Rashi Nigam 07-09-2018

Bike Sharing Program

Overview

- A bicycle-sharing system, is a service in which bicycles are made available for shared use to individuals on a short-term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" and return it at another dock belong to the same system.
- One of the such leading Bike Sharing System is Capital Bikeshare System and Client for this project.

Why the Need

- Increase personal mobility, providing people with better access to destinations throughout the City
- Integrate bike share as an extension of public transit network
- Develop an innovative transportation system that improves livability and economic competitiveness
- Reduce the environmental impact of transportation and help achieve goal of 'Go Green'
- Develop a system that serves users in minority and low-income communities and improves their access to key destinations, such as jobs and recreation
- safe mode of transportation that promotes active and healthy living
- Create a system that is financially sustainable, transparently operated, and accountable to the public.

Project Objective

The client Capital Bikeshare System needed a research analysis on their dataset to optimize their service & operations that

- Predict the Bike Rental volume/count
- Factors or features that influence Bike Rental Count

About the Data

- Capital Bikeshare posts quarterly data reports of bike trip times, start and end locations, and type of user (registered or casual).
- Each trip is on one line of data. Dataset contains 17379 rows with 17 features.
- These data are readily and publicly available at https://www.kaggle.com/c/bike-sharing-demand/data

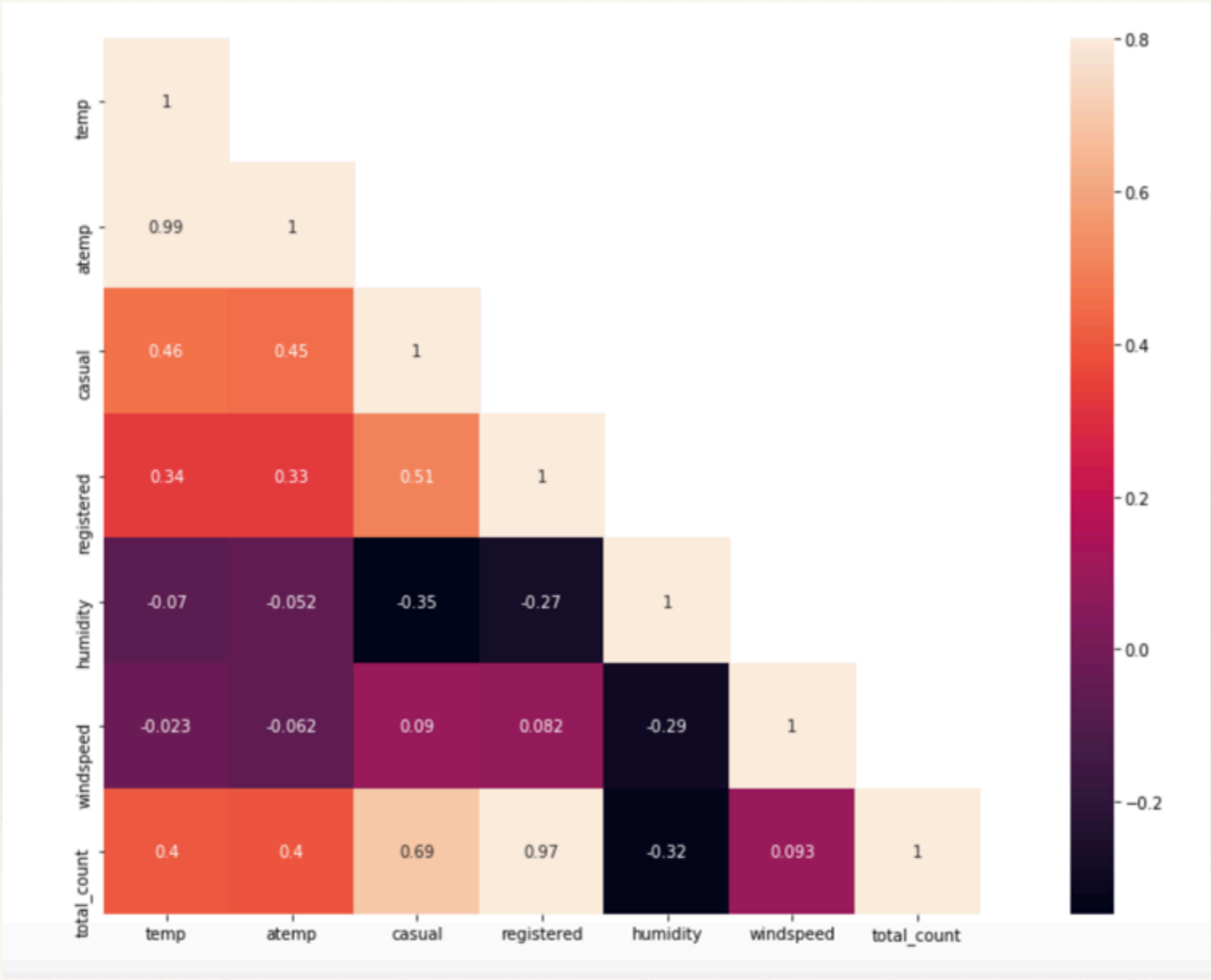
Data Wrangling Steps

The Capital bikeshare dataset required data wrangling in terms of

- extracting dataset followed by identifying meaningful dataset
- renaming a few columns based on preference
- feature transforming date timestamp to day, month, year

