In [77]:
```
import pandas as pd
import numpy as np
import pickle
import datetime
import calendar
import warnings
warnings.filterwarnings('ignore')
```

In [78]:
```
a=pd.read_csv("C:\\Users\\reshma_koduri\\OneDrive\\Documents\\uber.csv")
a
```

Out[78]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitu
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744
...	...	...	...	...	...	
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720

200000 rows × 9 columns

In [79]:
```
a.describe()
```

Out[79]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_lat
count	2.000000e+05	200000.000000	200000.000000	200000.000000	199999.000000	199999.0
mean	2.771250e+07	11.359955	-72.527638	39.935885	-72.525292	39.9
std	1.601382e+07	9.901776	11.437787	7.720539	13.117408	6.7
min	1.000000e+00	-52.000000	-1340.648410	-74.015515	-3356.666300	-881.9
25%	1.382535e+07	6.000000	-73.992065	40.734796	-73.991407	40.7

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_lat
50%	2.774550e+07	8.500000	-73.981823	40.752592	-73.980093	40.7
75%	4.155530e+07	12.500000	-73.967154	40.767158	-73.963658	40.7
max	5.542357e+07	499.000000	57.418457	1644.421482	1153.572603	872.6

In [80]:
```
a.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200000 entries, 0 to 199999
Data columns (total 9 columns):
 #   Column             Non-Null Count   Dtype  
---  ------             --------------   -----  
 0   Unnamed: 0         200000 non-null  int64  
 1   key                200000 non-null  object 
 2   fare_amount        200000 non-null  float64
 3   pickup_datetime    200000 non-null  object 
 4   pickup_longitude   200000 non-null  float64
 5   pickup_latitude    200000 non-null  float64
 6   dropoff_longitude  199999 non-null  float64
 7   dropoff_latitude   199999 non-null  float64
 8   passenger_count    200000 non-null  int64  
dtypes: float64(5), int64(2), object(2)
memory usage: 13.7+ MB
```

In [81]:
```
a.isna().sum()
```

Out[81]:
```
Unnamed: 0           0
key                  0
fare_amount          0
pickup_datetime      0
pickup_longitude     0
pickup_latitude      0
dropoff_longitude    1
dropoff_latitude     1
passenger_count      0
dtype: int64
```

In [82]:
```
from datetime import datetime
a['year'] = pd.DatetimeIndex(a['key']).year
a['time'] = pd.DatetimeIndex(a['key']).hour
a['month'] = pd.DatetimeIndex(a['key']).month
```

In [83]:
```
a
```

Out[83]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitu
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitu
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744
...	...	...	...	...	...	
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720

200000 rows × 12 columns

In [84]:
```
list(a)
```

Out[84]:
```
['Unnamed: 0',
 'key',
 'fare_amount',
 'pickup_datetime',
 'pickup_longitude',
 'pickup_latitude',
 'dropoff_longitude',
 'dropoff_latitude',
 'passenger_count',
 'year',
 'time',
 'month']
```

In [85]:
```
b=a.drop(['Unnamed: 0','pickup_datetime','dropoff_longitude','pickup_latitude','pick
b
```

Out[85]:

	key	fare_amount	passenger_count	year	time	month
0	2015-05-07 19:52:06.0000003	7.5	1	2015	19	5
1	2009-07-17 20:04:56.0000002	7.7	1	2009	20	7
2	2009-08-24 21:45:00.00000061	12.9	1	2009	21	8
3	2009-06-26 08:22:21.0000001	5.3	3	2009	8	6
4	2014-08-28 17:47:00.000000188	16.0	5	2014	17	8
...	...	...	...	...	...	...
199995	2012-10-28 10:49:00.00000053	3.0	1	2012	10	10
199996	2014-03-14 01:09:00.0000008	7.5	1	2014	1	3

	key	fare_amount	passenger_count	year	time	month
199997	2009-06-29 00:42:00.00000078	30.9	2	2009	0	6
199998	2015-05-20 14:56:25.0000004	14.5	1	2015	14	5
199999	2010-05-15 04:08:00.00000076	14.1	1	2010	4	5

200000 rows × 6 columns

In [86]:
```
b.to_csv('result2023.csv')
```

In [53]:
```
list(b)
```

Out[53]:
```
['key', 'fare_amount', 'passenger_count', 'date', 'time', 'year', 'month']
```

In [62]:
```
b.groupby('year').count()
```

Out[62]:

year	key	fare_amount	passenger_count	date	time	month
2009	30536	30536	30536	30536	30536	30536
2010	30194	30194	30194	30194	30194	30194
2011	31945	31945	31945	31945	31945	31945
2012	32396	32396	32396	32396	32396	32396
2013	31195	31195	31195	31195	31195	31195
2014	29968	29968	29968	29968	29968	29968
2015	13766	13766	13766	13766	13766	13766

In [87]:
```
#assigning values and creating csv file
```

In [66]:
```
year = [2009,2010,2011,2012]
passenger_count = [30536,30194,3194,32396]
month = [5, 6, 7, 9]
Year = {'Year': year, 'Passenger_count': passenger_count, 'Month': month}
df=pd.DataFrame(Year)
```

In [67]:
```
Year
```

Out[67]:
```
{'Year': [2009, 2010, 2011, 2012],
 'Passenger_count': [30536, 30194, 3194, 32396],
 'Month': [5, 6, 7, 9]}
```

In [75]:
```
df.to_csv('Year.csv')
```

In [76]:
```
df.to_csv("C:\\Users\\reshma_koduri\\OneDrive\\Documents\\Year.csv")
```

In [57]:
```
b.groupby('month').count()
```

Out[57]:

month	key	fare_amount	passenger_count	date	time	year
1	17668	17668	17668	17668	17668	17668
2	16695	16695	16695	16695	16695	16695
3	18763	18763	18763	18763	18763	18763
4	18606	18606	18606	18606	18606	18606
5	18859	18859	18859	18859	18859	18859
6	17787	17787	17787	17787	17787	17787
7	15095	15095	15095	15095	15095	15095
8	14221	14221	14221	14221	14221	14221
9	15266	15266	15266	15266	15266	15266
10	16212	16212	16212	16212	16212	16212
11	15312	15312	15312	15312	15312	15312
12	15516	15516	15516	15516	15516	15516

In [39]:
```
#c=pd.get_dummies(b,dtype=int)
#c
```

In [ ]:
```

```