**Group 1: Regular Data Science Questions**

1. Calculate the mean and median salary in the dataset.

Mean: $60,798; Median: $59,000

1. Determine the age distribution (in terms of frequency) in 10-year intervals (e.g., 20-29, 30-39, etc.).

Ages 20-29: 50 individuals; Ages 30-39: 48 individuals; Ages 40-49: 1 individual; Ages 50-59: 0 individuals; Ages 60-69: 0 individuals; Ages 70-79: 0 individuals; Ages 80-89: 0 individuals

1. Find the city with the highest average weight of its inhabitants.

Pheonix with an average weight of 81.0 kg

**Group 2: Multistep Hard Data Science Questions**

1. For each gender, calculate the average height. Then, determine which gender has a higher average height.

Average height for females is approximately 163.56 cm.; Average height for males is approximately 177.64 cm; Males have a higher average height.

1. Identify the top 3 cities with the highest average salary and calculate the average age of people in these cities.

Boston: Average Salary $69,875, Average Age 32 years.; Phoenix: Average Salary $67,833, Average Age 29.17 years.; Chicago: Average Salary $67,778, Average Age 32 years.

1. Create a new column “BMI” (Body Mass Index) using the formula: BMI = Weight(kg) / (Height(m))^2. Determine the average BMI for each city.

The average BMI values range from 21.47 in Los Angeles to 25.55 in Chicago

1. Find the correlation between age and salary. Is there a positive, negative, or no correlation?

correlation coefficient is 0.207

1. Calculate the standard deviation of height for each city. Which city has the greatest variability in height?

The greatest variability in height is in Houston, with a standard deviation of approximately 6.75

1. Perform a group by operation on 'City' and find the median salary for each city. Then, rank these cities based on median salary.

The median salaries range from $52,000 in Atlanta to $69,000 in Boston and Chicago.; Cities are ranked based on their median salaries, with Boston and Chicago sharing the top rank.

**Group 3: Multistep Data Analysis and Machine Learning Questions**

1. Using linear regression, predict the salary based on age, height, and weight. What is the model's R-squared value?

0.887

1. Implement a K-means clustering algorithm to group individuals based on age, height, and weight. Determine the optimal number of clusters.

2 clusters

1. Using a decision tree classifier, predict the Gender based on age, height, weight, and salary. What is the accuracy of the model?

100%

1. Apply a random forest regressor to predict individual salaries based on age, height, weight, and city. Evaluate the model using Mean Squared Error (MSE).

5,653,652

1. Develop a neural network to classify individuals into different salary brackets (<$50k, $50k-$100k, >$100k). Report the model's precision and recall.

For the salary bracket '50k-100k':Precision: 96.55%; Recall: 100%. For the salary bracket '<50k': Precision: 0% (The model did not predict any samples in this class; Recall: 0%. The overall accuracy of the model is 96.55%

1. Use a support vector machine (SVM) to classify individuals based on BMI (Underweight, Normal, Overweight, Obese). What is the F1 score of the model?

0.804