

# NYC OpenData

## Data Dictionary - Dataset Information

|   |   |
|---|---|
| <b>Dataset Name</b>   | Yellow Taxi Trip Data   |
| <b>Dataset URL</b>  | <a href="https://data.cityofnewyork.us/browse?Data-Collection_Data-Collection=TLC%20Trip%20Data&amp;sortBy=alpha">https://data.cityofnewyork.us/browse?Data-Collection_Data-Collection=TLC%20Trip%20Data&amp;sortBy=alpha</a>   |
| <b>Data Provided by</b><br><i>The name of the NYC agency providing this data to the public.</i>   | Taxi and Limousine Commission (TLC)   |
| <b>Each row is a...</b><br><i>The unit of analysis/level of aggregation of the dataset</i>  | Taxi trip record  |
| <b>Publishing Frequency</b><br><i>How often changed data is published to this dataset. For an automatically updated dataset, this is the frequency of that automation</i>   | Historical data   |
| <b>Data Change Frequency</b><br><i>How often the data underlying this dataset is changed</i>  | As needed   |
| <b>Frequency Details</b><br><i>Additional details about the publishing or data change frequency, if needed</i>  | As this is historical data, the dataset is only published once. The data will only be changed rarely if there are corrections needed.   |
| <b>Dataset Description</b><br><i>Overview of the information this dataset contains, including overall context and definitions of key terms. This field may include links to supporting datasets, agency websites, or external resources for additional context.</i>   | These records are generated from the trip record submissions made by yellow taxi Technology Service Providers (TSPs). Each row represents a single trip in a yellow taxi. The trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off taxi zone locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts.  |
| <b>Why is this data collected?</b><br><i>Purpose behind the collection of this data, including any legal or policy requirements for this data by NYC Executive Order, Local Law, or other policy directive.</i>   | In partnership with the New York City Office of Technology and Innovation (OTI), TLC has published millions of trip records from both yellow medallion taxis and green Street Hail Livery (SHLs). Publicizing trip record data through an open platform permits instant access to records which previously were available only through a formal process (FOIL request.). Internally, TLC uses similar data to guide and evaluate policy decisions.  |
| <b>How is this data collected?</b><br><i>The methods used to create and update this dataset, including what cleaning or processing was involved prior to dataset publication.<br/>If data collection includes interpreting physical information this field includes technical details.<br/>If data collection includes fielding applications, requests, or complaints, this field includes details about the forms, applications, and processes used.</i> | 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.  |
| <b>How can this data be used?</b><br><i>Examples of and/or links to projects or agency operations that have used this dataset.<br/>Where relevant, includes links to online projects, agency websites, visualizations, maps, or dashboards.<br/>What are some questions one might answer using this dataset?</i>  | TLC internally uses similar data to provide internal and external public dashboards, please see: <a href="https://www.nyc.gov/site/tlc/about/data-and-research.page">https://www.nyc.gov/site/tlc/about/data-and-research.page</a> . This data is heavily used as an example dataset for data engineering and data science tasks. For an excellent third-party set of visualizations, see <a href="https://toddwischneider.com/dashboards/nyc-taxi-ridehailing-uber-lyft-data/">https://toddwischneider.com/dashboards/nyc-taxi-ridehailing-uber-lyft-data/</a> . |
| <b>What are the unique characteristics or limitations of this dataset?</b><br><i>Unique characteristics of this dataset to be aware of, specifically, constraints or limitations to the use of the data.</i>  | As the trip data was provided by technology providers to TLC, there may be some noise. This may occur in the form of unexpected categories or numbers out of expected ranges in some columns.   |
| <b>Additional geospatial information</b><br><i>For any datasets with geospatial data, specify the coordinate reference system or projection used and other relevant details.</i>  | Please see our website under "Taxi Zone Maps and Lookup Tables": <a href="https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page">https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page</a> . There is a shapefile to associate LocationID zones to geographic coordinates.  |