

Walmart Sales Data Analysis (Case Study)

SQL + BI tools

Data Wrangling:

This is the first step where inspection of data is done to make sure **NULL** values and missing values are detected and data replacement methods are used to replace, missing or **NULL** values.

- - - Build a database
- - - Create table and insert the data.
- - - Select columns with null values in them.

There are no null values in our database as in creating the tables, we set **NOT NULL** for each field, hence null values are filtered out.

Feature Engineering:

- Add a new column named `time_of_day` to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.

```
---CREATE A NEW COLUMN(TIME OF DATE)---
ALTER TABLE walmartSales
ADD time_of_day VARCHAR(10);

UPDATE walmartSales
SET time_of_day = CASE
    WHEN DATEPART(hour,Time) BETWEEN 6 AND 11 THEN
        'Morning'
    WHEN DATEPART(hour,Time) BETWEEN 12 AND 17 THEN
        'Afternoon'
    WHEN DATEPART(hour,Time) BETWEEN 18 AND 23 THEN
        'Evening'
    ELSE 'Night'
END;
```

```
(1000 rows affected)
```

```
Completion time: 2024-05-29T10:30:14.4391113+05:30
```

- Add a new column named `day_name` that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.

```
---CREATE A NEW COLUMN(DAY NAME)---
ALTER TABLE walmartSales
ADD day_name VARCHAR(20);

UPDATE walmartSales
SET day_name = LEFT(DATENAME(weekday, Date), 3);
```

```
(1000 rows affected)
```

```
Completion time: 2024-05-29T10:30:14.4391113+05:30
```

- Add a new column named `month_name` that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.

```
---CREATE A NEW COLUMN(MONTH NAME)---
ALTER TABLE walmartSales
ADD month_name VARCHAR(20);

UPDATE walmartSales
SET month_name = LEFT(DATENAME(MONTH, Date), 3);
```

```
(1000 rows affected)
```

```
Completion time: 2024-05-29T10:30:14.4391113+05:30
```

-----REVENUE AND PROFIT CALCULATION-----

--COGS(COST OF GOODS SERVICE) = Units Price * quantity ↗
--VAT(VALUE ADDED TAX)= 5% * COGS ↗
--VAT is added to the COGS and this is what is billed to the customer.
--Total (gross_sales) = VAT + COGS ↗
--Gross Profit = Total - COGS ↗
--Gross Margin is gross profit expressed in percentage of the
--total(gross profit/revenue)
--Gross Margin = gross income / total revenue ↗

-----EXAMPLE-----

--Data given:
--Unite Price = 45.79 \$
--Quantity = 7 \$

--COGS = 45.79 * 7 = 320.53 \$ ↗

--VAT = 5% * COGS = 5% * 320.53 = 16.0265 \$ ↗

--Total = VAT + COGS = 16.0265 + 320.53 = 336.5565 \$ ↗

--Gross profit=Total-COGS=336.5565-320.53 ↗

--Gross Margin Percentage = gross income / total revenue = 16.0265 /
--336.5565 ↗
--= 0.047619 approx
--= 4.7619 %

Generic Questions

- How many unique cities does the data have?

```
SELECT count(DISTINCT CITY) AS 'UNIQUE CITIES' FROM walmartSales;
```

	UNIQUE CITIES
1	3

- In which city is each branch?

```
SELECT CITY,BRANCH, COUNT(BRANCH) AS COUNT_OF_BRANCHES
FROM walmartSales
GROUP BY CITY,BRANCH
ORDER BY BRANCH;
```

	CITY	BRANCH	COUNT_OF_BRANCHES
1	Yangon	A	340
2	Mandalay	B	332
3	Naypyitaw	C	328

PRODUCT ANALYSIS

1. How many unique product lines does the data have?

```
SELECT COUNT(DISTINCT product_line) AS TOTAL_UNIQUE_PRODUCT_LINES
FROM walmartSales;
```

	TOTAL_UNIQUE_PRODUCT_LINES
1	6

2. What is the most common payment method?

```
SELECT TOP 1 PAYMENT, COUNT(PAYMENT) AS COUNT_OF_PAYMENT
FROM walmartSales
GROUP BY PAYMENT
ORDER BY COUNT_OF_PAYMENT DESC;
```

