



AMAZON SALES ANALYSIS

- S10117

- ARIF SHAIK

OBJECTIVE OF PROJECT:

The primary objective of this project is to analyze Amazon's sales data to gain insights into the various factors influencing sales across different branches.

OVERVIEW OF PROJECT:

The dataset comprises sales records from three branches in Myanmar: Naypyitaw, Yangon, and Mandalay, covering the first quarter of 2019. It contains a total of 1,000 rows and 17 columns.

APPROACH USED:

DATA WRANGLING:

This initial step involves inspecting the data to identify NULL and missing values. Appropriate data replacement methods are then applied to address these issues, ensuring the dataset is clean and ready for analysis.

FEATURE ENGINEERING:

This process involves creating new columns derived from existing ones, enhancing the dataset's richness and enabling more insightful analysis.

ANSWERING BUSINESS QUESTIONS

Q1: What is the count of distinct cities in the dataset?

```
9 • SELECT COUNT(distinct(City)) AS Distinct_Cities  
0 FROM amazon;
```

Result Grid	Filter Rows:	Export:	Wrap
Distinct_Cities			
3			

Q2: For each branch, what is the corresponding city?

```
86 • SELECT DISTINCT City, Branch  
87 FROM amazon;
```

Result Grid	Filter Rows:
City	Branch
Yangon	A
Naypyitaw	C
Mandalay	B

Q3: What is the count of distinct product lines in the dataset?

```
93 • SELECT COUNT(DISTINCT(`Product line`)) as Distinct_Pl  
94 FROM amazon;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
Distinct_Pl			
6			

Q4: Which payment method occurs most frequently?

```
100 • SELECT Payment, COUNT(Payment) as Occurs  
101 FROM amazon  
102 GROUP BY Payment  
103 ORDER BY Occurs DESC;
```

Result Grid	Filter Rows:	Export:
Payment	Occurs	
Ewallet	345	
Cash	344	
Credit card	311	

Q5: Which product line has the highest sales?

```

109 • SELECT `Product line`, SUM(Quantity) Quantity_Sold, ROUND(SUM(Quantity * `Unit price`), 3) as Total_Sales
110 FROM amazon
111 GROUP BY `Product line`
112 ORDER BY Total_Sales DESC
113 LIMIT 1;
114
115 -- 'Food and beverages' has generated Highest Total_Sales Amount (53471.28) with 952 Quantity Sold.
116 -- Where 'Electronic accessories' has Highest Quantity Sold with 971.

```

Product line	Quantity_Sold	Total_Sales
Food and beverages	952	53471.28

Q7: In which month did the cost of goods sold reach its peak?

```

129 • SELECT Month_Name, SUM(cogs) Total_Cogs
130 FROM amazon
131 GROUP BY Month_Name
132 ORDER BY Total_Cogs
133 LIMIT 1;

```

Month_Name	Total_Cogs
February	92589.88

Q6: How much revenue is generated each month?

```

120 • SELECT Month_Name, ROUND(SUM(Total), 3) Monthly_Revenue
121 FROM amazon
122 GROUP BY Month_Name
123 ORDER BY Month_Name;

```

Month_Name	Monthly_Revenue
February	97219.374
January	116291.868
March	109455.507

Q8: Which product line generated the highest revenue?

```

139 • SELECT `Product line`, ROUND(SUM(Total), 3) as Revenue
140 FROM amazon
141 GROUP BY `Product line`
142 ORDER BY Revenue DESC
143 LIMIT 1;

```

Product line	Revenue
Food and beverages	56144.844

Q9: In which city was the highest revenue recorded?

```

149 • SELECT City, ROUND(SUM(Total), 3) Revenue
150 FROM amazon
151 GROUP BY City
152 ORDER BY Revenue DESC
153 LIMIT 1;
154

```

Result Grid	Filter Rows:	Export:
City	Revenue	
Naypyitaw	110568.706	

Q10: Which product line incurred the highest Value Added Tax?

