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;
UPDATE amazon
SET date = STR_TO_DATE(date, '%d-%m-%Y')
SET SQL_SAFE_UPDATES = 1;
ALTER TABLE amazon
MODIFY date DATE;
alter table amazon
modify Time Time;
ALTER TABLE amazon
MODIFY `Invoice ID` VARCHAR(30);
```

Limit to 1000 rows

```
39
40
41  #Feature Engineering:
42  #Add new columns for timeofday, dayname, and monthname
43  alter Table amazon
44  add timeofday varchar(10);
45
46  alter Table amazon
47  add dayname varchar(10);
48
49  alter Table amazon
50  add monthname varchar(10);
```

```
52  update amazon
53  set timeofday=case
54  when Time(time) BETWEEN '06:00:00' AND '11:59:59' THEN 'Morning'
55  when Time(time) BETWEEN '12:00:00' AND '17:59:59' THEN 'Afternoon'
56  when Time(time) BETWEEN '18:00:00' AND '23:59:59' THEN 'Evening'
57  ELSE 'Night'
58  END;
59
60  update amazon
61  set dayname=Dayname(date);
62  update amazon
63  set monthname=Monthname(date);
64  select timeofday,dayname,monthname from amazon;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

timeofday	dayname	monthname
Afternoon	Saturday	January
Morning	Friday	March
Afternoon	Sunday	March
Evening	Sunday	January
Morning	Friday	February

Business Questions To Answer:

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

```
66  #Business Questions To Answer:
67
68  #1.What is the count of distinct cities in the dataset?
69  select Distinct City from amazon;
70
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

City
Yangon
Naypyitaw
Mandalay

2.For each branch, what is the corresponding city

```
71  #2.For each branch, what is the corresponding city?
72  select distinct branch,City from amazon;
73
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

branch	City
A	Yangon
C	Naypyitaw
B	Mandalay

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

```
74    #3.What is the count of distinct product lines in the dataset?
75 •  select distinct `Product line` from amazon;
76
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Product line
Health and beauty
Electronic accessories
Home and lifestyle
Sports and travel
Food and beverages
Fashion accessories

4.Which payment method occurs most frequently?

```
77    #4.Which payment method occurs most frequently?
78 •  select Payment,count(*) as frequency
79    from amazon
80    group by Payment
81    order by frequency Desc
82    limit 1;
83
```

Result Grid | Filter Rows: | Export: | Wrap Cel

Payment	frequency
Ewallet	345

5.Which product line has the highest sales?

```
84    #5.Which product line has the highest sales?
85 •  select `Product line`,count(*) as `highest sales` from amazon
86    group by `Product line`
87    order by `highest sales` desc
88    limit 1;
89
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fe

Product line	highest sales
Fashion accessories	178

6.How much revenue is generated each month?

```
90    #6.How much revenue is generated each month?
91 •  select distinct monthname from amazon;
92 •  select monthname, sum(Total)as revenue from amazon
93    group by monthname
94    order by revenue desc;
95
```

Result Grid | Filter Rows: | Export: | Wrap Cell C

monthname	revenue
January	116291.86800000005
March	109455.50700000004
February	97219.37399999997

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?
97 •  select monthname,sum(cogs)as totalcost_of_goods
98    from amazon
99    group by monthname
100   order by totalcost_of_goods desc
101   limit 1;
102
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fe

monthname	totalcost_of_goods
January	110754.16000000002

8.Which product line generated the highest revenue?

```
103   #8.Which product line generated the highest revenue?
104 •  select `Product line`,sum(total) as higest_revenue
105   from amazon
106   group by `Product line`
107   order by higest_revenue desc
108   limit 1;
109
```

Result Grid | Filter Rows: | Export: | Wrap Cell Cont

Product line	higest_revenue
Food and beverages	56144.844000000005

9.In which city was the highest revenue recorded?

```
#9.In which city was the highest revenue recorded?
select City,sum(total) as higest_revenue
from amazon
group by City
order by higest_revenue desc
limit 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Conte

City	higest_revenue
Naypyitaw	110568.70649999994

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

```
#10.Which product line incurred the highest Value Added Tax?
select `Product line`,sum(`Tax 5%`) as total_tax
from amazon
group by `Product line`
order by total_tax desc
limit 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fe

Product line	total_tax
Food and beverages	2673.5639999999994

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

```
UPDATE amazon a
JOIN (
    SELECT `Product line`,
           CASE
               WHEN SUM(Total) > (SELECT AVG(Total) FROM amazon) THEN 'Good'
               ELSE 'Bad'
           END AS sales_classification
    FROM amazon
    GROUP BY `Product line`
) AS subquery ON a.`Product line` = subquery.`Product line`
SET a.sales_classification = subquery.sales_classification;
select  `Product line`, sales_classification from amazon;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Product line	sales_classification
Health and beauty	Good
Electronic accessories	Good
Home and lifestyle	Good
Health and beauty	Good
Sports and travel	Good
Electronic accessories	Good

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

```
142    #12.Identify the branch that exceeded the average number of products sold.
143 •  select Branch, sum(Quantity) as total_quantity
144    from amazon
145    group by Branch
146    having total_quantity > (select avg(Quantity) from amazon);
147
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Branch	total_quantity
A	1859
C	1831
B	1820

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

```
148    #13.Which product line is most frequently associated with each gender?
149 •  select Gender,`Product line`,count(*) as frequency
150    from amazon
151    group by Gender,`Product line`
152    Order By Frequency Desc
153    limit 1;
154
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Gender	Product line	frequency
Female	Fashion accessories	96

14.Calculate the average rating for each product line.

