Data Dictionary Table of Contents:

2. User Guide

8. Yellow Taxi Data Dictionary

9. Green Taxi Data Dictionary (SHL)

11. FHV Data Dictionary (For-Hire Vehicles)

TLC Trip Records User Guide

Last Updated September 23, 2019

The TLC & Data

The New York City Taxi and Limousine Commission (TLC), created in 1971, is the agency responsible for licensing and regulating New York City's medallion (yellow) taxis, street hail livery (green) taxis, for-hire vehicles (FHVs), commuter vans, and paratransit vehicles. The TLC collects trip record information for each taxi and for-hire vehicle trip completed by our licensed drivers and vehicles. We receive taxi trip data from the technology service providers (TSPs) that provide electronic metering in each cab, and FHV trip data from the app, community livery, black car, or luxury limousine company, or base, who dispatched the trip.

In each trip record dataset, one row represents a single trip made by a TLC-licensed vehicle.

Data timeline

- **2009**: TLC begins to receive taxi trip data from taxi technology providers (now called technology service providers, or TSPs)
- **2013:** Green Taxis are added to the fleet. They may only pick up above W 110 St/E 96th St in Manhattan and in the boroughs
- **2015:** Taxi data for yellow and green trips is released online through the Open Data portal
- **2015:** TLC begins receiving FHV trip data from all bases, including app bases
- **2016**: TLC begins publishing FHV trip data from all bases, including app bases
    - Initial fields are dispatching base, pickup date/time, pickup location
- **2017**: TLC begins to receive and publish drop-off date/time and location from all FHV bases
- **2019**: TLC creates new license class, High Volume For Hire Services (HVFHS), for companies doing 10,000 + trips per day through their bases[1]. The license class has additional reporting requirements that create parity with the information received from taxi trips. As of Feb 2019, the companies doing 10k+ trips per day are Uber, Lyft, Via, and Juno

Downloading the data

There are two ways to download the taxi trip data—from the TLC website in CSV format, or from the NYC Open Data Portal in multiple formats.

TLC WEBSITE

The TLC publishes trip records on our website at this address:

https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

The data is separated by year, month, and type (yellow/green/FHV), and is available as a comma-separated value (CSV) file. When you click one of the links corresponding to a dataset,

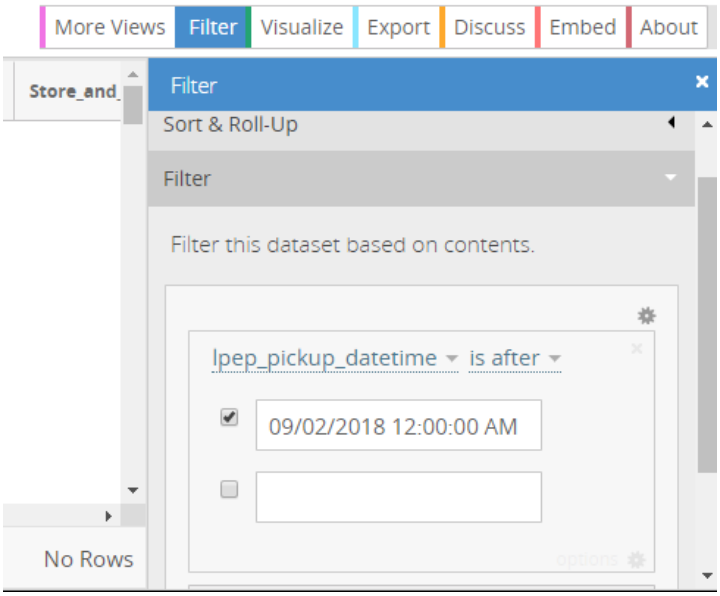
[1] A single company may have multiple bases that they dispatch trips through. See a complete list of app company bases on page 5.

the download should start automatically. Keep in mind that a whole month of trip data may include millions of rows, which could take up a sizable amount of disk space (and a while to download depending on your internet connection). If you do not need a whole month of data, or if you intend to filter the trip data before downloading (e.g. to only get trips after a certain date, or between two neighborhoods, or where the fare was greater than $10), we recommend that you use the NYC Open Data option below.

The TLC currently updates trip records every six months, so you should expect files for January-June by the end of August and July-December by the end of February.

NYC OPEN DATA

NYC Open Data is a citywide platform where all agencies share data for free, with everyone, to increase transparency and foster civic innovation. The TLC publishes numerous datasets through the NYC Open Data portal, with about 60 currently available online. The portal allows users to customize the data they'd like to download before they download it, and to download it in the format that works best for them. The portal is here: https://opendata.cityofnewyork.us/ and you can search for the exact dataset you'd like (e.g. 2016 Yellow Taxi Trip Records) or filter by agency (Taxi & Limousine Commission).