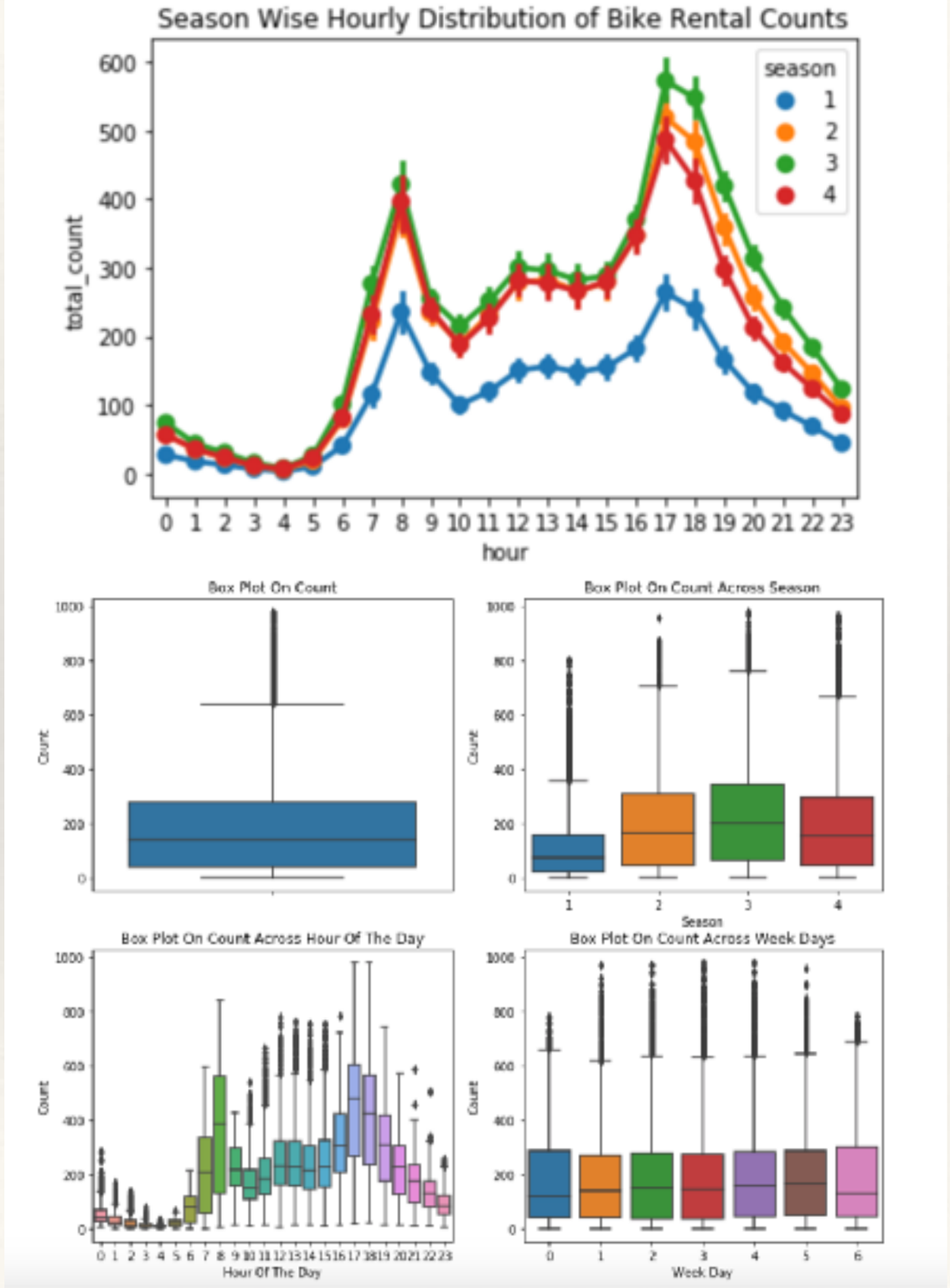
Exploratory Data Analysis

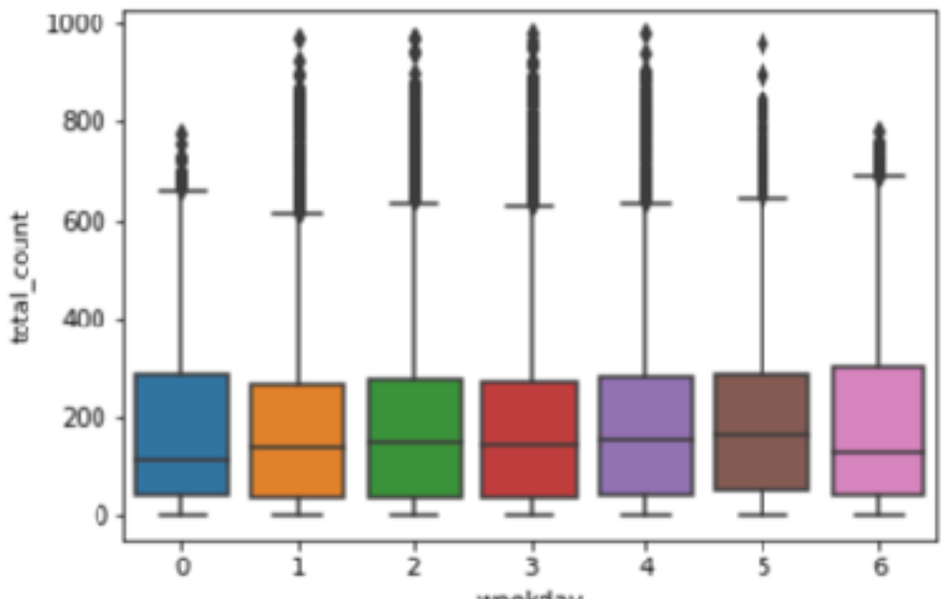
