DATA ANALYSIS OF WALMART SALES DATA

----------------Exploratory Data Analysis (EDA)---------------------

Generic Questions

-- 1.How many distinct cities are present in the dataset?

select distinct city from branch;

	city
▶	Yangon
	Naypyitaw
	Mandalay Mand

-- 2.In which city is each branch situated?

select distinct branch , city from branch ;

	branch	city
▶	A	Yangon
	C	Naypyitaw
	B	Mandalay

-- 3.How many distinct product lines are there in the dataset?

SELECT DISTINCT PRODUCT_LINE FROM PRODUCT ;

	product_line
▶	Health and beauty
	Electronic accessories
	Home and lifestyle
	Sports and travel
	Food and beverages
	Fashion accessories

4.What is the most common payment method?

SELECT PAYMENT, COUNT(PAYMENT) MOST_USING_PAYMENT_METHOD FROM PAYMENT_MODE

GROUP BY PAYMENT ORDER BY COUNT(PAYMENT) DESC ;

	payment	most_using_payment_method
▶	Ewallet	345
	Cash	344
	Credit card	311

-- 5.What is the most selling product line?

SELECT PRODUCT_LINE, ROUND(SUM(TOTAL),2) FROM PRODUCT

GROUP BY PRODUCT_LINE ORDER BY SUM(TOTAL) DESC LIMIT 1 ;

	product_line	round(sum(total),2)
▶	Food and beverages	56144.84

-- 6.What is the total revenue by month?

SELECT T.MONTH,ROUND(SUM(P.TOTAL),2) TOTAL_REVENUE FROM TIME_ZONE T INNER JOIN PRODUCT P

ON T.TRANSACTION_ID = P.TRANSACTION_ID GROUP BY T.MONTH ORDER BY SUM(P.TOTAL) DESC LIMIT 1 ;

	month	total_revenue
▶	January	116291.87

-- 7.Which month recorded the highest Cost of Goods Sold (COGS)?

SELECT T.MONTH,ROUND(SUM(PS.COGS),2) TOTAL_REVENUE FROM TIME_ZONE T INNER JOIN PRODUCT_SOLD PS

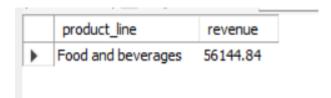ON T.TRANSACTION_ID = PS.TRANSACTION_ID GROUP BY T.MONTH ORDER BY SUM(COGS) DESC LIMIT 1 ;

	month	total_revenue
▶	January	110754.16

-- 8.Which product line generated the highest revenue?

SELECT PRODUCT_LINE , ROUND(SUM(TOTAL),2) AS REVENUE

FROM PRODUCT

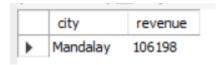GROUP BY PRODUCT_LINE ORDER BY SUM(TOTAL) DESC LIMIT 1 ;

	product_line	revenue
▶	Food and beverages	56144.84

-- 9.Which city has the highest revenue?

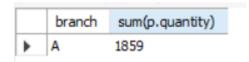SELECT B.CITY , ROUND(SUM(P.TOTAL)) AS REVENUE FROM BRANCH B INNER JOIN PRODUCT P

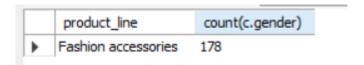ON B.TRANSACTION_ID = P.TRANSACTION_ID GROUP BY B.CITY ORDER BY SUM(P.TOTAL) LIMIT 1;

	city	revenue
▶	Mandalay	106198

-- 10.Retrieve each product line and add a column product_category, indicating 'Good' or 'Bad,'based on whether its sales are above the average.

