Analyze the NYC Taxi Data

September 6, 2019

1 Analyze the NYC Taxi Data

Queries databasees from "taxi_data" http://localhost:50070/explorer.html#/user/local/temp/taxi_data (FDFS) with tables : trips, pefer, licenses

```
In [36]: import pandas as pd
         import numpy as np
         import os
         import matplotlib.pyplot as plt
         import seaborn as sns
         %matplotlib inline
         sns.set()
```

```
In [2]: from hdfs3 import HDFileSystem
        hdfs = HDFileSystem(host='localhost', port=9000);
        print(hdfs.ls("/user/local/temp/taxi_data"))
```

```
['/user/local/temp/taxi_data/data_for_1week', '/user/local/temp/taxi_data/data_for_2days', '/use
```

```
In [10]: with hdfs.open('/user/local/temp/taxi_data/data_for_1week/trip_data_week1.csv') as twk:
              trips_week = pd.read_csv(twk)
```

```
In [4]: with hdfs.open('/user/local/temp/taxi_data/data_for_1week/fare_data_week1.csv') as fwk:
            fare_week = pd.read_csv(fwk)
```

```
In [5]: with hdfs.open('/user/local/temp/taxi_data/data_for_2days/trips_sunday.csv') as tsd:
            trips_sunday = pd.read_csv(tsd)
```

```
In [6]: with hdfs.open('/user/local/temp/taxi_data/data_for_2days/trips_wednesday.csv') as twd:
            trips_wednesday = pd.read_csv(twd)
```

```
In [7]: with hdfs.open('/user/local/temp/taxi_data/data_for_2days/fares_sunday.csv') as fsd:
            fare_sunday = pd.read_csv(fsd)
```

```
In [8]: with hdfs.open('/user/local/temp/taxi_data/data_for_2days/fares_wednesday.csv') as fwd:
            fare_wednesday = pd.read_csv(fwd)
```

```
In [9]: with hdfs.open('/user/local/temp/taxi_data/vehicle_data/licenses.csv') as lcs:
            licenses = pd.read_csv(lcs)

In [105]: trips_week.head(5)

Out[105]:                          medallion                      hack_license  \
          0  AA70234AB7643A84903E4B0705352D8A  2A3BEA5321E55025D86D65269A67DDD7
          1  E28E2AD14EF5E6D4E3019702A243E982  D99683EBE31E9B9B26A5A04570E1F7F8
          2  1AF573B78F7BEAFF25E721041D80D2D2  F4AEEE8C03292EF4BE4C3E0BC4FF66EE
          3  942EDD26C4E3337133058C095AD23289  CA0BCAF81499737B35644B5F9AB19CCC
          4  6BA29E9A69B10F218C1509BEDD7410C2  4FE29988ED28B24418058814A371F326

            vendor_id  rate_code store_and_fwd_flag      pickup_datetime  \
          0       VTS          2                NaN  2013-08-05 23:07:00
          1       VTS          1                NaN  2013-08-05 10:37:00
          2       VTS          1                NaN  2013-08-06 01:05:00
          3       VTS          2                NaN  2013-08-05 23:09:00
          4       VTS          1                NaN  2013-08-07 08:35:00

                dropoff_datetime  passenger_count  trip_time_in_secs  trip_distance  \
          0  2013-08-05 23:07:00                3                  0           0.00
          1  2013-08-05 10:47:00                2                600           5.64
          2  2013-08-06 01:27:00                6               1320           6.54
          3  2013-08-05 23:10:00                1                 60           0.00
          4  2013-08-07 08:43:00                6                480           1.57

             pickup_longitude  pickup_latitude  dropoff_longitude  dropoff_latitude
          0          0.000000         0.000000           0.000000          0.000000
          1        -73.789703        40.641075         -73.729752         40.677792
          2          0.000000         0.000000           0.000000          0.000000
          3        -73.776718        40.645199         -73.776718         40.645199
          4        -73.955681        40.779476         -73.969414         40.762016

In [107]: fare_week.head(5)

Out[107]:                          medallion                      hack_license  \
          0  AA70234AB7643A84903E4B0705352D8A  2A3BEA5321E55025D86D65269A67DDD7
          1  E28E2AD14EF5E6D4E3019702A243E982  D99683EBE31E9B9B26A5A04570E1F7F8
          2  1AF573B78F7BEAFF25E721041D80D2D2  F4AEEE8C03292EF4BE4C3E0BC4FF66EE
          3  942EDD26C4E3337133058C095AD23289  CA0BCAF81499737B35644B5F9AB19CCC
          4  6BA29E9A69B10F218C1509BEDD7410C2  4FE29988ED28B24418058814A371F326

            vendor_id      pickup_datetime payment_type  fare_amount  surcharge  \
          0       VTS  2013-08-05 23:07:00          CRD         52.0        0.0
          1       VTS  2013-08-05 10:37:00          CSH         17.5        0.0
          2       VTS  2013-08-06 01:05:00          CRD         21.5        0.5
          3       VTS  2013-08-05 23:09:00          CSH         52.0        0.0
          4       VTS  2013-08-07 08:35:00          CSH          8.0        0.0
```

```
   mta_tax  tip_amount  tolls_amount  total_amount
0      0.5        10.4           0.0          62.9
1      0.5         0.0           0.0          18.0
2      0.5         4.4           0.0          26.9
3      0.5         0.0           0.0          52.5
4      0.5         0.0           0.0           8.5
```