What impacts the Bike Rental Count

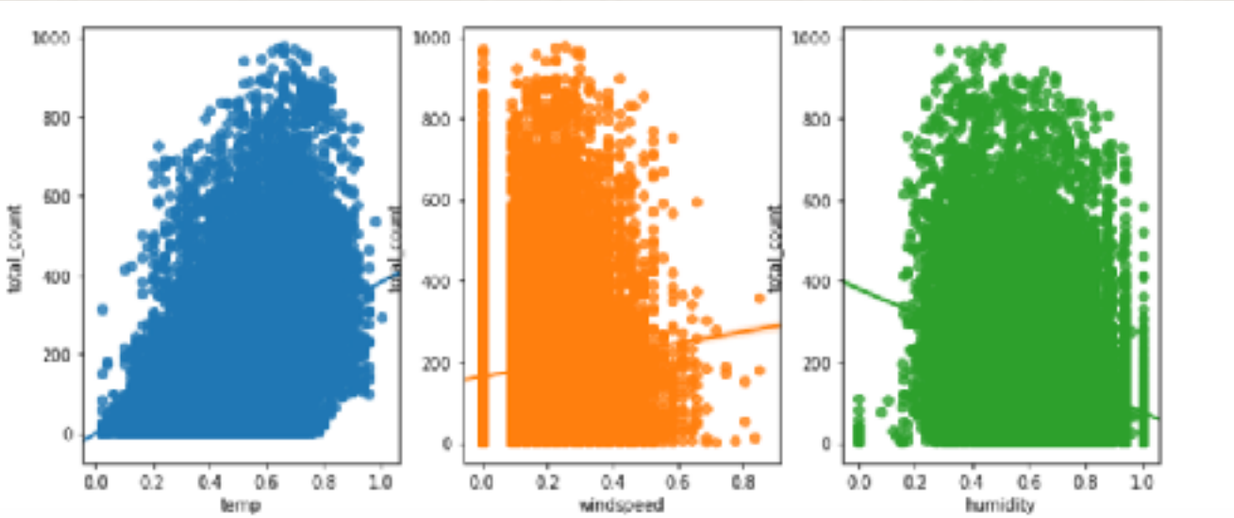


Exploratory Data Analysis

When impacts the Bike Rental Count

