased by nearly 50% from 2017 to 2018, so target needs to be ready for Future Growth by Stepping up Inventory, Logistics, Human Resources & Infrastructures

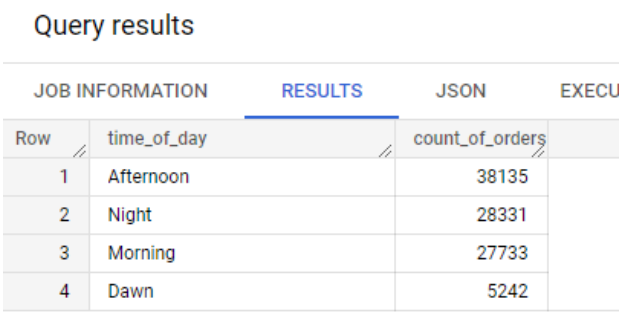
Question 2.2

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

Query:

```
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"
   end as time_of_day,
   count(distinct order_id) as count_of_orders
from `CustomersSales.orders`
group by 1
order by 2 desc;
```

Output:



Query results

JOB INFORMATION | RESULTS | JSON | EXECU

Row	time_of_day	count_of_orders
1	Afternoon	38135
2	Night	28331
3	Morning	27733
4	Dawn	5242

Insight's

Brazilians most preferred part of Day for Purchase is Afternoon i.e., between 13 ~ 18 Hrs
Next preferred time is Morning & Night [Both almost Equally Preferred]
Least Preferred time is Dawn i.e., between 0 ~ 6 Hrs

Recommendation:

*Target can use Flash deals in Dawn time to increase Sales during in Dawn also
*Target can Maximize its advertising time in Afternoon to reach to maximum customers

Question 3.1

Get month on month orders by states

Query:

```
select A.State,A.month,(order_count -
 (lag(order_count,1) over (partition by A.State order by A.month))) as State_MOM_orders
from
(select customer_state as State,extract(month from order_purchase_timestamp) as month,count(order_id)
as order_count
from `CustomersSales.orders` lefty
left join `CustomersSales.customers` righty
on lefty.customer_id = righty.customer_id
group by 1,2) A
order by 1,2;
```

Output:

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXE

Row	State	month	State_MOM_ord
1	AC	1	null
2	AC	2	-2
3	AC	3	-2
4	AC	4	5
5	AC	5	1
6	AC	6	-3
7	AC	7	2
8	AC	8	-2
9	AC	9	-2
10	AC	10	1
11	AC	11	-1

Insight's :

Peaks of order count are in the month : November, May, March
Valleys of Order count are in the Month : December

Recommendation:

* Target needs Increase Inventory, Logistics, Human Resources & Infrastructures management for the peak month of November, January, March, May
* Use Valley month to restructure Logistics, HR , Finances and do PDCA to perform better in next peak month

Question 3.2

Distribution of customers across the states in Brazil

Query:

```
select State,customer_count,round((customer_count/sum(customer_count) over())*100,2) as customer_distribution_percentage
from
(select customer_state as State,count(distinct lefty.customer_id) as customer_count
from `CustomersSales.orders` lefty
left join `CustomersSales.customers` righty
on lefty.customer_id = righty.customer_id
group by 1)
order by 2 desc;
```

Output:

Row	State	customer_count	customer_distrit
1	SP	41746	41.98
2	RJ	12852	12.92
3	MG	11635	11.7
4	RS	5466	5.5
5	PR	5045	5.07
6	SC	3637	3.66
7	BA	3380	3.4
8	DF	2140	2.15
9	ES	2033	2.04
10	GO	2020	2.03

Insight's :

• Target has highest customer base of 42% in SP , followed by 13% In RJ and 11.7% in MG

•Target has least customer base of 0.05% in RR, followed by 0.07% in AP and 0.08% in AC

Recommendation:

- Target has to focus 20% [like more marketing, logistics] in SP,RJ & MG kind of states to increase furthermore Sales from these states as they are many prospective customers
- Rest 80% Focus to be kept on Least customer base states like RR,AP,AC by optimizing Freight Costs & reducing delivery time and giving exciting deals for customers in these states

