Yulu E-Cycle Demand Analysis

Problem Statement:

Yulu is India's leading micro-mobility service provider, which offers unique vehicles for the daily commute. Starting off as a mission to eliminate traffic congestion in India, Yulu provides the safest commute solution through a user-friendly mobile app to enable shared, solo and sustainable commuting.

Yulu zones are located at all the appropriate locations (including metro stations, bus stands, office spaces, residential areas, corporate offices, etc) to make those first and last miles smooth, affordable, and convenient!

Yulu has recently suffered considerable dips in its revenues. They have contracted a consulting company to understand the factors on which the demand for these shared electric cycles depends. Specifically, they want to understand the factors affecting the demand for these shared electric cycles in the Indian market.

Data and Methodology:

- **Dataset:** yulu_data.csv
- **Features:**
    - **Temporal:**
        - datetime (datetime format)
        - season (categorical: 1-spring, 2-summer, 3-fall, 4-winter)
        - holiday (binary: 1-holiday, 0-not holiday)
        - workingday (binary: 1-working day, 0-weekend/holiday)
    - **Weather:**
        - weather (categorical: 1-Clear/Few clouds, 2-Mist, 3-Light Rain/Snow, 4-Heavy Rain/Snow)
        - temp (continuous: temperature in Celsius)
        - atemp (continuous: feeling temperature in Celsius)
        - humidity (continuous: %)
        - windspeed (continuous: wind speed)
    - **User type:**
        - casual (count of casual users)
        - registered (count of registered users)

- **Target variable:** count (total rentals: sum of casual and registered rentals)

Steps:

1. **Import Libraries and Load Data:** Import necessary libraries (e.g., pandas, matplotlib) and load the yulu_data.csv dataset.
2. **Exploratory Data Analysis (EDA):**
   - Check data structure and identify missing values.
   - Analyze descriptive statistics (mean, median, standard deviation) for numerical features.
   - Visualize the distribution of numerical features using histograms, boxplots, etc.
   - Analyze frequency tables and explore relationships between categorical features.
   - Visualize relationships between features (e.g., scatter plots, heatmaps) and the target variable (count).
3. **Data Cleaning (if necessary):**
   - Handle missing values using appropriate techniques (e.g., imputation, deletion).
   - Address outliers if they significantly impact analysis.
4. **Feature Engineering (if necessary):**
   - Create new features based on existing ones (e.g., temperature categories, high/low humidity).
   - Encode categorical features using techniques like one-hot encoding or label encoding.
5. **Bi-variate Analysis:**
   - Analyze relationships between independent variables (e.g., working day, weather, season) and the dependent variable (count) using techniques like:
     - **2-sample t-test:** Compare rental counts between different groups (e.g., weekdays vs. weekends).
     - **ANOVA:** Analyze differences in rental counts across multiple categories (e.g., seasons).

        - **Chi-square test:** Determine if categorical variables (e.g., weather) are associated with rental counts.

6. **Hypothesis Testing:**
    - Formulate null hypotheses (H0) assuming no significant difference in rental counts between groups.
    - Define alternative hypotheses (H1) assuming a difference exists.
    - Apply relevant statistical tests (e.g., t-test, ANOVA, chi-square) considering assumptions (normality, equal variance). Interpret results visually (histograms, Q-Q plots) and statistically (Levene's test, Shapiro-Wilk test).
    - Draw conclusions based on hypothesis testing and visual analysis.

Expected Outcomes:

- Identify key factors influencing Yulu's e-cycle rental demand.
- Gain insights into how weather, seasonality, working days/holidays, user types, and their interactions affect rental behaviour.