In [109]:
```
trips_sunday.head(5)
```

Out[109]:
```
                          medallion                       hack_license  \
0  4B37DE7600AEF9C61F784B05FDEEE0E9  1D7E4CD01ED1C7A6E662A2A9A4B7153F
1  EEC9C6596BD11B4F213367BEF164ED40  902B96BCB437D747BA50888778132BE4
2  B009310787A68502FFD50A2F9CB5CE26  A44307E7C864D631E9A26B49C25DD23B
3  A408F138216DE3E432BBF2FD88665A88  82EA6A085709BE93AA9DA363A85A04FF
4  C46A099283B423340CD9CC2837E73761  4339B58CF42D6B3011479B8D9731CA7F

  vendor_id  rate_code store_and_fwd_flag      pickup_datetime  \
0       CMT          1                  N  2013-08-04 00:03:59
1       CMT          1                  N  2013-08-04 00:16:29
2       CMT          1                  N  2013-08-04 00:11:00
3       CMT          1                  N  2013-08-04 00:04:26
4       CMT          1                  N  2013-08-04 00:09:55

      dropoff_datetime  passenger_count  trip_time_in_secs  trip_distance  \
0  2013-08-04 00:12:46                2                527            1.3
1  2013-08-04 00:21:41                1                311            1.2
2  2013-08-04 00:21:57                1                656            4.3
3  2013-08-04 00:18:27                1                840            3.2
4  2013-08-04 00:18:53                1                538            1.0

   pickup_longitude  pickup_latitude  dropoff_longitude  dropoff_latitude
0         -74.008743        40.738098         -73.992302         40.743961
1         -73.955505        40.776752         -73.942024         40.786846
2         -73.993073        40.698143         -73.979675         40.657543
3         -73.954834        40.765484         -73.999046         40.760777
4         -73.988869        40.723156         -74.001343         40.731052
```

In [111]:
```
trips_wednesday.head(5)
```

Out[111]:
```
                          medallion                       hack_license  \
0  6BA29E9A69B10F218C1509BEDD7410C2  4FE29988ED28B24418058814A371F326
1  CBBE30BB243B09EADA18DAFB28035441  22457AF4EC023E0E8BBF351791A5C811
2  39EBD6484D03EF51127B8B7D6A14C172  5509A943094052ED58D903C69F3DCD9C
3  FCFFBC6FFBA23178D9C569CAE435020F  A39C4955A7F7B10AC4F27130DD752B19
4  925FAAB0AEC05317DCD15C9EA48B26B3  52DF74CF19155D44C87E64FF9E715F43

  vendor_id  rate_code store_and_fwd_flag      pickup_datetime  \
0       VTS          1                NaN  2013-08-07 08:35:00
1       VTS          1                NaN  2013-08-07 08:41:00
```

```
2       VTS          1                NaN  2013-08-07 08:48:00
3       VTS          1                NaN  2013-08-07 08:40:00
4       VTS          2                NaN  2013-08-07 08:08:00

      dropoff_datetime  passenger_count  trip_time_in_secs  trip_distance  \
0  2013-08-07 08:43:00                6                480           1.57
1  2013-08-07 08:52:00                1                660           1.93
2  2013-08-07 08:50:00                6                120           0.53
3  2013-08-07 08:52:00                1                720           2.77
4  2013-08-07 08:52:00                1               2640          19.01

   pickup_longitude  pickup_latitude  dropoff_longitude  dropoff_latitude
0        -73.955681        40.779476         -73.969414         40.762016
1        -73.998047        40.725616         -74.003555         40.742741
2        -73.984123        40.743126         -73.988693         40.737148
3        -73.990112        40.740761         -74.014198         40.715439
4        -73.786743        40.644646         -73.959900         40.809681
```