	PAYMENT	COUNT_OF_PAYMENT
1	Ewallet	345

3. What is the most selling product line?

```
SELECT TOP 1 PRODUCT_LINE, COUNT(QUANTITY) AS COUNT_OF_QUANTITY  
FROM walmartSales  
GROUP BY PRODUCT_LINE  
ORDER BY COUNT_OF_QUANTITY DESC;
```

	PRODUCT_LINE	COUNT_OF_QUANTITY
1	Fashion accessories	178

4. What is the total revenue by month?

```
SELECT MONTH_NAME,ROUND(SUM(TOTAL),2) AS 'TOTAL REVENUE'  
FROM walmartSales  
GROUP BY MONTH_NAME  
ORDER BY CASE  
        WHEN MONTH_NAME = 'Jan' THEN 1  
        WHEN MONTH_NAME = 'Feb' THEN 2  
        WHEN MONTH_NAME = 'Mar' THEN 3  
        WHEN MONTH_NAME = 'Apr' THEN 4  
        WHEN MONTH_NAME = 'May' THEN 5  
        WHEN MONTH_NAME = 'Jun' THEN 6  
        WHEN MONTH_NAME = 'Jul' THEN 7  
        WHEN MONTH_NAME = 'Aug' THEN 8  
        WHEN MONTH_NAME = 'Sep' THEN 9  
        WHEN MONTH_NAME = 'Oct' THEN 10  
        WHEN MONTH_NAME = 'Nov' THEN 11  
        WHEN MONTH_NAME = 'Dec' THEN 12  
END;
```

	MONTH_NAME	TOTAL REVENUE
1	Jan	116291.87
2	Feb	97219.37
3	Mar	109455.51

5. What month had the largest COGS?

```
SELECT TOP 1 MONTH_NAME,ROUND(MAX(COGS),2) AS "LARGEST_COGS"  
FROM walmartSales  
GROUP BY MONTH_NAME  
ORDER BY LARGEST_COGS DESC;
```

	MONTH_NAME	LARGEST_COGS
1	Feb	993

6. What product line had the largest revenue?

```
SELECT TOP 1 PRODUCT_LINE,ROUND(MAX(TOTAL),2) AS TOTAL_REVENUE  
FROM walmartSales  
GROUP BY PRODUCT_LINE  
ORDER BY TOTAL_REVENUE DESC;
```

	PRODUCT_LINE	TOTAL_REVENUE
1	Fashion accessories	1042.65

7. What is the city with the largest revenue?

```
SELECT TOP 1 CITY,ROUND(SUM(TOTAL),2) AS TOTAL_SALES  
FROM walmartSales  
GROUP BY CITY  
ORDER BY TOTAL_SALES DESC;
```

	CITY	TOTAL_SALES
1	Naypyitaw	110568.71

8. What product line had the largest VAT?

```
SELECT TOP 1 PRODUCT_LINE,ROUND(MAX(TAX_5),2) AS LARGEST_VAT  
FROM walmartSales  
GROUP BY PRODUCT_LINE  
ORDER BY LARGEST_VAT DESC;
```

	PRODUCT_LINE	LARGEST_VAT
1	Fashion accessories	49.65

9. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales

```
--CREATE A NEW COLUMN(REVIEW)--  
ALTER TABLE walmartSales  
ADD REVIEW VARCHAR(10);  
  
UPDATE walmartSales  
SET REVIEW = CASE  
    WHEN TOTAL>(SELECT ROUND(AVG(TOTAL),2) FROM walmartSales) THEN 'GOOD'  
    ELSE 'BAD'  
END;
```

(1000 rows affected)

Completion time: 2024-05-29T08:56:46.1166103+05:30

10. Which branch sold more products than average product sold?

```
SELECT TOP 1 BRANCH, SUM(QUANTITY) AS TOTAL_QUANTITY  
FROM walmartSales  
WHERE QUANTITY > (SELECT AVG(QUANTITY) FROM walmartSales)  
GROUP BY BRANCH  
ORDER BY TOTAL_QUANTITY DESC;
```

	BRANCH	TOTAL_QUANTITY
1	C	1385

11. What is the most common product line by gender?

```
select Gender, PRODUCT_LINE, count(product_line) AS  
COUNT_PRODUCT_LINE  
from walmartSales  
group by Gender, PRODUCT_LINE  
ORDER BY COUNT_PRODUCT_LINE DESC;
```

