Amazon Sales data Analysis Report

By Himanshi Agarwal

Overview

This dataset contains sales transactions from three different branches of Amazon, respectively located in Mandalay, Yangon and Naypyitaw. The data contains 17 columns and 1000 rows

Objective

- •Understand what affects sales performance at three different branches.
- •Identify trends in sales and customer behavior.
- •Generate insights to improve business strategies.
- Answer Business Questions

Approach Used

- **Data Wrangling**: Build a database, Create the table, import the data from dataset and filter the NULL values.
- **Feature Engineering**: Create new columns like "timeofday" (Morning, Afternoon, Evening), "dayname" (days of the week), and "monthname" (months) to analyze sales patterns.
- Exploratory Data Analysis (EDA): Perform EDA to answer key business questions and achieve project goals.

Feature Engineering

Created new columns like "timeofday" (Morning, Afternoon, Evening), "dayname" (days of the week), and "monthname" (months)

```
#check data type and convert date, time and id column with correct datatype
SET SQL SAFE UPDATES = 0;
                                                                                                                 54
                                                                                                                 55
UPDATE amazon
                                                                                                                 56
SET date = STR_TO_DATE(date, '%d-%m-%Y')
                                                                                                                 57
SET SQL SAFE UPDATES = 1;
                                                                                                                 58
ALTER TABLE amazon
                                                                                                                 59
                                                                                     Limit to 1000 rows
MODIFY date DATE;
                                          39
                                                                                                                 61
alter table amazon
                                          40
modify Time Time;
                                                  #Feature Engineering:
                                                                                                                 63
ALTER TABLE amazon
                                                  #Add new columns for timeofday, dayname, and monthname
MODIFY 'Invoice ID' VARCHAR(30);
                                                  alter Table amazon
                                                  add timeofday varchar(10);
                                          44
                                          45
                                                  alter Table amazon
                                          47
                                                  add dayname varchar(10);
                                          48
                                                  alter Table amazon
                                                  add monthname varchar(10);
```

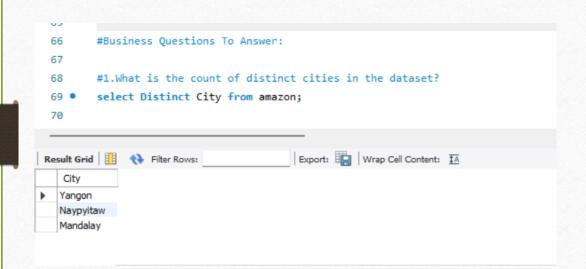
```
update amazon

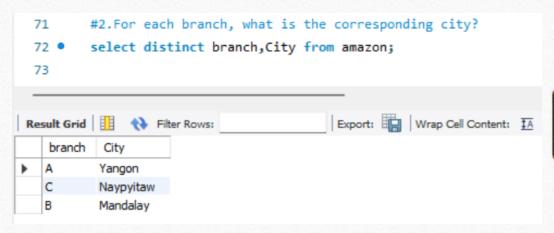
⊖ set timeofday=case

        when Time(time) BETWEEN '06:00:00' AND '11:59:59' THEN 'Morning'
        when Time(time) BETWEEN '12:00:00' AND '17:59:59' THEN 'Afternoon'
        when Time(time) BETWEEN '18:00:00' AND '23:59:59' THEN 'Evening'
        ELSE 'Night'
         END;
        update amazon
         set dayname=Dayname(date);
        update amazon
         set monthname=Monthname(date);
         select timeofday, dayname, monthname from amazon;
                                           Export: Wrap Cell Content: TA Fetch rows:
Result Grid
              Filter Rows:
   timeofday
            dayname
                        monthname
  Afternoon
            Saturday
                        January
            Friday
                        March
  Morning
            Sunday
                        March
  Afternoon
            Sunday
                        January
  Evening
            Friday
  Morning
                        February
```

Business Questions To Answer:

1. What is the count of distinct cities in the dataset? 2. For each branch, what is the corresponding city





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



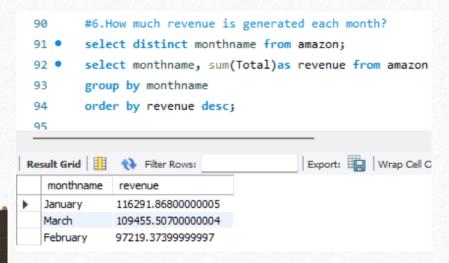
5. Which product line has the highest sales?

