

# 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 exceed 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 utilize a data set of Lyft rides to see what factors cause such price surges. We will, then, use the determined factors to forecast the price of any Lyft ride.

## Methods

The Lyft data set was retrieved from Kaggle in the form of zip files (<https://www.kaggle.com/sliu65/data-mining-project-boston>).

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](https://github.com/jessicapadilla/mta_finances/blob/master/code.R)).

## Results