New York Taxi DataSet

In this exercise, we will be exploring a dataset containing the taxi trips made in New York City in 2013. We will analyze a subset of this dataset containing 0.5% of all trips (about 850,000 rides). Compressed, this subset data represents a little less than 100MB.

NYC Taxi and Limousine Commission (TLC) is the technology provider for the trip data. More about the project: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml (Links to an external site.)
The Data Dictionary for the yellow taxi trip data: https://www1.nyc.gov/assets/tlc/downloads/pdf/data_dictionary_trip_records_yellow.pdf (Links to an external site.)
An interactive web application based on this dataset: http://hubcab.org (Links to an external site.)

Assignment Details:

Please complete the following:

- Download and unzip the NYC taxi dataset from Cyrille Rossant on GitHub: https://github.com/ipython-books/minibook-2nd-data (Links to an external site.)
- Open the notebook file attached below. You will be adding your code (make sure you add headers and comments) to the existing code, and make sure your code is well organized.
- Please upload the data and display data columns, number of rows, variable types, and numeric statistics + categorical variable frequencies.
- Display a scatter plot of pick up locations. For which vendor is it easiest to find a cab?
- Display a histogram of trip distances. What is the most common trip distance?
- Display a histogram of the fare total amounts. What can you say about the data?
- How many unusually long trips (of greater than 100 miles) do you see?

please upload your notebook as either PDF or ipynb file format. Convert your notebook to pdf from your browser ***File > Export as PDF*** option.

NY TAXI EZ Source Code