|  |
| --- |
| Text  Description automatically generated  Electrical & Computer Engineering & Computer Science (ECECS) |

|  |  |
| --- | --- |
|  |  |



Contents

|  |  |
| --- | --- |
| Project Name | 2 |
| Summary | 2 |
| Abstract | 3 |
| Highlights | 3 |
| Methodology | 4 |
| CRISP DM | 14 |
| Conclusion | 15 |
| References | 15 |

|  |
| --- |
|  |

|  |  |  |
| --- | --- | --- |
| Summary Like public transportation, Uber and Lyft have variable ride costs. The availability and demand for transportation significantly impact them at any time. The project's main idea is to know the effect of each field on price with every other field of the dataset by applying different machine learning models to complete the analysis. The data is collected from Kaggle. Then data is processed and analyzed by using various libraries in python like Pandas, Skit learn. A question is framed based on the collected data and is processed through various operations to analyze the situation and obtain an answer. After the data is read, cleaned, analyzed/visualized, and represented, this analyzed data helps get the desired output, which is the project's objective. Based on the output conclusion for the project is determined | | |
|  | | |
| **Team Members:**  **Name 1 Harshavardhan Reddy**  **Name 2 Sai Mohan Majoju**  **Name 3 Sharath P**  **Name 4 Raghavendra Reddy** |  | Submitted on: 12-12-2022 |

|  |
| --- |
| **Price Prediction by analyzing Uber and lyft data** |

## 

## Abstract

Using platforms that are in competition, such as ride-sharing services Uber and Lyft; as a result, a superior business wants to provide customers with a better bargain than the opposition. By examining the price trends of Uber and Lyft, we can better understand the pricing for our company strategy. It's important to consider if we will accomplish this first or whether competitors have already used data analysis to personalize prices. And "Can we beat competitors' prices?"

Highlights of the project

The steps involved in the project are:

1. Data Collection.

2. Analyzing

3. Visualization.

4. Evaluation.

5. Deployment.

CRISP DM Diagram

Description automatically generated

Data Collection:

To comprehend how other fields affect the price column, we need data. An established method called web scraping can be used to collect the necessary data. First, the project's data needs to be chosen from a variety of sources, including data sets, Kaggle, and other platforms. In our case, we made a choice from the kaggle. After choosing the data, we used data collection methods to compile it into a CSV file. Web scraping is the method used to collect data from many platforms for data analysis.

Analyzing:

Data analysis tools like Pandas, NumPy, and matplotlib, among others, are used to examine the data acquired from various sources in various ways. Analyzing and cleaning involves reading the collected data. The undesirable data is cleaned up and made exact for data visualization throughout this phase. The data study on the collected data is carried out based on this analysis, and the pertinent output or outcomes are identified.

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated with medium confidence

Feature Engineering:

The act of choosing, modifying, and converting unprocessed data into features that can be applied in supervised learning is known as feature engineering.

Correlation:

The concept of correlation describes the connections between one or more variables. These factors could be characteristics of the raw data used to forecast our target variable.

