Final Project Proposal

Group Members

Ruwinda Rowel – s15654	Sithmi Wijesinghe – s15494
Vihanga Anjana – s15627	Darshi Yashodha – s15584

Description of the problem

Since the inception of ride sharing apps such as Uber, Lyft and even entities such as PickMe in Sri Lanka, the ride sharing industry has amassed to be a global industry worth $47.62 billion in 2024 and is expected to grow to $86.99 billion by 2029. This has been further driven by the further advancements in the telecommunication industry, growth in smartphones and many other factors. The pricing strategy behind many of these ride sharing apps is a method called dynamic pricing which was first introduced by Duke University and has been the standard practice of many companies such as Uber and Lyft since as early as 2015. In dynamic pricing, the model mimics metropolitan areas, areas where ride sharing is mostly frequent as many smaller hubs. The goal of dynamic pricing is to find the optimum price point for riders based on their situation in one of these hubs as well as their ride. (Ref : https://www.fuqua.duke.edu/duke-fuqua-insights/algorithms-behind-pricing-your-ride)

The ride-sharing company in this problem is looking to enhance its pricing strategy by implementing a dynamic approach that adjusts fares based on real-time market conditions. Currently, the company determines fares solely based on ride duration. To achieve this, they plan to utilize data-driven techniques to analyze past data and develop a predictive model that can adaptively set prices according to various factors as they change over time.

Thus, the key objective of the project is to

1. Identify key drivers to be included in this dynamic pricing model and distinguish any hidden patterns, relationships between them as well as their influence on the price.

2. Construct a dynamic pricing model and leverages machine learning to predict and set prices for the ride.

Description of the Dataset

The "Dynamic Pricing Dataset" comprises 1,000 data points with 10 variables, of which 6 are numerical and the remaining are categorical. This dataset is based on a hub and spoke system where each location is a part of a wider metropolitan area the ride sharing app is frequently used. Ref: https://www.kaggle.com/datasets/arashnic/dynamic-pricing-dataset/data

Variable Name	Description
Number of riders	Historical number of riders who have used the app in a specific area
Number of drivers	Historical number of drivers in that specific area
Location category	The category that specific location falls under
Customer loyalty status	Loyalty rating given to customers in that area
Number of past riders	The historical count of rides a customer has taken
Average ratings	Average ratings given to driver in that specific location by riders from that location
Time of booking	Time period of the day when most of the rides are booked
Vehicle type	Main type of vehicle requested in that area
Expected ride duration	Expected duration of the ride from pick up to drop off
Historical cost of ride	Cost of the ride borne by the rider, including any penalties incurred