Chicago Taxi Trips Data Analysis

Gulay Samatli

Introduction

DATA Acquisition
January-October, 2021
Soda API
CHICAGO DATA PORTAL
↓
Data Storage
SQLAlchemy
↓
Data Process
Python
↓
Web App
Streamlit

Question: Demand and fare by time and location; Chicago Taxi Cab Revenue Optimization

- Districts
- Month
- Days of the week
- Competitive market

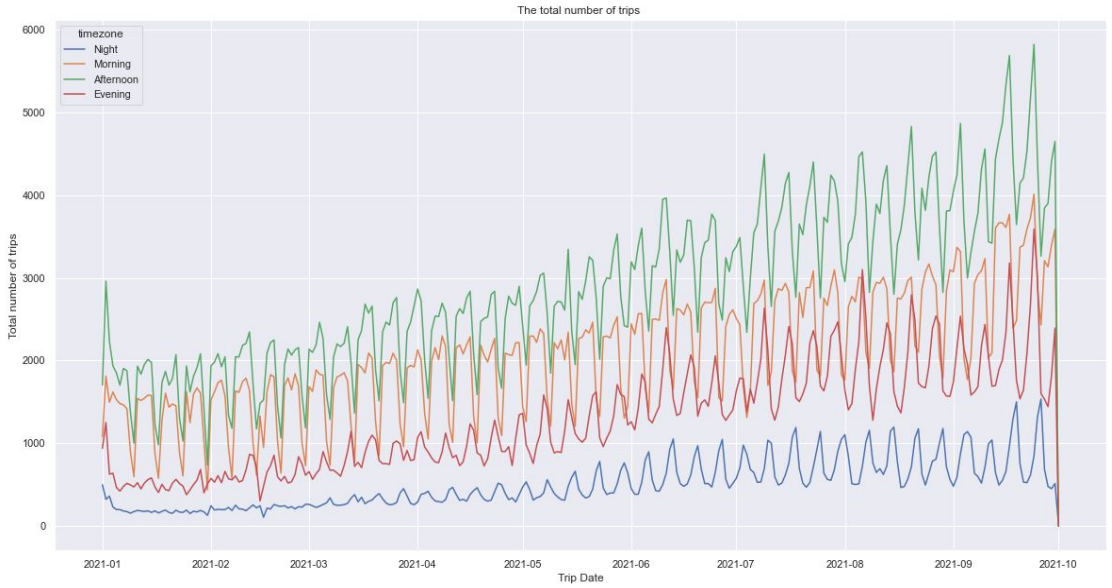
Motivation:

- to help taxi cab companies to schedule their staff

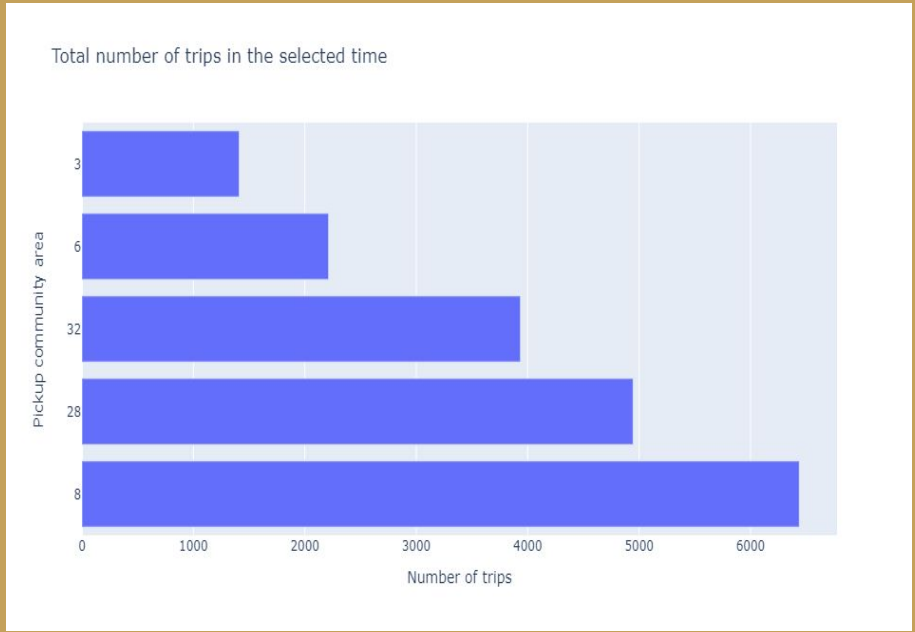
Data, EDA

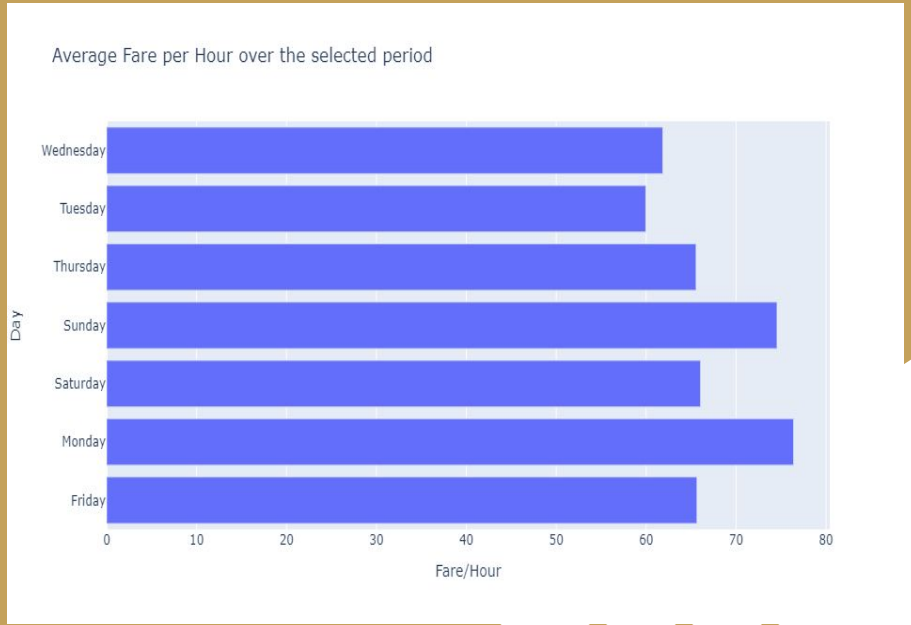
- Features:

