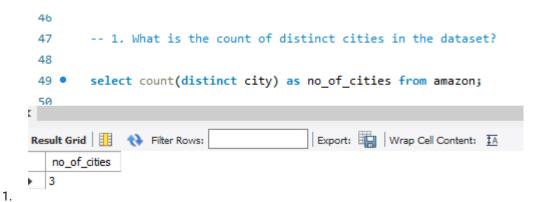
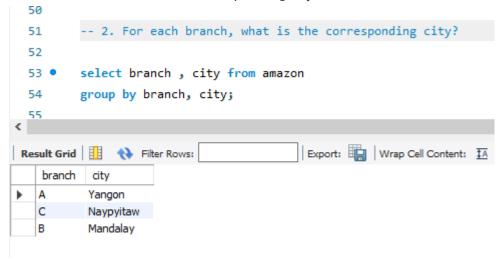
## SQL Capstone project - Amazon sales data analysis

## **Business Questions:**

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



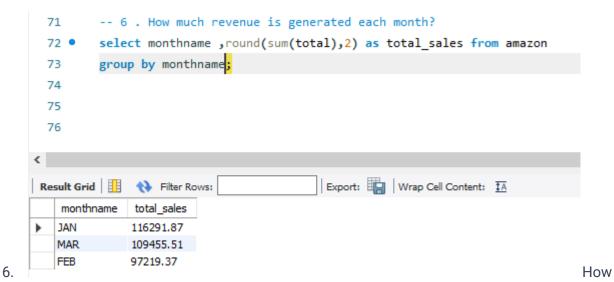
2. For each branch, what is the corresponding city?



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

4. Which payment method occurs most frequently?

5. Which product line has the highest sales?



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
         group by monthname
 77
         order by total_cogs desc
 78
         limit 1;
 79
                                           Export: Wrap Cell Content: A Fetch
Result Grid
              Filter Rows:
   monthname
              total_cogs
  JAN
              110754.16
```

8. Which product line generated the highest revenue?

```
-- 8. Which product line generated the highest revenue?
       select product_line ,
                               round(sum(gross_income),2) as total_revenue from amazon
82 •
       group by product line
83
       order by total_revenue desc
84
85
       limit 1;
86
87
                                 Export: Wrap Cell Content: 🔼 Fetch rows:
total_revenue
  product_line
 Food and beverages
               2673.56
```

9. In which city was the highest revenue recorded?

```
-- 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
      order by total_revenue desc
90
      limit 1;
91
92
                               Export: Wrap Cell Content: A Fetch rows:
total_revenue
Naypyitaw
         5265.18
```

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

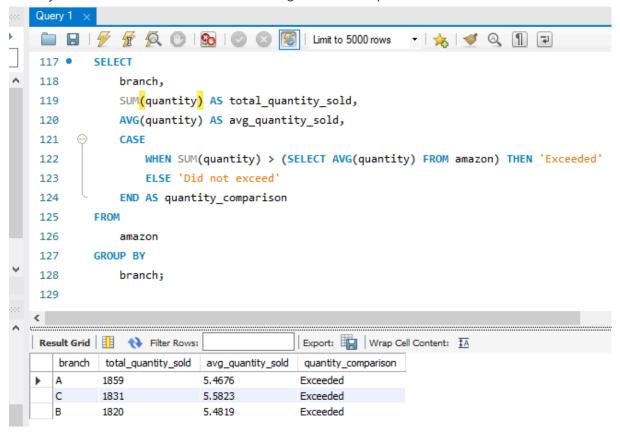
```
-- 10. Which product line incurred the highest Value Added Tax?
       select product_line , round(sum(vat),2) as total_vat from amazon
95
       group by product_line
       order by total_vat desc
96
97
       limit 1;
98
99
                                 Export: Wrap Cell Content: 🚻 | Fetch rows:
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."

```
-- 11. For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."
         SELECT
100 •
101
102
             CASE
103
                 WHEN Total > avg_sales THEN 'Good'
                ELSE 'Bad'
104
             END AS sales performance
105

→ FROM (
106
107
             SELECT
108
                Product_line,
                 (SELECT round(AVG(Total),2) FROM amazon) AS avg_sales
110
111
112
                 amazon
         ) AS subquery_alias;
113
Export: Wrap Cell Content: IA
                      Total
                                           sales_performance
   Product_line
                                avg_sales
  Health and beauty
                      548.9715
                                322.97
  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
```

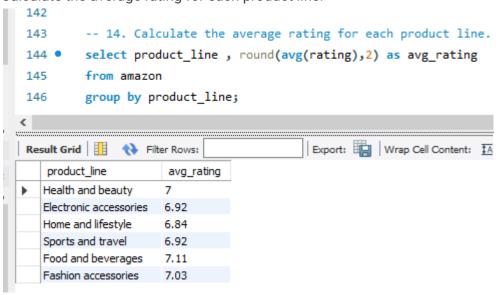
12. Identify the branch that exceeded the average number of products sold.



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

```
129
        -- 13. Which product line is most frequently associated with each gender?
        select *
 130 •
      ⊖ from(
 131
        select gender, product_line, count(*)as frequent
 132
        from amazon
 133
        where gender = 'male'
 134
        group by product line
 135
        order by frequent desc
 136
       limit 1) as male
 137
        union
 138
      139
        from amazon
 140
        where gender = 'female'
 141
        group by product line
 142
        order by frequent_female desc
 143
 144
        limit 1);
<
                                     Export: Wrap Cell Content: IA
product_line
                         frequent
   gender
   Male
          Health and beauty
                        88
   Female
        Fashion accessories
                        96
```

14. Calculate the average rating for each product line.



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.
        SELECT timeofday, COUNT(invoice_id)
156 •
        FROM amazon
157
        WHERE dayname NOT IN ('Sat', 'Sun')
158
        group by timeofday;
159
160
161
Export: Wrap Cell Content: IA
   timeofday COUNT(invoice_id)
  Morning
           141
           185
   Evening
   Afternoon
           377
```

16. Identify the customer type contributing the highest revenue.

