


SQL Capstone project - Amazon sales data analysis

Business Questions:

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

```
46
47 -- 1. What is the count of distinct cities in the dataset?
48
49 • select count(distinct city) as no_of_cities from amazon;
50
```




The screenshot shows a SQL query result grid. The query is: `select count(distinct city) as no_of_cities from amazon;`. The result grid has a single column labeled `no_of_cities` and a single row with the value `3`.

no_of_cities
3

- 1.
2. For each branch, what is the corresponding city?

```
50
51 -- 2. For each branch, what is the corresponding city?
52
53 • select branch , city from amazon
54 group by branch, city;
55
```



The screenshot shows a SQL query result grid. The query is: `select branch , city from amazon group by branch, city;`. The result grid has two columns: `branch` and `city`. There are three rows: `A` (Yangon), `C` (Naypyitaw), and `B` (Mandalay).

branch	city
A	Yangon
C	Naypyitaw
B	Mandalay

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

```

55
56 -- 3. What is the count of distinct product lines in the dataset?
57 • select count(distinct product_line) as distinct_product_lines from amazon;
58
59
60
61

```

Result Grid

distinct_product_lines
6

4. Which payment method occurs most frequently?

```

59 -- 4. Which payment method occurs most frequently?
60 • select payment, count(payment) as count_of_payment from amazon
61 group by payment
62 order by payment desc
63 limit 1;
64

```

Result Grid

payment	count_of_payment
Ewallet	345

5. Which product line has the highest sales?

```

65 -- 5. Which product line has the highest sales?
66 • select product_line , round(sum(total),2) as total_sales from amazon
67 group by product_line
68 order by total_sales desc
69 limit 1;
70

```

Result Grid

product_line	total_sales
Food and beverages	56144.84

```

71 -- 6 . How much revenue is generated each month?
72 • select monthname ,round(sum(total),2) as total_sales from amazon
73 group by monthname;
74
75
76

```

Result Grid		
	monthname	total_sales
▶	JAN	116291.87
	MAR	109455.51
	FEB	97219.37

6.

How

much revenue is generated each month?

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

```

75 -- 7. In which month did the cost of goods sold reach its peak?
76 • select monthname , round(sum(cogs),2) as total_cogs from amazon
77 group by monthname
78 order by total_cogs desc
79 limit 1;
80

```

Result Grid		
	monthname	total_cogs
▶	JAN	110754.16

8. Which product line generated the highest revenue?

```

81 -- 8. Which product line generated the highest revenue?
82 • select product_line , round(sum(gross_income),2) as total_revenue from amazon
83 group by product_line
84 order by total_revenue desc
85 limit 1;
86
87

```

Result Grid		
	product_line	total_revenue
▶	Food and beverages	2673.56

9. In which city was the highest revenue recorded?

```
87 -- 9. In which city was the highest revenue recorded?
88 • select city , round(sum(gross_income),2) as total_revenue from amazon
89 group by city
90 order by total_revenue desc
91 limit 1;
92
```

Result Grid

city	total_revenue
Naypyitaw	5265.18

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

```
93 -- 10. Which product line incurred the highest Value Added Tax?
94 • select product_line , round(sum(vat),2) as total_vat from amazon
95 group by product_line
96 order by total_vat desc
97 limit 1;
98
99
```

Result Grid

product_line	total_vat
Food and beverages	2673.56

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

```

99  -- 11. For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."
100 • SELECT
101      *,
102      CASE
103          WHEN Total > avg_sales THEN 'Good'
104          ELSE 'Bad'
105      END AS sales_performance
106  FROM (
107      SELECT
108          Product_line,
109          Total,
110          (SELECT round(AVG(Total),2) FROM amazon) AS avg_sales
111      FROM
112          amazon
113  ) AS subquery_alias;

```

Product_line	Total	avg_sales	sales_performance
Health and beauty	548.9715	322.97	Good
Electronic accessories	80.22	322.97	Bad
Home and lifestyle	340.5255	322.97	Good
Health and beauty	489.048	322.97	Good
Sports and travel	634.3785	322.97	Good

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

Query 1

```

117 • SELECT
118     branch,
119     SUM(quantity) AS total_quantity_sold,
120     AVG(quantity) AS avg_quantity_sold,
121     CASE
122         WHEN SUM(quantity) > (SELECT AVG(quantity) FROM amazon) THEN 'Exceeded'
123         ELSE 'Did not exceed'
124     END AS quantity_comparison
125  FROM
126     amazon
127  GROUP BY
128     branch;
129

```

branch	total_quantity_sold	avg_quantity_sold	quantity_comparison
A	1859	5.4676	Exceeded
C	1831	5.5823	Exceeded
B	1820	5.4819	Exceeded

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

```

129  -- 13. Which product line is most frequently associated with each gender?
130  •  select *
131      from(
132          select gender, product_line, count(*)as frequent
133          from amazon
134          where gender = 'male'
135          group by product_line
136          order by frequent desc
137          limit 1) as male
138      union
139      (select gender, product_line, count(*)as frequent_female
140      from amazon
141      where gender = 'female'
142      group by product_line
143      order by frequent_female desc
144      limit 1) ;

```

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

	gender	product_line	frequent
▶	Male	Health and beauty	88
	Female	Fashion accessories	96

14. Calculate the average rating for each product line.

```

142
143  -- 14. Calculate the average rating for each product line.
144  •  select product_line , round(avg(rating),2) as avg_rating
145      from amazon
146      group by product_line;

```

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

	product_line	avg_rating
▶	Health and beauty	7
	Electronic accessories	6.92
	Home and lifestyle	6.84
	Sports and travel	6.92
	Food and beverages	7.11
	Fashion accessories	7.03

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

```

155 -- 15. Count the sales occurrences for each time of day on every weekday.
156 • SELECT timeofday, COUNT(invoice_id)
157 FROM amazon
158 WHERE dayname NOT IN ('Sat', 'Sun')
159 group by timeofday;
160
161

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
timeofday	COUNT(invoice_id)			
Morning	141			
Evening	185			
Afternoon	377			

16. Identify the customer type contributing the highest revenue.