	Gender	PRODUCT_LINE	COUNT_PRODUCT_LINE
1	Female	Fashion accessories	96
2	Female	Food and beverages	90
3	Male	Health and beauty	88
4	Female	Sports and travel	88
5	Male	Electronic accessories	86
6	Female	Electronic accessories	84
7	Male	Food and beverages	84
8	Male	Fashion accessories	82
9	Male	Home and lifestyle	81
10	Female	Home and lifestyle	79
11	Male	Sports and travel	78
12	Female	Health and beauty	64

12. What is the average rating of each product line?

```
SELECT PRODUCT_LINE, ROUND(AVG(Rating), 2) AS AVG_RATING  
FROM walmartSales  
GROUP BY PRODUCT_LINE  
ORDER BY AVG_RATING DESC;
```

	PRODUCT_LINE	AVG_RATING
1	Food and beverages	7.11
2	Fashion accessories	7.03
3	Health and beauty	7
4	Electronic accessories	6.92
5	Sports and travel	6.92
6	Home and lifestyle	6.84

Customer Analysis

1. How many unique customer types does the data have?

```
SELECT COUNT(DISTINCT(CUSTOMER_TYPE)) AS TOTAL_UNIQUE_CUSTOMER  
FROM walmartSales;
```

	TOTAL_UNIQUE_CUSTOMER
1	2

2. How many unique payment methods does the data have?

```
SELECT COUNT(DISTINCT(PAYMENT)) AS TOTAL_UNIQUE_PAYMENT_METHOD  
FROM walmartSales;
```

	TOTAL_UNIQUE_PAYMENT_METHOD
1	3

3. Which customer type buys the most?

```
SELECT TOP 1 CUSTOMER_TYPE, SUM(QUANTITY) AS BUYS_THE_MOST  
FROM walmartSales  
GROUP BY CUSTOMER_TYPE  
ORDER BY BUYS_THE_MOST DESC;
```

	CUSTOMER_TYPE	BUYS_THE_MOST
1	Member	2785

4. What is the gender of most of the customers?

```
SELECT TOP 1 GENDER, COUNT(CUSTOMER_TYPE) AS  
COUNT_OF_CUSTOMER_TYPE  
FROM walmartSales  
GROUP BY GENDER  
ORDER BY COUNT_OF_CUSTOMER_TYPE DESC;
```

	GENDER	COUNT_OF_CUSTOMER_TYPE
1	Female	501

5. What is the gender distribution per branch?

```
SELECT GENDER, BRANCH, COUNT(GENDER) AS distribution_per_branch  
FROM walmartSales  
GROUP BY GENDER, BRANCH  
ORDER BY BRANCH;
```

	GENDER	BRANCH	distribution_per_branch
1	Female	A	161
2	Male	A	179
3	Female	B	162
4	Male	B	170
5	Female	C	178
6	Male	C	150

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

```
SELECT TOP 1 TIME_OF_DAY, COUNT(RATING) AS MAX_RATING  
FROM walmartSales  
GROUP BY TIME_OF_DAY  
ORDER BY MAX_RATING DESC;
```

	TIME_OF_DAY	MAX_RATING
1	Afternoon	528

7. Which time of the day do customers give most ratings per branch?

```
WITH RatingsCount AS (  
    SELECT BRANCH, TIME_OF_DAY, ROUND(AVG(RATING), 2) AS RATING_COUNT  
    FROM walmartSales  
    GROUP BY BRANCH, TIME_OF_DAY),  
    MaxRatings AS (SELECT BRANCH, MAX(RATING_COUNT) AS  
        MAX_RATING_COUNT  
    FROM RatingsCount  
    GROUP BY BRANCH)  
    SELECT rc.BRANCH, rc.TIME_OF_DAY, rc.RATING_COUNT AS MAX_RATING  
    FROM RatingsCount rc  
    INNER JOIN  
    MaxRatings mr  
    ON rc.BRANCH = mr.BRANCH  
    AND rc.RATING_COUNT = mr.MAX_RATING_COUNT  
    ORDER BY rc.RATING_COUNT DESC;
```

	BRANCH	TIME_OF_DAY	MAX_RATING
1	C	Afternoon	7.1
2	A	Afternoon	7.06
3	B	Morning	6.89

