PYTHON PROGRAMMING FOR DATA SCIENCE

INTRODUCTION TO CENTRAL TENDENCY AND DATA VARIATION

This script performs statistical analysis on the age column of a dataset containing 2000 records of employee data. The key steps and functionalities include:

1. **Data Loading**: The dataset is loaded from a CSV file.
2. **Statistical Calculations**:
    1. **Mean**: The average age.
    2. **Median**: The middle value of the age data.
    3. **Mode**: The most frequently occurring age.
    4. **Variance**: The measure of age data spread.
    5. **Standard Deviation**: The amount of variation in age data.
    6. **Skewness**: The asymmetry of the age distribution.
    7. **Range**: The difference between the maximum and minimum ages.
    8. **Quartiles (Q1 and Q3)**: The 25th and 75th percentiles.
    9. **Quartile Deviation**: Half the interquartile range.
    10. **Mean Deviation**: The average absolute deviation from the mean.
3. **Visualizations**:
    1. **Bar Graph with Age Intervals**: The ages are grouped into intervals (18-24, 25-34, 35-44, 45-54, 55-64) and their frequencies are plotted.
    2. **Box Plot**: A visual representation of the age distribution, including median and quartiles.
4. **Frequency Table**:
    1. **Frequency**: The number of occurrences of each age.
    2. **Relative Frequency**: The proportion of each age relative to the total.
    3. **Percentage Frequency**: The relative frequency expressed as a percentage.
    4. **Cumulative Frequency**: The running total of frequencies.
    5. **Cumulative Relative Frequency**: The running total of relative frequencies.

Visualizations

1. **Bar Graph with Age Intervals**: Helps in understanding the distribution of ages across defined age ranges.
2. **Box Plot**: Provides a summary of the age distribution, showing central tendency and dispersion.