DATA 606 Capstone in Data Science (04.2488)

Group D
Team Members:
1. Tejaswini Penneni
2. Shiva Shankar Kanike
3. Soniya Koduru

Project Title:
Data-driven Fare and Duration prediction with Peak hour classification for NYC Taxi Trips.

Description:
Our project focuses on using technology to develop a system that can forecast taxi fares and trip durations while categorizing trips into peak and off hours based on data, from New York City taxi rides. We will use regression and classification models to deliver predictions and valuable information that will be integrated into a user web platform utilizing Streamlit for an engaging and immediate user experience.

Type of Analysis:
- Prediction
- Classification

Dataset: (Size : > 3 GB)
Link:
https://catalog.data.gov/dataset/2023-yellow-taxi-trip-data
https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page

Features:
VendorID, pickup_datetime, dropoff_datetime, passenger_count, trip_distance, RatecodeID, store_and_fwd_flag, PULocationID, DOLocationID, payment_type, fare_amount extra, mta_tax, tip_amount, tolls_amount, improvement_surcharge, total_amount, congestion_surcharge, Airport_fee

Performance evaluation metrics:
- Prediction (RMSE, R2 scores)
- Classification (Accuracy, F1 Scores)

We will be integrating Streamlit framework to input the parameters (Pick and Drop location, Time) and the below output will be displayed.

- Estimated Time duration
- Price