8. Which day for the week has the best avg ratings?

```
select top 1 day_name,round(avg(rating),2) as average_rating  
FROM walmartSales  
group by day_name  
order by average_rating desc;
```

	day_name	average_rating
1	Mon	7.15

9. Which day of the week has the best average ratings per branch?

```
WITH RatingsCount AS (  
    SELECT BRANCH,day_name,cast(AVG(RATING) as decimal(10,2)) AS  
RATING_COUNT  
    FROM walmartSales  
    GROUP BY BRANCH,day_name  
)  
,  
MaxRatings AS (  
    SELECT BRANCH,MAX(RATING_COUNT) AS MAX_RATING_COUNT  
    FROM RatingsCount  
    GROUP BY BRANCH  
)  
SELECT rc.BRANCH,rc.day_name,rc.RATING_COUNT AS MAX_RATING  
FROM RatingsCount rc  
INNER JOIN  
    MaxRatings mr  
ON rc.BRANCH = mr.BRANCH  
AND rc.RATING_COUNT = mr.MAX_RATING_COUNT  
ORDER BY rc.BRANCH ;
```

	BRANCH	day_name	MAX_RATING
1	A	Fri	7.31
2	B	Mon	7.34
3	C	Fri	7.28

Sales Analysis

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

```
SELECT DAY_NAME, TIME_OF_DAY, COUNT(TOTAL) AS NUMBER_OF_SALES
FROM walmartSales
GROUP BY DAY_NAME, TIME_OF_DAY
ORDER BY CASE
WHEN DAY_NAME= 'SUN' THEN 1
WHEN DAY_NAME= 'MON' THEN 2
WHEN DAY_NAME= 'TUE' THEN 3
WHEN DAY_NAME= 'WED' THEN 4
WHEN DAY_NAME= 'THU' THEN 5
WHEN DAY_NAME= 'FRI' THEN 6
ELSE 7
END,
CASE
WHEN TIME_OF_DAY= 'MORNING' THEN 1
WHEN TIME_OF_DAY= 'AFTERNOON' THEN 2
ELSE 3
END;
```

	DAY_NAME	TIME_OF_DAY	NUMBER_OF_SALES
1	Sun	Morning	22
2	Sun	Afternoon	70
3	Sun	Evening	41
4	Mon	Morning	21
5	Mon	Afternoon	75
6	Mon	Evening	29
7	Tue	Morning	36
8	Tue	Afternoon	71
9	Tue	Evening	51
10	Wed	Morning	22
11	Wed	Afternoon	81
12	Wed	Evening	40
13	Thu	Morning	33
14	Thu	Afternoon	76
15	Thu	Evening	29
16	Fri	Morning	29
17	Fri	Afternoon	74
18	Fri	Evening	36
19	Sat	Morning	28
20	Sat	Afternoon	81
21	Sat	Evening	55

2. Which of the customer types brings the most revenue?

```
SELECT TOP 1 CUSTOMER_TYPE,ROUND(SUM(TOTAL),2) AS MOST_REVENUE  
FROM walmartSales  
GROUP BY CUSTOMER_TYPE  
ORDER BY MOST_REVENUE DESC;
```

	CUSTOMER_TYPE	MOST_REVENUE
1	Member	164223.44

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

```
SELECT TOP 1 CITY,MAX(TAX_5) AS MAX_VAT  
FROM walmartSales  
GROUP BY CITY  
ORDER BY MAX_VAT DESC;
```

	CITY	MAX_VAT
1	Naypyitaw	49.6500015258789

4. Which customer type pays the most in VAT?

```
SELECT TOP 1 CUSTOMER_TYPE,ROUND(MAX(TAX_5),2) AS MAX_VAT_PAYS  
FROM walmartSales  
GROUP BY CUSTOMER_TYPE  
ORDER BY MAX_VAT_PAYS DESC;
```

