# SQL CAPSTONE PROJECT

**Data Analysis:** There is 17 rows and 1000 columns in Dataset with no null values.

**Product Analysis:** There is 6 total distinct Product line.

1. Health and Beauty
2. Electronic Accessories
3. Home and Analysis
4. Sports and Travel
5. Food and Beverages
6. Fashion and Accessories

There is 3 distinct Cities.

1. Yangon
2. Naypyitaw
3. Mandalay

There is 3 distinct Payment methods.

1. E-Wallet
2. Cash
3. Credit Cards

**Sales Analysis:** Food and Beverages are the Highest sales in Jan month and rather Health and Beauty has lowest sales. As Female customer are more in dataset so we can focus on the gender wise on them and in best way we can increase the Health and Beauty product sales.

**Customer Analysis:** So there is Member and Normal Customer type with Male and Female Gender format.

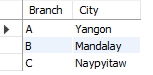
1-- What is the count of distinct cities in the dataset?

select distinct City from sales\_mst;



2-- For each branch, what is the corresponding city?

select Branch, City from sales\_mst group by Branch, City order by Branch;



3-- What is the count of distinct product lines in the dataset?

select count(distinct product\_line) as count\_of\_product\_Line from sales\_mst;



4-- Which payment method occurs most frequently?

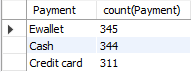
select Payment, count(Payment) from sales\_mst where Payment='Ewallet'

union

select Payment, count(Payment) from sales\_mst where Payment='Cash'

union

select Payment, count(Payment) from sales\_mst where Payment='Credit card';



5-- Which product line has the highest sales?

select Product\_line, Total from sales\_mst order by Total desc limit 1;

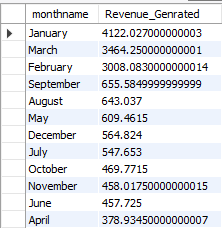


6-- How much revenue is generated each month?

select monthname(str\_to\_date(Date, '%d/%m/%Y')) as monthname,

sum(gross\_income) as Revenue\_Genrated

from sales\_mst group by monthname order by Revenue\_Genrated desc;



7-- In which month did the cost of goods sold reach its peak?

select monthname(str\_to\_date(Date, '%d/%m/%Y')) as monthname,

cogs

from sales\_mst order by cogs desc limit 1;



8-- Which product line generated the highest revenue?

select Product\_line,

sum(gross\_income) as Revenue\_Genrated

from sales\_mst group by Product\_line order by Revenue\_Genrated desc limit 1;



9-- In which city was the highest revenue recorded?

select City,

sum(gross\_income) as Revenue\_Genrated

from sales\_mst group by City order by Revenue\_Genrated desc limit 1;



10-- Which product line incurred the highest Value Added Tax?

select Product\_line,

Tax

from sales\_mst order by Tax desc limit 1;



11-- For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."

select Product\_line, Average\_Sales, Total,

case

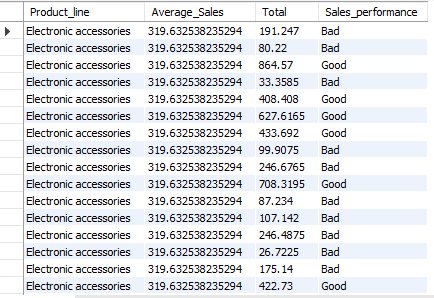
when Total > Average\_Sales then 'Good'

else 'Bad'

end as Sales\_performance from (select \*,

avg(Total) over (partition by Product\_line) as Average\_Sales

from sales\_mst) as avg\_sales;



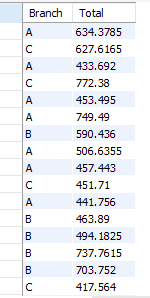
12-- Identify the branch that exceeded the average number of products sold?

select Branch, Total from sales\_mst where Total >

(select

avg(Total) as Average\_Sales

from sales\_mst );



13-- Which product line is most frequently associated with each gender?

with gender\_product\_line as (

select Gender, Product\_line, count(\*) as Count,

Row\_Number() over (Partition by Gender Order by count(\*)desc) as rnb

from sales\_mst

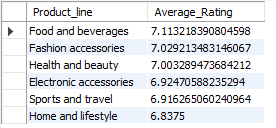
group by Gender, Product\_line

)select Gender, Product\_line, Count from gender\_product\_line where rnb=1;



14-- Calculate the average rating for each product line.

select Product\_line, avg(Rating) as Average\_Rating from sales\_mst group by Product\_line order by Average\_Rating desc;



15-- Count the sales occurrences for each time of day on every weekday?

select Case

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 0 then 'Sunday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 1 then 'Monday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 2 then 'Tuesday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 3 then 'Wednesday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 4 then 'Thursday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 5 then 'Friday'