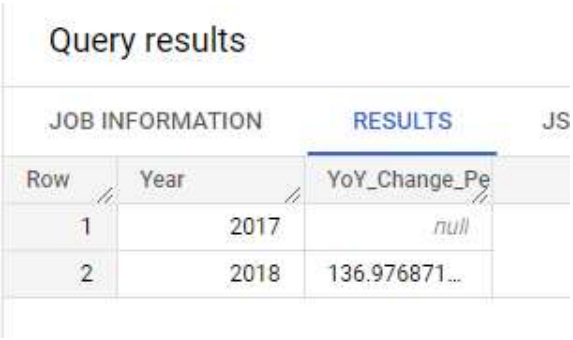
Question 4.1

Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use “payment_value” column in payments table

Query:

```
select Year,
    ((totalpaymentinyear -
 lag(totalpaymentinyear,1) over (order by Year))/lag(totalpaymentinyear,1) over (order by Year)) *100 as YoY
_Change_Percent
from
(select extract(year from order_purchase_timestamp) as Year,
     sum(payment_value) as totalpaymentinyear
from `CustomersSales.orders` lefty
inner join `CustomersSales.payments` righty
on lefty.order_id = righty.order_id
where extract(month from order_purchase_timestamp) < 9 and extract(year from order_purchase_timesta
mp) in (2017,2018)
group by 1)
order by 1;
```

Output:



Insight's

- Target has got 136% Growth in sales value from 2017 to 2018

Recommendations

Sales Growth is exponential in Brazil , so target needs to be ready for Future Growth by better managing of all operations by stepping up Inventory, Logistics, Human Resources & Infrastructures

Question 4.2

Mean & Sum of price and freight value by customer state

Query:

```
select customer_state as State,
      round(avg(price)) as price_avg,
      round(sum(price)) as price_sum,
      round(avg(freight_value))as freight_avg,
      round(sum(freight_value)) as freight_sum,
      round(sum(freight_value)/sum(price),3) as freight_to_price_ratio,
from `CustomersSales.order_items` lefty
left join `CustomersSales.orders` righty
on lefty.order_id = righty.order_id
inner join `CustomersSales.customers` righty2
on righty.customer_id = righty2.customer_id
group by 1
order by 6 desc
```

Output:

Row	State	price_avg	price_sum	freight_avg	freight_sum	freight_to_price_
1	RR	151.0	7829.0	43.0	2235.0	0.285
2	MA	145.0	119648.0	38.0	31524.0	0.263
3	RO	166.0	46141.0	41.0	11417.0	0.247
4	AM	135.0	22357.0	33.0	5479.0	0.245
5	PI	160.0	86914.0	39.0	21218.0	0.244
6	SE	153.0	58921.0	37.0	14111.0	0.239
7	TO	158.0	49622.0	37.0	11733.0	0.236
8	AC	174.0	15983.0	40.0	3687.0	0.231
9	RN	157.0	83035.0	36.0	18860.0	0.227
10	PE	146.0	262788.0	33.0	59450.0	0.226

Insight's

- Cost of freight to Cost of Price is very high for State - RR ,
- Cost of freight to cost of Price is least for State – SP
- It Seems one of the reason for least freight cost for State-SP due to channelized Logistics considering high Qty of orders received from this SP state . Similarly due to low Qty of orders from state – RR , may be Freight cost is not much optimised

Recommendations:

* Target must optimize the Freight Cost for States like RR, so that overall cost burden will be reduced on Customers…there by making Target a best option to Buy products