Chart, treemap chart

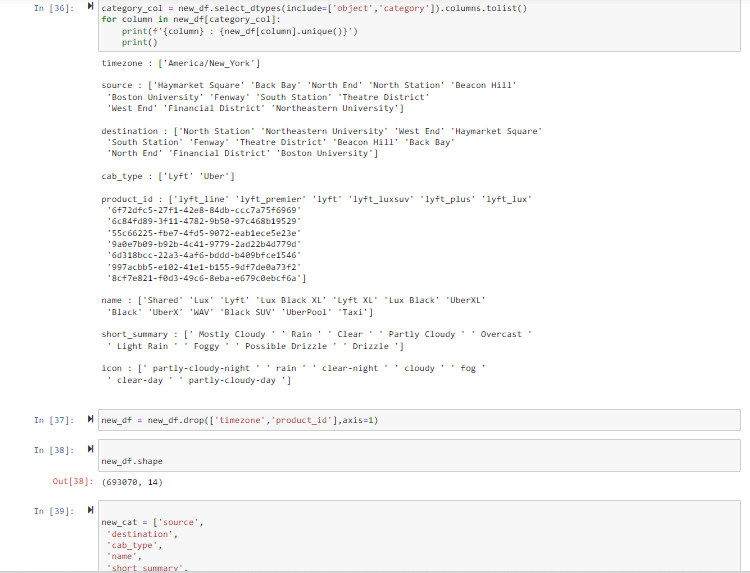
Description automatically generated

Graphical user interface, application, Teams

Description automatically generated

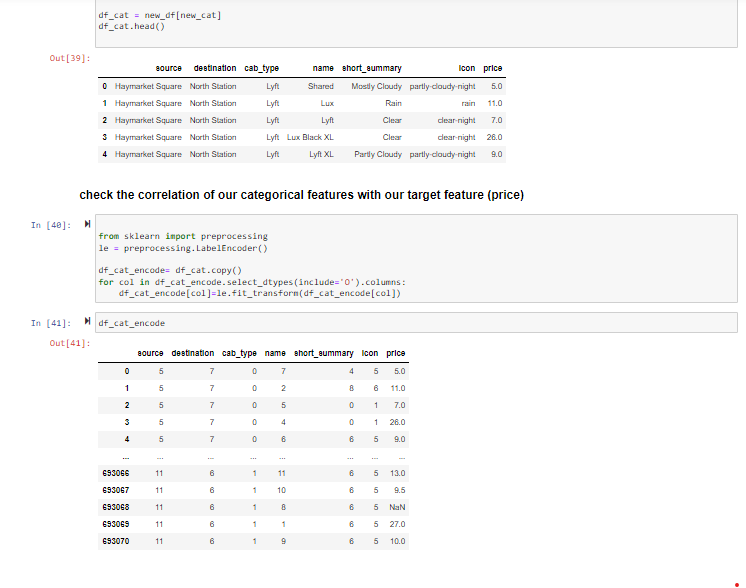
Removing Unwanted columns:

Based on the correlation the columns are removed and modified



categorical values:

You might have certain features that are discrete values without an ordered relationship. Examples include words, postal codes, and dog breeds. These characteristics are categorized, and each value is referred to as a category. Although categorical values can be expressed as strings or even numbers, they cannot be added or subtracted from one another.



Visualization:

The analyzed data based on the need may be required to be visualized and represented with visualization. We normally use libraries like Matplotlib and Bokeh for data visualization. Visualization is one kind of technique that converts raw data into an easily understandable format. This may be the representation of the data in the form of bar graphs, pie charts, scatter plots, etc. This process helps in understanding the data and representing the data in a simpler format that can be used to draw conclusions quite easily.

Bar chart

Description automatically generated

Table

Description automatically generated

Evaluation:

Based on the results of the analysis and the visualization techniques, the results that are obtained will be evaluated to arrive at a model for prediction.

