**Intsructions to deploy the Cab Fare Prediction model**

The industry today relies heavily on data analytics to make predictions. These predictions lead to successful business models that incentivise heavily from machine learning. Popular taxi services such as Uber and Lyft provide their users with a prediction of taxi fare before the customer is mapped to a driver. We can provide a similar solution using the open dataset provided. The intention is to process voluminous data in streams from public data repository and perform parallel feature engineering and deploy a prediction engine on top of it.

In this project we can implement a data analytics pipeline to process over million records of public dataset historical data from a public repository and predicted taxi fares. We contribute to parallel data preprocessing on AWS EMR using PySpark and Pandas and added machine learning models on top of it. Also implemented a Flask web application as an interface for users to query (serving layer) the trained models.