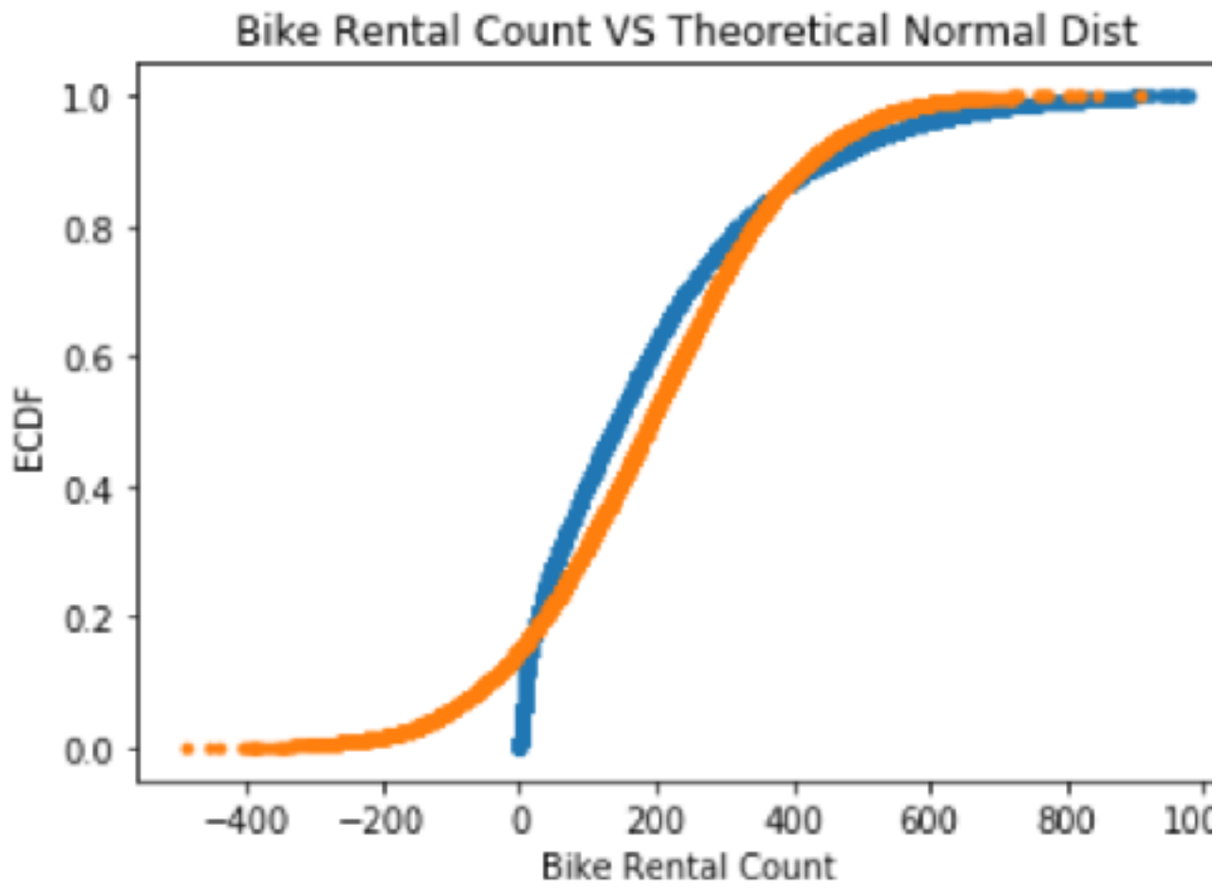


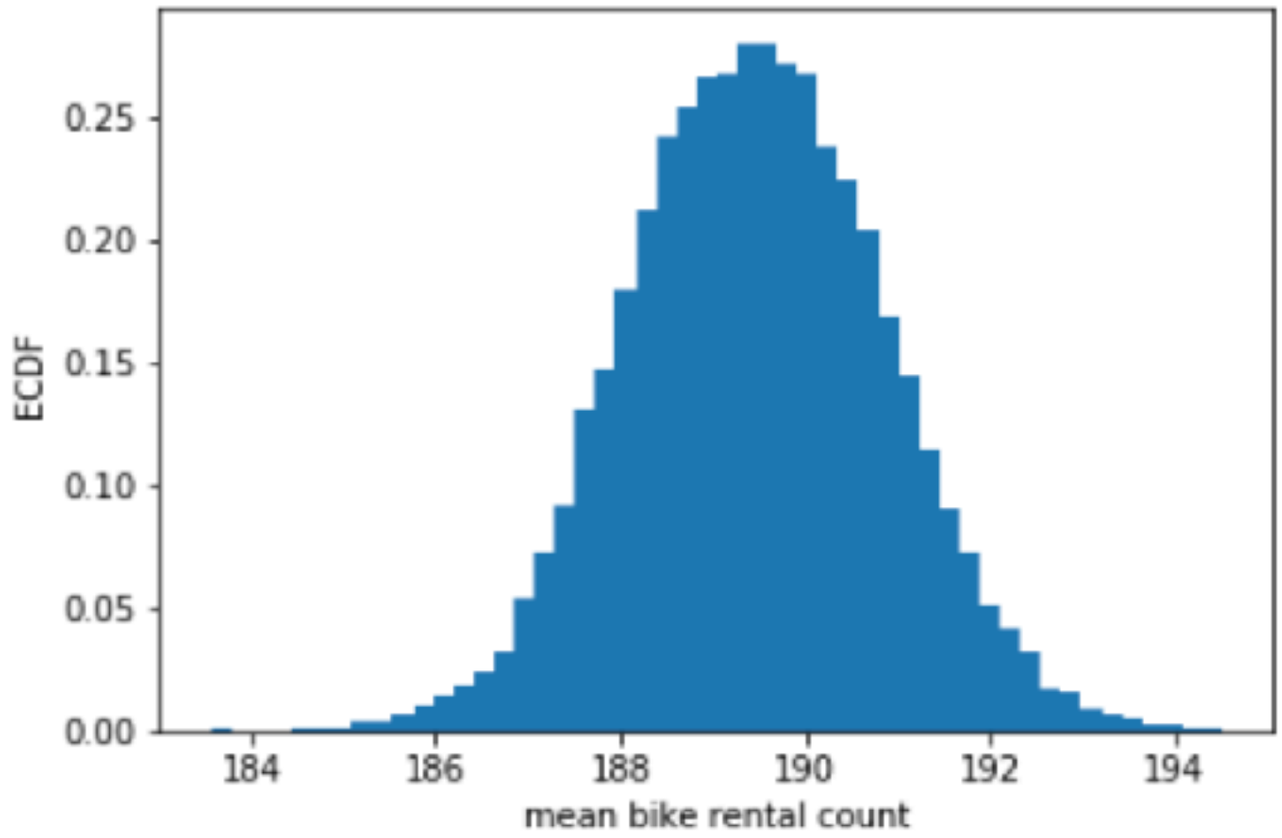




EDA/Inferential Statistics