4. Which payment method occurs most frequently?

```
#4.Which payment method occurs most frequently?
    78 •
             select Payment, count(*) as frequency
             from amazon
    80
             group by Payment
             order by frequency Desc
    81
             limit 1;
    82
                                                 Export: Wrap Cel
   Result Grid Filter Rows:
      Payment frequency
  Ewallet
                345
        #5.Which product line has the highest sales?
        select `Product line`,count(*) as `highest sales` from amazon
        group by 'Product line'
 86
        order by 'highest sales' desc
 87
        limit 1;
 88
                                       Export: Wrap Cell Content: A Fel
Result Grid Filter Rows:
                  highest
   Product line
                  sales

    Fashion accessories
```

6. How much revenue is generated each month?



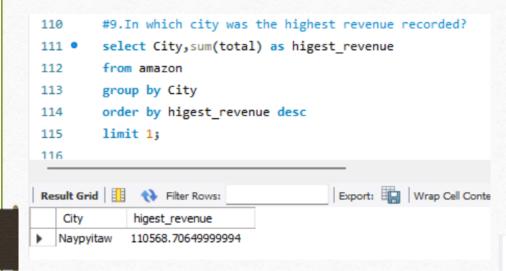
7.In which month did the cost of goods sold reach its peak?

```
96
         #7. In which month did the cost of goods sold reach its peak?
         select monthname, sum(cogs)as totalcost_of_goods
         from amazon
 98
         group by monthname
 99
         order by totalcost of goods desc
100
         limit 1;
101
102
               ♦ Filter Rows:
                                            Export: Wrap Cell Content: TA Fe
Result Grid
   monthname
              totalcost of goods
 January
              110754.160000000002
```

8. Which product line generated the highest revenue?

```
#8.Which product line generated the highest revenue?
103
         select `Product line`,sum(total) as higest revenue
         from amazon
105
         group by 'Product line'
106
         order by higest revenue desc
107
108
         limit 1;
                                           Export: Wrap Cell Cont
Result Grid | Filter Rows:
    Product line
                     higest_revenue
Food and beverages
                    56144.844000000005
```

9.In which city was the highest revenue recorded?

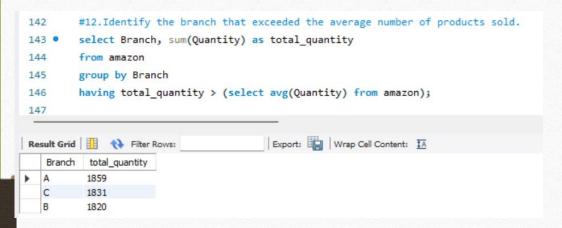


11.For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."

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

```
#10.Which product line incurred the highest Value Added Tax?
     117
               select `Product line`,sum(`Tax 5%`) as total tax
     118 •
     119
               from amazon
               group by 'Product line'
      120
               order by total_tax desc
     121
               limit 1;
      122
                                                    Export: Wrap Cell Content: TA Fe
     Result Grid
                    Filter Rows:
         Product line
                            total tax
        Food and beverages
                           2673.5639999999994
        UPDATE amazon a
      O JOIN (
129
130
            SELECT 'Product line',
                   CASE
131
132
                       WHEN SUM(Total) > (SELECT AVG(Total) FROM amazon) THEN 'Good'
133
134
                   END AS sales classification
135
            FROM amazon
            GROUP BY 'Product line'
136
137
        ) AS subquery ON a. Product line = subquery. Product line
138
        SET a.sales classification = subquery.sales classification;
139 •
        select `Product line`, sales classification from amazon;
140
Export: Wrap Cell Content: IA Fetch rows:
   Product line
                    sales classification
  Health and beauty
                    Good
  Electronic accessories
                    Good
  Home and lifestyle
  Health and beauty
                    Good
  Sports and travel
                    Good
  Electronic accessories
                   Good
```

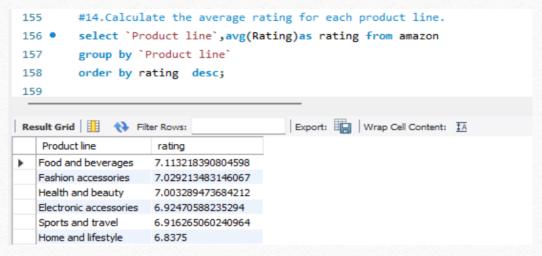
12. Identify the branch that exceeded the average number of products sold.