```

159 • SELECT `Product line`, MAX(`Tax 5%`) Highest_Vat,
160 ROUND(SUM(`Tax 5%`), 3) Vat_Amount
161 FROM amazon
162 GROUP BY `Product line`
163 ORDER BY Highest_Vat DESC
164 LIMIT 1;

```

Result Grid	Filter Rows:	Export:	Wrap Cell
Product line	Highest_Vat	Vat_Amount	
Fashion accessories	49.65	2585.995	

Q11: For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."

```

170 • SELECT `Product line`, sum(`Unit price` * Quantity) AS Total_Sales,
171 CASE
172 WHEN SUM(Quantity * `Unit price`) >
173 (SELECT SUM(`Unit price` * Quantity)/COUNT(DISTINCT(`Product line`))
174 FROM amazon) THEN 'Good'
175 ELSE 'Bad'
176 END Sales_Performance
177 FROM amazon
178 GROUP BY `Product line`
179 ORDER BY Total_Sales DESC;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
Product line	Total_Sales	Sales_Performance	
Food and beverages	53471.280000000006	Good	
Sports and travel	52497.930000000002	Good	
Electronic accessories	51750.029999999998	Good	
Fashion accessories	51719.899999999997	Good	
Home and lifestyle	51297.059999999998	Good	
Health and beauty	46851.179999999998	Bad	

Q12: Identify the branch that exceeded the average number of products sold.

```

186 • SELECT Branch, SUM(Quantity) as Total_Units_Sold
187 FROM amazon
188 GROUP BY Branch
189 HAVING Total_Units_Sold >
190 (SELECT SUM(Quantity) / COUNT(DISTINCT(Branch)) AS Avg_Quantity
191 FROM amazon);

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
Branch	Total_Units_Sold		
A	1859		

Q13: Which product line is most frequently associated with each gender?

```
197 • SELECT `Product line`,
198 SUM(CASE WHEN Gender = 'Female' THEN 1 ELSE 0 END) AS Female,
199 SUM(CASE WHEN Gender = 'Male' THEN 1 ELSE 0 END) AS Male
200 FROM amazon
201 GROUP BY `Product line`
202 ORDER BY Male DESC;
203
```

Product line	Female	Male
Health and beauty	64	88
Electronic accessories	84	86
Food and beverages	90	84
Fashion accessories	96	82
Home and lifestyle	79	81
Sports and travel	88	78

Q14: Calculate the average rating for each product line.

```
209 • SELECT `Product line`, ROUND(AVG(Rating), 2) Avg_Rating
210 FROM amazon
211 GROUP BY `Product line`
212 ORDER BY Avg_Rating DESC;
```

Product line	Avg_Rating
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

Q15: Count the sales occurrences for each time of day on every weekday.

```
218 • SELECT Day_Name, Time_of_day, COUNT(*) AS Sales_Count
219 FROM amazon
220 GROUP BY Day_Name, Time_of_day
221 ORDER BY FIELD(Day_Name, 'Sunday', 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday'),
222 FIELD(Time_of_day, 'Morning', 'Afternoon', 'Evening'); -- used Field for custom sorting
```

Day_Name	Time_of_day	Sales_Count
Sunday	Morning	22
Sunday	Afternoon	70
Sunday	Evening	41
Monday	Morning	21
Monday	Afternoon	75
Monday	Evening	29
Tuesday	Morning	36
Tuesday	Afternoon	71
Tuesday	Evening	51
Wednesday	Morning	22
Wednesday	Afternoon	81

Q16: Identify the customer type contributing the highest revenue.

```
229 • select `Customer type`, ROUND(SUM(Total), 3) as Revenue
230 FROM amazon
231 GROUP BY `Customer type`
232 ORDER BY Revenue DESC;
233
```

Customer type	Revenue
Member	164223.444
Normal	158743.305

Q17: Determine the city with the highest VAT percentage.

