E-COMMERCE DATABASE PROJECT

DATA INTRODUCTION:

This Project consist of one Table of E-commerce website for 5 months October 2020-February 2021.

The table consisted of details of users using the website .

The abbreviations used are below:

UTC-Universal Time Coordinated

IST – Indian Time Standard

PROBLEM STATEMENT:

Despite the huge usage of ecommerce website a section of the society still prefers offline stores. Why?

Objective:

- To analyze the usage of ecommerce website.

- To provide suggestions for improving the business on basis of Findings.

The data for the analysis has been attached below:

events.csv.zip

The data in CSV file suggested that it had more than 8 Lakhs records or rows and nine fields or columns namely .

My Work File:

Final Project SQL script.sql

MY WORK ON DATA :

Firstly I disable safe update mode.

```
SET SQL_SAFE_UPDATES =0;
```

Then I removing UTC from event_time.

```
UPDATE events
SET event_time = trim('UTC' from event_time);
```

After removing UTC from event time then, I updated UTC to IST.

```
UPDATE events
SET event_time = convert_tz(event_time,'+00:00','+05:30');
```

After that I updated price datatype into float.

```
alter table events
 modify price float(10,2);
```

After modifying data I started to analyse by listing various factors affecting business and started to write query.

- **Below are the factors which are affecting business basically:**

1. Month of Sales

2.Top Time of Visit

3.Top brands by Sale

4.Demand for Items

5.Frequency of Purchase

6.Actual Time purchased

Q1(First_Query)

Month of Sales:

Here students are required to find months in decreasing order of sales and also highest sales in which month and year and lowest sales in which month and year with the help of SQL query.

Query:

- Sales by Desc.

```
select year(event_time) as year,month(event_time) as month,
sum(price) as sales
from events group by year(event_time),month(event_time)
order by sales desc;
```

Output:

year	month	sales
2021	1	33508355.90
2021	2	29372958.90
2020	11	23188583.16
2020	12	21567050.10
2020	10	18326437.54
2020	9	3023831.60
2021	3	532570.51

Explanation:

Here we need to calculate heighst sales monthly and yearly

In descending order.

So we need to extract year and month from event_time and as

“Year” and “month” respectively and then we need to take **sum of price as “sales”**

From events .

And then we need to make group of year(event_time) and month(event_time)

So we use **Group by clause here**.

And then we need firstly descending order of “sales” and highest sales.

So we need to use **Order by clause here.**

Q1(Second_Query)

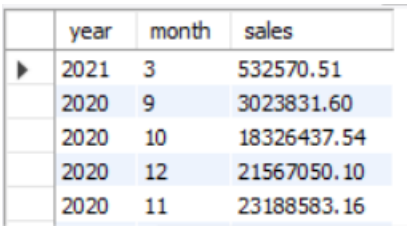
Here students are required to find months in decreasing order of sales and also highest sales in which month and year and lowest sales in which month and year with the help of SQL query.

- Sales by asc.

Query:

```
select year(event_time) as year,month(event_time) as month,
sum(price) as sales
from events group by year(event_time),month(event_time)
order by sales asc;
```

Output:



year	month	sales
2021	3	532570.51
2020	9	3023831.60
2020	10	18326437.54
2020	12	21567050.10
2020	11	23188583.16

Explanation:

Here we need to calculate lowest sales monthly and yearly

In ascending order.

So we need to extract year and month from event_time and as

“Year” and “month” respectively and then we need to take **sum of price as “sales”**

From events .

And then we need to make group of year(event_time) and month(event_time)

So we use **Group by clause here**.

And then we need firstly Ascending order of “sales” and Lowest sales.

So we need to use **Order by clause here.**

Q2(First_Query)

Top Time of Visit:

Here Students are required to calculate the time for viewing, adding

to cart and Purchase time with help of SQL query and need to

confirm if mostly viewed and added to cart time is similar to Purchase time or not.

select * from events

Query:

```
with CTE as
(select *,case
WHEN TIME (event_time) between '00:00:00' and '04:00:00' then 'late_night'
WHEN TIME (event_time) between '04:00:00' and '08:00:00' then 'early_morning'
WHEN TIME (event_time) between '08:00:00' and '12:00:00' then 'morning'
WHEN TIME (event_time) between '12:00:00' and '16:00:00' then 'afternoon'
WHEN TIME (event_time) between '16:00:00' and '20:00:00' then 'evening'
WHEN TIME (event_time) between '20:00:00' and '24:00:00' then 'night'
end as event_time_catagory
     from events)
select * FROM CTE;
```

Output:

