Uber Question1

April 17, 2022

[15]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv('/Users/xiaoyan/Documents/oa/uber data analyst/
 ↪green_tripdata_2014-06.csv')
df.VendorID = pd.to_datetime(df.VendorID)
print(df.shape)
```

```
(1337759, 20)
```

[16]:
```
df = df.loc[~((df['Trip_distance'] == 0))]
print(df.shape)
```

```
(747227, 20)
```

[17]:
```
dgroup = df.groupby(pd.Grouper(key='VendorID', freq='1D')).sum()
dgroup.head(5)
```

[17]:

VendorID	Lpep_dropoff_datetime	Store_and_fwd_flag	RateCodeID	\
2014-06-01	24278	-1.785572e+06	9.840058e+05	
2014-06-02	20095	-1.479732e+06	8.157017e+05	
2014-06-03	22547	-1.660762e+06	9.154031e+05	
2014-06-04	23572	-1.736499e+06	9.571132e+05	
2014-06-05	25274	-1.858440e+06	1.024297e+06	

VendorID	Pickup_longitude	Pickup_latitude	Dropoff_longitude	\
2014-06-01	-1.787628e+06	9.851251e+05	36860	
2014-06-02	-1.480894e+06	8.163005e+05	28663	
2014-06-03	-1.661186e+06	9.155912e+05	32183	
2014-06-04	-1.737366e+06	9.575405e+05	33806	
2014-06-05	-1.859525e+06	1.024850e+06	36620	

VendorID	Dropoff_latitude	Passenger_count	Trip_distance	Fare_amount	\
2014-06-01	78990.91	313256.06	12174.59	12091.5	
2014-06-02	58184.85	236495.00	14237.00	10012.5	

```
2014-06-03          60592.51        255955.05       16984.50      11235.5
2014-06-04          68554.68        280135.00       16986.00      11751.0
2014-06-05          72474.30        298157.50       17765.00      12581.5

               Extra  MTA_tax  Tip_amount  Tolls_amount  Ehail_fee  \
VendorID
2014-06-01  27680.22  2376.32         0.0     367586.84      38321
2014-06-02  19582.49  1658.80         0.0     281996.79      32503
2014-06-03  22628.83  1646.32         0.0     308459.34      35904
2014-06-04  25711.32  2049.09         0.0     336635.51      37218
2014-06-05  27014.22  2069.75         0.0     357613.57      39910

            Total_amount  Payment_type  Trip_type
VendorID
2014-06-01       24210.0           0.0        0.0
2014-06-02       20054.0           0.0        0.0
2014-06-03       22503.0           0.0        0.0
2014-06-04       23536.0           0.0        0.0
2014-06-05       25197.0           0.0        0.0
```

[18]:
```
print(dgroup["Passenger_count"].max())
print(dgroup["Fare_amount"].max())
```

```
406589.5
16969.5
```