On each dataset's landing page, you can opt to ***View Data***, which shows you the rows and columns and allows you to set filters according to the columns. For example, if you are only looking for trips that occurred after a certain date, or where the fare was above a certain amount, you can filter based upon those conditions (see left).

The Open Data portal also allows you to export each dataset in multiple formats (CSV, JSON, RDF, XML, SHP (if the dataset is geographic, etc.) and you may connect to the dataset via an API if that works better. There is documentation on connecting via API here: https://dev.socrata.com/foundry/data.cityofnewyork.us/pqfs-mqru.

Finally, each dataset on open data has linked metadata, or information about the data, available. Information like when the dataset was last updated, what it contains, and often a link to a data dictionary, is stored here. It is important to always read the metadata so you can understand exactly what the values in each column are supposed to represent, and whether the data is up-to-date.

Taxi Data

YELLOW

Trips made by New York City's iconic yellow taxis have been recorded and provided to the TLC since 2009. Yellow taxis are traditionally hailed by signaling to a driver who is on duty and seeking a passenger (street hail), but now they may also be hailed using an e-hail app like Curb or Arro. Yellow taxis are the only vehicles permitted to respond to a street hail from a passenger in all five boroughs.

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 records were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology service providers. The trip data was not created by the TLC, and TLC cannot guarantee their accuracy.

Data dictionary download is here, in PDF format.

GREEN

Green taxis, also known as boro taxis and street-hail liveries, were introduced in August of 2013 to improve taxi service and availability in the boroughs. Green taxis may respond to street hails, but only in the areas indicated in green on the map (i.e. above W 110 St/E 96th St in Manhattan and in the boroughs).

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. As with the yellow taxi data, these records were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology service providers. The trip data was not created by the TLC, and TLC cannot guarantee their accuracy.

Data dictionary is here, in PDF format.

FHV Data

FHV data includes trip data from high-volume for-hire vehicle bases (bases for companies dispatching 10,000+ trip per day, meaning Uber, Lyft, Via, and Juno), community livery bases, luxury limousine bases, and black car bases.

The TLC began receiving FHV trip data from bases in 2015, but the amount of information that has been provided has changed over time. In 2015, only the dispatching base number, pickup datetime, and the location of the pickup (see section on matching zone IDs below) were provided

to the TLC. In summer of 2017, the TLC mandated that the companies provide the drop-off date/time and the drop-off location. In 2017, the TLC also started to receive information on shared rides, like those offered in services like Lyft Line and Uber Pool. A trip is only considered shared if it was reserved specially with one of these services. See note below for more information on shared rides. After the high volume license type was created in Feb 2019, a high-volume license number was added. This is an overall identifier for app companies who may have multiple base licenses.

Data dictionary* is here, in PDF format.

Data dictionary* for High Volume Data from Feb 1 2019 on is here, in PDF format.

*for FHV, the dictionary varies by year. The dictionary linked above is the most recent/inclusive of all fields

Dispatching_base_num—MATCHING TO FHV BASES

In order to determine which base in the FHV trip records dispatched the trip, you need to join the base number (field= "dispatching_base_num") to the license number (field= "License Number"). However, in the case of High Volume For Hire Services bases, which dispatch 10,000 + trips per day, the name of the base is not necessarily the name of the company people usually recognize. Currently, four companies either have a High Volume For-Hire Service license, or have started the application process for that license: Juno, Lyft, Uber, and Via. Here is a key for each of the HVFHS companies' bases.

High Volume License Number	License Number	Base Name	App Company Affiliation
HV0002	B02914	VULCAN CARS LLC	Juno
HV0002	B02907	SABO ONE LLC	Juno
HV0002	B02908	SABO TWO LLC	Juno
HV0002	B03035	OMAHA LLC	Juno
HV0005	B02510	TRI-CITY,LLC	Lyft
HV0005	B02844	ENDOR CAR & DRIVER,LLC.	Lyft
HV0003	B02877	ZWOLF-NY, LLC	Uber
HV0003	B02866	ZWEI-NY,LLC	Uber
HV0003	B02882	ZWANZIG-NY,LLC	Uber
HV0003	B02869	ZEHN-NY,LLC.	Uber
HV0003	B02617	WEITER LLC	Uber
HV0003	B02876	VIERZEHN-NY, LLC	Uber
HV0003	B02865	VIER-NY,LLC	Uber
HV0003	B02512	UNTER LLC	Uber
HV0003	B02888	SIEBZEHN-NY,LLC	Uber
HV0003	B02864	SIEBEN-NY,LLC	Uber
HV0003	B02883	SECHZEHN-NY, LLC	Uber
HV0003	B02875	SECHS-NY, LLC	Uber
HV0003	B02682	SCHMECKEN LLC	Uber
HV0003	B02880	NEUNZEHN-NY, LLC	Uber
HV0003	B02870	NEUN-NY,LLC	Uber
HV0003	B02404	KUCHEN,LLC	Uber
HV0003	B02598	HINTER LLC	Uber

