**Problem Statement**

In this case study you’ll be learning Exploratory Data Analytics with the help of a dataset on yellow taxi rides in New York City. Taxis play a crucial role in New York City’s urban transport network. With the city’s dynamic environment, taxi companies need to continuously adapt and optimise their operations to meet changing demand patterns, ensure profitability, and enhance customer satisfaction.

As an analyst at an upcoming taxi operation in NYC, you are tasked to use the 2023 taxi trip data to uncover insights that could help optimise taxi operations. The goal is to analyse patterns in the data that can inform strategic decisions to improve service efficiency, maximise revenue, and enhance passenger experience.

Let's understand the business objective and the dataset structure. You can download the datasets and data description from [this link](https://drive.google.com/file/d/1IWK7v7FL3Zg9AIh6PvG7b2rqri7rFKil/view?usp=drive_link).

The yellow taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts.

The data is stored in Parquet format (*.parquet*). The dataset is from 2009 to 2024. However, for this assignment, **we will only be using the data from 2023**. The data for each month is present in a different parquet file. You will get twelve files for each of the months in 2023.

Along with the trip records, you also have geometric data on various parts of the city divided as taxi zones. These taxi zones are present as a shapefile (*.shp*). You will find more about dealing with this while attempting the assignment.

The data was collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers like vendors and taxi hailing apps.

Next, you can download the starter Jupyter notebook which contains all the tasks and related instructions. Perform all analyses in the starter notebook only.