```
select `customer type` , round(sum(total),2) as total_revenue
163 •
164
         from amazon
165
         group by `customer type`;
166
Result Grid
                                         Export: Wrap Cell Content: IA
              Filter Rows:
   customer
                total_revenue
   type
  Member
                164223.44
  Normal
               158743.31
```

17. Determine the city with the highest VAT percentage.

```
-- 17. Determine the city with the highest VAT percentage.
       select city , round(sum(vat),2) as vat_total
168 •
       from amazon
169
170
      group by city
       order by vat total desc
171
172
       limit 1;
173
174
175
                                   Export: Wrap Cell Content: TA Fetch rov
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
       select `Customer type`, round(sum(vat),2) as total_vat_payments
176 •
177
       from amazon
178
       group by `Customer type`
       order by total_vat_payments DESC
179
180
       limit 1;
181
<
                                                                    4
                                 Export: Wrap Cell Content: 🔀 | Fetch rows:
Customer
             total_vat_payments
   tvpe
  Member
             7820.16
```

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?

```
191
        -- 21. Which customer type occurs most frequently?
192
        select `customer type` , count(*) as frequent
193 •
194
        from amazon
195
        group by `customer type`
        order by frequent desc
196
197
        limit 1 ;
                                    Export: Wrap Cell Content: 🖽 Fetch r
Result Grid
             Filter Rows:
   customer
              frequent
   type
  Member
              501
```

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

```
-- 22. Identify the customer type with the highest purchase frequency.
199
200
        select `customer type` , count(invoice_id) as frequent
201 •
202
        from amazon
        group by `customer type`
203
        order by frequent desc
204
        limit 1;
205
206
                                       Export: Wrap Cell Content: TA Fetch rows:
customer
               frequent
   type
  Member
              501
```

23. Determine the predominant gender among customers.

```
207
         -- 23.Determine the predominant gender among customers.
         select gender , count(*) as predominant_gender
 208 •
         from amazon
 209
 210
         group by gender
         order by predominant_gender desc
 211
 212
         limit 1;
 213
 214
 <
                                  Export: Wrap Cell Content: A Fetch rows:
 predominant_gender
    gender
  Female
          501
```

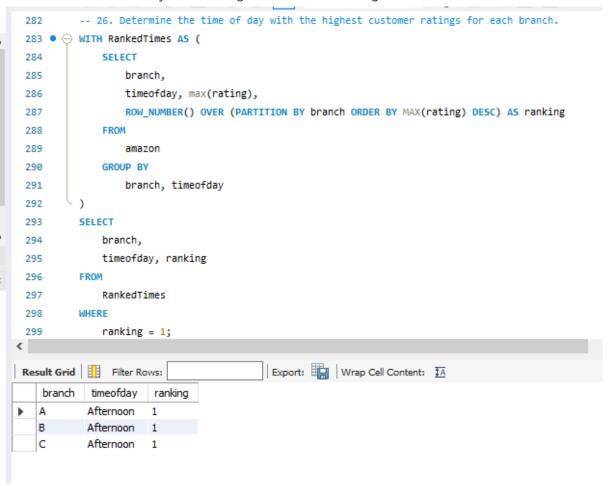
24. Examine the distribution of genders within each branch.

```
214
        -- 24 . Examine the distribution of genders within each branch.
215
        select branch, gender , count(*) as gender_count
216 •
217
        from amazon
        group by branch, gender
218
219
        order by branch;
220
                                     Export: Wrap Cell Content: TA
Result Grid
            Filter Rows:
                gender_count
   branch
          gender
         Male
                179
   В
         Female
                162
   В
         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
       select timeofday, count(*) as rating_count
223 •
       from amazon
224
       group by timeofday
225
       order by rating_count desc;
226
Export: Wrap Cell Content: IA
   timeofday rating_count
  Afternoon
          528
  Evening
          281
  Morning
          191
```

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



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

```
-- 27. Identify the day of the week with the highest average ratings.
235
236 •
       SELECT
237
           dayname,
238
           ROUND(AVG(Rating),2) AS average_rating
       FROM
239
240
           amazon
       GROUP BY
241
242
           dayname
243
       ORDER BY
244
           average_rating DESC
245
       LIMIT 1;
                                                                   -
                                 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.

```
-- 28. Determine the day of the week with the highest average ratings for each branch.
248
          SELECT *
249

→ FROM (
250
              SELECT
251
                  branch,
252
253
                  dayname,
254
                  ROUND(AVG(rating),2) AS average_rating,
                  RANK() OVER (PARTITION BY branch ORDER BY AVG(rating) DESC) AS ranked
255
256
              FROM
                  amazon
258
              GROUP BY
259
                  branch,
                  dayname
260
          ) AS subquery
261
          WHERE
262
263
              ranked = 1;
                                                Export: Wrap Cell Content: IA
Result Grid
                Filter Rows:
   branch
            dayname
                      average_rating
                                      ranked
           FRI
                      7.31
                                     1
   В
           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