Bike Rental Count Distribution

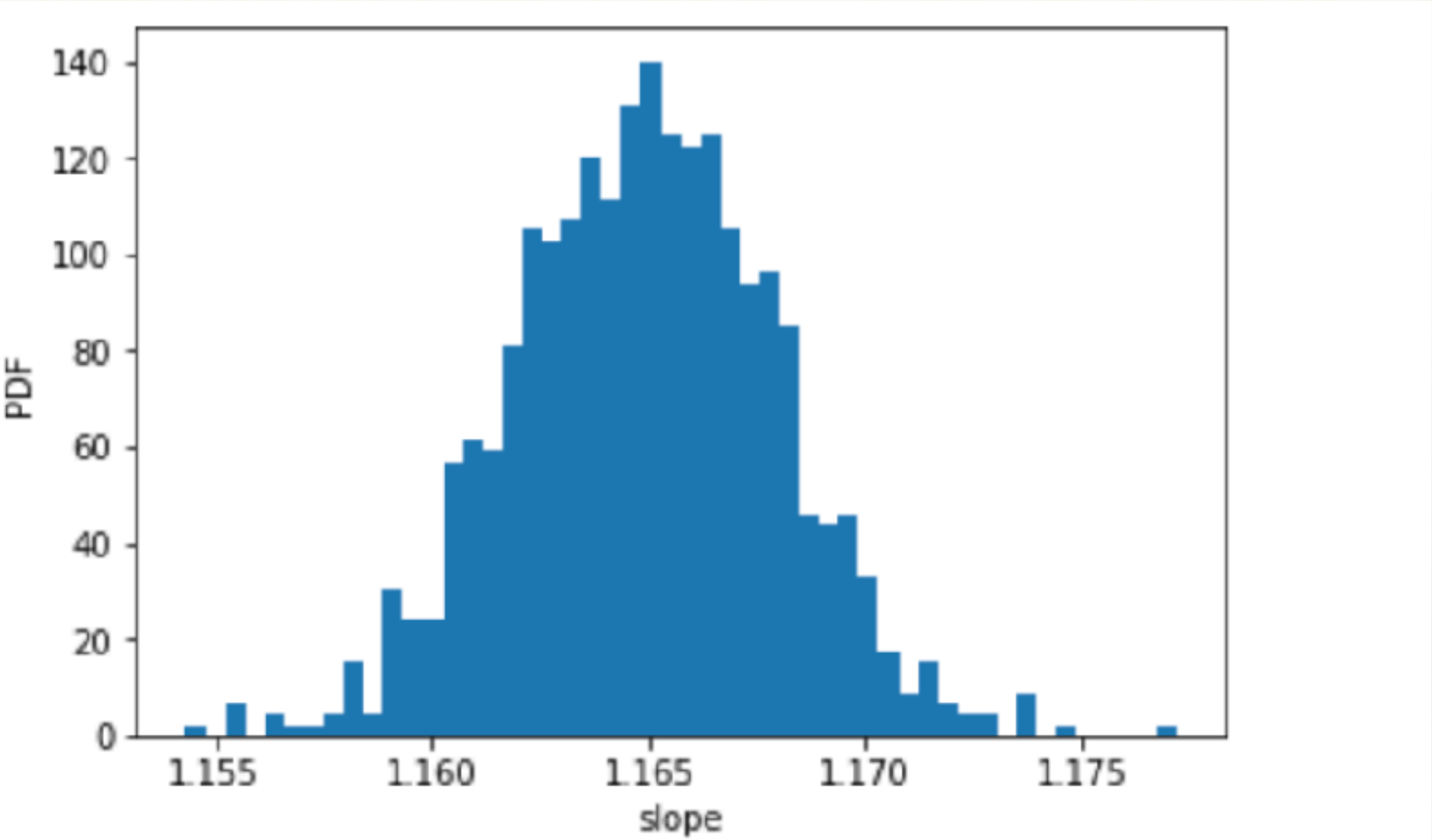




This is bootstrap estimate of the probability distribution function of the mean Bike Rental Count at the Capital Bikeshare System. It assumes 95% Confidence Interval.