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

event_time	event_type	product_id	category_id	category_code	brand	price	user_id	user_session	event_time_catagory
2020-09-24 22:57:06.000000	view	1996170	2.14442E+18	electronics.telephone		31.90	1.51592E+18	LJuJVLEjPT	night
2020-09-24 22:57:26.000000	view	139905	2.14442E+18	computers.components.cooler	zalman	17.16	1.51592E+18	tdicluNnRY	night
2020-09-24 22:57:27.000000	view	215454	2.14442E+18			9.81	1.51592E+18	4TMArHtXQy	night
2020-09-24 22:57:33.000000	view	635807	2.14442E+18	computers.peripherals.printer	pantum	113.81	1.51592E+18	aGFYrNgC08	night
2020-09-24 22:57:36.000000	view	3658723	2.14442E+18		cameronsino	15.87	1.51592E+18	aa4mmk0kwQ	night
2020-09-24 22:57:59.000000	view	664325	2.14442E+18	construction.tools.saw	carver	52.33	1.51592E+18	vnkdP81DDW	night
2020-09-24 22:58:23.000000	view	3791349	2.14442E+18	computers.desktop		215.41	1.51592E+18	J1t6sIYXiV	night
2020-09-24 22:58:24.000000	view	716611	2.14442E+18	computers.network.router	d-link	53.14	1.51592E+18	kVBeYDPcBw	night

Result 75

Explanation:

Firstly we need to divided event_time in to time category

So we make time category temporary column as event_time_category.

We make category as follows.

12am-4am →Late Night

4am -8am→Early Morning

8am-12pm→ Morning

12pm - 4pm →Afternoon

4pm-8pm →Evening.

8pm →12pm Night.

We use CTE with statements for making common table expression.

So select all from CTE.

Our new column “event_time_catgory”is ready.

So now we can easily analyse data.

Q2(Second_Query)

Top Time of Visit:

Here Students are required to calculate the time for viewing, adding

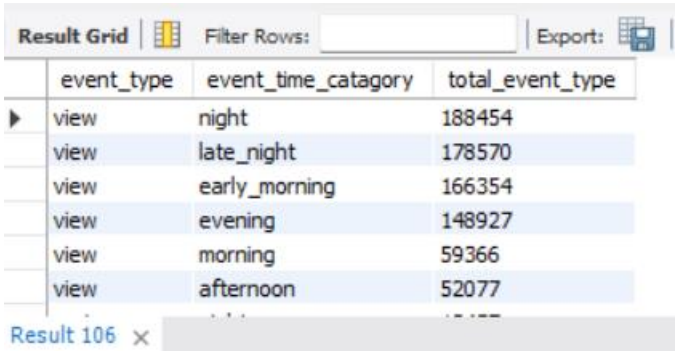
to cart and Purchase time with help of SQL query and need to

confirm if mostly viewed and added to cart time is similar to Purchase time or not.

select * from events

```
with CTE as
(select *,case
WHEN TIME (event_time) between '00:00:00' and '04:00:00' then 'late_night'
WHEN TIME (event_time) between '04:00:00' and '08:00:00' then 'early_morning'
WHEN TIME (event_time) between '08:00:00' and '12:00:00' then 'morning'
WHEN TIME (event_time) between '12:00:00' and '16:00:00' then 'afternoon'
WHEN TIME (event_time) between '16:00:00' and '20:00:00' then 'evening'
WHEN TIME (event_time) between '20:00:00' and '24:00:00' then 'night'
end as event_time_catagory
     from events)
select event_type,event_time_catagory ,count(*) as total_event_type
from CTE
group by event_time_catagory,event_type
order by total_event_type desc;
```

Output:



Result Grid | Filter Rows: | Export:

event_type	event_time_catagory	total_event_type
view	night	188454
view	late_night	178570
view	early_morning	166354
view	evening	148927
view	morning	59366
view	afternoon	52077

Result 106

Explanation:

So here need to check total events occurred in event_time_catagory.

Here we select event_type,event_type category,count all as event type category.

From CTE .Then group by event_time_catagory and event_type.Order by total_event type.