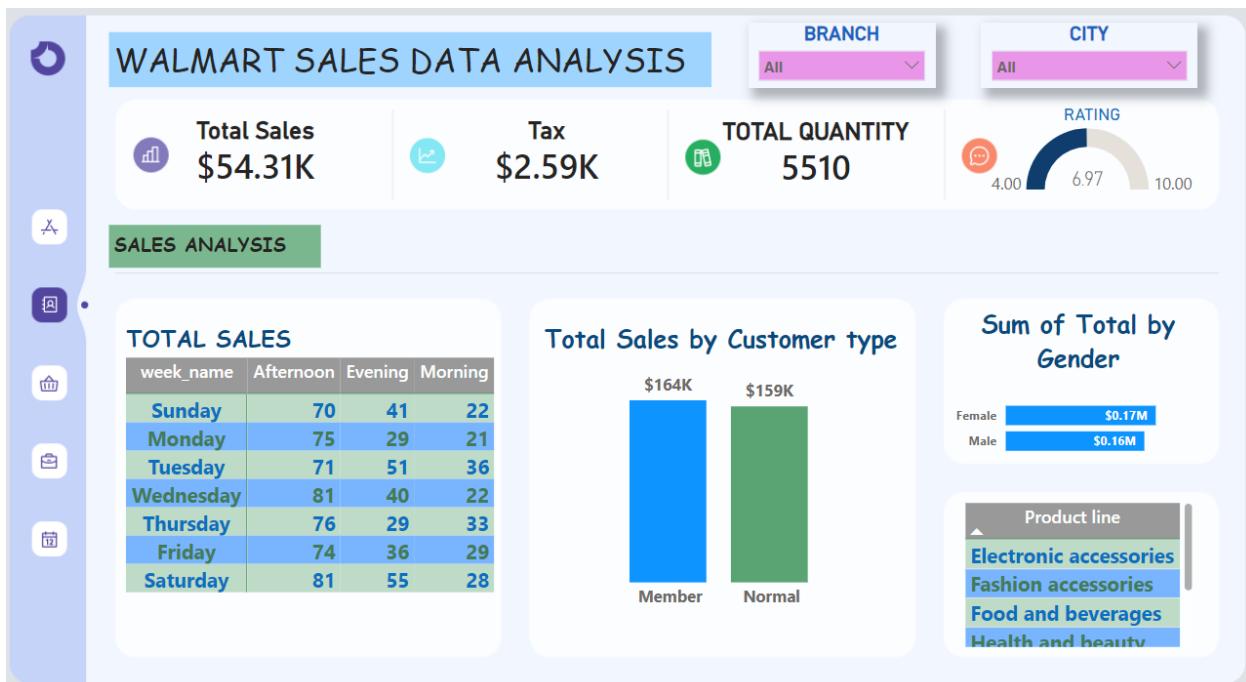
	CUSTOMER_TYPE	MAX_VAT_PAYS
1	Member	49.65

WALMART SALES REPORT – POWER B.I (BUSINESS INTELLIGENCE)

OVERVIEW



SALES ANALYSIS



PRODUCT ANALYSIS

WALMART SALES DATA ANALYSIS

Total Sales \$54.31K Tax \$2.59K Branch A Product Line Fashion accessories

PRODUCT ANALYSIS

Products purchased

Product line	Male	Female
Sports and travel	17.23%	16.77%
Home and lifestyle	16.43%	19.16%
Health and beauty	16.83%	17.96%
Food and beverages	17.64%	12.77%
Fashion accessories	16.23%	15.77%
Electronic accessories	15.63%	17.56%

Quick Summary

Fashion accessor...	\$1,039.29
Product line	Maximum Sales
Yangon	\$16,332.5085
City	Total Sales
A	263
Branch	Total Quantity

WEEKS

Sunday	Tuesday
Monday	Wednesday

City

- Yangon
- Mandalay
- Naypyitaw

WALMART SALES DATA ANALYSIS

REVIEW

REVIEW ● BAD ● GOOD

Category	BAD (%)	GOOD (%)
Home and lifestyle	48.34%	51.66%
Sports and travel	49.20%	50.80%
Health and beauty	50.43%	49.57%
Food and beverages	46.29%	53.71%
Electronic accessories	48.07%	51.93%
Fashion accessories	54.09%	45.91%

