

# Yulu

❖ Topic: Hypothesis Testing

❖ Duration: 1 week

---

## Introduction:

Yulu, India's pioneering micro-mobility service provider, has embarked on a mission to revolutionize daily commutes by offering unique, sustainable transportation solutions.

However, recent revenue setbacks have prompted Yulu to seek the expertise of a consulting company to delve into the factors influencing the demand for their shared electric cycles, specifically in the Indian market.

## Why this case study?

### From Yulu's Perspective:

- Strategic Expansion: Yulu's decision to enter the Indian market is a strategic move to expand its global footprint. Understanding the demand factors in this new market is essential to tailor their services and strategies accordingly.
- Revenue Recovery: Yulu's recent revenue decline is a pressing concern. By analyzing the factors affecting demand for shared electric cycles in the Indian market, they can make informed adjustments to regain profitability.

### From Learners' Perspective:

- Real-World Problem-Solving: It presents an opportunity to apply machine learning and data analysis techniques to address a real-world business problem.
  - Market Insights: Analyzing factors affecting demand in the Indian market equips learners with market research skills. This knowledge is transferable to various industries.
  - Consulting Skills: Learners can develop their ability to act as consultants, providing data-driven insights to organizations.
-

Dataset:

Link: <https://drive.google.com/file/d/1o94fXnmvrx6jRgl6S-SeZ3tfnKjCDY0i/view?usp=sharing>

### Column Profiling:

- **datetime**: datetime
  - **season**: season (1: spring, 2: summer, 3: fall, 4: winter)
  - **holiday** : whether day is a holiday or not
  - **workingday** : if day is neither weekend nor holiday is 1, otherwise is 0.
  - **weather**:
    - o 1: Clear, Few clouds, partly cloudy
    - o 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
    - o 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
    - o 4: Heavy Rain + Ice Pellets + Thunderstorm + Mist, Snow + Fog
  - **temp**: temperature in Celsius
  - **atemp**: feeling temperature in Celsius
  - **humidity**: humidity
  - **windspeed**: wind speed
  - **casual**: count of casual users
  - **registered**: count of registered users
  - **count**: count of total rental bikes including both casual and registered
- 

### How to get started?

To complete the case study, begin by downloading the CSV files from the provided link. Afterward, proceed to upload them onto Google Colab / Jupyter Notebook for further analysis.

If you are using Google Colab, you can directly start working on the notebook on [Colab](#).

Install Anaconda using the [link](#). Once Anaconda has been installed on your system, open Jupyter Notebook. Refer [link](#).

Now, the netflix CSV file needs to be uploaded/imported in the Colab/Jupyter notebook respectively.

Once the file have been successfully uploaded/imported, you can conveniently access them within the notebook using the [read\\_csv\(\)](#) method.

---

## What is expected?

The company wants to know:

- Which variables are significant in predicting the demand for shared electric cycles in the Indian market?
- How well those variables describe the electric cycle demands.

## Submission Process:

- Type your insights and recommendations in the text editor.
- Convert your jupyter notebook into PDF (Save as PDF using Chrome browser's Print command), upload it in your Google Drive (set the permission to allow public access), and paste that link in the text editor.
- Optionally, you may add images/graphs in the text editor by taking screenshots or saving matplotlib graphs using `plt.savefig(...)`.
- After submitting, you will not be allowed to edit your submission.

## General Guidelines:

- Evaluation will be kept lenient, so make sure you attempt this case study.
  - It is understandable that you might struggle with getting started on this. Just brainstorm, discuss with peers, or get help from TAs.
  - There is no right or wrong answer. We have to get used to dealing with uncertainty in business. This is exactly the skill we want to develop.
- 

## What does 'good' look like?

### 1. Define the Problem Statement, Import the required Libraries and perform Exploratory Data Analysis.

- a. Examine dataset structure, characteristics, and statistical summary.
  - i. **Hint:** You can use `.shape`, `.info()`, `.describe()`
- b. Identify missing values and perform Imputation using an appropriate method.

- i. **Hint:** You can use `.isnull()` or `.isna()`
- c. Identify and remove duplicate records.
  - i. **Hint:** You can use `.duplicated()`
- d. Analyze the distribution of Numerical & Categorical variables, separately
 

**Hint:**

  - i. For Numerical features use Histogram, Distplot, etc.
  - ii. Hint: For Categorical features use Countplot, Pie Chart, etc.
- e. Check for Outliers and deal with them accordingly.
 

**Hint:**

  - i. You can use Boxplot, Interquartile Range (IQR)
  - ii. Remove/Clip existing outliers as necessary.