EDA/Inferential Statistics

Who impacts Bike Rental Count Distribution

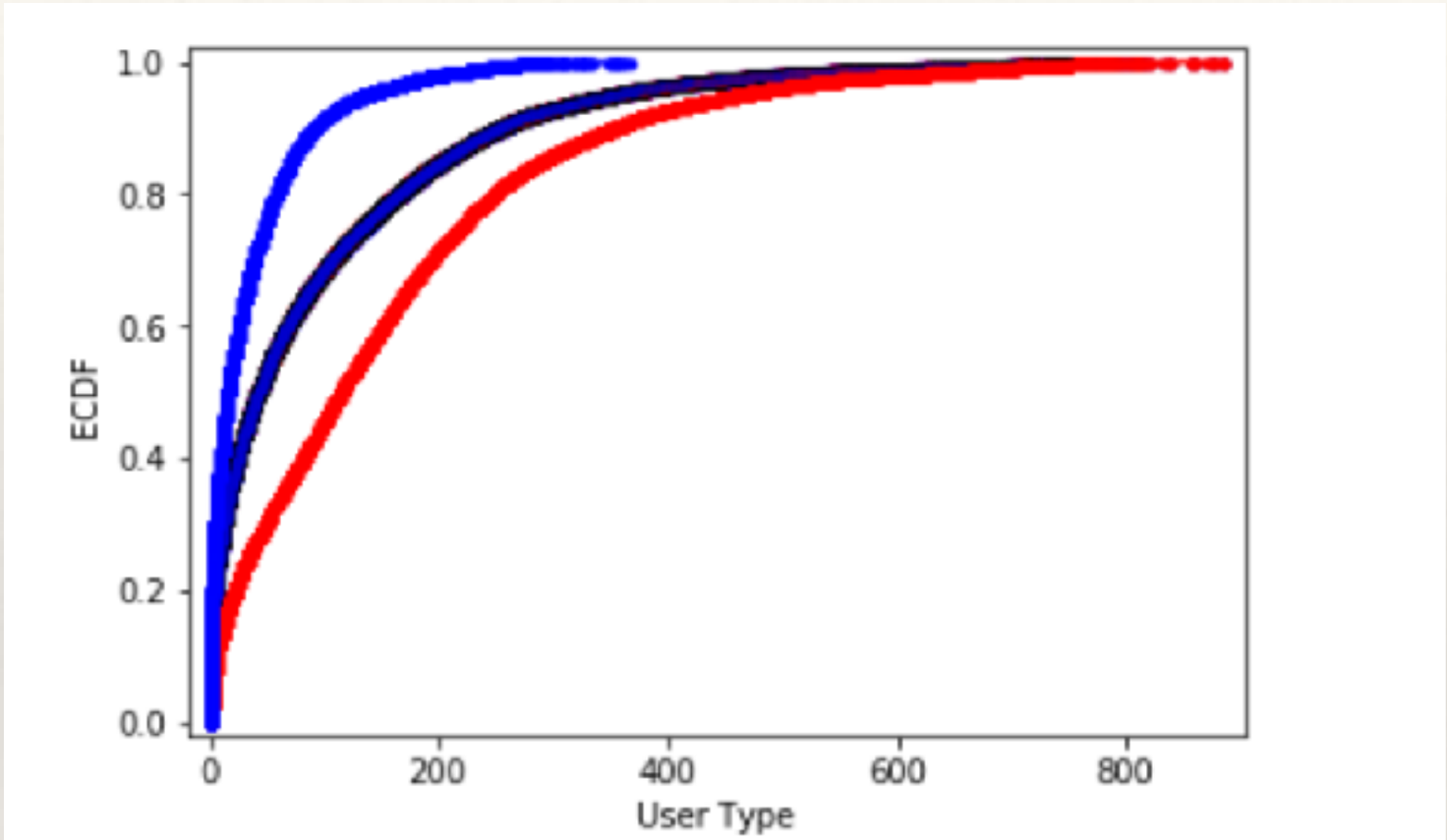


Extending Confidence Interval Concept to Pairs Bootstrap between Bike Rental Count and Registered User Type