```

163 • select `customer type` , round(sum(total),2) as total_revenue
164 from amazon
165 group by `customer type`;
166

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
customer type	total_revenue			
Member	164223.44			
Normal	158743.31			

17. Determine the city with the highest VAT percentage.

```

167 -- 17. Determine the city with the highest VAT percentage.
168 • select city , round(sum(vat),2) as vat_total
169 from amazon
170 group by city
171 order by vat_total desc
172 limit 1;
173
174
175

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:	Fetch row
city	vat_total				
Naypyitaw	5265.18				

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

```
174 -- 18. Identify the customer type with the highest VAT payments.
175
176 • select `Customer type`, round(sum(vat),2) as total_vat_payments
177 from amazon
178 group by `Customer type`
179 order by total_vat_payments DESC
180 limit 1;
181
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
	Customer type	total_vat_payments			
▶	Member	7820.16			

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

```
183 -- 19. What is the count of distinct customer types in the dataset?
184 • select count(distinct `customer type`) as unique_customer_type
185 from amazon;
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	unique_customer_type			
▶	2			

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

```
187 -- 20. What is the count of distinct payment methods in the dataset?
188 • select count(distinct payment) as unique_payment_method
189 from amazon;
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	unique_payment_method			
▶	3			

21. Which customer type occurs most frequently?


```
191 -- 21. Which customer type occurs most frequently?
192
193 • select `customer type` , count(*) as frequent
194 from amazon
195 group by `customer type`
196 order by frequent desc
197 limit 1 ;
```

<

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

	customer type	frequent
▶	Member	501

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

```
199 -- 22. Identify the customer type with the highest purchase frequency.
200
201 • select `customer type` , count(invoice_id) as frequent
202 from amazon
203 group by `customer type`
204 order by frequent desc
205 limit 1;
206
```

<

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

	customer type	frequent
▶	Member	501

23. Determine the predominant gender among customers.

```

207 -- 23.Determine the predominant gender among customers.
208 • select gender , count(*) as predominant_gender
209 from amazon
210 group by gender
211 order by predominant_gender desc
212 limit 1 ;
213
214

```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
	gender	predominant_gender				
▶	Female	501				

24. Examine the distribution of genders within each branch.

```

214 -- 24 . Examine the distribution of genders within each branch.
215
216 • select branch, gender , count(*) as gender_count
217 from amazon
218 group by branch, gender
219 order by branch ;
220

```

Result Grid				Filter Rows:	Export:	Wrap Cell Content:
	branch	gender	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.

```

221 -- 25. Identify the time of day when customers provide the most ratings.
222
223 • select timeofday, count(*) as rating_count
224 from amazon
225 group by timeofday
226 order by rating_count desc;

```

timeofday	rating_count
Afternoon	528
Evening	281
Morning	191

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

```

282 -- 26. Determine the time of day with the highest customer ratings for each branch.
283 • WITH RankedTimes AS (
284     SELECT
285         branch,
286         timeofday, max(rating),
287         ROW_NUMBER() OVER (PARTITION BY branch ORDER BY MAX(rating) DESC) AS ranking
288     FROM
289         amazon
290     GROUP BY
291         branch, timeofday
292 )
293 SELECT
294     branch,
295     timeofday, ranking
296 FROM
297     RankedTimes
298 WHERE
299     ranking = 1;

```

branch	timeofday	ranking
A	Afternoon	1
B	Afternoon	1
C	Afternoon	1

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
237     dayname,
238     ROUND(AVG(Rating),2) AS average_rating
239 FROM
240     amazon
241 GROUP BY
242     dayname
243 ORDER BY
244     average_rating DESC
245 LIMIT 1;

```

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

	dayname	average_rating
▶	MON	7.15

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

```

247 -- 28. Determine the day of the week with the highest average ratings for each branch.
248
249 • SELECT *
250 FROM (
251     SELECT
252         branch,
253         dayname,
254         ROUND(AVG(rating),2) AS average_rating,
255         RANK() OVER (PARTITION BY branch ORDER BY AVG(rating) DESC) AS ranked
256     FROM
257         amazon
258     GROUP BY
259         branch,
260         dayname
261 ) AS subquery
262 WHERE
263     ranked = 1;


```

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

	branch	dayname	average_rating	ranked
▶	A	FRI	7.31	1
	B	MON	7.34	1
	C	FRI	7.28	1

Business insights and recommendations:

After through analysis of the given amazon data, i have come up with the few business insights and recommendations which will implore the customer loyalty and inturn increase the revenue generated

 Business Insights and Recommendations