Assignment Question: Seoul Bike Data - Exploratory Data Analysis (EDA) and Data Preprocessing

You have been provided with the "Seoul Bike Sharing Demand" dataset, which contains historical records of bike rentals in Seoul, South Korea. Your task is to perform Exploratory Data Analysis (EDA) and preprocess the data to prepare it for further analysis and modeling. The dataset is available in a CSV file, and it includes the following columns:

1. Date: Date of bike rental (in YYYY-MM-DD format).

2. Rented Bike Count: The number of bikes rented on a given date (target variable).

3. Hour: Hour of the day when the rental occurred.

4. Temperature: Temperature in Celsius at the time of rental.

5. Humidity: Humidity level at the time of rental.

6. Wind Speed: Wind speed in m/s at the time of rental.

7. Visibility: Visibility distance in meters at the time of rental.

8. Dew Point Temperature: Dew point temperature in Celsius at the time of rental.

9. Solar Radiation: Solar radiation in MJ/m2 at the time of rental.

10. Rainfall: Rainfall measured in mm at the time of rental.

11. Snowfall: Snowfall measured in cm at the time of rental.

12. Seasons: Season of the year (Spring, Summer, Autumn, Winter).

13. Holiday: Binary variable indicating whether it was a holiday (1) or not (0).

14. Functioning Day: Binary variable indicating whether the bike rental station was in functioning condition (1) or not (0) on that day.

Your assignment has two parts:

\*\*Part 1: Exploratory Data Analysis (EDA)\*\*

1. Perform preliminary data inspection to understand the structure, shape, and basic statistics of the dataset.

2. Identify missing values and decide on an appropriate strategy to handle them.

3. Conduct univariate analysis for the numerical variables (e.g., rented bike count, temperature) and categorical variables (e.g., seasons, holiday) to understand their distributions and characteristics.

4. Visualize the relationship between variables (e.g., correlation between temperature and bike count) using appropriate plots like scatter plots, bar charts, or box plots.

\*\*Part 2: Data Preprocessing\*\*

1. Handle missing values based on the strategy decided in Part 1.

2. Encode categorical variables (e.g., seasons, holiday) into numerical format, if necessary.

3. Normalize or scale numerical features, if required, to ensure they are on similar scales for modeling.

4. Consider whether any feature engineering is necessary, such as creating new features based on existing ones.

5. Split the dataset into training and testing sets, ensuring a proper time-based split (e.g., use earlier data for training and later data for testing).

Finally, document your findings, the steps you took for data preprocessing, and the rationale behind your decisions in a comprehensive report. Your report should include relevant visualizations and explanations to support your analysis.

Note: The goal of this assignment is to understand the dataset, clean and preprocess the data, and prepare it for further analysis or modeling. You do not need to build predictive models in this assignment.