Final Project Report
Introduction to Data Analytics

Project Title:
Bike Sharing System

Prepared by:
Siddhesh Nikam(N015367555)
Ishita Arora(N01543414)

ITE 5201 – Winter 2023
Humber College

1. Problem Statement

=> Analysis of a bike-sharing system that lets people use bikes for a short period of time for a cost or for free.

2. Dataset Description

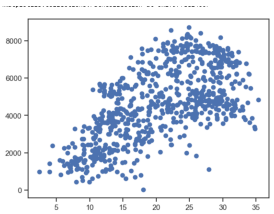
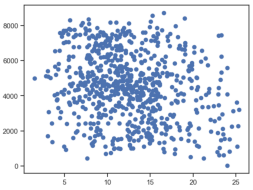
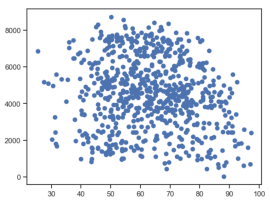
- A sizable dataset on daily bike demand across the American market has been gathered by BoomBikes. Based on a number of parameters, the dataset contains information regarding the demand for shared bikes.
- The dataset might additionally contain extra factors like place, hour of the day, day of the week, and holidays.
- Variables such as hum, temp, windspeed, season,holidays,weathersit, cnt are used for the analysis and predictions.

3. Dataset Analysis and Observations

=> For dataset analysis, we used pairplots for univariate and bivariate analysis and heatmap for finding correlation coefficient.

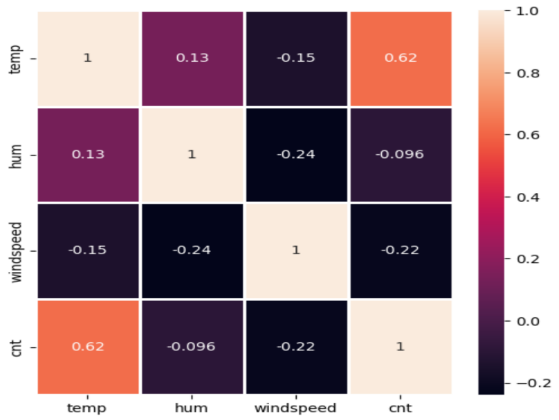
=> The data can reveal, for instance, that weekdays are better than weekends for shared bike demand or that bad weather decreases demand.

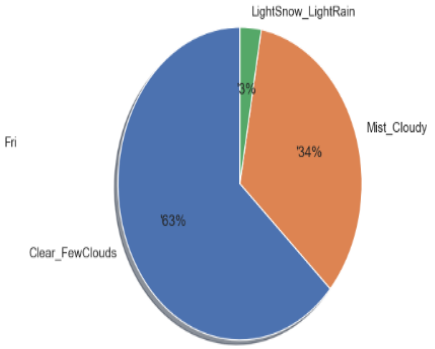
-**Observation**

	Cnt VS Temp	Cnt VS Weather	Cnt VS Hum
PairPlot			
○ Direction ○ Form ○ Strength ○ Outlier	Positive Liner Weak Yes	Negative Non Linear Weak Yes	Neither Non Linear Weak Yes

=> This heatmap plot using spearman method which donates relationship between two data variables and returns its Correlation Coefficient rank ®

- -R = 1 Strong Positive relationship.
- -R = 0 Not Linearly correlated.
- -R = -1 Strong negative relationship.
- From this heatmap, we concluded that temp & windspeed is strongly correlated with cnt.





4. Proposed Analytical/Prediction Model

- In this analysis, the dependent variable is client demand, and the independent variables are things like temperature, weather, humidity, and wind speed.
- We have created a linear regression model to forecast how the demand will change as a result of these attributes.
- We divided the data into training and testing sets and trained the model on the training data using the LinearRegression() method.
- The model's performance was then assessed using both the testing data and the projected data. Our goal is to demonstrate how successfully the model has been trained to forecast client demand.

5. Results and Discussions

- In this analysis, we have used the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) metrics to evaluate the performance of our model in predicting customer demand based on the input variables. We also created a scatter plot to compare the actual and predicted demand values, using the color red to represent cases where the actual demand value is less than the predicted value, and green for cases where the actual value is greater than or equal to the predicted value. Our model achieved an MAE of 0.134, an MSE of 0.0268, and an RMSE of 0.163.
- Both the train and test dataset could conclude that the above variables can well explain almost 80% of bike demand.
- Coefficients of the variables explain the factors affecting the bike demand.