**About Dataset**

This dataset captures Ola Bike ride requests over time, along with associated environmental and user metrics. Each record represents a specific hour and includes details such as the season, weather situation, temperature, humidity, and windspeed at that time.

Libraries used include ;

1.**pandas** - For loading the dataset, cleaning data, handling missing values, and creating new time features

2. **numpy** - For numerical operations, especially mean imputation and aggregations

3 . **matplotlib.pyplot** - For basic plotting (line plots, formatting visuals).

4. **seaborn** - For advanced and attractive visualizations (bar plots, heatmaps,

**1. Data Cleaning**

Observations:

Dataset has 10,886 rows and 9 columns.

Columns temp, humidity, and windspeed have missing values.

datetime is of type object (needs to be converted to datetime).

No column names contain typos or inconsistent casing.

**Actions Taken:**

1. **Converted** datetime column to proper datetime format.
2. **Extracted time features**: hour, day, month, year, dayofweek.
3. **Checked for duplicates**: **0 duplicate rows** found.
4. **Missing values**:
   * temp: 1,632 missing
   * humidity: 1,632 missing
   * windspeed: 1,632 missing

I used mean imputation to deal with the missing data instead of dropping them as I faced a risk of losing 15% of my data to deal with the missing data

Dataset is now clean and ready for analysis and visualization

**Exploratory Data Analysis & Visualization Report**

This step involved analyzing trends, patterns, and relationships in the Ola Bike dataset using visual tools. Below is a summary of each visualization and the key insights drawn:

Customized the Visual Style in **Seaborn**

**Visualizations**

**1. Total Ride Count Over Time**

* **Description**: A line plot of total ride count across the full datetime range.
* **Insight**: The graph shows an upward trend in ride demand over time, with noticeable fluctuations across seasons and years. This may indicate increasing user adoption or seasonal effects influencing ridership.

**2. Average Ride Count by Hour of Day**

* **Description**: A bar plot showing average rides for each hour of the day (0–23).
* **Insight**: Peak hours are typically **8 AM** and **5–6 PM**, aligning with standard **commuting times**. There's also a visible dip during late night and early morning hours (1–5 AM), which suggests fewer rides during off-peak hours.

**3. Average Ride Count by Day of Week**

* **Description**: A bar chart of average ride counts across weekdays (0 = Monday, 6 = Sunday).
* **Insight**: Ridership is slightly higher on **weekdays**, especially **Thursday and Friday**, likely due to work-related travel. **Weekends** (Saturday and Sunday) show a decline, possibly indicating fewer commuting-based rides.

**4. Feature Correlation Heatmap**

* **Description**: A correlation matrix displaying relationships between numerical features such as temp, humidity, windspeed, casual, registered, and count.
* **Insight**:
  + - count is strongly positively correlated with registered users (as expected).
    - There is a moderate positive correlation between temp and count, suggesting that warmer weather encourages more rides.
    - humidity and windspeed show weak or slightly negative correlation with count, indicating they have less direct impact.

**Summary**

* The **demand for rides** follows daily patterns typical of urban commuting.
* **Weather conditions** like temperature do influence ridership.
* **Time-based features** (hour, weekday) are important predictors of user behavior.
* This analysis prepares the ground for predictive modeling, feature selection, and business recommendations.