```

:38 • SELECT City, ROUND(SUM(`Unit price` * Quantity), 2) as Before_Tax,
:39     ROUND(SUM(Total), 2) AS After_Tax,
:40     ROUND(SUM(`Tax 5%`), 2) as Vat_Amount,
:41     ROUND((SUM(`Tax 5%`) / SUM(Total)) * 100, 2) AS Vat_Percentage --
:42 From amazon
:43 GROUP BY City
:44 ORDER BY Vat_Percentage DESC;
:45

```

City	Before_Tax	After_Tax	Vat_Amount	Vat_Percentage
Yangon	101143.21	106200.37	5057.16	4.76
Naypyitaw	105303.53	110568.71	5265.18	4.76
Mandalay	101140.64	106197.67	5057.03	4.76

Q18: Identify the customer type with the highest VAT payments.

```

250 • SELECT `Customer type`, ROUND(SUM(`Unit price` * Quantity), 2) as Before_Tax,
251     ROUND(SUM(Total), 2) AS After_Tax, ROUND(SUM(`Tax 5%`), 2) as VAT
252 FROM amazon
266 • SELECT COUNT(DISTINCT(Payment)) Distinct_Pm
267 FROM amazon;

```

Distinct_Pm
3

Q19: What is the count of distinct customer types in the dataset?

```

260 • SELECT COUNT(DISTINCT(`Customer type`)) as Distinct_Cust_Types
261 FROM amazon;

```

Distinct_Cust_Types
2

Q20: What is the count of distinct payment methods in the dataset?

```

266 • SELECT COUNT(DISTINCT(Payment)) Distinct_Pm
267 FROM amazon;

```

Distinct_Pm
3

Q21: Which customer type occurs most frequently?

```
272 • SELECT `Customer type`, COUNT(`Customer type`) AS Count_Cust
273 FROM amazon
274 GROUP BY `Customer type`;
```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	Customer type	Count_Cust			
▶	Member	501			
	Normal	499			

Q22: Identify the customer type with the highest purchase frequency.

```
0 • SELECT `Customer type`, COUNT(*) Purchase_Frequency,
1 SUM(Quantity) Quantity_Purchased, ROUND(SUM(Total), 2) Revenue_Generated
2 FROM amazon
3 GROUP BY `Customer type`
4 ORDER BY Revenue_Generated DESC;
```

Result Grid				Filter Rows:	Export:	Wrap Cell Content:
Customer type	Purchase_Frequency	Quantity_Purchased	Revenue_Generated			
Member	501	2785	164223.44			
Normal	499	2725	158743.31			

Q23: Determine the predominant gender among customers.

```
289 • SELECT Gender, COUNT(Gender) Dominant
290 FROM amazon
291 GROUP BY Gender
292 ORDER BY Dominant DESC
293 LIMIT 1;
```

Result Grid			Filter Rows:	Export:
	Gender	Dominant		
▶	Female	501		

Q24: Examine the distribution of genders within each branch.

```
298 • SELECT Branch,
299 SUM(CASE WHEN Gender = 'Female' THEN 1 ELSE 0 END) AS Female,
300 SUM(CASE WHEN Gender = 'Male' THEN 1 ELSE 0 END) AS Male
301 FROM amazon
302 GROUP BY Branch
303 ORDER BY Female, Male DESC;
304
```

Result Grid				Filter Rows:	Export:	Wrap Cell Content:
	Branch	Female	Male			
▶	A	161	179			
	B	162	170			
	C	178	150			

Q25: Identify the time of day when customers provide the most ratings.

```
309 • SELECT Time_of_day, COUNT(Rating) Most_Ratings
310 FROM amazon
311 GROUP BY Time_of_day
312 ORDER BY Most_Ratings DESC;
```

Result Grid	Filter Rows:	Export:	Wrap
Time_of_day	Most_Ratings		
Afternoon	528		
Evening	281		
Morning	191		

Q26: Determine the time of day with the highest customer ratings for each branch.

