Title: New York City Taxi Fare Prediction

Problem: The goal of this project is to predict the fare amount of a taxi ride in New York City given the pickup and dropoff locations, as well as some additional details about the ride. In a competition-driven market, it's important for a taxi company to be able to predict a fare amount that is both:

1. Competitive, or customers could divert to competitors
2. Profitable, because money matters.

These criteria have likely been incorporated in previous fare calculations, so the goal of this project, is to automate this process, whereby a model can predict or propose a fare amount that satisfies the criteria above based on existing data.

Client: While this project specifically applies to the NYC Taxi and Limousine Commission, fare prediction is common in the transport industry in general. Airlines, trains, car pooling services, buses, etc, are all interested in satisfying the conditions above. And with the precedence of data, and the speed of technology, users expect an accurate and immediate price estimation for their travels.

Data: The data will be acquired from the NYC Taxi and Limousine Commission, which has been providing NYC taxi trip information since 2009. I will be training my model from entries in 2017 and testing it against records in 2018. I have chosen the entire year of 2017 to account for any seasonal changes that may affect the fare. Some of the probable features I will be analyzing include the pickup and dropoff locations, times and dates, the trip distance, and the passenger counts.

Methodology: This is a supervised problem as we will be training the model using existing data, which it will then use to predict future fare amounts. Specifically, it is a linear regression problem, and we are looking at out how specific features in the dataset will affect the fare amount. In addition to using regression techniques, I will use random forests and xgboost as supplementary methods.

Deliverables:

1. Code (notebooks) for various stages of the project lifecycle.
2. A final paper, highlighting any results and key findings from the project.
3. A presentation on the capstone project.