

Description of the application goals regarding trend analysis

The goals that this application intends to accomplish through trend analysis are catered to both the drivers and passengers of NYC taxicabs as well as data analysts.

Taxicab Drivers: this application will in determining the best ways to maximize profit by analyzing past trends in the taxi records from observations such as when would be the best day and time to do trips in a certain area of NYC to attain the most passengers.

Taxicab Passengers: This application will help passengers determine how to make their trip cost is minimized dependent on the day, time, and place through trend analysis of various factors from the taxi records.

Data Analysts: This application will display complex trends that data analysts can use for various purposes. For instance, the effects that ride sharing apps have contributed to the possible decline of taxi cabs in New York City.

Description of the real-world data forming the basis of the application and the complex trend queries

For the real-world data surrounding the basis of our application, we will be utilizing the real-world datasets provided by the New York City Taxi & Limousine Commission website that holds information on New York City yellow and green taxicab trip records from the years 2009 to 2022. Each month has more than 2 million records of taxi trip records. Therefore, there are more than 336 million records in total to choose from and operate on.

Each record will include 24 attributes of each taxicab's trip such as the location that the taxi meter was disengaged, location where the taxi meter was engaged, extra fees and surcharges, fare amount, improvement surcharge, MTA tax, rate code, store and forward flag, trip distance, date and time when the meter was engaged, date and time when the meter was disengaged, month, year, passenger count, payment method, tip amount, toll amount, total amount, vendor ID, start longitude and latitude, and end longitude and latitude.

Utilizing this data, we will be able to formulate complex queries to filter and analyze any interesting trends that can appeal to multiple types of users. The queries will all focus on the various data attributes listed above and observe their changes that occurred over time. The real-world data will all be raw data and not be manipulated by any algorithms in any way to maximize the efficiency of the trend analysis.