Graphical user interface, text, application, email

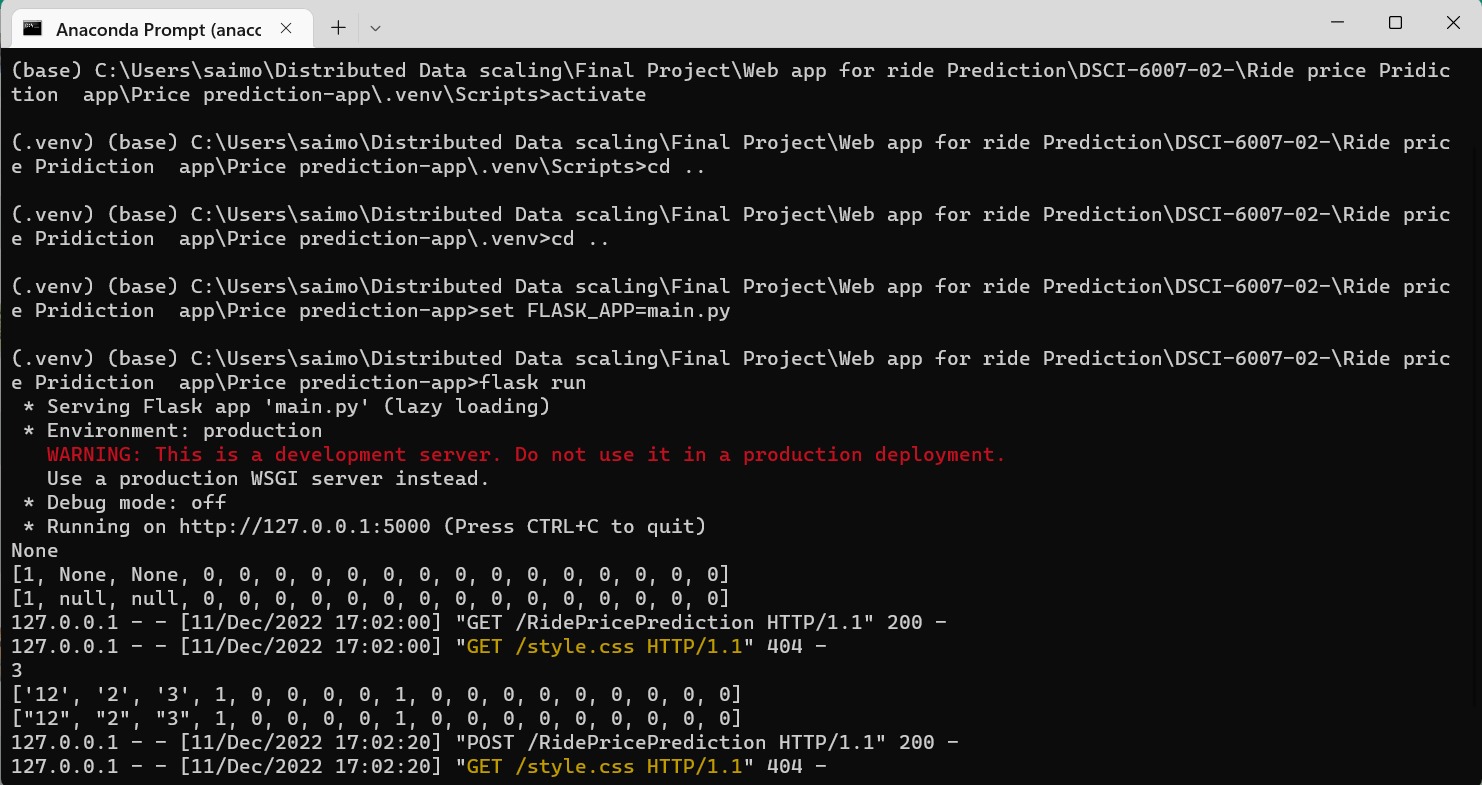
Description automatically generated

Split Data:

Create training and testing sets by dividing the data set into two halves. This entails taking a random sample without replacement of roughly 75% of the rows and adding them to your training set. You add the final 25 percent to your test set. Keep in mind that the colors in "Features" and "Target" denote the destination of their data.

## Result Section:

## ****Deployment on Flask****



Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

.

## CRISP DM

Business understanding: What is the primary requirement in this project providing a fair price for the rides to improve business

Data understanding: What Type of data we are dealing with and how to understand the data plays very important role. In this project we are dealing with structured data which contains previous ride information and price.

Data preparation: Finding the relation between the features and cleaning the data is part of this step. We use various techniques to prepare the data to train a particular model

Modeling: After Analysis the model is selected for training. Training and test data splits are made to train and test the model.

Evaluation: The trained model is evaluated based depending upon the accuracy of the model deployed.

Deployment: The trained model is deployed in the server and ready for consumption

Conclusion:

## It can be tough to determine whether ride is worth the given price at that moment. Distance, Size of car , Weather conditions and other characteristics can all have an impact on a ride’s true value. Thus the  use of  machine learning algorithms will help  in developing models for predicting used car prices based on given features of car.

Reference:

<https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city>

<https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma>

Git link:

<https://github.com/majojusaimohan/DSCI-6007-02->