# Data set description

| | |
|---|---|
| **Data set name** | New York city Taxi trip data |
| **Data set size** | 258MB |
| **Date of release** | 2017/07/20 |
| **No of attributes** | 11 |
| **No of data records** | Training set -1458644 <br><br> Testing set - 625134 |
| **Data source provider** | https://www.kaggle.com/c/nyc-taxi-trip-duration/data |
| **Data privacy** | Publicly available (To download the data, it is needed to be joined to the competition and accept the rules.) |
| **Notes** | The data set is based on the 2016 NYC Yellow Cab trip record data made available in Big Query on Google Cloud Platform. The data was originally published by the NYC Taxi and Limousine Commission (TLC). |
| **Prepared by** | Team-ETA |
| **Point of contact** | B.K.S. Ishanka <br> E.G.D. Sandaruwan <br> B.K.D.V.Vimarshana |
| **Team members** | B.K.S. Ishanka (209334C) <br> E.G.D. Sandaruwan (209375C) <br> B.K.D.V.Vimarshana |

# Attribute description

CN- Categorical Nominal
    ID- Identification
    DI- Dichotomous
MC- Metric Continuous
MD- Metric Discrete

| Attribute name | Data type | Data sub type | Description | Examples | Additional notes |
|---|---|---|---|---|---|
| id | CN | ID | a unique identifier for each trip | id2875421 | - |
| vendor_id | CN | DI | a code indicating the provider associated with the trip record | 2 | - |
| pickup_datetime | MC | Date | date and time when the meter was engaged | 2016-03-14 17:24:55 | - |
| dropoff_datetime | MC | Date | date and time when the meter was disengaged | 2016-03-14 17:32:30 | - |
| passenger_count | MD | - | the number of passengers in the vehicle (driver entered value) | 1 | - |
| pickup_longitude | MC | - | the longitude where the meter was engaged | -73.9821548461914 | - |
| pickup_latitude | MC | - | the latitude where the meter was engaged | 40.767936706543 | - |
| dropoff_longitude | MC | - | the longitude where the meter was disengaged | -73.9646301269531 | - |
| dropoff_latitude | MC | - | the latitude where the meter was disengaged | 40.7656021118164 | - |
| store_and_fwd_flag | CN | DI | his flag indicates whether the trip record was held in vehicle memory before sending to the vendor because the vehicle did not have a connection to the server - Y=store and forward; N=not a store and forward trip N | N | - |
| trip_duration | MD | - | duration of the trip in seconds | 455 | - |