Exploratory Data Analysis

How many impacts the Bike Rental Count

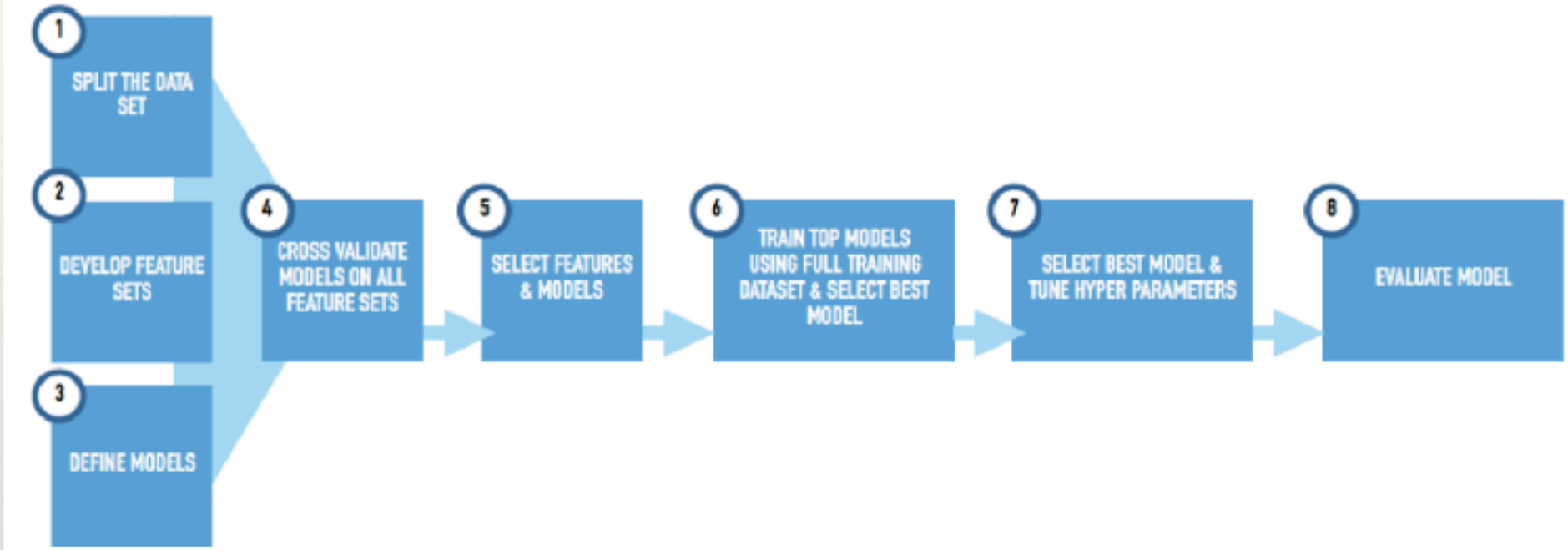
Null Hypothesis- There is no significant difference between registered and casual user type mean on Bike Rental Count.



Registered & Casual User Type Means are not identically distributed and do not influence data in similar way. So Null Hypothesis is rejected.

Supervised Regression

Steps in Machine Learning Modeling



Supervised Regression

Machine Learning Algorithms Analyzed

Regressor	Hyperparameters
Linear Regression	R2, CV
Ridge Regression	R2, CV
Lasso Regression	Min Samples Split = 10, Max Leaf Nodes = 500, criterion = MSE, Max
Decision Tree Regression	R2, CV, Pipeline

Supervised Regression

Model Evaluation

Regressor	R^2/MSE	Grid Search Cross Validation R squared/Elastic Net using Pipeline
Linear Regression	0.387/ -4525.62	0.389, 1.0
Ridge Regression	0.390	0.389, 1.0
Lasso Regression	0.378	0.389, 1.0
Decision Tree Regression	0.886/-97.24	0.858

Recommendations & Future Work

With this predictive model, Client may benefit in better prediction of Bike Rental Count:

- Demand for bike share program is maximum between 7 AM to 9 AM and 4 PM to 6 PM
- Weather conditions have direct correlation with Bike Rental Count
- Registered and Casual users are in equation to total bike rental count
- Bike Rental Count distribution by Registered and Casual Users is good predictive model

Future Work

There is lot of potential to enhance the model by:

- Collection of more features in the dataset like Gender and Age to help customer know if bike rental preference is by any age or gender group

- Model improvement using other Regression models like Random Forest, Support Vector

References

- https://www.washingtonpost.com/local/trafficandcommuting/capital-bikeshare-gears-up-for-another-expansion/2017/10/02/bcf81b4a-a2fe-11e7-ade1-76d061d56efa_story.html?utm_term=.7690885fb8e3
- http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1196&context=jpt
- https://en.wikipedia.org/wiki/Data_wrangling
- https://www.datawatch.com/what-is-data-wrangling/

Thank You