```
155    #14.Calculate the average rating for each product line.
156 •  select `Product line`,avg(Rating)as rating from amazon
157    group by `Product line`
158    order by rating  desc;
159
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Product line	rating
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.

```
160     #15.Count the sales occurrences for each time of day on every weekday.
161 •   select timeofday,dayname,count(Total)as totalsales
162     from amazon
163     group by timeofday,dayname
164     order by totalsales desc;
165
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

timeofday	dayname	totalsales
Afternoon	Saturday	81
Afternoon	Wednesday	81
Afternoon	Thursday	76
Afternoon	Monday	75
Afternoon	Friday	74
Afternoon	Tuesday	71

16.Identify the customer type contributing the highest revenue.

```
166     #16.Identify the customer type contributing the highest revenue.
167 •   select `Customer type`,sum(Total) as higest_revenue
168     from amazon
169     group by `Customer type`
170     order by higest_revenue desc;
171
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Customer type	higest_revenue
Member	164223.44400000002
Normal	158743.30500000005

17.Determine the city with the highest VAT percentage.

```
172     #17.Determine the city with the highest VAT percentage.
173 •   select City,avg(`Tax 5%`) as total_vat
174     from amazon
175     group by City
176     order by total_vat desc
177     limit 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch ro

City	total_vat
Naypyitaw	16.05236737804879

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

```
179    #18.Identify the customer type with the highest VAT payments.
180    select `Customer type`,sum(`Tax 5%`) as total_vat
181    from amazon
182    group by `Customer type`
183    order by total_vat desc
184    limit 1;
185
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Customer type	total_vat
Member	7820.164000000002

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

```
186    #19.What is the count of distinct customer types in the dataset?
187    select count(distinct `Customer type`)
188    from amazon;
189
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(distinct `Customer type`)
2

Result Grid

Customer type
Member
Normal

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

Result Grid

Payment
Ewallet
Cash
Credit card

```
190    #20.What is the count of distinct payment methods in the dataset?
191    select count(distinct Payment)
192    from amazon;
193
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(distinct Payment)
3

21.Which customer type occurs most frequently?

```
196     #21.Which customer type occurs most frequently?
197 •   select  `Customer type`,count(*) as frequently
198     from amazon
199     group by  `Customer type`
200     order by frequently desc
201     limit 1;
202
```

Result Grid | Filter Rows: | Export: | Wrap Cell Co

	Customer type	frequently
▶	Member	501

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

```
203     #22.Identify the customer type with the highest purchase frequency.
204 •   select  `Customer type`,count(Quantity) as  purchase_count
205     from amazon
206     group by  `Customer type`
207     order by  purchase_count desc
208     limit 1;
209
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

	Customer type	purchase_count
▶	Member	501

23.Determine the predominant gender among customers.

```
210     #23.Determine the predominant gender among customers.
211 •   SELECT gender, COUNT(*) AS count
212     FROM amazon
213     GROUP BY gender
214     ORDER BY count DESC
215     limit 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Conter

	gender	count
▶	Female	501

24.Examine the distribution of genders within each branch.

```
217    #24.Examine the distribution of genders within each branch.
218 •  SELECT Branch,gender, COUNT(*) AS count
219    FROM amazon
220    GROUP BY Branch,gender
221    order by Branch,gender;
222
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Branch	gender	count
A	Female	161
A	Male	179
B	Female	162
B	Male	170
C	Female	178
C	Male	150

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

```
223    #25.Identify the time of day when customers provide the most ratings.
224 •  select timeofday,count(*)as rating_counts
225    from amazon
226    group by timeofday
227    order by rating_counts desc;
228
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

timeofday	rating_counts
Afternoon	528
Evening	281
Morning	191

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

```
229    #26.Determine the time of day with the highest customer ratings for each branch.
230 •  select timeofday,branch,avg(rating) as avg_rating
231    from amazon
232    group by timeofday,branch
233    order by  avg_rating desc;
234
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

timeofday	branch	avg_rating
Afternoon	C	7.0955801104972345
Evening	C	7.0920454545454525
Afternoon	A	7.0567567567567595
Morning	A	7.005479452054794
Evening	A	6.979268292682928
Morning	C	6.974576271186442

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

```
235    #27.Identify the day of the week with the highest average ratings.
236    select dayname,avg(rating) as avg_rating
237    from amazon
238    group by dayname
239    order by  avg_rating desc;
240
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

dayname	avg_rating
Monday	7.153599999999999
Friday	7.076258992805756
Sunday	7.011278195488723
Tuesday	7.003164556962025
Saturday	6.901829268292688
Thursday	6.88985507246377

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

```
242    #28.Determine the day of the week with the highest average ratings for each branch.
243    select dayname,branch,avg(rating) as avg_rating
244    from amazon
245    group by dayname,branch
246    order by  avg_rating desc;
247
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

dayname	branch	avg_rating
Monday	B	7.335897435897434
Friday	A	7.3119999999999985

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 best-selling 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