# Problem Statement formation

NYC Yellow Taxi trip data analysis

# Criteria for success

Analyze the most popular hours and off-peak hours for Yellow Taxi trip.

# Data sources

NYC Yellow Taxi Trip Data - <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

# Scope of solution space

Download CSV files. Write Python script to retrieve data from the CSV files, clean data and load to database.

# Proposed architecture for the solution and rationale

ETL for smaller data sets.

# Choice of technology

Python – To load CSV files

MS SQL Server – To store data

Python - Data validation and cleanup