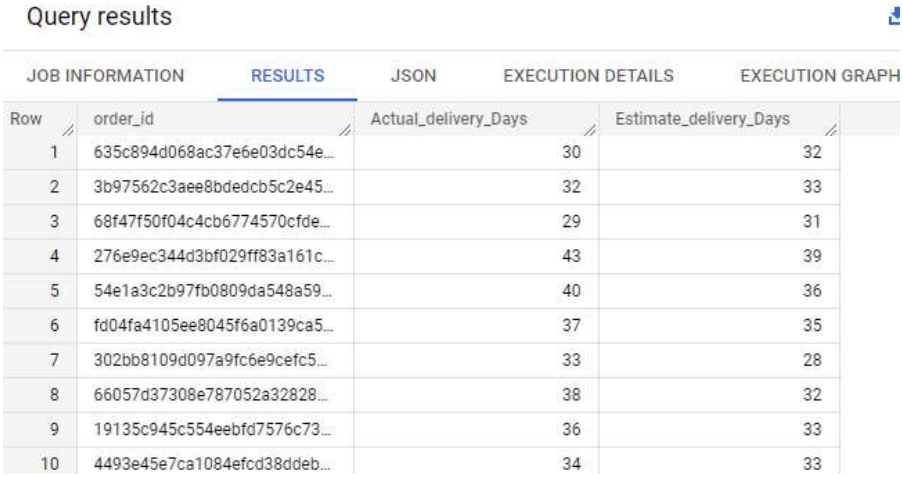
Question 5.1

Calculate days between purchasing, delivering and estimated delivery

Query:

```
select order_id, date_diff(order_delivered_customer_date,order_purchase_timestamp,day)
as Actual_delivery_Days,
    date_diff(order_estimated_delivery_date,order_purchase_timestamp,day) as
Estimate_delivery_Days
from `CustomersSales.orders`
where order_status = "delivered"
```

Output:



Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	order_id	Actual_delivery_Days	Estimate_delivery_Days
1	635c894d068ac37e6e03dc54e...	30	32
2	3b97562c3aee8bdedcb5c2e45...	32	33
3	68f47f50f04c4cb6774570cfde...	29	31
4	276e9ec344d3bf029ff83a161c...	43	39
5	54e1a3c2b97fb0809da548a59...	40	36
6	fd04fa4105ee8045f6a0139ca5...	37	35
7	302bb8109d097a9fc6e9cefc5...	33	28
8	66057d37308e787052a32828...	38	32
9	19135c945c554eebfd7576c73...	36	33
10	4493e45e7ca1084efcd38ddeb...	34	33

Insight's

Most of the orders were delivered quite before estimated delivery date

Recommendations:

Target can observe the Actual Delivery Days and revise their Estimated delivery date state wise & city wise, so that it gives additional factor for customer to buy from Target due to less estimated delivery date

Question 5.2

Find time_to_delivery & diff_estimated_delivery. Formula for the same given below:

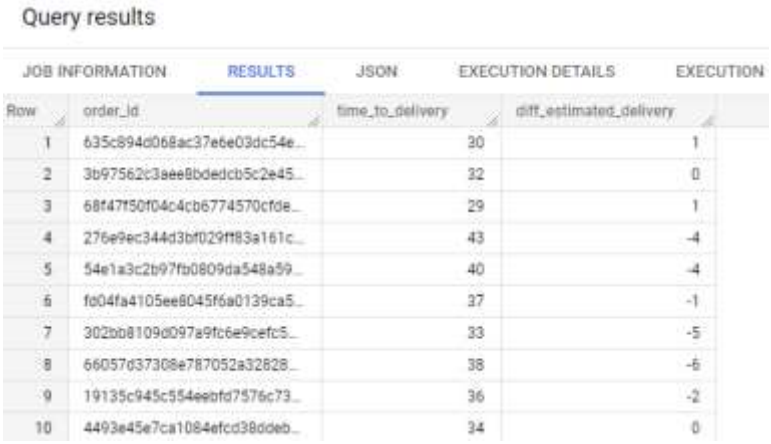
time_to_delivery = order_delivered_customer_date-order_purchase_timestamp

diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date

Query:

```
select order_id, date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as time_to_delivery,
    date_diff(order_estimated_delivery_date,order_delivered_customer_date,day) as diff_estimated_delivery
from `CustomersSales.orders`
where order_status = "delivered";
```

Output:



Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION

Row	order_id	time_to_delivery	diff_estimated_delivery
1	635c894d068ac37e6e03dc54e...	30	1
2	3b97562c3aee8bdedcb5c2e45...	32	0
3	68f47f50f04c4cb6774570cfde...	29	1
4	276e9ec344d3bf029ff83a161c...	43	-4
5	54e1a3c2b97fb0809da548a59...	40	-4
6	fe04fa4105ee8045f6a0139ca5...	37	-1
7	302bb8109d097a9fc6e9cefc5...	33	-5
8	66057d37308e787052a32828...	38	-6
9	19135c945c554eebfd7576c73...	36	-2
10	4493e45e7ca1084efcd38ddeb...	34	0

Insight's

- Mean Time taken for delivery of orders is 12 days
- Deliveries are done 11 days before Estimated delivery date

Recommendations:

