**Problem Statement:**

The company aims to analyze and identify key factors that significantly influence the demand for shared electric cycles in the Indian market. By leveraging historical data, the objective is to build a predictive model that accurately estimates electric cycle demand based on various factors such as weather conditions, time of day, seasonality, and user behavior.

**EDA:**

**No missing values** in the dataset.

**No duplicate records** detected.

**Season, holiday, working day, and weather** are categorical variables.

**Temp, humidity, windspeed, casual, registered, and count** are numerical.

The **count (total rentals)** has a wide range (min = 1, max = 977).

The histograms show the distribution of numerical variables:

* **Temperature (temp, atemp)**: Normally distributed, centered around 20-25Â°C.
* **Humidity**: Right-skewed, with most values between 40-80%.
* **Windspeed**: Shows many values at 0, possibly indicating missing or incorrectly recorded data.
* **Bike Rentals (count)**: Right-skewed, indicating more lower rental counts and fewer high-demand instance

The Plot count show the distribution of categorical values:

Â·         **Season Distribution (season) column**might contains equal values all four seasons (1: Spring, 2: Summer, 3: Fall, 4: Winter).

* **Holiday (holiday) column**expect a much higher count for **non-holidays (0)** compared to holidays (1). This suggests that most bike rentals occur on regular days rather than holidays.
* **Working Day (workingday) column,**More rentals might happen on working days (1) compared to non-working days (0).ï‚·
* Categories: 1 (Clear), 2 (Cloudy), 3 (Light Rain/Snow), 4 (Heavy Rain/Snow).
* **Weather Conditions (weather)**might have highest rentals on **clear weather days (1)** than **cloudy day (2)** and  fewer rentals on **rainy/snowy days (3 or 4)**.

**Checking and handling Outliers by Boxplots:**

1. **Windspeed & Humidity**might have many 0 values, which could be **incorrect data or missing values** recorded as zero. Humidity may have some **extreme values (close to 0 or 100)**, indicating potential outliers.
2. **Bike Rentals (count)** has many outliers, it suggests **extremely high or low rental numbers on certain days**.
3. **Temperature (temp & atemp)**,by looking graph do not show any outliers.

**Insights from the Correlation Heatmap:**

**High Positive Correlations:**

* temp & atemp: Close to **1.0**, indicating **high collinearity** (one should be removed).
* temp & count: Strong correlation (~0.4 - 0.6), meaning temperature **affects bike demand**.
* atemp & count: Similar to temp, but we should keep only one (temp is usually preferred).

**Negative Correlations:**

* windspeed & count: Slight negative correlation (bike rentals drop as wind speed increases).
* humidity & count: Weak to moderate negative correlation (higher humidity may reduce rides).

**Weak/No Correlation:**

* If a feature has close to 0 correlation with count, it **might not be useful** for predicting demand.

**Checking is there any significant difference between the no. of bike rides on Weekday and Weekends?**

**we are comparing the means of two independent groups (Weekdays vs. Weekends), we use the Independent Two-Sample T-test.**

Null Hypothesis (Hâ‚€): There is no significant difference in the number of bike rentals between weekdays and weekends.

Alternative Hypothesis (Hâ‚): There is a significant difference in the number of bike rentals between weekdays and weekends.

After testing we found: p valueÃ¨ 0.2264 and T staticalÃ¨ 1.209

We fail to reject Null Hypothesis

**Recommendation**

Suggests a **consistent rental pattern** across all days.

Business strategy should focus on daily consistent availability rather than weekday-weekend variations.

**Checking if the demand of bicycles on rent is the same for different Weather conditions?**

**checking skewness of count column¨**

Value of skewness is 1.24, which is greater than 0. Hence we can conclude that it is right or positively skewed.

**Shapiro-Wilk Test to check normal distribution¨**

#Ho: Data is Gaussian

#Ha: Data is  not Gaussian

test stat- 0.88 and p value 5.369X10^-68. We have considered alpha is 0.05, hence we reject the null Hypothesis.  Given count column is not Gaussian.

Also we have checked by Q-Q plot. Graphically also it is not showing normally distributed.

**#checking variance between count and weathers by Homogeneity of Variance (Levene Test)**

# H0: Variances are equal

# Ha: Variances are not equal

levene stat 54.85 and p value 3.5049X10^-35. We have considered alpha is 0.05, hence rejected H0. Variances are not equal.

**we are comparing the means of more than two groups (weather categories), we use a One-Way ANOVA Test. But is not Gaussian, hence we will test it by Kruskal Wallis Test**

Null Hypothesis (Hâ‚€): There is no significant difference in bike demand across different weather conditions.

Alternative Hypothesis (Hâ‚): There is a significant difference in bike demand across different weather conditions.

f\_stats 205.04 and p\_value 3.50X10^-44, Hence we reject H0. meaning weather significantly affects bike demand.

**Recommendation**

Optimize bike availability based on weather forecasts.

Offer discounts during unfavorable weather to maintain demand. Business strategy should focus on daily consistent availability rather than weekday-weekend variations.

**Checking if the demand of bicycles on rent is the same for different Seasons?**

**#checking variance between count and weathers by Homogeneity of Variance (Leveneâ€™s Test)Ã¨**

# H0: Variances are equal

# Ha: Variances are not equal

levene stat 187.77 and p value 1.014X10^-118. We have considered alpha is 0.05, hence rejected H0. Variances are not equal.

**we are comparing the means of more than two groups (weather categories), we use a One-Way ANOVA Test. But is not Gaussian and variances are not equal, we will test it by Kruskal Wallis Test**

Null Hypothesis (Hâ‚€): There is no significant difference in bicycle demand across different seasons.

Alternative Hypothesis (Hâ‚): There is a significant difference in bicycle demand across different seasons.

f\_stats 699.28 and p\_value 2.479X10^-151, Hence we reject H0. meaning weather significantly affects bike demand.

**Recommendation**

Adjust bike availability based on seasonality trends.

Offer seasonal promotions or discounts to balance demand.

Plan maintenance schedules in lower-demand seasons.

**Checking if the Weather conditions are significantly different during different Season.**

Null Hypothesis (Hâ‚€): Weather conditions and seasons are independent (i.e., weather conditions are evenly distributed across all seasons).

Alternative Hypothesis (Hâ‚): Weather conditions depend on seasons (i.e., some weather conditions are more likely in specific seasons).

Since both **"Weather"** and **"Season"** are categorical variables, we use the **Chi-Square Test for Independence**

chi stat is 49.158 and p value 1.54X10^-07. Hence rejected the H0, they are associated.

**Recommendation**

**Seasonal Planning:** Prepare different rental strategies based on weather expectations for each season.

**Safety Measures:** Offer protective gear (raincoats, helmets) in seasons with frequent adverse weather.

**Demand Forecasting:** Adjust bike availability based on seasonal weather patterns