Business Case: Yulu - Hypothesis Testing

About Yulu

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.

How you can help here?

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

Concept Used:

- Bi-Variate Analysis
- 2-sample t-test: testing for difference across populations
- ANNOVA
- Chi-square

How to begin:

- Import the dataset and do usual exploratory data analysis steps like checking the structure & characteristics of the dataset
- Try establishing a relation between the dependent and independent variable (Dependent "Count" & Independent: Workingday, Weather, Season etc)
- Select an appropriate test to check whether:
  - Working Day has effect on number of electric cycles rented
  - No. of cycles rented similar or different in different seasons
  - No. of cycles rented similar or different in different weather
  - Weather is dependent on season (check between 2 predictor variable)
- Set up Null Hypothesis (H0)
- State the alternate hypothesis (H1)
- Check assumptions of the test (Normality, Equal Variance). You can check it using Histogram, Q-Q plot or statistical methods like levene's test, Shapiro-wilk test (optional)
  - 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
- Set a significance level (alpha)
- Calculate test Statistics.
- Decision to accept or reject null hypothesis.
- Inference from the analysis

Yulu DATA ANALYSIS

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!

As per the business problem, 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.

It was a simple data frame which consisted data of number of electric vehicle rental purchases at each hour from 2011 and 2012 years. I found the data very clean and properly structured. There were no missing values found.

With the given data, I was able to draft certain observations which can help improve Yulu's business. Please find them given below:

- Starting with the total electric vehicle rental count, we can see that there has been a huge increase. It increased from 80k to 130k. If Yulu keeps the momentum going, there will be a huge increase in revenue.
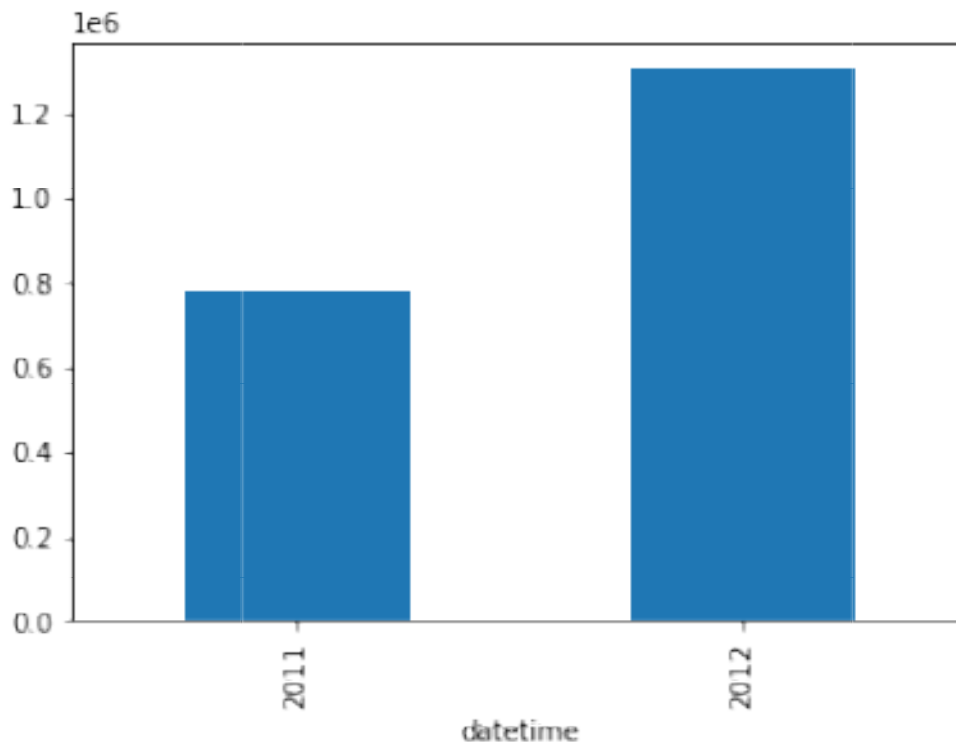
- From the below graph, we can see that the demand is higher at 8am and 4-7pm. These are office start and end hours. Hence, we need to keep more vehicles available to meet the demand. Vehicle unavailability at peak times can make a customer lose interest in future purchases.
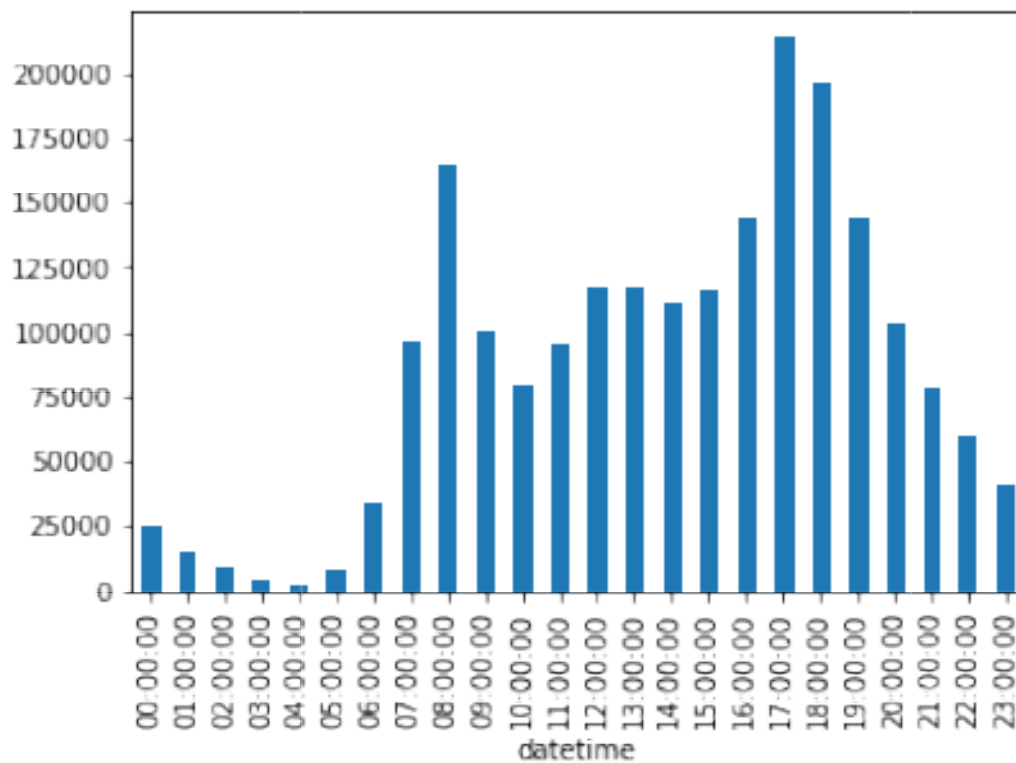
- I conducted a 2- Sample T-Test to check if Working Day has an effect on the number of electric vehicles rented. I could see that the count data for working and non working days were able to follow a normal distribution as per the qq plot. The data also satisfied all the assumptions required to conduct a t test. After, conducting the test, I was not able to find any difference in means. Hence, we can conclude that working day does not have any effect on rental purchase count of electric vehicles.