HV0003	B02765	GRUN LLC	Uber
HV0003	B02879	FUNFZEHN-NY, LLC	Uber
HV0003	B02867	FUNF-NY, LLC	Uber
HV0003	B02878	ELF-NY,LLC	Uber
HV0003	B02887	EINUNDZWANZIG-NY, LLC	Uber
HV0003	B02872	EINS-NY,LLC	Uber
HV0003	B02836	DRINNEN-NY LLC	Uber
HV0003	B02884	DREIZEHN-NY, LLC	Uber
HV0003	B02835	DREIST NY LLC	Uber
HV0003	B02764	DANACH-NY,LLC	Uber
HV0003	B02889	ACHTZEHN-NY, LLC	Uber
HV0003	B02871	ACHT-NY,LLC	Uber
HV0003	B02395	ABATAR LLC	Uber
HV0004	B03136	GREENPOINT TRANSIT LLC	Via
HV0004	B02800	FLATIRON TRANSIT LLC	Via

A NOTE ON SHARED TRIPS

In the 2018 FHV records, there is a field called SR_Flag, which is supposed to indicate if the trip was a part of a shared ride chain offered by a High Volume FHS company (e.g. Uber Pool, Lyft Line). For shared trips, the value is 1. For non-shared rides, this field is null.

NOTE: For most High Volume FHS companies, only shared rides that were requested AND matched to another shared-ride request over the course of the journey are flagged. However, Lyft (hvfhs_license_num='HV0005') also flags rides for which a shared ride was requested but another passenger was not successfully matched to share the trip—therefore, trips records with SR_Flag=1 from those two bases could indicate EITHER a trip in a shared trip chain OR a trip for which a shared ride was requested but never matched. Users should anticipate an overcount of successfully shared trips completed by Lyft.

Note also that Juno does not offer shared trips at this time.

PULocationid & DOLocationID—matching zone numbers to the map

Each of the trip records contains a field corresponding to the location of the pickup or drop-off of the trip (or in FHV records before 2017, just the pickup), populated by numbers ranging from 1-263. These numbers correspond to taxi zones, which may be downloaded as a table or map/shapefile and matched to the trip records using a join. The data is currently available on the Open Data Portal at https://data.cityofnewyork.us/Transportation/NYC-Taxi-Zones/d3c5-ddgc, or on the trip records page on the TLC website, https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page, under Taxi Zone Maps and Lookup Tables, see below:

Taxi Zone Maps and Lookup Tables

- **Taxi Zone Lookup Table** (CSV)
- **Taxi Zone Shapefile** (CSV)
- **Taxi Zone Map – Bronx** (JPG)
- **Taxi Zone Map – Brooklyn** (JPG)
- **Taxi Zone Map – Manhattan** (JPG)
- **Taxi Zone Map – Queens** (JPG)
- **Taxi Zone Map – Staten Island** (JPG)

You can download the shapefile from this link: https://s3.amazonaws.com/nyc-tlc/misc/taxi_zones.zip

or the lookup table in CSV format from this link: https://s3.amazonaws.com/nyc-tlc/misc/taxi+_zone_lookup.csv

Note that the Taxi Zones are roughly based on NYC Department of City Planning's Neighborhood Tabulation Areas (NTAs) and are meant to approximate neighborhoods, so you can see which neighborhood a passenger was picked up in, and which neighborhood they were dropped off in.

ANY OTHER QUESTIONS? WANT MORE TLC INFO?

Please email us @ research@tlc.nyc.gov with any questions or comments about the state of our data. You can also submit feedback or ask questions about a specific dataset through the Open Data Portal.

Also check out our blog @ https://medium.com/@NYCTLC to see our ongoing projects!