Total Sales and gross margin

Total sales by month

Month	Total Sales
January	\$116K
February	\$97K
March	\$109K

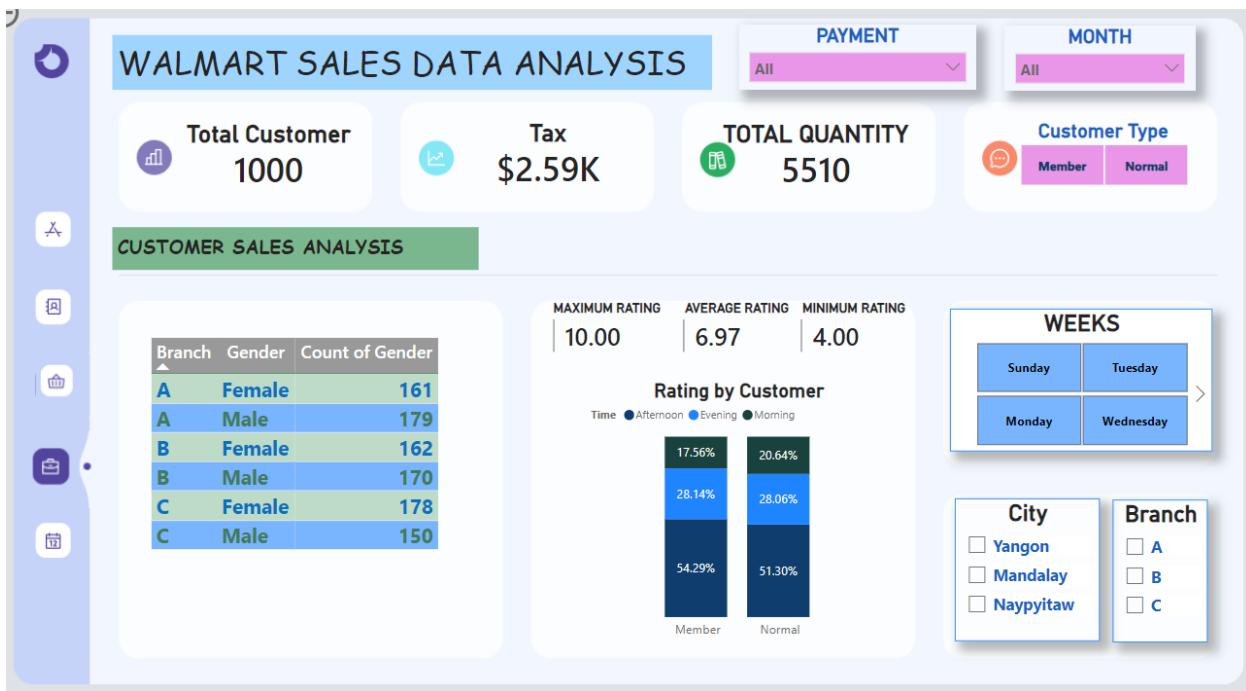
Average Rating

Product line	Rating
Electronic accessories	5.71
Home and lifestyle	5.69
Health and beauty	5.62
Sports and travel	5.54
Food and beverages	5.47
Fashion accessories	5.07

Gross Income

Product line	Gross Income
Health and beauty	2.34K
Home an...	2.56K
Fashion accessories	2.59K
Food and beverages	2.67K
Sports an...	2.62K
Electronic accessories	2.59K

CUSTOMER ANALYSIS



SUMMARIZE

1. Overall Sales Performance

- Total Sales: \$54.31K, highlighting the overall revenue generated.
- Total Tax Collected: \$2.59K, indicating the tax revenue.

2. Branch and Product Line Focus

- Branch A: The primary branch of interest.
- Fashion Accessories: Noted for significant sales within this branch.

3. Product Purchases by Gender

- Male vs. Female: Sales are analyzed across different product lines (e.g., Sports and travel, Home and lifestyle, Health and beauty), showcasing gender preferences in purchasing.

4. Sales Breakdown by City

- Cities: Yangon (\$16,332.51), Mandalay, and Naypyitaw, providing geographical insights into sales distribution.

5. Weekly Sales Trends

- **Sales by Day:** Sales performance segmented by each day of the week (Sunday to Saturday), offering insights into daily sales patterns.

6. Customer Reviews and Ratings

- **Good vs. Bad Reviews:** Percentage of positive and negative reviews for each product line, giving a sense of customer satisfaction.
- **Average Ratings:** Average ratings for different product lines, indicating overall customer satisfaction levels.

7. Sales and Gross Margin Trends

- **Sales and Gross Margin Over Time:** A line graph showing trends, helping in understanding sales and profitability over different periods.

8. Monthly Sales Analysis

- **Sales by Month:** Total sales figures for January, February, and March, helping to identify monthly trends.

9. Gross Income by Product Line

- **Gross Income Distribution:** A pie chart showing the contribution of different product lines to the gross income.