- ❖ ALTER TABLE PRODUCT ADD COLUMN PRODUCT_CATEGORY VARCHAR (25);
- ❖ UPDATE PRODUCT SET PRODUCT_CATEGORY =

  (SELECT IF (TOTAL >= (SELECT AVG(TOTAL)),"GOOD","BAD"));

- ❖ SELECT IF (TOTAL >= (SELECT AVG(TOTAL) FROM PRODUCT),"GOOD","BAD") FROM PRODUCT;
- ❖ SELECT PRODUCT_LINE, PRODUCT_CATEGORY, COUNT(PRODUCT_CATEGORY) FROM PRODUCT

  GROUP BY PRODUCT_LINE, PRODUCT_CATEGORY ORDER BY COUNT(PRODUCT_CATEGORY) DESC LIMIT 1;

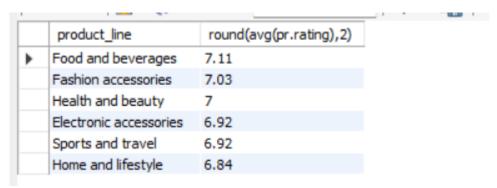|	product_line	product_category	count(product_category)
▶	Fashion accessories	Good	178

-- 11.Which branch sold more products than average product sold?

❖ SELECT B.BRANCH , SUM(P.QUANTITY) FROM BRANCH B INNER JOIN PRODUCT P
ON B.TRANSACTION_ID = P.TRANSACTION_ID GROUP BY B.BRANCH HAVING SUM(P.QUANTITY) >
AVG(P.QUANTITY)
ORDER BY SUM(P.QUANTITY) DESC LIMIT 1 ;

branch	sum(p.quantity)
A	1859

-- 12.What is the most common product line by gender?

❖ SELECT P.PRODUCT_LINE, COUNT(C.GENDER) FROM CUSTOMER C INNER JOIN PRODUCT P ON
C.TRANSACTION_ID = P.TRANSACTION_ID GROUP BY P.PRODUCT_LINE ORDER BY COUNT(C.GENDER) DESC
LIMIT 1;

product_line	count(c.gender)
Fashion accessories	178

-- 13.What is the average rating of each product line?

❖ SELECT P.PRODUCT_LINE,ROUND(AVG(PR.RATING),2) FROM PRODUCT P INNER JOIN
PRODUCT_RATING PR
ON P.TRANSACTION_ID = PR.TRANSACTION_ID GROUP BY P.PRODUCT_LINE ORDER BY
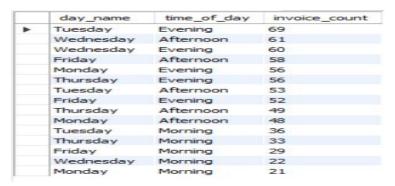ROUND(AVG(PR.RATING),2) DESC;

product_line	round(avg(pr.rating),2)
Food and beverages	7.11
Fashion accessories	7.03
Health and beauty	7
Electronic accessories	6.92
Sports and travel	6.92
Home and lifestyle	6.84

Sales Analysis

-- 14.Number of sales made in each time of the day per weekday

Method 1:
❖ SELECT DAY_NAME, TIME_OF_DAY, COUNT(BRANCH.INVOICE_ID) AS INVOICE_COUNT FROM
TIME_ZONE T INNER JOIN BRANCH ON T. TRANSACTION_ID = BRANCH.TRANSACTION_ID
WHERE DAY_NAME NOT IN ('SATURDAY', 'SUNDAY') GROUP BY DAY_NAME, TIME_OF_DAY
ORDER BY INVOICE_COUNT DESC;

day_name	time_of_day	invoice_count
Tuesday	Evening	69
Wednesday	Afternoon	61
Wednesday	Evening	60
Friday	Afternoon	58
Monday	Evening	56
Thursday	Evening	56
Tuesday	Afternoon	53
Friday	Evening	52
Thursday	Afternoon	49
Monday	Afternoon	48
Tuesday	Morning	36
Thursday	Morning	33
Friday	Morning	29
Wednesday	Morning	22
Monday	Morning	21

Method 2:

❖ SELECT day_name, time_of_day, COUNT(*) AS total_sales
FROM sales WHERE day_name NOT IN ('Saturday','Sunday') GROUP BY day_name, time_of_day;

-- 15.Identify the customer type that generates the highest revenue.