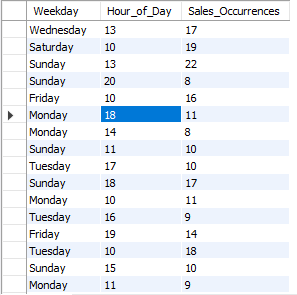
else 'Saturday'

end as Weekday,

substring(Time, 1, 2) as Hour\_of\_Day,

count(\*) as Sales\_Occurrences

from sales\_mst group by Weekday, Hour\_of\_Day;



16-- Identify the customer type contributing the highest revenue?

select Customer\_type, sum(gross\_income) as Revenue\_Genrated

from sales\_mst

group by Customer\_type

order by Revenue\_Genrated desc limit 1;



17-- Determine the city with the highest VAT percentage.

select City, sum(Tax) as Tax\_Genrated

from sales\_mst

group by City order by Tax\_Genrated desc limit 1;



18-- Identify the customer type with the highest VAT payments.

select Customer\_type, sum(Tax) as Total\_Vat

from sales\_mst

group by Customer\_type

order by Total\_Vat desc limit 1;



19-- What is the count of distinct customer types in the dataset?

select distinct Customer\_type, count(Customer\_type) as count\_Cust\_Type

from sales\_mst

group by Customer\_type;

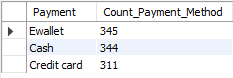


20-- What is the count of distinct payment methods in the dataset?

select distinct Payment, count(Payment) as Count\_Payment\_Method

from sales\_mst

group by Payment;



21-- Which customer type occurs most frequently?

select Customer\_type, count(\*) as Frequency

from sales\_mst

group by Customer\_type

order by Frequency desc limit 1;



22-- Identify the customer type with the highest purchase frequency.

select Customer\_type, Total

from sales\_mst

group by Customer\_type, Total

order by Total desc limit 1;



23-- Determine the predominant gender among customers.

select Gender, count(\*) as Gender\_count

from sales\_mst

group by Gender

order by Gender\_count desc limit 1;



24-- Examine the distribution of genders within each branch.

select Branch,

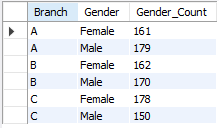
Gender,

count(\*) as Gender\_Count

from sales\_mst

group by Gender, Branch

order by Branch, Gender;



25-- Identify the time of day when customers provide the most ratings.

select Rating, case

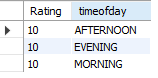
when cast(Substring\_index(Time, ':', 1) as UNSIGNED) < 12 then 'MORNING'

when cast(Substring\_index(Time, ':', 1) as UNSIGNED) < 18 then 'AFTERNOON'

else 'EVENING' end as timeofday from sales\_mst

group by Rating, timeofday

order by Rating desc limit 3;



26-- Determine the time of day with the highest customer ratings for each branch.

With Highest\_rating\_Branch\_wise as

(

select Branch, Rating, case

when cast(Substring\_index(Time, ':', 1) as UNSIGNED) < 12 then 'MORNING'

when cast(Substring\_index(Time, ':', 1) as UNSIGNED) < 18 then 'AFTERNOON'

else 'EVENING' end as timeofday,

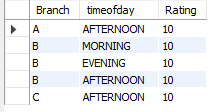
Rank() Over(Partition By Branch Order by Rating desc) as Rating\_Rank

from sales\_mst

group by Branch, Rating, timeofday

)

select Branch, timeofday, Rating from Highest\_rating\_Branch\_wise where Rating\_Rank=1;



27-- Identify the day of the week with the highest average ratings.

select Case

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 0 then 'Sunday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 1 then 'Monday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 2 then 'Tuesday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 3 then 'Wednesday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 4 then 'Thursday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 5 then 'Friday'

else 'Saturday'

end as dayname,

avg(Rating) as AVG\_Rating

from sales\_mst

group by dayname

order by AVG\_Rating desc limit 1;



28-- Determine the day of the week with the highest average ratings for each branch.

With Ranking\_Branchwise\_avgRating as

(

select Branch, Case

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 0 then 'Sunday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 1 then 'Monday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 2 then 'Tuesday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 3 then 'Wednesday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 4 then 'Thursday'

when Date\_Format(STR\_TO\_DATE(Date, '%d/%m/%Y'), '%w') = 5 then 'Friday'

else 'Saturday'

end as dayname,

avg(Rating) as AVG\_Rating,

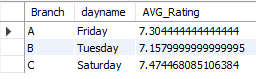
Rank() Over(Partition by Branch Order by avg(Rating) desc) as Ranking

from sales\_mst

group by Branch, dayname

)

select Branch, dayname, AVG\_Rating from Ranking\_Branchwise\_avgRating where Ranking=1;



----------------------------------------------------\*\*\*\*\*\*\*--------------------------------------------------------

Stu\_ID: S7475

Name: Lokesh Mohanlal Joshi

Place: Bhatapara(C.G)