QUESTIONS(adult):

1. How many unique values are there in each column of the adult census income dataset? Use a bar chart to visualize the frequency of each value in each column.

2. What is the distribution of the 'income' variable in the adult census income dataset? Use a pie chart to visualize the percentage of individuals who make more or less than $50,000 per year. What is the proportion of individuals who make more than $50,000 per year in the dataset?

3.Identify and describe any data quality issues or inconsistencies within the data set. What steps would you take to clean and pre-processes the data to ensure its accuracy for further analysis.

4. What are the most common values for the 'workclass' and 'occupation' variables in the adult census income dataset? Use a stacked bar chart to visualize the frequency of individuals in each workclass and occupation category. How do these variables relate to the income level of the individuals?

5. Using a scatter plot, examine the relationship between the 'age' and 'hours-per-week' variables in the adult census income dataset. What is the correlation coefficient between these two variables and what does it indicate?

6. Consider first 10 rows of the dataset and perform a t-test to compare the mean 'capital-gain' of the individuals who make more and those who make less than $50,000 per year in the adult census income dataset. What are the null and alternative hypotheses, the test statistic, the p-value, and the confidence interval? What is your interpretation of the results?

7. Using a box plot, compare the distribution of the 'education-num' variable across the different 'income' categories in the adult census income dataset. What are some differences and similarities that you observe? How does the education level affect the income level of the individuals?

8. Explore the distribution of individuals based on their 'education-level' and 'occupation' features. Utilize a heatmap to visualize the frequency of education levels across different occupations. What insights can you derive from this heatmap regarding the educational attainment associated with various occupations?

9. Using a normal probability plot, assess the normality of the ‘age’ variable in the adult census income dataset. Describe the plot, state the null and alternative hypotheses, and report the test statistic and p-value. What is your conclusion?

10. How does the ‘native-country’ variable affect the income level of the individuals in the adult census income dataset? Use a suitable visual data representation method to visualize the average income of each country in the dataset. Which country has the highest and lowest average income?