This data dictionary describes yellow taxi trip data. For dictionaries describing green taxi and FHV data, please visit http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

Field Name	Description
VendorID	A code indicating the TPEP provider that provided the record. **1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.**
tpep_pickup_datetime	The date and time when the meter was engaged.
tpep_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value.
Trip_distance	The elapsed trip distance in miles reported by the taximeter.
Pickup_longitude	Longitude where the meter was engaged.
Pickup_latitude	Latitude where the meter was engaged.
RateCodeID	The final rate code in effect at the end of the trip. **1= Standard rate** **2=JFK** **3=Newark** **4=Nassau or Westchester** **5=Negotiated fare** **6=Group ride**
Store_and_fwd_flag	This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka “store and forward,” because the vehicle did not have a connection to the server. **Y= store and forward trip** **N= not a store and forward trip**
Dropoff_longitude	Longitude where the meter was disengaged.
Dropoff_ latitude	Latitude where the meter was disengaged.
Payment_type	A numeric code signifying how the passenger paid for the trip. **1= Credit card** **2= Cash** **3= No charge** **4= Dispute** **5= Unknown** **6= Voided trip**
Fare_amount	The time-and-distance fare calculated by the meter.
Extra	Miscellaneous extras and surcharges. Currently, this only includes the $0.50 and $1 rush hour and overnight charges.
MTA_tax	$0.50 MTA tax that is automatically triggered based on the metered rate in use.
Improvement_surcharge	$0.30 improvement surcharge assessed trips at the flag drop. The improvement surcharge began being levied in 2015.
Tip_amount	Tip amount – This field is automatically populated for credit card tips. Cash tips are not included.
Tolls_amount	Total amount of all tolls paid in trip.
Total_amount	The total amount charged to passengers. Does not include cash tips.

This data dictionary describes SHL trip data. For dictionaries describing yellow taxi and FHV data, please visit http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

Field Name	Description
VendorID	A code indicating the LPEP provider that provided the record. **1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.**
lpep_pickup_datetime	The date and time when the meter was engaged.
lpep_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value.
Trip_distance	The elapsed trip distance in miles reported by the taximeter.
Pickup_longitude	Longitude where the meter was engaged.
Pickup_latitude	Latitude where the meter was engaged.
RateCodeID	The final rate code in effect at the end of the trip. **1= Standard rate** **2=JFK** **3=Newark** **4=Nassau or Westchester** **5=Negotiated fare** **6=Group ride**
Store_and_fwd_flag	This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka “store and forward,” because the vehicle did not have a connection to the server. **Y= store and forward trip** **N= not a store and forward trip**
Dropoff_longitude	Longitude where the meter was timed off.
Dropoff_ latitude	Latitude where the meter was timed off.
Payment_type	A numeric code signifying how the passenger paid for the trip. **1= Credit card** **2= Cash** **3= No charge** **4= Dispute** **5= Unknown** **6= Voided trip**
Fare_amount	The time-and-distance fare calculated by the meter.
Extra	Miscellaneous extras and surcharges. Currently, this only includes the $0.50 and $1 rush hour and overnight charges.
MTA_tax	$0.50 MTA tax that is automatically triggered based on the metered rate in use.
Improvement_surcharge	$0.30 improvement surcharge assessed on hailed trips at the flag drop. The improvement surcharge began being levied in 2015.
Tip_amount	Tip amount – This field is automatically populated for credit card tips. Cash tips are not included.

Tolls_amount	Total amount of all tolls paid in trip.
Total_amount	The total amount charged to passengers. Does not include cash tips.
Trip_type	A code indicating whether the trip was a street-hail or a dispatch that is automatically assigned based on the metered rate in use but can be altered by the driver. **1= Street-hail** **2= Dispatch**

This data dictionary describes FHV trip data. Each row represents a single trip in an FHV.

For a dictionary describing yellow and green taxi data, or a map of the TLC Taxi Zones, please visit http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

Field Name	Description
Dispatching_base_num	The TLC Base License Number of the base that dispatched the trip
Pickup_datetime	The date and time of the trip pick-up
DropOff_datetime	The date and time of the trip dropoff
PULocationID	TLC Taxi Zone in which the trip began
DOLocationID	TLC Taxi Zone in which the trip ended
SR_Flag	Indicates if the trip was a part of a shared ride chain offered by a High Volume FHV company (e.g. Uber Pool, Lyft Line). For shared trips, the value is **1**. For non-shared rides, this field is **null**. **NOTE:** For most High Volume FHV companies, only shared rides that were requested AND matched to another shared-ride request over the course of the journey are flagged. However, Lyft (base license numbers **B02510** + **B02844**) also flags rides for which a shared ride was requested but another passenger was not successfully matched to share the trip—therefore, trips records with SR_Flag=1 from those two bases could indicate EITHER a first trip in a shared trip chain OR a trip for which a shared ride was requested but never matched. Users should anticipate an overcount of successfully shared trips completed by Lyft.