So our result is ready. We got total_event_type

Q2(Third query)

Query:

Top Time of Visit:

Here Students are required to calculate the time for viewing, adding

to cart and Purchase time with help of SQL query and need to

confirm if mostly viewed and added to cart time is similar to Purchase time or not.

select * from events

```sql
with CTE as
 (select *,case
 WHEN TIME (event_time) between '00:00:00' and '04:00:00' then 'late_night'
 WHEN TIME (event_time) between '04:00:00' and '08:00:00' then 'early_morning'
 WHEN TIME (event_time) between '08:00:00' and '12:00:00' then 'morning'
 WHEN TIME (event_time) between '12:00:00' and '16:00:00' then 'afternoon'
 WHEN TIME (event_time) between '16:00:00' and '20:00:00' then 'evening'
 WHEN TIME (event_time) between '20:00:00' and '24:00:00' then 'night'
 end as event_time_catagory
      from events),
      CTEM as
      (select event_type,event_time_catagory ,count(*) as total_event_type from CTE
      group by event_time_catagory,event_type
      order by total_event_type desc)select event_type,event_time_catagory , total_event_type
      from CTEM
      where total_event_type in(select max(total_event_type) from CTEM group by event_type)
```

Output:

	event_type	event_time_catagory	total_event_type
▶	view	night	188454
	cart	night	13457
	purchase	night	9787

Explanation:

Mostly viewed and added to cart time is similar to Purchase time .

We have find **max of total event type in every event_type and the event time category**

Is coming same which is **night**.

Q3

Top brands by Sale:

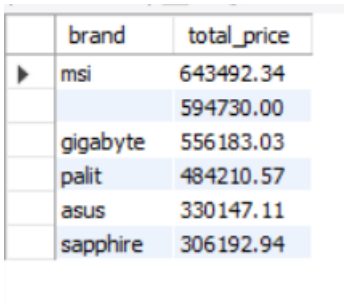
Here Students are required to find Top 6 brand w.r.t

Sales which is the Top Brand followed by other brand with the help of SQL query.

Query:

```
select brand,sum(price) as total_price
from events
where event_type='purchase'
group by brand order by total_price desc
limit 6
```

Output:



brand	total_price
msi	643492.34
	594730.00
gigabyte	556183.03
palit	484210.57
asus	330147.11
sapphire	306192.94

Explanation:

Firstly we need to select brand,sum of price as total_price.

And here we needs to use **where condition**

In where condition we write **event_type =’purchase’**

After where condition we need to made brand **Group**

And order by brand need to make it in **Descending order**

At the end of query select and show top 6 brand

So here we use **LIMIT** function so here we write **LIMIT6.**

Q4

Demand for Items:

Here Students are required to find Top 6 Category which was

sold most number of times and should also show the count as well.

select * from events

Query:

```
select category_code,count(category_id) as sold_time
from events where event_type='purchase'
group by category_code order by sold_time desc
limit 6
```

Output:

category_code	sold_time
	7568
computers.components.videocards	6888
electronics.telephone	4119
stationery.cartrige	2739
computers.peripherals.printer	2557
computers.notebook	1297

Explanation:

Here we need to find top 6 category .

Firstly we required to select category_code,

Count of category_id as sold times because category_id is mention infront each event_type.

Then we use **where condition** because we want purchase time quantity as sold time

So we use **event_type = 'purchase'**

After that we made groups by category_code

We need top 6 category .

So we use sold_time is in **descending order .**

We need only top 6 category .

So we use **LIMIT** function.

LIMIT 6.

Q5

Frequency of Purchase:

Here Students are required to find count of Users who has purchase at each time i.e.

number of users purchasing once,twice,thrice etc once in a given time span of 5 months

and also Maximum number of Times one person has purchased in a given time span of 5 months

select * from events

Query:

```
select year(event_time)year,month(event_time)month,count(*)as user_purchases,user_id
from events where event_type ='purchase'
group by  year,month,user_id
order by user_purchases Desc
```

Output:

year	month	user_purchases	user_id
2021	1	8208	1.51592E+18
2021	2	7657	1.51592E+18
2020	11	7553	1.51592E+18
2020	12	6621	1.51592E+18
2020	10	6251	1.51592E+18
2020	9	938	1.51592E+18
2021	3	118	1.51592E+18

Explanation :

Here we need to find **Frequency of Purchase** in given time span of **Five_months.**

Firstly we required to extract year and month from event_time .

Then count all from events as user_purchases.

After that we need to use **where condition** for select only **purchase event_type.**

So we write **event_type = 'purchase'.**

And made **group of year and month**,**user_id**

At the last we make our results in to **desending order.**

Q6

Actual Time purchased :

Here Students have to find the number of times that the item has actually been purchased

after the users have viewed the items i.e. the query should give result for actual number of times purchased.

Also query should be written to find the number of times the item has been viewed.

Query:

```
select event_type,count(*)
from events
group by event_type;
```

Output:-

	event_type	count(*)
▶	view	793748
	cart	54035
	purchase	37346

Explanation of query:

Here we need to find actual time purchase after the users have viewed the items

So firstly we need to **select event_type and count all from events**.

And then we need to **make groups by event_types.**

Conclusion

So I analyse data and make some conclusions below

1. **Month of Sales:**
   We calculated monthly sales of our E-commerce website.

