In [1]:
```
import numpy as np
```

In [2]:
```
import pandas as pd
```

Pre-Processing

In [3]:
```
data=pd.read_csv(r"C:\Users\user\Downloads\7_uber.csv")
data
```

Out[3]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40
...	...	...	...	...	...	...
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40

200000 rows × 9 columns

In [4]:
```
data.head()
```

Out[4]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitud
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.73835
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.72822
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.74077
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.79084
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.74408

In [5]:
```
data.tail()
```

Out[5]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_la
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.7
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.7
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.7
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.7
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.7

In [6]:
```
data.describe()
```

Out[6]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropof
count	2.000000e+05	200000.000000	200000.000000	200000.000000	199999.000000	1999
mean	2.771250e+07	11.359955	-72.527638	39.935885	-72.525292	3
std	1.601382e+07	9.901776	11.437787	7.720539	13.117408	
min	1.000000e+00	-52.000000	-1340.648410	-74.015515	-3356.666300	-8
25%	1.382535e+07	6.000000	-73.992065	40.734796	-73.991407	4
50%	2.774550e+07	8.500000	-73.981823	40.752592	-73.980093	4
75%	4.155530e+07	12.500000	-73.967154	40.767158	-73.963658	4
max	5.542357e+07	499.000000	57.418457	1644.421482	1153.572603	8

In [8]:
```
print(np.shape(data))
```

```
(200000, 9)
```

In [9]:
```
print(np.size(data))
```

```
1800000
```

In [7]:
```
print(data.isna())
```

```
        Unnamed: 0    key  fare_amount  pickup_datetime  pickup_longitude  \
0            False  False        False            False             False
1            False  False        False            False             False
2            False  False        False            False             False
3            False  False        False            False             False
4            False  False        False            False             False
...            ...    ...          ...              ...               ...
199995       False  False        False            False             False
199996       False  False        False            False             False
199997       False  False        False            False             False
199998       False  False        False            False             False
199999       False  False        False            False             False

        pickup_latitude  dropoff_longitude  dropoff_latitude  passenger_count
0                 False              False             False            False
1                 False              False             False            False
2                 False              False             False            False
3                 False              False             False            False
4                 False              False             False            False
...                 ...                ...               ...              ...
199995            False              False             False            False
199996            False              False             False            False
199997            False              False             False            False
199998            False              False             False            False
199999            False              False             False            False

[200000 rows x 9 columns]
```

In [8]:
```
data.fillna(value=0)
```

Out[8]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40
...	...	...	...	...	...	
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40

200000 rows × 9 columns

In [9]:
```
data.dropna()
```

Out[9]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40
...	...	...	...	...	...	
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40

199999 rows × 9 columns

Visualization

In [10]:
```
import matplotlib.pyplot as pp
```

In [11]:
```
data
```

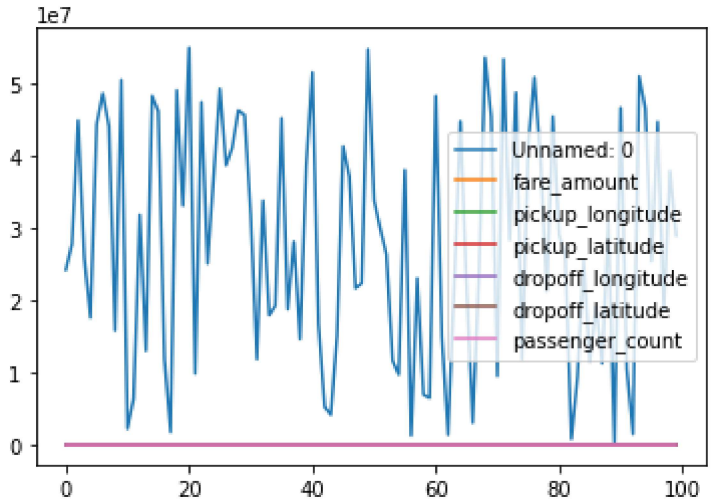
Out[11]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40
...	...	...	...	...	...	...
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40

200000 rows × 9 columns

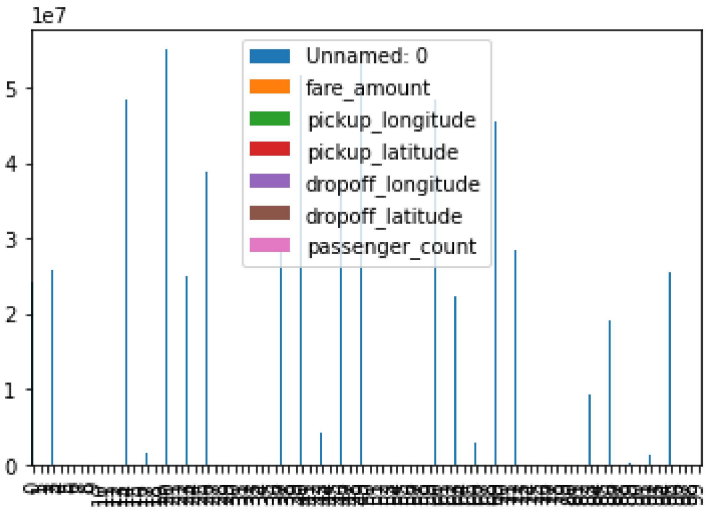
In [12]:
```
da=data.head(100)
da.plot.line()
```

