**Total Records: 36256549**

**Input data source:**

[TLC Trip Record Data - TLC (nyc.gov)](https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page)

**Analysis Cases:**

1. Find out which Hours that have highest no of taxi rides started
2. Find the areas for the following cases,
   * Location where the highest rides are starts from
   * Location where the highest rides are ended
3. Average fair of the taxi rides
4. Average Trip distance covered
5. Ratio of taxi vendors

**Can be implemented,**

1. Rather than batch file, process the live data using Spark Streaming
2. Visualize the result of the analysis