trip_id
trip_start_timestamp
trip_seconds
trip_miles
dropoff_community_area
fare
trip_total
payment_type
company
dropoff_centroid_latitude
dropoff_centroid_longitude
pickup_community_area
pickup_centroid_latitude
pickup_centroid_longitude

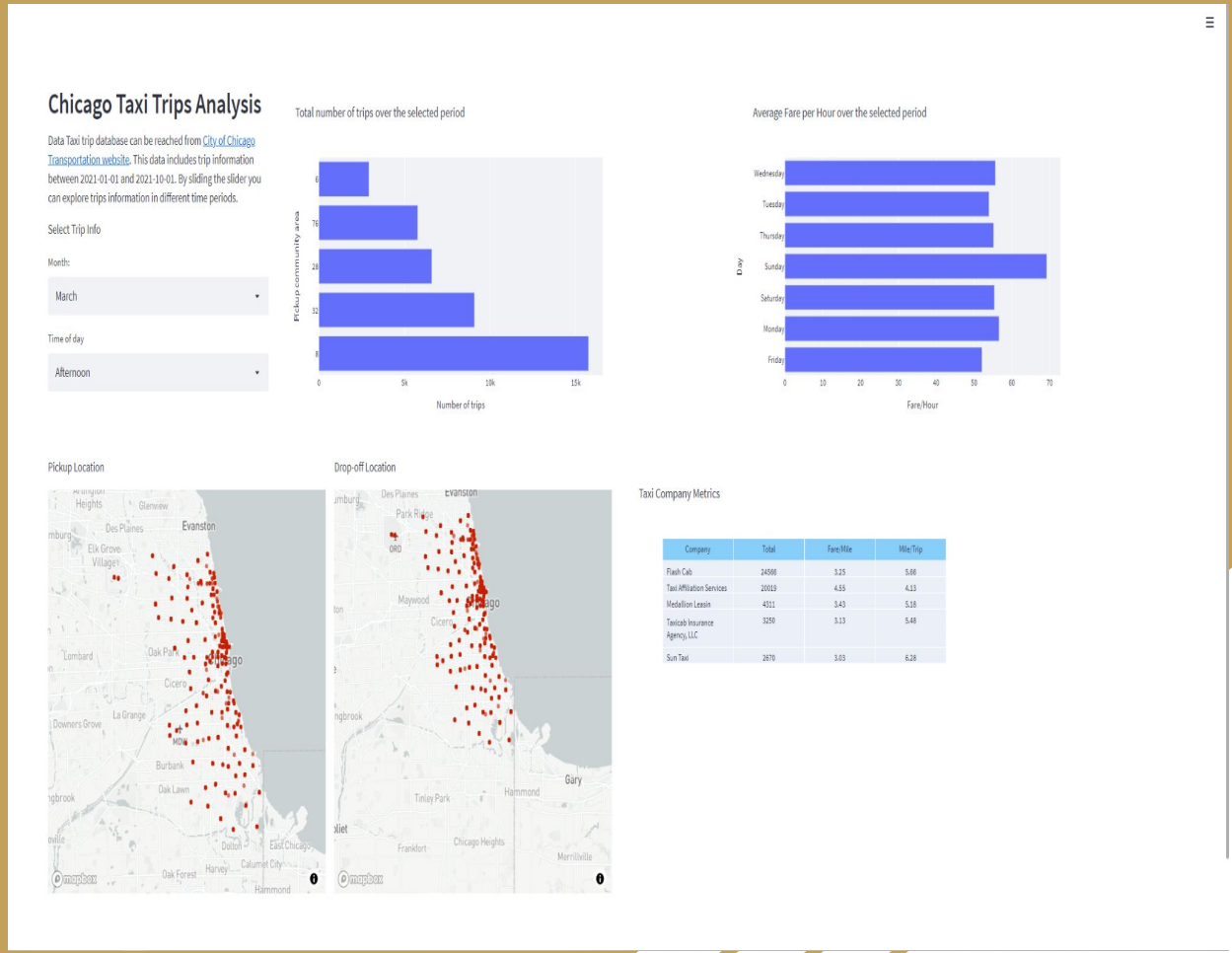


EDA


Total number of trips in the selected time
Pickup community area
3
6
32
28
8
0
1000
2000
3000
4000
5000
6000
Number of trips


Average Fare per Hour over the selected period
Day
Wednesday
Tuesday
Thursday
Sunday
Saturday
Monday
Friday
0
10
20
30
40
50
60
70
80
Fare/Hour

- For the selected month and time of day
  - Districts with top 5 demand
  - Popular pickup/Dropoff locations
  - Average fare per hour with respect to day of the week
  - Top 5 best performing companies



Future Works

- Build a predictive model
  - taxi trip duration, fare, demand
- Provide more filters on app
- Provide more graphs/charts on app

Thank you