10. Customer Analysis

- **Total Customers:** 1000 customers, giving an overview of the customer base size.
- **Gender Distribution:** Customer gender distribution across different branches.
- **Ratings by Customer Type and Time:** Differences in ratings between member and normal customers, and ratings across different times of the day (Morning, Afternoon, Evening).

11. Sales Analysis

- **Sales by Weekday and Time:** Detailed sales data by weekday and time of day.
- **Sales by Customer Type:** Total sales figures broken down by member and normal customers.
- **Sales by Gender:** Sum of total sales categorized by gender.

Key Insights

- **Highest Sales Product Line:** Fashion accessories.
- **Top Performing City:** Yangon.
- **Customer Preferences:** Gender-based preferences in product purchases.
- **Daily and Monthly Trends:** Identifying peak sales days and months.
- **Customer Satisfaction:** Overall good reviews and high average ratings in most product lines.
- **Sales Distribution:** Detailed analysis showing sales trends by various demographics and times.

SUGGESTION

1. Enhance Popular Product Lines

- **Fashion Accessories:** Since this product line has significant sales, consider expanding the range of fashion accessories, introducing exclusive or limited-edition items, and running targeted promotions.
- **Product Bundling:** Create attractive bundles that include fashion accessories and other related products to increase average transaction values.

2. Focus on High-Performing Branches and Cities

- **Yangon:** Given its high sales figures, increase marketing efforts in Yangon. Invest in local promotions, community events, and partnerships to further enhance brand presence and customer loyalty.
- **Branch-Specific Promotions:** Tailor promotions specific to branches showing potential for growth. For instance, offer branch-specific discounts or loyalty rewards.

3. Improve Customer Experience Based on Reviews

- **Address Negative Reviews:** Focus on improving product lines that have lower ratings or higher negative reviews. This could involve quality checks, customer feedback loops, and enhancements based on customer suggestions.
- **Encourage Positive Reviews:** Incentivize customers to leave reviews by offering discounts or loyalty points for every review submitted, especially positive ones.

4. Leverage Gender-Based Insights

- **Targeted Marketing Campaigns:** Develop gender-specific marketing campaigns to appeal to male and female customers based on their purchasing preferences. For example, advertise health and beauty products more heavily to female customers.
- **Expand Popular Categories:** Since certain categories are popular with specific genders, consider expanding these categories and introducing new products that align with these preferences.

5. Optimize Sales Timing

- **Peak Days and Times:** Increase marketing efforts, promotions, and staffing on days and times with higher sales. For example, if weekends show higher sales, consider special weekend deals or events.
- **Time-Based Discounts:** Offer time-specific discounts to attract customers during slower periods, such as early morning or late evening specials.

6. Enhance Membership Program

- **Increase Membership Benefits:** Make the membership program more attractive by adding exclusive benefits, such as special discounts, early access to sales, or members-only events.
- **Promote Membership Sign-Ups:** Run campaigns to increase the number of members. Offer incentives for sign-ups, such as a discount on the first purchase or bonus loyalty points.

7. Expand Digital and Online Sales Channels

- **E-commerce Integration:** Improve the online shopping experience by ensuring a seamless and user-friendly e-commerce platform. Highlight popular and highly-rated products.
- **Digital Marketing:** Invest in digital marketing strategies, including social media advertising, email marketing, and search engine optimization, to reach a broader audience.

8. Geographical Expansion

- **New Markets:** Consider expanding to new geographical areas showing potential for growth. Conduct market research to identify such regions.
- **Local Partnerships:** Form partnerships with local businesses in new markets to increase brand visibility and reach.

9. Customer Loyalty Programs

- **Loyalty Rewards:** Enhance the existing loyalty programs to encourage repeat purchases. Offer rewards points, exclusive discounts, and special deals for frequent shoppers.
- **Feedback Programs:** Implement programs where customers can earn rewards for providing feedback, helping improve products and services while increasing customer engagement.

10. Decision Making

- **Regular Analysis:** Continuously analyze sales data to identify trends, opportunities, and areas for improvement. Use this data to make informed decisions regarding inventory, marketing, and customer service strategies.
- **Predictive Analytics:** Utilize predictive analytics to forecast demand, manage stock levels efficiently, and personalize marketing efforts to individual customer preferences.