❖ SELECT CUSTOMER_TYPE , SUM(TOTAL) AS TOTAL_REVENUE
FROM CUSTOMER INNER JOIN PRODUCT
ON CUSTOMER.TRANSACTION_ID = PRODUCT.TRANSACTION_ID
GROUP BY CUSTOMER_TYPE ORDER BY SUM(TOTAL) DESC ;
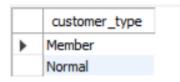
	customer_type	Total_revenue
▶	Member	164223.44400000002
	Normal	158743.30500000005

-- 16.Which city has the largest tax percent/ VAT (Value Added Tax)?

❖ SELECT BRANCH.CITY,ROUND(SUM(`TAX_5%`),2) FROM BRANCH INNER JOIN PRODUCT
ON BRANCH.TRANSACTION_ID = PRODUCT.TRANSACTION_ID
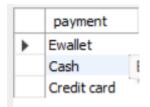GROUP BY BRANCH.CITY ORDER BY SUM(`TAX_5%`) DESC ;

	city	round(sum(`tax_5%`),2)
▶	Naypyitaw	5265.18
	Yangon	5057.16
	Mandalay	5057.03

CUSTOMER ANALYSIS

-- 17.How many unique customer types does the data have?

❖ SELECT DISTINCT CUSTOMER_TYPE FROM CUSTOMER;

	customer_type
▶	Member
	Normal

-- 18.How many unique payment methods does the data have?

❖ SELECT DISTINCT PAYMENT FROM PAYMENT_MODE;

	payment
▶	Ewallet
	Cash
	Credit card

-- 19.Which is the most common customer type?

❖ SELECT CUSTOMER_TYPE , COUNT(CUSTOMER_TYPE) FROM CUSTOMER
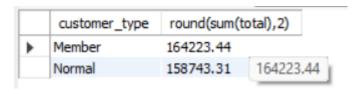GROUP BY CUSTOMER_TYPE ORDER BY COUNT(CUSTOMER_TYPE) DESC LIMIT 1;

	customer_type	count(customer_type)
▶	Member	501

-- 20.Which customer type buys the most?

SELECT CUSTOMER_TYPE, ROUND(SUM(TOTAL),2) FROM CUSTOMER INNER JOIN PRODUCT
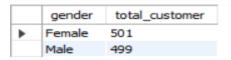
ON CUSTOMER.TRANSACTION_ID = PRODUCT.TRANSACTION_ID

GROUP BY CUSTOMER_TYPE ORDER BY SUM(TOTAL) DESC ;

customer_type	round(sum(total),2)	
Member	164223.44	
Normal	158743.31	164223.44

-- 21.What is the gender of most of the customers?

SELECT GENDER , COUNT(*) AS TOTAL_CUSTOMER FROM CUSTOMER

GROUP BY GENDER ORDER BY COUNT(*) DESC;

gender	total_customer
Female	501
Male	499

-- 22.What is the gender distribution per branch?

SELECT GENDER,BRANCH,COUNT(GENDER) FROM CUSTOMER INNER JOIN BRANCH
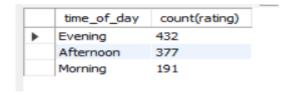
ON CUSTOMER.TRANSACTION_ID = BRANCH.TRANSACTION_ID

GROUP BY BRANCH,GENDER ORDER BY COUNT(GENDER) DESC ;

gender	branch	count(gender)
Male	A	179
Female	C	178
Male	B	170
Female	B	162
Female	A	161
Male	C	150

-- 23.Which time of the day do customers give most ratings?

SELECT TIME_OF_DAY, COUNT(RATING) FROM TIME_ZONE INNER JOIN PRODUCT_RATING

ON TIME_ZONE.TRANSACTION_ID = PRODUCT_RATING.TRANSACTION_ID

GROUP BY TIME_OF_DAY ORDER BY COUNT(*) DESC ;

time_of_day	count(rating)
Evening	432
Afternoon	377
Morning	191

-- 24.Which day of the week has the best avg ratings?

SELECT DAY_NAME, COUNT(RATING) FROM TIME_ZONE INNER JOIN PRODUCT_RATING

ON TIME_ZONE.TRANSACTION_ID = PRODUCT_RATING.TRANSACTION_ID

GROUP BY DAY_NAME ORDER BY COUNT(RATING) DESC LIMIT 1;

day_name	count(rating)
Saturday	164