```
317 • SELECT Branch, Time_of_day, Most_Ratings
318 FROM (
319     SELECT Branch, Time_of_day, COUNT(Rating) AS Most_Ratings,
320            ROW_NUMBER() OVER (PARTITION BY Branch ORDER BY COUNT(Rating) DESC) AS Rating
321     FROM amazon
322     GROUP BY Branch, Time_of_day
323 ) AS ranked
324 WHERE Rating = 1;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
Branch	Time_of_day	Most_Ratings	
A	Afternoon	185	
B	Afternoon	162	
C	Afternoon	181	

Q27: Identify the day of the week with the highest average ratings.

```
330 • SELECT Day_Name, ROUND(AVG(Rating), 3) Avg_Rating
331 FROM amazon
332 GROUP BY Day_Name
333 ORDER BY Avg_Rating DESC;
334 -- LIMIT 1;
```

Result Grid	Filter Rows:	Export:	Wrap Cell
Day_Name	Avg_Rating		
Monday	7.154		
Friday	7.076		
Sunday	7.011		
Tuesday	7.003		
Saturday	6.902		
Thursday	6.89		
Wednesday	6.806		

Q28: Determine the day of the week with the highest average ratings for each branch.

```
40 • SELECT Day_Name, Branch, Avg_Rating FROM
41 (SELECT Branch, Day_Name, ROUND(AVG(Rating), 3) AS Avg_Rating,
42     ROW_NUMBER() OVER(PARTITION BY Branch ORDER BY AVG(Rating) DESC) AS High_Rating
43 FROM amazon
44 GROUP BY Branch, Day_Name
45 ORDER BY Avg_Rating DESC) AS bb
46 WHERE High_Rating = 1;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
Day_Name	Branch	Avg_Rating	
Monday	B	7.336	
Friday	A	7.312	
Friday	C	7.279	

INSIGHTS

PRODUCT ANALYSIS:

HIGHEST REVENUE PRODUCT LINE: Food and Beverages Generating Highest Revenue (\$56144.96)

LOWEST REVENUE PRODUCT LINE: Health and beauty Generating Lowest Revenue (\$49193.84)

HIGHEST SALES PRODUCT LINE: Electronic Accessories (Units Sold:971)

LOWEST SALES PRODUCT LINE: Health and Beauty (units Sold: 854)

SALES ANALYSIS:

MONTH WITH HIGHEST REVENUE: January has Generated Highest Revenue (\$ 116292.11)

MONTH WITH LOWEST REVENUE: February has Generated Lowest Revenue (\$ 97219.58)

CITY WITH HIGHEST REVENUE: Naypyitaw has Generated Highest Revenue (\$ 110568.86)

CITY WITH LOWEST REVENUE: Mandalay has Generated Lowest Revenue (\$ 106198.00)

PEAK SALES TIME OF DAY: Afternoon has recorded Highest Sales while Morning with Lowest Sales

PEAK SALES DAY OF WEEK: Saturday has recorded Highest Sales

CUSTOMER ANALYSIS:

PREDOMINANT GENDER: Females contributed more to Revenue, although difference between contribution is not much.

PREDOMINANT CUSTOMER TYPE: Member

HIGHEST REVENUE GENDER: Female (\$ 167883.26)

HIGHEST REVENUE CUSTOMER TYPE: Member (\$ 164223.81)

MOST POPULAR PRODUCT LINE (MALE): Health and Beauty

MOST POPULAR PRODUCT LINE (FEMALE): Fashion and Accessories

RECOMMENDATIONS:

- Health and Beauty products underperform across metrics and require a comprehensive plan to enhance the product line.
- Leverage the high sales and revenue in January by offering diverse customer options and implementing effective, targeted strategies.
- Prioritize new product launches or campaigns during the peak afternoon sales hours.
- Develop a plan to increase memberships, as member customers contribute more, potentially due to incentives or offers. This can reduce Acquisition Cost and drive higher revenue.