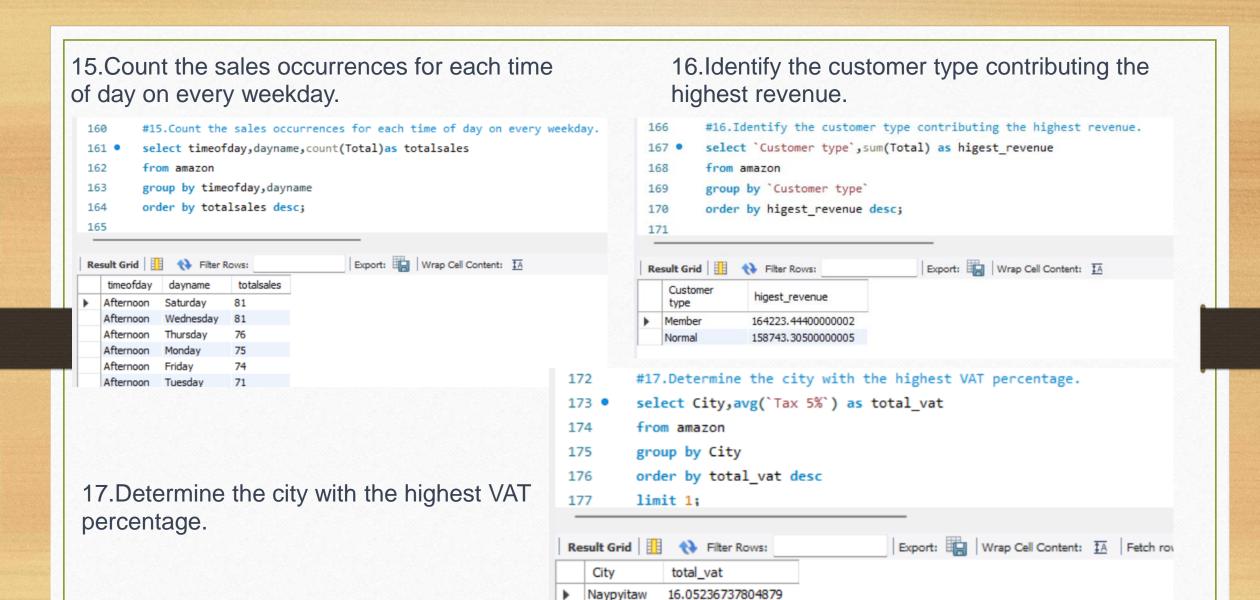


13. Which product line is most frequently associated with each gender?

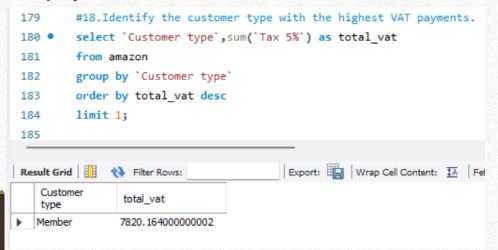
```
#13. Which product line is most frequently associated with each gender?
148
149 •
         select Gender,`Product line`,count(*) as frequency
150
         from amazon
        group by Gender, `Product line`
151
152
        Order By Frequency Desc
153
        limit 1;
154
                                          Export: Wrap Cell Content: A Fetch rows:
Result Grid Filter Rows:
          Product line
                           frequency
 Female Fashion accessories
```

14. Calculate the average rating for each product line.

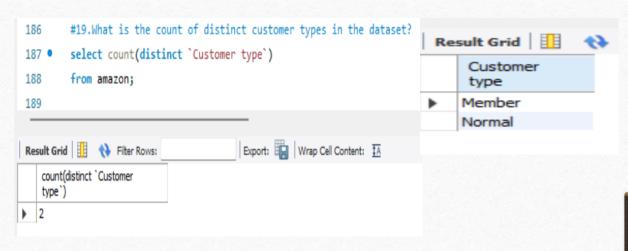




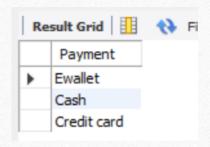
18. Identify the customer type with the highest VAT payments.



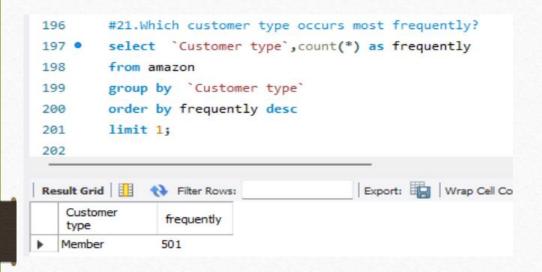
19. What is the count of distinct customer types in the dataset?



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



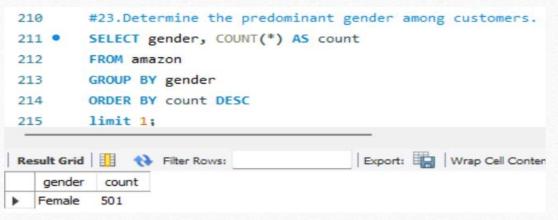
21. Which customer type occurs most frequently?