In [113]:
```
fare_sunday.head(5)
```

Out[113]:
```
                          medallion                     hack_license  \
0  4B37DE7600AEF9C61F784B05FDEEE0E9  1D7E4CD01ED1C7A6E662A2A9A4B7153F
1  EEC9C6596BD11B4F213367BEF164ED40  902B96BCB437D747BA50888778132BE4
2  B009310787A68502FFD50A2F9CB5CE26  A44307E7C864D631E9A26B49C25DD23B
3  A408F138216DE3E432BBF2FD88665A88  82EA6A085709BE93AA9DA363A85A04FF
4  C46A099283B423340CD9CC2837E73761  4339B58CF42D6B3011479B8D9731CA7F

  vendor_id      pickup_datetime payment_type  fare_amount  surcharge  \
0       CMT  2013-08-04 00:03:59          CRD          7.5        0.5
1       CMT  2013-08-04 00:16:29          CRD          6.5        0.5
2       CMT  2013-08-04 00:11:00          CRD         14.5        0.5
3       CMT  2013-08-04 00:04:26          CRD         12.5        0.5
4       CMT  2013-08-04 00:09:55          CRD          7.5        0.5

   mta_tax  tip_amount  tolls_amount  total_amount
0      0.5        2.10           0.0         10.60
1      0.5        1.87           0.0          9.37
2      0.5        3.00           0.0         18.50
3      0.5        3.37           0.0         16.87
4      0.5        1.50           0.0         10.00
```

In [117]:
```
fare_wednesday.head(5)
```

Out[117]:
```
                          medallion                     hack_license  \
0  6BA29E9A69B10F218C1509BEDD7410C2  4FE29988ED28B24418058814A371F326
1  CBBE30BB243B09EADA18DAFB28035441  22457AF4EC023E0E8BBF351791A5C811
2  39EBD6484D03EF51127B8B7D6A14C172  5509A943094052ED58D903C69F3DCD9C
3  FCFFBC6FFBA23178D9C569CAE435020F  A39C4955A7F7B10AC4F27130DD752B19
4  925FAAB0AEC05317DCD15C9EA48B26B3  52DF74CF19155D44C87E64FF9E715F43
```

```
  vendor_id      pickup_datetime payment_type  fare_amount  surcharge  \
0       VTS  2013-08-07 08:35:00          CSH          8.0        0.0
1       VTS  2013-08-07 08:41:00          CRD          9.5        0.0
2       VTS  2013-08-07 08:48:00          CSH          4.0        0.0
3       VTS  2013-08-07 08:40:00          CRD         11.5        0.0
4       VTS  2013-08-07 08:08:00          CRD         52.0        0.0

   mta_tax  tip_amount  tolls_amount  total_amount
0      0.5        0.00          0.00           8.5
1      0.5        1.90          0.00          11.9
2      0.5        0.00          0.00           4.5
3      0.5        1.50          0.00          13.5
4      0.5       12.17          5.33          70.0
```

```
In [11]: licenses.head(5)
```

Out[11]:

```
                          medallion                    name      types  \
0  D7479C954DF136B1545CBF8491361A8D          TORRES,HAROLD  MEDALLION
1  F9FA751DAECA69F0AF703177916A29F8         BELIARD,LECLERC  MEDALLION
2  D2842D4D37F17851610C2991B8B3B164  ST.SURIN, HENRY CLAUDE  MEDALLION
3  9383E6D0EB4524BBC576072BFBE2E053          PIERRE, GERALD  MEDALLION
4  160248D54674DD1633EC5CE7896215E8         KANG,SUNG,CHOON  MEDALLION

  current_status DMV_license_plate vehicle_VIN_number vehicle_type  \
0            CUR             3T50A  2FAFP70W26X156009          CNG
1            CUR             3D30A  3VWPL7AJ3BM637581          DSE
2            CUR             4A52A  3VWPL7AJ4BM637329          DSE
3            CUR             8C72A  3VWPL7AJ1BM691252          DSE
4            CUR             5C61A  JTEBW3EH1A2041875          HYB

   model_year    medallion_type  agent_number  agent_name  \
0        2006  OWNER MUST DRIVE             1  TIRU CABS
1        2011      NAMED DRIVER             1  TIRU CABS
2        2011      NAMED DRIVER             1  TIRU CABS
3        2011  OWNER MUST DRIVE             1  TIRU CABS
4        2010      NAMED DRIVER             1  TIRU CABS

  agent_telephone_number  agent_website             agent_address  \
0          (718)937-2550            NaN  10 METROTECH NEW YORK NY
1          (718)937-1122            NaN  10 METROTECH NEW YORK NY
2          (718)937-1123            NaN  10 METROTECH NEW YORK NY
3          (718)937-2551            NaN  10 METROTECH NEW YORK NY
4          (718)937-1124            NaN  10 METROTECH NEW YORK NY

  last_updated_date last_updated_time
0         1/31/2015             13:20
1         1/31/2015             13:20
```