## 2. Try establishing a Relationship between the Dependent and Independent Variables.

**Hint:**

- i. Plot a Correlation Heatmap and draw insights.
- ii. Remove the highly correlated variables, if any.

## 3. Check if there any significant difference between the no. of bike rides on Weekdays and Weekends?

- a. Formulate Null Hypothesis (H0) and Alternate Hypothesis (H1)
- b. Select an appropriate test -
  - i. **Hint:** 2- Sample Independent T-test
- c. Set a significance level
  - i. **Hint:**  $\alpha=5\%$  is recommended
- d. Calculate test Statistics / p-value
- e. Decide whether to accept or reject the Null Hypothesis.

**Hint:**

- i. If the p-value is less than or equal to the predetermined level of significance ( $\alpha$ ), we have evidence to reject the null hypothesis.
  - ii. If the p-value is greater than the predetermined level of significance ( $\alpha$ ), we do not have sufficient evidence to reject the null hypothesis.
- f. Draw inferences & conclusions from the analysis and provide recommendations.

## 4. Check if the demand of bicycles on rent is the same for different Weather conditions?

- a. Formulate Null Hypothesis (H0) and Alternate Hypothesis (H1)
- b. Select an appropriate test -

- i. **Hint:** One-way ANOVA test
- c. Check assumptions of the test
  - i. Normality
    - Hint:**
      - 1. Use Histogram, Q-Q Plot, Skewness & Kurtosis
      - 2. Shapiro-Wilk's test
  - ii. Equality Variance
    - Hint:**
      - 1. Levene's test
  - iii. Please continue doing the analysis even if some assumptions fail (Levene's test or Shapiro-wilk test) but double check using visual analysis and report wherever necessary.
- d. Set a significance level and Calculate the test Statistics / p-value.
  - i. **Hint:** alpha=5% is recommended
- e. Decide whether to accept or reject the Null Hypothesis.
  - Hint:**
    - i. If the p-value is less than or equal to the predetermined level of significance (alpha), we have evidence to reject the null hypothesis.
    - ii. If the p-value is greater than the predetermined level of significance (alpha), we do not have sufficient evidence to reject the null hypothesis.
- f. Draw inferences & conclusions from the analysis and provide recommendations.

**5. Check if the demand of bicycles on rent is the same for different Seasons?**

- a. Formulate Null Hypothesis (H0) and Alternate Hypothesis (H1)
- b. Select an appropriate test -
  - i. **Hint:** One-way ANOVA test
- c. Check assumptions of the test
  - i. Normality
    - Hint:**
      - 1. Use Histogram, Q-Q Plot, Skewness & Kurtosis
      - 2. Shapiro-Wilk's test
  - ii. Equality Variance
    - Hint:**
      - 1. Levene's test
  - iii. Please continue doing the analysis even if some assumptions fail (Levene's test or Shapiro-wilk test) but double check using visual analysis and report wherever necessary.

- d. Set a significance level and Calculate the test Statistics / p-value.
  - i. **Hint:** alpha=5% is recommended
- e. Decide whether to accept or reject the Null Hypothesis.  
**Hint:**
  - i. If the p-value is less than or equal to the predetermined level of significance (alpha), we have evidence to reject the null hypothesis.
  - ii. If the p-value is greater than the predetermined level of significance (alpha), we do not have sufficient evidence to reject the null hypothesis.
- f. Draw inferences & conclusions from the analysis and provide recommendations.

**6. Check if the Weather conditions are significantly different during different Seasons?**

- a. Formulate Null Hypothesis (H0) and Alternate Hypothesis (H1)
    - i. **Hint:** The values in 'Weather' & 'Season' columns should be treated as distinct Categories. Encode them accordingly.
  - b. Select an appropriate test -
    - i. **Hint:** Chi-square test
  - c. Create a Contingency Table against 'Weather' & 'Season' columns
    - i. **Hint:** You can use a Pandas' CrossTab
  - d. Set a significance level and Calculate the test Statistics / p-value.
    - i. **Hint:** alpha=5% is recommended
  - e. Decide whether to accept or reject the Null Hypothesis.  
**Hint:**
    - i. If the p-value is less than or equal to the predetermined level of significance (alpha), we have evidence to reject the null hypothesis.
    - ii. If the p-value is greater than the predetermined level of significance (alpha), we do not have sufficient evidence to reject the null hypothesis.
  - f. Draw inferences & conclusions from the analysis and provide recommendations.
- 

## FAQs

**Q. Which platform am I supposed to use?**

You may use either Google Colab or Jupyter notebook.

**Q. I am having issues setting up Jupyter notebook**

Install Anaconda using the [link](#). Once Anaconda has been installed on your system, open Jupyter Notebook. Refer [link](#).