Target can observe trend in the Actual Delivery Days and revise their Estimated delivery date state wise & city wise, so that it gives additional factor for customer to buy from Target due to less estimated delivery date

Question 5.3

Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_deliver

Query:

```
select customer_state as State,
    round(avg(freight_value))as freight_avg,
    round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
Mean_time_to_delivery,
    round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day))) as
Mean_diff_estimated_delivery
from `CustomersSales.orders` lefty
left join `CustomersSales.order_items` righty
on lefty.order_id = righty.order_id
inner join `CustomersSales.customers` righty2
on lefty.customer_id = righty2.customer_id
where order_status = "delivered"
group by 1
order by 1;
```

Output:

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTIO

Row	State	freight_avg	Mean_time_to_d	Mean_diff_estim
1	AC	40.0	20.0	20.0
2	AL	36.0	24.0	8.0
3	AM	33.0	26.0	19.0
4	AP	34.0	28.0	17.0
5	BA	26.0	19.0	10.0
6	CE	33.0	21.0	10.0
7	DF	21.0	13.0	11.0
8	ES	22.0	15.0	10.0
9	GO	23.0	15.0	11.0
10	MA	38.0	21.0	9.0

Insight's :

•Best Parameter are for State-SP : it has Least Average freight value of 15$ & fastest average time to delivery of 8 days

•Worst Parameter are for State-RR : It has Highest Average Freight Value of 43$ & Slowest Average time to delivery of 28 Days

Recommendation:

* Target must optimize the Logistics for States like RR, so that overall freight performance and delivery time will Improve, thereby making Target a best option to Buy products

Question 5.5

Top 5 states with highest/~~lowest~~ average freight value - sort in desc~~/asc~~ limit 5

Query:

```
with freight_deliverytime_table as
(select customer_state as State,
     round(avg(freight_value))as freight_avg,
     round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
Mean_time_to_delivery,
     round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day))) as
Mean_diff_estimated_delivery
from `CustomersSales.orders` lefty
left join `CustomersSales.order_items` righty
on lefty.order_id = righty.order_id
inner join `CustomersSales.customers` righty2
on lefty.customer_id = righty2.customer_id
where order_status = "delivered"
group by 1)
select State,freight_avg
from freight_deliverytime_table
order by 2 desc
limit 5
```

Output:

JOB INFORMATION | RESULTS | JSON | EXECU

Row	State	freight_avg
1	RR	43.0
2	PB	43.0
3	RO	41.0
4	AC	40.0
5	PI	39.0

Insight's

State : RR & PB have highest Avg Freight Value of 43
State : SP has lowest Avg Freight value of 15

Recommendations:

* Target must optimize the Freight Cost for States like RR, so that overall cost burden will be reduced on Customers…there making Target a best option to Buy products

Question 5.6

Top 5 states with highest~~/lowest~~ average time to delivery

Query:

```
with freight_deliverytime_table as
(select customer_state as State,
      round(avg(freight_value))as freight_avg,
      round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
Mean_time_to_delivery,
      round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day))) as
Mean_diff_estimated_delivery
from `CustomersSales.orders` lefty
left join `CustomersSales.order_items` righty
on lefty.order_id = righty.order_id
inner join `CustomersSales.customers` righty2
on lefty.customer_id = righty2.customer_id
where order_status = "delivered"
group by 1)
select State, Mean_time_to_delivery
from freight_deliverytime_table
order by 2 desc
limit 5;
```

Output:

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION D

Row	State	Mean_time_to_delivery
1	RR	28.0
2	AP	28.0
3	AM	26.0
4	AL	24.0
5	PA	23.0

Insight's

State – RR has highest Average time to deliver i.e., 28 Days
State – SP has lowest Average time to deliver i.e., 8 Days

Recommendation:

* Target must optimize the Logistics for States like RR, so that overall delivery time will Improve, thereby making Target a best option to Buy products

Question 5.7

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

Query:

```
with freight_deliverytime_table as
(select customer_state as State,
      round(avg(freight_value))as freight_avg,
      round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
Mean_time_to_delivery,
      round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day))) as
Mean_diff_estimated_delivery
from `CustomersSales.orders` lefty
left join `CustomersSales.order_items` righty
on lefty.order_id = righty.order_id
inner join `CustomersSales.customers` righty2
on lefty.customer_id = righty2.customer_id
where order_status = "delivered"
group by 1)
select State,Mean_diff_estimated_delivery
from freight_deliverytime_table
order by 2 desc
limit 5;
```