- To check if Weather is dependent on the season, I performed a Chi-square test. It is a non parametric test with only one assumption which is the data in the cells should be frequencies, or counts rather than percentages or some other transformation of the data. I was able to get this using pandas cross tab. After the test, I was able to come to a conclusion that there is an association or dependency between weather and season.

- I also conducted a 2 way ANNOVA to check if No. of cycles rented is similar or different in different 1. weather 2. season. I was able to find a clear and huge dependency between weather vs count and season vs count. Number of cycles rented is different in different Weather and season.

- Finally, I did plot bar graphs to check how temp, atemp, humidity and windspeed affect vehicle purchases. Humans ideally feel comfortable between temperatures of 22 to 27 degC and a relative humidity of 40 to 60%. From the graphs given below, we can see that rental purchase count is higher if the there is ideal temp or humidity. If the temp and relative humidity increases and decreases beyond the ideal, then the electric cycle count starts to decrease. The same goes with windspeed too.
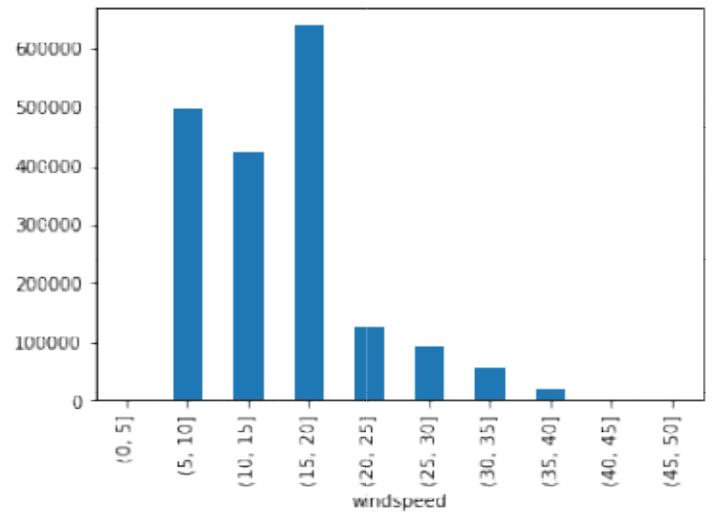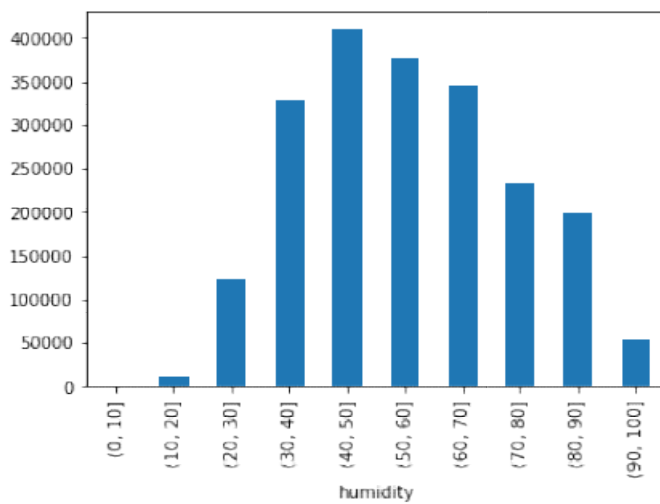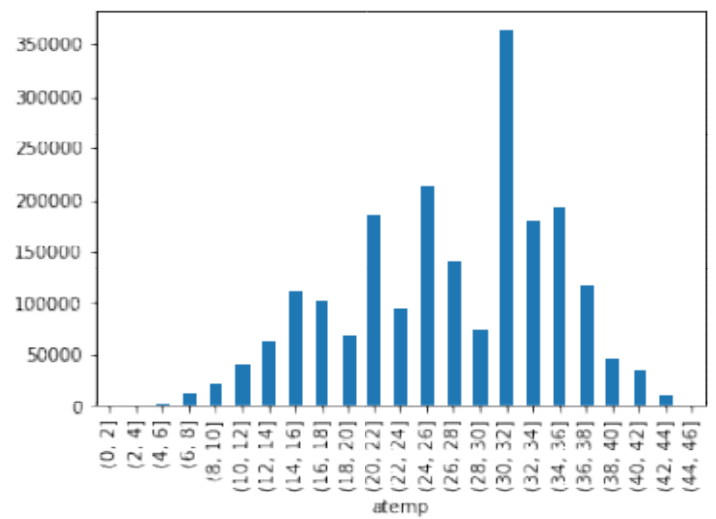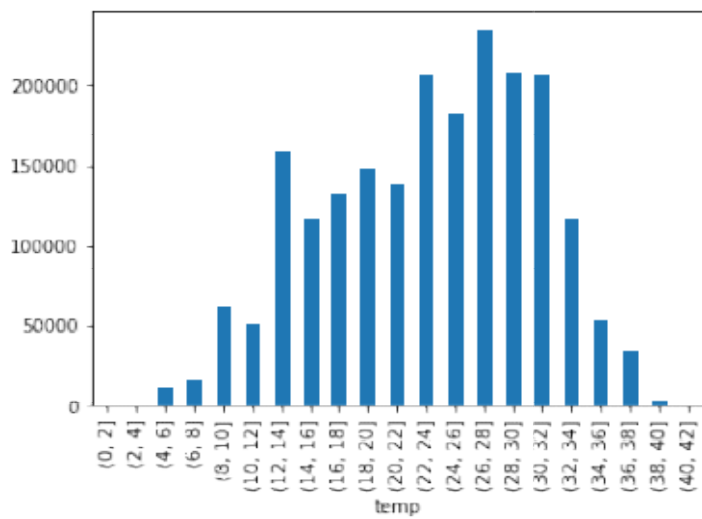
To whoever reads this, I hope my insights from this case study were meaningful.

Thank you,
Krishna