

## Google Colab Link:

<https://colab.research.google.com/drive/10tzYgZQsW2BZdLC0dV2jEAg5zN8QkQva?usp=sharing>

1. Define Problem Statement and perform Exploratory Data Analysis (10 points)
  - Definition of problem (as per given problem statement with additional views)

Assumptions: Since the count attribute is not following the Gaussian Distribution, I have considered sample of count, which will always follow Gaussian Distribution according to CLT.

- Yulu has recently suffered considerable dips in its revenues.
  - They would like to know the factors on which the demand for these shared electric cycles depends, with the help of data provided.
    - Observations on shape of data, data types of all the attributes, conversion of categorical attributes to 'category' (If required) , missing value detection, statistical summary.
1. The data consists of 10886 rows, 18 columns.
  2. There are no missing values in the given dataset.
  3. Total count of the rental vehicles in the dataset is 2085476.
  4. The date attribute is given in object datatype which has been converted to datetime datatype.
  5. It consists of the data from 2011-01-01 to 2012-12-19, which corresponds to almost 2 years of data.
  6. **The data from 20<sup>th</sup> to 30<sup>th</sup> dates for all the months seems missing, as no record has these dates.**
    - Univariate Analysis (distribution plots of all the continuous variable(s) barplots/countplots of all the categorical variables)
- From the barplot of the year column, it is clearly visible that the count of the rental vehicles has increased by almost 66% in 2012 , compared to that of 2011.
  - There are almost same no.of records corresponding to all the seasons.
  - There are very less no.of government holidays.
  - The no.of records corresponding to working days are almost double than that of non-working days.
  - Highest no.of records corresponds to weather 1, followed by 2,3. There are almost no records corresponding to weather 4.
  - Most no.of records corresponds to temperature range of 5 to 37 degree Celsius.
  - Most no.of records corresponds to feeling temperature range of 5 to 40 degree Celsius.
  - Most no.of records corresponds to Humidity range of 20 to 100.

- There are almost no records for windspeed above 30.
  - Bivariate Analysis (Relationships between important variables such as workday and count, season and count, weather and count).
- The count of the rentals increased by almost 66 percent from 2011 to 2012.
- The count of casual rentals has increased slightly whereas the count of registered rentals has almost increased by 50 percent from 2011 to 2012.
- The rental count is more at 7 – 8 AM and from 16 to 19 hours in a day.
- The rental count is more in the 3<sup>rd</sup> season followed by 2<sup>nd</sup>, 4<sup>th</sup> and 1<sup>st</sup> and all of them have outliers.
- The mean count on a holiday is less than that of non-holiday and the boxplot on non-holiday has more no.of outliers, whereas box plot on holiday don't have any outliers, and the median of the count is almost same for a holiday and a non-holiday.
- The median value of the count of rentals is same on working day and a non-working day, where the spread of the box plot is more for a non-working day, and the boxplot of the working day has more no.of outliers compared to that of non-working day.
- The count of rental vehicles is more for weather 1 and 2 compared to that of 3. The count of rentals is almost zero for weather 4.
- 3<sup>rd</sup> season has more temperature followed by 2,4,1.
- For all the seasons and weathers, the rental count shows a similar pattern, and for weather 4, the count of rental vehicles is almost 0.
- The count of rentals is more for temperature range: 13-37 degree Celsius.
- The count of rental vehicles is more for humidity range: 18 to 89.
- The count of rental vehicles is more for windspeed range: 0 to 47.
- The count of rentals for January and February is less compared to other months.
- The count of rentals is almost same for all the days of a month.
- The median and spread of the box plot for count Vs Day of the week is almost same with similar median values. On the weekends, the casual rentals increased and the registered rentals decreased compared to other days.
  - Illustrate the insights based on EDA
    - Comments on range of attributes, outliers of various attributes
    - Comments on the distribution of the variables and relationship between them
    - Comments for each univariate and bivariate plots

- Almost all the boxplots of all the attributes Vs count have outliers.
- All the attributes are showing similar pattern for the years 2011 and 2012, except that the proportion of all the attributes has increased significantly for 2012 compared to 2011.
- On the rainy days the rental count is almost zero.
- On Government holidays, the rental count is very low.
- On weekends, the count of rentals is similar to that of weekdays, with more no. of outliers.

## 2. Hypothesis Testing (30 Points):

- **2- Sample T-Test** to check if Working Day has an effect on the number of electric cycles rented (10 points)
  - The median value for the count of vehicles on working day and non- working day is almost similar. So we can do a Ttest\_ind to know the dependency of count on the working day.
  - $H_0$  : Working day has no effect on the no. of rented vehicles.  
 $H_a$  : Working days has effect on the no. of rented vehicles.
  - As we got a Pvalue(0.59) > alpha(0.05), we can say that working day has no effect on the count of rental vehicles.
- **ANNOVA** to check if No. of cycles rented is similar or different in different 1. weather 2. season (10 points)

Since the qqplot of count is not aligned with that of Gaussian plot, we can conclude that count is not gaussian distribution, we can go for Kruskal test instead of Annova.

ANNOVA to check if No. of cycles rented is similar or different in different **weather**

- $H_0$  : The count of vehicles is similar in all weathers.  
 $H_a$  : The count of vehicles is not similar in all weathers
- Since Pvalue(3.5e-44) < alpha (0.05), we can reject  $H_0$ . So, the count of vehicles is not similar in all weathers.

ANNOVA to check if No. of cycles rented is similar or different in different **seasons**.

- $H_0$ : Seasons have no effect on the count of rental vehicles.  
 $H_a$ : Seasons have effect on the count of rental vehicles.
- Since Pvalue(0.0) < alpha(0.05), we can reject  $H_0$ . So, we can conclude that seasons have impact on the count of rental vehicles.

- **Chi-square test** to check if Weather is dependent on the season (10 points)
  - $H_0$  : Seasons and weather are independent of each other.

- $H_a$  : Seasons and weather are dependent on each other.
- Since  $P\text{value}(1.54e-07) < \alpha (0.05)$  ,Reject  $H_0$ , So, we can say that seasons have no impact on weather.

### 3. Notebook Quality (10 points):

- Structure & Flow
- Well commented code

What good looks like (distribution of 10 points):

- Visual analysis (1)
- Hypothesis formulation (1)
- Select the appropriate test (1)
- Check test assumptions (2)
- Find the p-value(1)
- Conclusion based on the p-value (2)

### Insights:

- There is almost 66 percent hike in the total rental count in 2012 compared to 2011.
- There is slight increase in the casual rentals but the registered rentals has seen almost 70 percent hike.
- All the corresponding patterns shape of Count Vs time, season, holiday, working day, temperature, humidity, windspeed are almost same for year 2011 and 2012 except that there is huge increase in 2012 compared to 2011(pattern shape is same and all the counts increased proportionately).
- So, the dip likely corresponds to 2011 revenue, as we can see there is significant growth in count in 2012 due to rapid increase in the registered users.
- The dip likely corresponds to lack of awareness of the services and the app among the people for the year 2011 and may be due to some other external factors.

### Recommendations:

- An awareness campaign about the services and the app can be run across various social networking sites and OTT platforms, so that more people can know about the services that the company is providing.
- Coupons can be sent to the newly registered customers for first 3 to 4 rides, so that they can understand the service better and the probability of converting them into regular users can be increased.
- Feedback can be collected from those users on improvement of the services on using the app and rental vehicles.

- Feedback also helps in analyzing the growing demand to get future ready to handle the demand and expand the business, which can be win-win situation for both users and common people in the current scenario, where traffic, pollution, expenditure on gasoline can be reduced without the compromise on comfort for the people.