Automatidata Project Lab

Activity: Course 2 Automatidata

Instructions

In this lab portion of the end-of-course project, you will open a Jupyter Notebook and follow instructions to enter code and written responses where prompted.

Data dictionary

This project uses a dataset called 2017_Yellow_Taxi_Trip_Data.csv. It data gathered by the New York City Taxi & Limousine Commission and published by the city of New York as part of their NYC Open Data program. In order to improve the learning experience and shorten runtimes, a sample was drawn from the 113 million rows in the 2017 Yellow Taxi Trip Data table.

The dataset contains:

22,699 rows – each row represents a different trip

18 columns

Column name	Description
ID	Trip identification number
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.
PULocationID	TLC Taxi Zone in which the taximeter was engaged
DOLocationID	TLC Taxi Zone in which the taximeter was disengaged
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 being sent 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**
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.

Remember, you can access and download the data for any Jupyter notebook activity from within the notebook itself by navigating to the Lab Files dropdown menu at the top of the page, clicking into the /home/jovyan/work folder, selecting the relevant data file, and clicking Download.

Refer to NYC Open Data for more information related to this dataset.