FUNCTIONAL and TECHNICAL REQUIREMENTS DOCUMENT

Uber and Lyft Exploratory Data Analysis

Developed & Implementation by

Umar Khayam

TABLE OF CONTENTS

Table of Contents

1. General Information

1.1 Purpose

The purpose of this document is to provide information to the DevNation authorities to allow their endorsement of performing exploratory data analysis on Uber data analysis. This document explains the high-level technical and functional requirements, and provides information about this project.

1.2 Scope

The scope of this project is to analyze the data of Uber versus Lyft and we will also perform some operations. our main focus on analysing Uber data.

- Cleaning of data
- Exploration and analysis
- Data visualization

1.2.1 Questions we will cover in this EDA:

1. Which is the cheapest one, Uber or Lyft?
    - Total rides by Uber and Lyft
    - Total earning by Uber and Lyft
    - Total distance covered by Uber and Lyft
    - Uber vs Lyft by distance and price
2. Which is the cheapest and most expensive from Uber cab type and Lyft cab type.
3. Average Rides per day? Busiest day in a week?
4. Which is the best Cab for a long ride?
5. Which is the best cab for Shortest ride?
6. Most rides in weather?
    - Weather's effects on price
7. Surge multiplier with distance
    - Surge multiplier on days
    - Surge multiplier by times of day
8. Rush Hours
9. Which date has the most rides and why?
10. Which cab is popular at what time?
11. Most ride from source to destination?
12. A model that takes source, destination, distance, and weather that predicts the price of the ride.

1.3 Project References

- https://github.com/Unnati0104/Uber-Data-Analysis
- https://www.kaggle.com/theoddwaffle/uber-data-analysis

1.4 Acronyms and/or Definitions

- **Python** is programming language
- **Skit-learn** is open source library used for provide machine learning algorithm
- **Jupyter Notebook** is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.

- **NumPy** a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices
- **Pandas** is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.
- **Plotly** is python library used for plotting graph

2.0 Current summary

2.1 Summary

Uber and Lyft is a platform where those who drive and deliver can connect with riders, eaters, and restaurants. In cities where Uber and Lyft are available we will analyze the different time series, and average hours of working and growth of Uber and Lyft and will calculate the price of distance travel and also will analyse Lyft companies growth with Uber and check which one is best.

3.0 Functional requirement and user impact

3.1 Summary of functions

- EDA of Uber services

3.1.1 Functional requirements

- Perform EDA
- Data cleaning
- Data manipulations
- Machine learning algorithm

3.2 Summary of Impacts

3.2.1 Organizational Impacts

In this analysis Uber and Lyft companies will come to know where their services are or less and other companies have a grip and difference between their prices and other companies' prices.

3.2.2 Operational Impacts

By the EDA Uber companies can make necessary steps to improve the quality of their services in the places where they provide less.

4.0 Data source

In this project I used the dataset available on Kaggle.
https://www.kaggle.com/brllrb/uber-and-lyft-dataset-boston-ma