



SQL Capstone Project: Amazon

Name- Amol shinde

Introduction:

- **Amzon:**
- **Dataset:**
 - originally dataset has 17 columns and 1000 rows.
 - data is from three major cities of Myanmar name as Mandalay, Yangon and Naypyitaw

Objective:

- **The major aim of this project:**
 1. To **gain insight into the sales data** of Amazon and
 2. To understand the different factors that affect sales of the different branches.

- **Analysis to be done:**
 1. *Product Analysis*- Products lines performing best and the product lines that need to be improved.

 2. *Sales Analysis*- Sales trends of product, sales strategy the business applies and what modifications are needed to gain more sales.

 3. *Customer Analysis*- To uncover the different customer segments, purchase trends and the profitability of each customer segment.

Approach Used:

1- Data Wrangling:

- Created database in MySQL workbench and imported amazon.csv file in database
- checking missing values or null values

2- Feature Engineering: Added 3 new column for better sales analysis

- timeofday - Morning, Afternoon, Evening
- dayname - Sunday, Monday, Tuesday etc.
- monthname - January, February and so on
- Change of datatype of Time and Date.

Analysis:

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

Ans: 3

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

Ans: A - Yangon
B - Mandalay
C - Naypyitaw

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

Ans: 6

- 4. Which payment method occurs most frequently?

Ans: Ewallet - 345 times

- 5. Which product line has the highest sales?

Ans: Electronic Accessories - 971 (according to Quantity of goods)

- 6. How much revenue is generated each month?

Ans: January - 116292

March - 109456

February - 97219

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

Ans: January - 110754 (total_cogs_sold)

- 8. Which product line generated the highest revenue?

Ans: Food and beverages - 56145(total_revenue)

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

Ans: Naypyitaw - 110569 (Highest_revenue)

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

Ans: Food and beverages - 2673.56 (Highest_VAT)

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

| | | |
|-------------------------|-----------|----------|
| Fashion accessories- | 75 (Good) | 103(bad) |
| Sports and travel - | 74 (Good) | 92(bad) |
| Food and beverages- | 67 (Good) | 107(bad) |
| Electronic accessories- | 67 (Good) | 103(bad) |
| Home and lifestyle - | 63 (Good) | 97(bad) |
| Health and beauty- | 60 (Good) | 92(bad) |

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

Ans: Branch -A 1859 1836.6667

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

Ans: Female - Fashion accessories -96

Male - Health and beauty - 88

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

Ans: Food and beverages- 7.113218390804598

Fashion accessories - 7.029213483146067

Health and beauty- 7.003289473684212

Electronic accessories- 6.92470588235294

Sports and travel - 6.916265060240964

Home and lifestyle -6.8375

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

Ans: Wednesday - Afternoon - 71

Saturday - Afternoon - 69

Friday - Afternoon - 68

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

Ans: Member - 164223 (Highest_revenue)

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

Ans: Mandalay- 4.761904761904766

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

Ans: Member- 7820

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

Ans: Member- 501

Normal -499

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

Ans: Ewallet - 345

Cash - 344

Credit card - 311

- 21. Which customer type occurs most frequently?

Ans: Member- 501

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

Ans: Member

- 23. Determine the predominant gender among customers.?

Ans: Female

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

Ans: A - Male - 179

Female - 161

B - Male - 170

Female - 162

C Male - 150

Female - 178

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

Ans: Afternoon - 454

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

Ans: A - Afternoon - 158

C - Afternoon - 154

B - Afternoon - 142

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

Ans: Monday - 7.153

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

Ans: C - Friday - 7.27

A - Friday - 7.31

B - Monday - 7.33

Conclusion:

1 . Product Analysis:

- product lines: 6
- best performing is food and beverages
- under performing is Health and beauty

2 . Sales Analysis:

- High sales and revenue generated are in January and low is february
- City with highest sales is Naypyitaw
- Ewallet is mostly used payment method
 - Female - Fashion accessories -96
 - Male - Health and beauty - 88

3. Customer Analysis:

- member - 501 and normal - 499
- frequent buyer and most profitable is members

Recommendations:

- Focus on High-Performing Product Lines
- Find cause of Underperforming Product Lines
- Customer Feedback Integration
- Seasonal Promotions
- Geographic Expansion
- Personalized Marketing
- Customer Satisfaction Improvement



Thank you



Want to make a presentation like this one?

Start with a fully customizable template, create a beautiful deck in minutes, then easily share it with anyone.

Create a presentation (It's free)