The demand for taxis in New York City is significant, with a high volume of passengers, reaching up to 200 million rides per year. Therefore, even a slight improvement in operational scheduling efficiency can result in significant benefits for the taxi industry. Specifically, this improvement can be seen in three areas: ...

We plan to conduct our analysis from **two main perspectives**:

**On one hand**, focusing on vehicle data, we will utilize the taxi trip data collected by the TLC (Taxi and Limousine Commission) in New York City. This dataset includes data from companies like Green and covers information such as locations, fares, pick-up, and drop-off times. For our study, we will use the (2019-2023, Green) dataset as an example.​

**On the other hand**, we will obtain relevant data from other sources, such as weather data, and combine it with the vehicle data for our analysis.​

The data we have obtained from public datasets regarding Green taxi is presented in the table below. You can review the specific descriptions corresponding to different fields in the table，for example PU/DO Location ID, Fare\_amount

With these data, we'd like to achieve these five aspects of decisions:​

1... 2.Fare components themselves, and other factors 3...

Based on the regional analysis we've acquired, we model the best scheduling strategy for vehicles across various regions during distinct time intervals, aiming to enhance operational efficiency and minimize costs.

Finally, let's dive into the methods of modeling and data exploration. By the way, all these methods are not covered in class. We gonna showcase them in order. ​

**In this page**, we display those **methods of data exploration**.​

1... 2.Secondly, as for FE... 3.Next, coz previously we want to do some area analysis, so here we use Geographic Information System to execute the spatial analysis and further visualize it via location-based taxi data.

**While in this page**, we will discuss those **methods we'll use for modelling**. ​

1. Since the data we previously analyzed is time-correlated, we aim to use time series analysis to identify potential patterns over time. 2. in->use->understand... 3. We employ optimization algorithms and reinforcement learning (RL) to continually learn and make the most optimal decisions for regional scheduling over time.​