   The highest sales is in 1st month 2021 Is **33508355.90**

year	month	sales
2021	1	33508355.90

   Lowest sales is in 3rd month 2021 Is 532570.51

	year	month	sales
▶	2021	3	532570.51

2.Top Time of Visit:

Rather than viewing it hour wise I made time category of 4 hrs each

Here we make time bins for event_time_catagory as follows

12am-4am →Late Night

4am -8am→Early Morning

8am-12am→ Morning

12pm - 4pm →Afternoon

4pm-8pm →Evening.

8pm →12pm Night.

Night time (8pm – 12pm)is the time where users are very frequently viewing and purchasing.

So this is time here we can really target users giving some offers and coupons.

Views in night(8pm – 12pm.)

So final conclusion is top time of visits and purchase is night(8pm – 12am)

event_type	event_time_catagory	total_event_type
view	night	188454
cart	night	13457
purchase	night	9787

3.Top brands by Sale:

We find top 6 brand by its total _sales as total price

Most of the products have not updated their brand name we can clearly see in our output so would have to update the brand name of the products so that it not would create any future error and it would to good look at or keep tack of.

Our top 6 selling brand is following:

brand	total_price
msi	643492.34
	594730.00
gigabyte	556183.03
palit	484210.57
asus	330147.11
sapphire	306192.94

4.Demand for Items:

Here we find our top 6 selling catgory

Most of the products have not updated their category_code and brand name we can clearly see in our output so would have to update the category_code and brand name of the products so that it not would create any future error and it would to good look at or keep tack of.

Our top 6 selling category as follows:

category_code	sold_time
	7568
computers.components.videocards	6888
electronics.telephone	4119
stationery.cartrige	2739
computers.peripherals.printer	2557
computers.notebook	1297

5.Frequency of Purchase:

Here we find Frequency of purchase of products by customer in the order of month.

So here only one user in data .

So frequency of user purchases is highest in 1st month of 2021.

And frequency of user purchases is lowest in 3rd month of 2021.

So final conclusion is the frequency of user purchases is average in 2nd,11th,12th,10th

But frequency of user purchases is very low in 9th and 3rd month in given data.

Specially we need focus on increase user_purchases frequency of 9th and 3rd month of the given data.

year	month	user_purchases	user_id
2021	1	8208	1.51592E+18
2021	2	7657	1.51592E+18
2020	11	7553	1.51592E+18
2020	12	6621	1.51592E+18
2020	10	6251	1.51592E+18
2020	9	938	1.51592E+18
2021	3	118	1.51592E+18

6.Actual Time purchased:

Here we need to find how many users are visited the site and after that how many users purchase the products.

So in this given data

Viewed product by user is **793748**

Actual time purchased by user is **37346**

In percentage form Actual time purchase percentage is 4.70%

So **Actual Time purchased** percentage is very low so our business is not going good so we need to create some offers and coupon for the customer .

event_type	count(*)
view	793748
cart	54035
purchase	37346

Answer of problem_statement:

Despite the huge usage of ecommerce website a section of the society still prefers offline stores. Why?

Answer:

Indian still prefer offline mode. India a developing country, which possess a large number of youth power & tremendous opportunity for world market . Recently mobile revolution have taken place in India & continuous growing in larger pace. In couple of few years things have changed many much , now these days one can see a low end or even a high end smartphone in the hand of a normal indian layman. A lots of telecom companies like Airtel, vodafone, bsnl, tata docomo, idea etc have covered almost every

corner of india with a net of telecommunication tower & cables. Now these days 2g/3g/4g, online payments, shopping etc are not remain alien words for a normal Indian layman. These days lots of people from India making online transactions with the help of phone or laptop etc. But still a lots of people actually a vary big segments of Indian populations doesn't believe in online world. They by no doubt prefer offline mode for financial & other transactions than online mode. Main reason behind this is lack of proper knowledge about internet & also fear of scam & scam like sites etc. You may find very easily a Indian standing in a Que for applying for reservation of train or paying bills ,although at that time he/she possess in its pocket a high end phone capable of doing all online stuff .

Reason behind same is lack of trust on online world & service providers. People from India still prefer to purchase from local market a local brand instead from a reputed online shopping site who may be offering a leading brand in affordable price with doorstep delivery. Indian prefer offline activity in place of online but things are changing .We are growing & learning. Later or sooner situation will improve & we will able to see a lots of Indian trusting online world.

Indians usually buy those things which they can first **see** and **touch** and then **bargain** according to their perception about the product. While buying online, one cannot do these things. Although the mindset is changing now and more and more people are trying online buying but still majority believes that it is better to get a feel before they buy anything.

Even if the market is now changing.