NYC Taxi Shiny App

By Kent Burgess

Data Retrieval/Cleaning

- nyc.gov
- Cleaning the data of negative or 0 amount totals and specific PU and DO locations that are coded as NA
- For building my app I just used a sample size of 1M of the rows for the month, for ease of checking, for the production I have 8.5M rows
- I had to manipulate the PU and DO into a POSIX class to able to calculate travel time, classify the day of the week, and time of day columns, and lastly joined a separate table of zone IDs to do lookups on zone names

www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

The yellow and green 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 used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP). The trip data was not created by the TLC, and TLC makes no representations as to the accuracy of these data.

The For-Hire Vehicle ("FHV") trip records include fields capturing the dispatching base license number and the pick-up date, time, and taxi zone location ID (shape file below). These records are generated from the FHV Trip Record submissions made by bases. Note: The TLC publishes base trip record data as submitted by the bases, and we cannot guarantee or confirm their accuracy or completeness. Therefore, this may not represent the total amount of trips dispatched by all TLC-licensed bases. The TLC performs routine reviews of the records and takes enforcement actions when necessary to ensure, to the extent possible, complete and accurate information.

Trip Sheet Data (CSV Format)

tpep_pickup_datetime	tpep_dropoff_datetime
2018-06-01 00:15:40	2018-06-01 00:16:46
2018-06-01 00:04:18	2018-06-01 00:09:18
2018-06-01 00:14:39	2018-06-01 00:29:46
2018-06-01 00:51:25	2018-06-01 00:51:29
2018-06-01 00:55:06	2018-06-01 00:55:10
2018-06-01 00:09:00	2018-06-01 00:24:01

traveltimeseconds	traveltimeminutes	PUBorough	PUZone	PUservicezone	DOBorough	DOZone	DOservicezone	timeofday	dayofweek
66	1.10	Queens	Long Island City/Hunters Point	Boro Zone	Queens	Long Island City/Hunters Point	Boro Zone	Night	Fri
300	5.00	Manhattan	Times Sq/Theatre District	Yellow Zone	Manhattan	Midtown Center	Yellow Zone	Night	Fri
907	15.12	Manhattan	Garment District	Yellow Zone	Manhattan	Yorkville West	Yellow Zone	Night	Fri