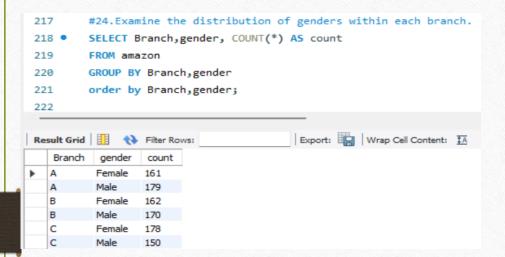
22.Identify the customer type with the highest purchase frequency.

```
#22. Identify the customer type with the highest purchase frequency.
203
        select `Customer type`,count(Quantity) as purchase_count
        from amazon
        group by 'Customer type'
206
        order by purchase count desc
207
        limit 1;
208
                                          Export: Wrap Cell Content: A Fetch rows:
              Filter Rows:
Result Grid
   Customer
                purchase count
   type
 Member
               501
```

23. Determine the predominant gender among customers.

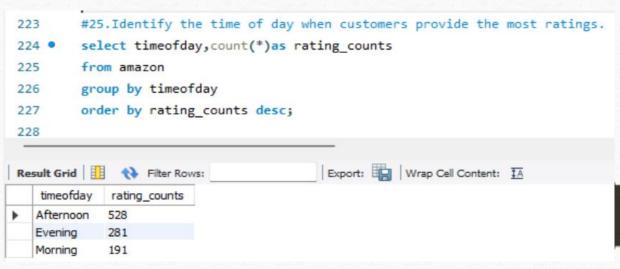


24.Examine the distribution of genders within each branch.



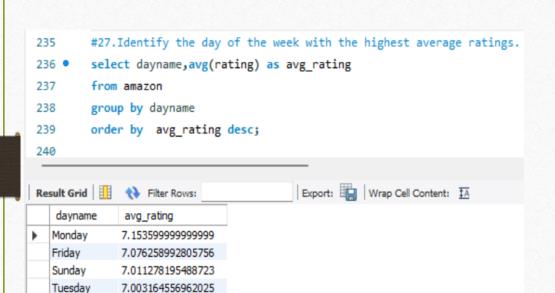
26.Determine the time of day with the highest customer ratings for each branch.

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



229 #26.Determine the time of day with the highest customer ratings for each branch. select timeofday, branch, avg(rating) as avg_rating 230 • 231 232 group by timeofday, branch 233 order by avg_rating desc; 234 Export: Wrap Cell Content: TA Result Grid Filter Rows: branch avg_rating timeofday Afternoon Evening 7.0920454545454525 Afternoon 7.0567567567567595 7.005479452054794 Evening 6.979268292682928 6.974576271186442

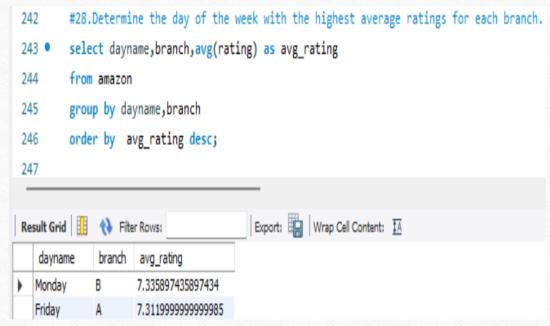
27. Identify the day of the week with the highest average ratings.



6.901829268292688

6.88985507246377

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



Analysis list

1. Product Analysis:

- •Available Products: Health and Beauty, Electronic Accessories, Home and Lifestyle, Sports and Travel, Food and Beverages, and Fashion Accessories are sold in three branches: A, B, and C.
- •Top Performer: Food and Beverages have the highest ratings, most sales, and generate the most revenue, making it the best product line.
- •Needs Improvement: Home and Lifestyle has the lowest ratings, so it needs improvement.
- •Gender Preferences: Fashion Accessories are mostly bought by females, while Health and Beauty products are mostly bought by males.
- •Focus on Females: Since most customers are female, improving products they like could increase sales. Enhancing Health and Beauty products could also attract more male customers.

2. Sales Analysis:

- •Highest Sales: Food and Beverages have the highest sales.
- •Best Month: January has the most sales, highest profits, and the highest revenue.
- •Best Day: Sunday sees the highest total sales
- •Best Time: Most sales happen in the afternoon.
- •Top Branch: Naypyitaw branch makes the most revenue.
- •Tax Focus: Yangon branch has the highest VAT percentage.

Branch Trends: Branch A sells more products than average. In Branch A and B, more male customers shop, while Branch C has more female customers.

3. Customer Analysis:

- •Popular Customer Type: Most customers are members.
- •Frequent Shoppers: Members shop the most often and pay the most VAT.
- •Gender Majority: Most customers are female.
- •Preferred Payment: Ewallet is the most frequently used payment method

Recommendations:

Improve Home and Lifestyle products to make customers happier with this category.

- •Focus on Food and Beverages, which are the bestselling and highest-rated products, to keep them performing well.
- •Enhance Health and Beauty products to attract more male customers.
- •Plan special sales and promotions in January and on Sundays, as these times have the highest sales.

Adjust strategies for each branch, especially focusing on Naypyitaw for its high revenue and Yangon for its VAT.

Thank You