Final Project Report

Amy Lovas

Hannah Abraham

Kyle Smith

Sagar D. Goswami

OR-568-004: Applied Predictive Analytics

George Mason University

October 23, 2022

Under Guidance of:

Dr. Vadim Sokolov

(George Mason University)

Final Project Report

Bike sharing has become a popular form of ground transportation in recent years as society has advocated for less congestion and reduced vehicle emissions, especially within metropolitan areas and cities with dense populations. During the COVID-19 Pandemic, the demand for bike sharing skyrocketed as people were less inclined to use public transportation as there is a higher risk of getting infected with COVID-19 due to the potential exposure in a closed space. People have turned to bike sharing as the most viable option for ground transportation as it is the safest option to travel to their destination, especially for people who do not have access to a personal vehicle. Hence, for bike-sharing companies to meet the supply of bikes to the public, it is crucial to have the capability to predict bike demands within their operating town/city.

Dataset

The dataset was sourced from the UCI Machine Learning Repository and contains 8760 observations of bike sharing demand for the years 2017-2018 in Seoul, South Korea. The dataset contains 14 variables, summarizing weather information, bike rental information, and Public Holidays information in Korea.

The names and datatypes for the variables are as follows:

Table 1 Variables in the Bike-Sharing Database

Variable	Data-Type
Date : year-month-day	Date
Rented Bike count - Count of bikes rented at each hour	Integer
Hour - Hour of the day	Integer
Temperature-Temperature in Celsius	Float
Humidity - %	Integer
Windspeed - m/s	Float
Visibility - 10m	Integer
Dew point temperature - Celsius	Float

Variable	Data-Type
Solar radiation - MJ/m2	Integer
Rainfall - mm	Float
Snowfall - cm	Float
Seasons - Winter, Spring, Summer, Autumn	Text (Categorical)
Holiday - Holiday/No holiday	Binary
Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)	Binary

Our goal for this project is to train predictive models that can accurately predict the bike demand based on the selected features. This model can be used by the bike sharing companies to maintain enough supply of bikes to meet the anticipated demand and plan a better resource allocation strategy. Moreover, the model can also be used to analyze the impact of each factor on the bike demand, and thereby analyze ways to make the bike-sharing system more efficient, accessible, and reliable

Using the above models and analyses, we hope to provide insights that can ensure there is enough supply of bicycles to meet public demand by using a range of factors such as weather information, rented bike counts, and bike rental hours to predict the number of bicycles rentals.

To assist us with our analysis, the key questions below will be answered:

1. What features can bike share companies use to create a dynamic pricing model to monitor bike demands?
    a. Based on linked features, can we predict how often bike share services will be used on certain days?
    b. Can we use these to provide decision insights to bike share companies to influence price?
2. What is the bike-sharing trends among the public in Seoul?
3. Is bike sharing more popular during holidays?
4. Does weather effect the demand for bike sharing?

This project will utilize predictive models such as linear regression and logistic regression to conduct our analysis and predict the bike rental demand in Seoul. Additional models and questions may be added depending on the result of our initial analysis. This project will provide us with informative insights that will be beneficial for bike-sharing companies to determine how many bicycles they should supply to minimize operational costs and maximize profit.