Output:

Row	State	Mean_diff_estimated_delivery
1	AC	20.0
2	RO	19.0
3	AM	19.0
4	AP	17.0
5	RR	17.0

Insight's

•On an Average , Orders are delivered by the Target quite before their Estimated Delivery Date

•In State- AC Delivery was done on Average 20 Days before estimated date and In State-AL, delivery was done on Average 8 days before

Recommendations:

Target can observe trend in the Actual Delivery Days and revise their Estimated delivery date state wise & city wise, so that it gives additional factor for customer to buy from Target due to less estimated delivery date

Question 6.1

Month over Month count of orders for different payment types

Query:

```
with payment_type_analysis as
(select payment_type ,extract(month from order_purchase_timestamp) as month,
      count(lefty.order_id) as Payments_count
from `CustomersSales.orders`lefty
left join `CustomersSales.payments` righty
on lefty.order_id = righty.order_id
group by 1,2)
select payment_type,month,Payments_count,Payments_count - lag(Payments_count) over (partition by
payment_type order by month) as MoM_payment_type_count
from payment_type_analysis
order by 1,2;
```

Output:

Row	payment_type	month	Payments_count	MoM_payment_type_count
1	*null*	9	1	*null*
2	UPI	1	1715	*null*
3	UPI	2	1723	8
4	UPI	3	1942	219
5	UPI	4	1783	-159
6	UPI	5	2035	252
7	UPI	6	1807	-228
8	UPI	7	2074	267
9	UPI	8	2077	3
10	UPI	9	903	-1174

Insight's :

- Nearly 74% Payments are done using Credit Card
- Next to Credit Card, 19% of payments done using UPI

❖Peaks of Payments are in the month : November, January, March, May

Recommendation:

* Target needs plan the payment infrastructure to handle the payment traffic in those months

Question 6.2

Count of orders based on the no. of payment installment

Query:

```
select No_of_installments,count(order_id) as Payments_count
from
(select distinct order_id,max(payment_installments) over (partition by order_id) as
No_of_installments
from `CustomersSales.payments`)
group by 1
order by 1
```

Output:

Row	No_of_installme	Payments_count
1	0	2
2	1	48268
3	2	12363
4	3	10429
5	4	7070
6	5	5227
7	6	3908
8	7	1622
9	8	4251
10	9	644

Insight's :

•Payments were done in Instalments maximum up to 24
• Nearly 48% payments were done in Single Instalment
•Nearly 99% payments were done with Maximum Instalment up to 10

Recommendations:

• Target can give No Cost EMI option for 12 Month & above to attract more customer and increase the sales for Expensive products

Actionable Insights

Targets Performance in Brazil:

- Target has Customers from 4310 Different Cities of 27 States with highest customer base of 42% in SP & least customer base of 0.05% in RR
- There is growing trend in Ecommerce in Brazil –Target's Order Quantity have increased from 2017 to 2018 by Approx. 50% and by 136% in orders Cost value
- Target delivers its products before their estimated delivery date and its delivery & Freight performance is best in State-SP and worst in state - RR

Brazilians Behaviour in E-Commerce Market:

- Brazilians most preferred part of Day for shopping is Afternoon & least Preferred part is Dawn.
- Brazilians Most preferred payment mode is Credit Card
- Brazilians Sales are high in the months of November, January, March, May

Recommendations

Targets Performance in Brazil:

- Target must find possibility to expand to its reach to furthermore as there are (5564-4310=)1254 untapped cities left in Brazil
- Sales Growth is exponential in Brazil ,So target needs to be ready for Future Growth by better managing of all operations by stepping up Inventory, Logistics, Human Resources & Infrastructures , also particularly in peak months like November, January, March, May
- Target must continue marketing & logistics activities in high selling states like SP,RJ & MG kind as they will be many future prospective customers considering past growth
- Target must increase Focus on Least customer base states like RR,AP,AC by optimizing Freight Costs & reducing delivery time and giving exciting deals for customers in these states to increase base in these stated
- Target can work on increasing accuracy of their estimated delivery date as they mostly deliver before estimated dates