Assignment 1

Note: Use Python with appropriate visualization libraries. Include code, visualizations, explanations, interpretations, and screenshots of outputs.

1. A city planning department wants to understand how weather, human activity, and resource usage interact in order to make better infrastructure and policy decisions. You are provided with a multi-source dataset containing the following information:

   - Date (Time Series Data)
   - Daily Average Temperature (°C)
   - Electricity Consumption (kWh)
   - Average Steps per Day (from wearable devices)
   - Resting Heart Rate (bpm)
   - Department-wise Monthly Salaries (Sales, IT, HR, Marketing)
   - Content Category and Monthly Viewing Hours (Drama, Comedy, Documentary, Action, Animation)

   Using this dataset, perform the following tasks:

   (a) Identify and classify the types of data present (categorical, numerical, time series) and justify the classification.

   (b) Describe the data visualization methodology you will follow, covering:

   - Data collection
   - Data preparation
   - Visualization
   - Reporting and sharing results

   (c) Assess the quality of the data by discussing at least four of the following aspects: completeness, accuracy, consistency, timeliness, validity, uniqueness, and integrity.

   (d) Create a line plot showing the relationship between Date, Temperature, and Electricity Consumption. Include a moving average and interpret overall trends.

   (e) Create a scatter plot showing the relationship between Average Steps per Day and Resting Heart Rate. Add a line of best fit, identify any outliers, and discuss correlation.

   (f) Create a bar chart comparing Monthly Viewing Hours across different content categories. Rank the categories and calculate the percentage contribution of the top two categories.

   (g) Create a violin plot showing the salary distribution for each department. Compare the spread, skewness, and variability between departments.

   (h) Based on all visualizations:

   - Explain at least three insights that can be drawn from the data.

- Discuss how different audiences (city officials, health researchers, business managers) might interpret these visualizations differently.

(i) Write a concise conclusion answering the following:

- Why is data visualization important for decision-making?
- Which visualization was most effective in conveying insights, and why?