Name Leah Sigana

Week 5 assignment

1. **Briefly explain any difficulties and 1 interesting thing you noticed about your chosen dataset.**

Difficulties:

While the analysis shows some income differences based on location and occupation, the margins are quite small. This makes it challenging to draw definitive conclusions about the impact of these factors on income in this dataset

Interesting

Higher concentration of software engineers in suburban areas: The table shows that there are 132 software engineers in suburban areas, compared to 119 in rural areas and 111 in urban areas. This might be surprising, as some might expect a higher concentration in urban areas closer to tech hubs.

2. **Find 2 cool facts hidden within the data (e.g., most popular interests).**

Instagram is the most popular social media platform for all three career groups (market managers, software engineers, and students). This could indicate a trend of females in these professions using Instagram more than other social media platforms, for professional networking, staying updated in their fields, or simply for personal use.

Social media usage seems to be similar across all three career groups. The data shows relatively close numbers across the board for Facebook, Instagram, and Youtube for each career group. This might suggest that social media use isn't strongly influenced by career choice in this dataset.

Question 3

i. Formulate 2 questions about the data (e.g., what are popular shows in different countries?).
ii. Write basic SQL queries (WHERE, ORDER BY) to find answers.
iii. Share what you learned from the answers.

Question 1: Which social media platform has the most female software engineers?

This question looks for the social media platform (Facebook, Instagram, Youtube) that has the highest number of female software engineers.

SQL
SELECT platform, COUNT(*) AS total_female_software_engineers

```
FROM your_table
WHERE career = 'Software Engineer' AND female = 'yes'
GROUP BY platform
ORDER BY total_female_software_engineers DESC;
```
Use code with caution.
content_copy

Explanation:

- SELECT platform, COUNT(*) AS total_female_software_engineers: This selects the platform name and calculates the count of female software engineers for each platform, naming the count as total_female_software_engineers.
- FROM your_table: This specifies the table where your data is stored. Replace "your_table" with the actual table name in your database.
- WHERE career = 'Software Engineer' AND female = 'yes': This filters the data to include only rows where the career is "Software Engineer" and the gender is "female".
- GROUP BY platform: This groups the data by the platform name.
- ORDER BY total_female_software_engineers DESC: This orders the results by the count of female software engineers in descending order (highest to lowest).

Question 2: Are there any social media platforms where there are more female marketers than female software engineers?

This question aims to identify any social media platforms where the number of female marketers is higher than the number of female software engineers.

SQL
```
SELECT platform
FROM your_table
WHERE female = 'yes'
GROUP BY platform
HAVING SUM(CASE WHEN career = 'Marketing Manager' THEN 1 ELSE 0 END) >
  SUM(CASE WHEN career = 'Software Engineer' THEN 1 ELSE 0 END);
```
Use code with caution.
content_copy

Explanation:

- SELECT platform: This selects the platform name.
- FROM your_table: This specifies the table where your data is stored. Replace "your_table" with the actual table name in your database.
- WHERE female = 'yes': This filters the data to include only rows where the gender is "female".
- GROUP BY platform: This groups the data by the platform name.
- HAVING SUM(CASE WHEN career = 'Marketing Manager' THEN 1 ELSE 0 END) > SUM(CASE WHEN career = 'Software Engineer' THEN 1 ELSE 0 END): This clause

filters the grouped data to show only platforms where the sum of female marketing managers (achieved through the CASE WHEN statement) is greater than the sum of female software engineers (also achieved through a separate CASE WHEN statement).