ML Take Home Assessment

Dataset

The following dataset consists of 3 types of taxi data. Yellow Taxi, FHVs (For Hire Vehicles) and Green Taxi: Street Hail Livery (SHL). **For this task we will only consider yellow taxi data between Jan - Mar 2022 & Jan - Mar 2023.**
The detailed information about the dataset can be found here

Data can be downloaded from https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page. (2022, 2023 data). The data used in the attached datasets were collected and provided by the NYC Taxi and Limousine Commission (TLC).

Links to the range of data to consider for the assignment

- **Train data** - JAN - MAR 2022
- **Test data** - JAN - MAR 2023

2022 - JAN	https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2022-01.parquet
2022 - FEB	https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2022-02.parquet
2022 - MAR	https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2022-03.parquet
2023 - JAN	https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2023-01.parquet
2023 - FEB	https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2023-02.parquet
2023 - MAR	https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2023-03.parquet

Problem statement

Given location coordinates(latitude & longitude) and time as input, the goal is to build a model which can predict the number of pickups by an yellow taxi in the query region and surrounding regions

(Hint : Time-series Forecasting and Regression)

Evaluation Criteria

- Code Cleanliness, Scalability & Explainability
- Problem - Solving Approach
- Exploratory Data Analysis & Feature Engineering Techniques
- Choice of algorithm & Performance of the model

Submission Format

- A small write-up with details of
    - End-End approach followed to solve the problem
    - List of models analyzed
    - Model finalized and why that specific model has been chosen
    - What are the performance metrics employed to evaluate the model with proper justifications (MAPE & RMSE Values are a must)
- Create a Github Repository with all the related files, submission documents, README and provide the link
- Bonus Points if you can come up with an API back-end service to serve requests for prediction of number of pickups for any request made (The API should be deployable on a new machine and should have all the necessary files to run and test)

Disclaimer

In case of any ambiguities please free to make reasonable assumptions and make sure to document them in the write-up