**Walmart SQL table and values setup and queues**

SELECT \* FROM walmart\_sales.sales;

-- we dont have any nulls

-- -------------------- feature engineering: will help use genderated some new columns from existing ones

-- 1) add new column named time of day to give insights of sales in the morning, afternon and evening.

select

time,

(Case

when `time` Between '00:00:00' And '12:00:00' Then 'Marning'

when `time` Between '12:01:00' And '16:00:00' Then 'Afternoon'

Else 'Evening'

End ) As tme\_of\_date

From Sales;

Alter Table sales Add Column time\_of\_day varchar(20);

UPDATE sales

SET time\_of\_day = (

CASE

WHEN `time` BETWEEN '00:00:00' AND '12:00:00' THEN 'Marning'

WHEN `time` BETWEEN '12:01:00' AND '16:00:00' THEN 'Afternoon'

ELSE 'Evening'

END

);

select \* from sales;

-- 2) add a new column dayName that contains the extracted day of the wek on which the given

-- transaction took place(mon, tue, wed, thur, fri)

Alter Table sales Change data date DATETIME;

Select

date,

dayname(date) As day\_name -- day when the date happened

From sales;

Alter Table sales Add Column day\_name Varchar(10);

-- insert day\_name

Update sales

set day\_name = dayname(date);

select \* from sales;

-- 3 ad new column named month\_name that cotains the extracted month of the year on which the given

-- transaction took place (Jan, feb, Mar)

-- month\_name

Select date,

monthname(date)

from sales;

Alter Table sales change column month\_name month\_name varchar(10) ;

UPDATE sales

Set month\_name = monthname(date);

-- ----------- Exloratory Data Analysis/EDA ------------------------

-- --------Generic Questons

-- 1) How many unque cities are in the data?

Select

DISTINCT city

from sales;

-- 3 distict cites

-- 2) how many branches do we have?

Select

DISTINCT branch

From Sales;

-- 3 a,b,

-- in which city is each branch?

Select

DISTINCT city,

branch

from sales;

-- a - yangon, naypyitaw -c and mandalay b

-- ---------- prodcut questions

-- 1) How many unique product lines does the data have?

Select

DISTINCT product\_line

from sales;

-- another way

select

COUNT(distinct product\_line)

from sales;

-- 6

-- 2) most common payment mothod

select payment\_method,

count(payment\_method) As cnt

from sales

group by payment\_method

order by cnt Desc

; -- cash: 344, credit card: 309, ewallet: 342

-- 3) what is the most selling product line?

Select product\_line,

Count(product\_line) As productLineCNT

from sales

group by product\_line

Order By productLineCNT Desc;

-- fashion accessiories highest at 178. health and beauty lowest at 151

-- 4) what is the total rev by month?

select

month\_name As month,

Sum(total) As totalRev

From sales

Group by month

Order by totalRev Desc

; -- Jan had the highest will 1,116,291 Feb had the least sales.

-- 5) what month had the largest cost of good sold-cogs

Select

Month\_name As month,

SUM(cogs) As totalCogs

From Sales

Group by month\_name

Order by totalCogs desc; -- jan had the highest total cost of goods. Feb had the lowest cost of goods sold

-- 6) What product line had the largest revenue?

Select

product\_line,

Sum(total) As totalRev

From sales

Group by product\_line

Order by totalRev Desc; -- food and beverage had the highest rev at 56,144.84. health and beauty was the least out of the 6 product lines

-- 7 ) what is the city with the largest rev?

Select city,

SUM(total) as totalRev

From Sales

Group by city

Order by totalRev desc; -- Naypyitaw had the highest rev, and monadolay had the least out of the 3 cities

-- 8) what product line had the largest VAT/tax

Select

product\_line,

avg(Vat) As avgTax

From Sales

Group by product\_line

Order by avgTax desc; -- home and lifestlye had the highest avg tax/val with 16. fashion accessories had the least avg value added tax

-- 9) Fetch each product line and add a column to those products line showing 'good', 'bad.

-- good if it great than the average sales

-- 10) which branch sold more products than the avg product?

Select

branch,

Sum(quantity) As qty

From sales

Group by branch

Having Sum(quantity) > (Select avg(quantity)From sales);

-- 11) What is the most common product line by gender - counting the gender

select \* from sales;

Select

gender,

product\_line,

count(gender) As gender\_total

From sales

Group by gender, product\_line

order by gender\_total desc;

-- 10) what is the avg rating of each product\_line

select product\_line,

Round(avg(rating),2) As avgRating

From Sales

Group by product\_line

Order by avgRating desc; -- food and beverages had the highest avg rating

-- -----------------------Sales anaylisis

-- 1)

-- # of sales made in each time of the day per weekday

Select

time\_of\_day,

Count(\*) as totalSales

From sales

Where day\_name in ('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday')

Group by time\_of\_day

Order by totalSales desc; -- during the week customers tend to shop more during the evening vs other times of day with 290 # of sales

-- 2) which of the customers types brings the most revenue

select \* from sales;

Select customer\_type,

Round(Sum(total),2) As totalRev

From sales

Group by customer\_type

Order by totalRev desc; -- members bring in the highest rev with 163,625.10

-- 3) which city has the largest tex/vat(value added tax)

Select city,

avg(vat) As vat

From Sales

Group by city

Order by vat desc; -- naypyitaw

-- 4) which customer type pays the most in vat

Select customer\_type,

round(avg(vat),2) As member\_avgVat

From Sales

Group by customer\_type

Order by member\_avgVat desc; -- member at 15.61

-- -----------customer

-- 1) how many inque customer types does the data have

Select

Count(DISTINCT customer\_type)

From sales; -- 2

-- 2) how many unique payment methods does the data have

Select

Count(DISTINCT payment\_method)

From sales; -- 3

Select customer\_type,

count(\*) As count

From Sales

group by customer\_type

Order by count; -- Member

-- 3) which customer type buys the most

Select customer\_type,

Count(\*) As customer\_count

From sales

GROUP BY customer\_type

Order by customer\_count desc; -- members

-- 4) what is the gender of the most of the customers

Select gender,

Count(\*) As gender\_count

From sales

Group by gender; -- male

-- 5 what is the gender distribution per branch

Select gender, branch,

Count(\*) As gender\_count

From sales

Group by gender, branch; -- male

-- 6 what time of day does customers give ratings?

Select

time\_of\_day,

avg(rating) as avg\_rating,

count(rating) as rating\_count

From sales

Group by time\_of\_day

Order by avg\_rating desc , rating\_count desc;

-- 7 which time of day do customers give most rating per branch

Select

time\_of\_day,

branch,

avg(rating) as avg\_rating,

count(rating) as rating\_count

From sales

Where branch = 'A'

Group by time\_of\_day

Order by avg\_rating desc , rating\_count desc;

-- 8 which day of the week ha the best avg rating

Select day\_name,

round(avg(rating), 2) As avg\_rating

From sales

Group by day\_name

Order by avg\_rating desc;

-- 9) which day of the week has the best average rating per branch

Select day\_name,

round(avg(rating), 2) As avg\_rating

From sales

Where branch = 'C' -- can use A, B too

Group by day\_name

Order by avg\_rating desc;