Forecasting Lyft Fares Using Regression

Jessica Padilla

December 13, 2019

Introduction

Lyft is one of the many rideshare services available to people today. Founded in 2012, Lyft allows users to order rides through the use of a phone app. There are 6 types of rides available: Shared Ride, Lyft, Lyft XL, Lux, Lux Black, and Lux Black XL (Lyft). This project will focus on the the basic Lyft service, which is the most commonly used ride.

Fares are generally determined by the distance for each ride. However, Lyft and other rideshare companies have adopted the practice of what is called price surging. This involves increasing the price of a ride because the demand for Lyft cars exceeds the number of Lyft cars that are actually in service (Lyft). This can result from many conditions such as extreme temperatures, bad weather, and busy commute hours within the day. In this project, we will utilize a data set of Lyft rides within Boston, MA to see what factors cause such price surges. We will, then, use the determined factors to forecast the price of any Lyft ride.

Methods

Lyft data for the city of Boston was retrieved from Kaggle (https://www.kaggle.com/sliu65/data-mining-project-boston) and uploaded onto GitHub (https://github.com/jessicapadilla/uber_and_lyft/blob/master/lyft.csv.zip?raw=true). All of the data was, then, imported and downloaded into R. The tidyverse package was used for data cleaning, regression analysis, and visuzalition. The psych package was also used to create scatterplot matrices to assess correlation between variables.

The R code for this project can be found within the Supplementary Materials section of this paper and on GitHub (https://github.com/jessicapadilla/mta_finances/blob/master/code.R).

Results

Lyft fare price surging is affected by what is termed the surge multiplier. If the surge multiplier is equal to 1, then it means that there are enough Lyft cars in the area to meet user demand and, therefore, the fare remains unchanged. If the surge multiplier is greater than 1, then there are more users than Lyft cars in service. As a result, the standard fare is increased by the surge multiplier.

Frequency of Price Surging
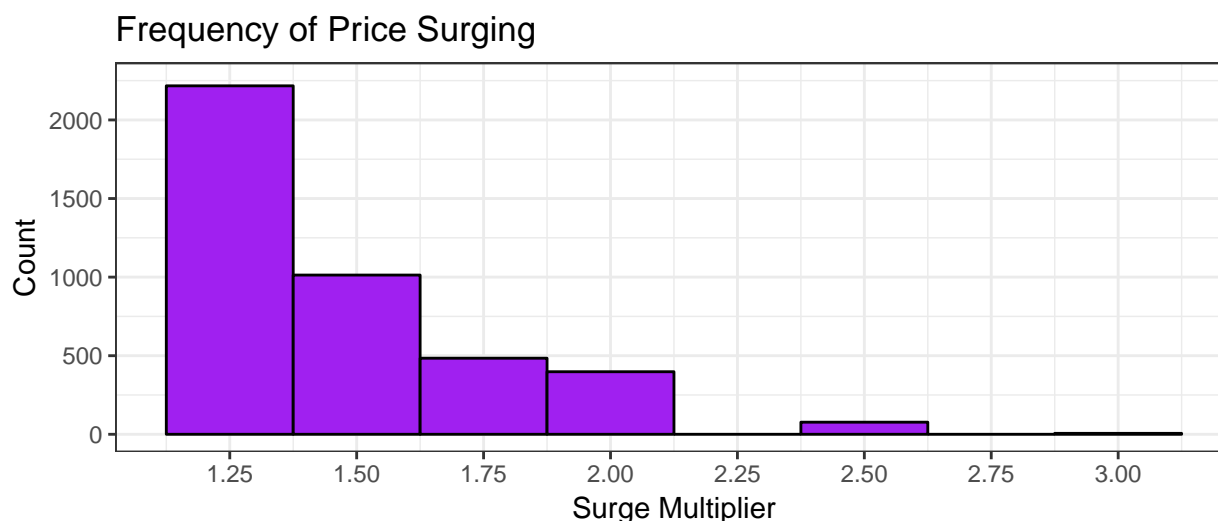
When surge multipliers are plotted against each Lyft ride's starting location, we see that the amount of surge multipliers is highest in the Boston district of Back Bay, which can be as high as roughly 3.00. Haymarket Square, on the other hand, has the fewest price surges and never goes above 1.75.

Surge Multiplier Based On Starting Location
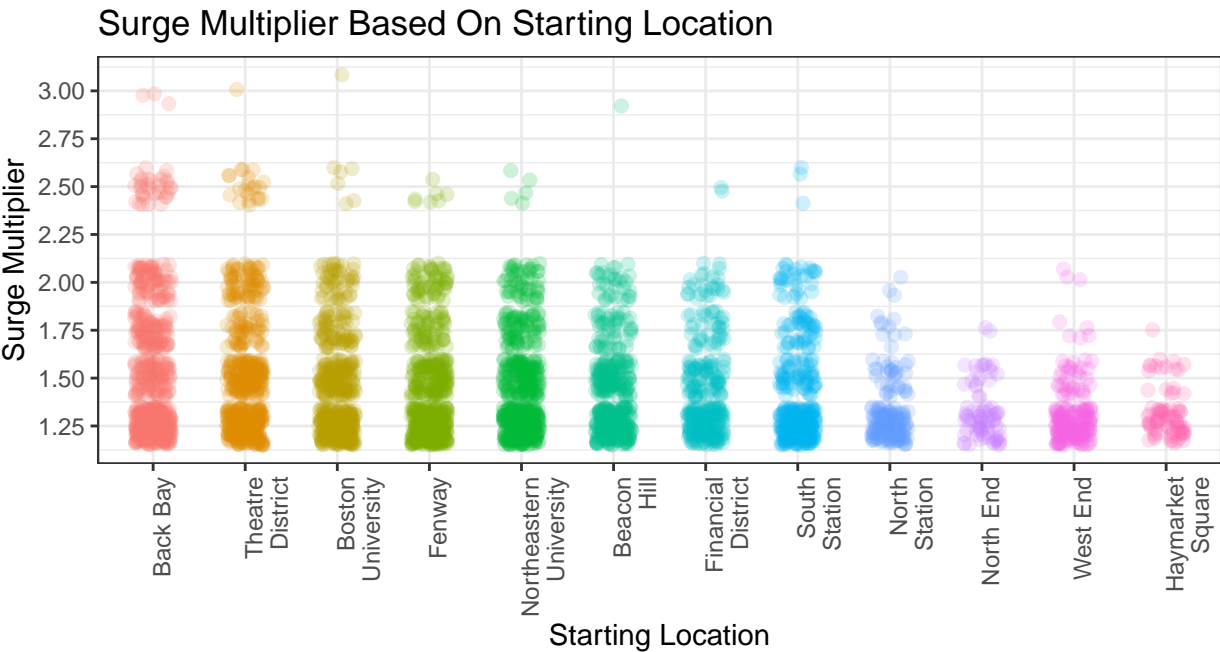
Thw weather can also affect surge multipliers. The number of rides that are affected by surge pricing are highest when weather conditions are cloudy. Furthermore, when the weather is either cloudy or rainy, surge multipliers can be up to 3.00. It is interesting to note that clear weather conditions have nearly the same pattern of surge pricing as rainy conditions, indicating that there are other variables that impact surge multipliers.

Surge Pricing During Different Weather Conditions