Out[12]: <AxesSubplot:>



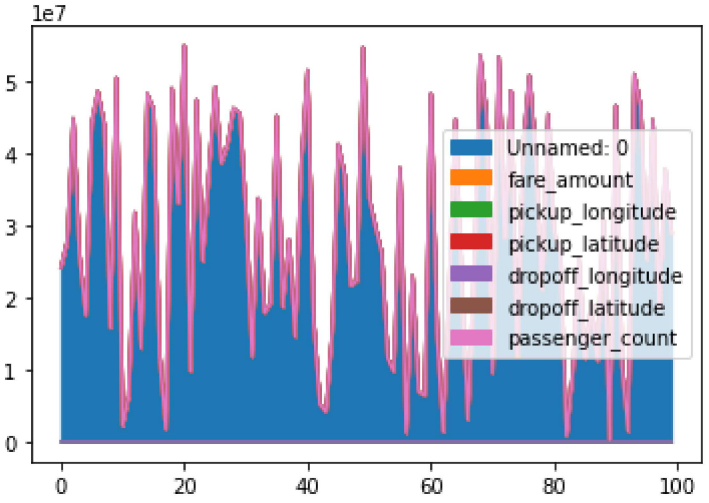
In [13]:
```
da.plot.bar()
```

Out[13]: <AxesSubplot:>



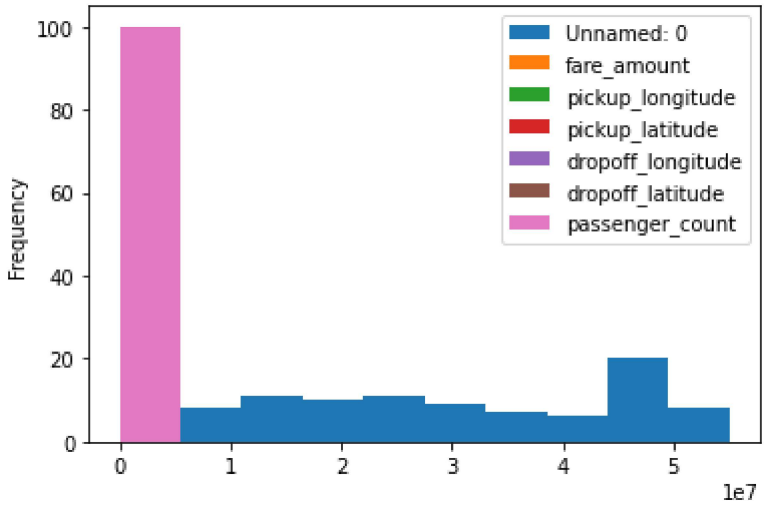
In [14]:
```
da.plot.area()
```

Out[14]: <AxesSubplot:>



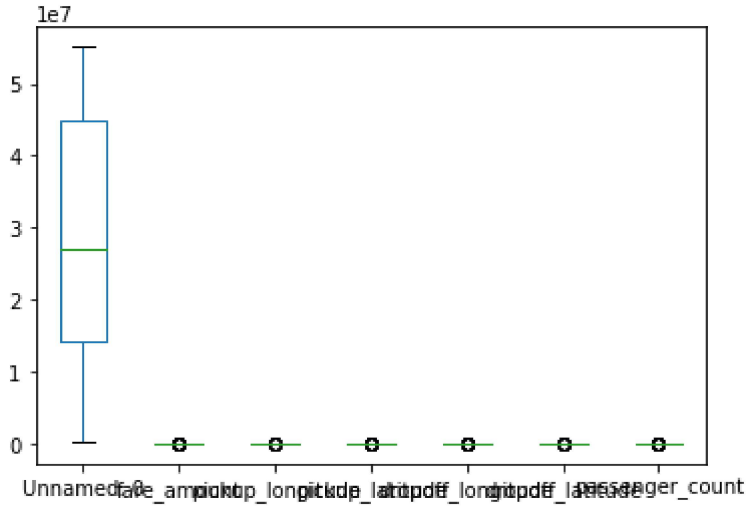
In [15]:
```
da.plot.hist()
```

Out[15]: <AxesSubplot:ylabel='Frequency'>



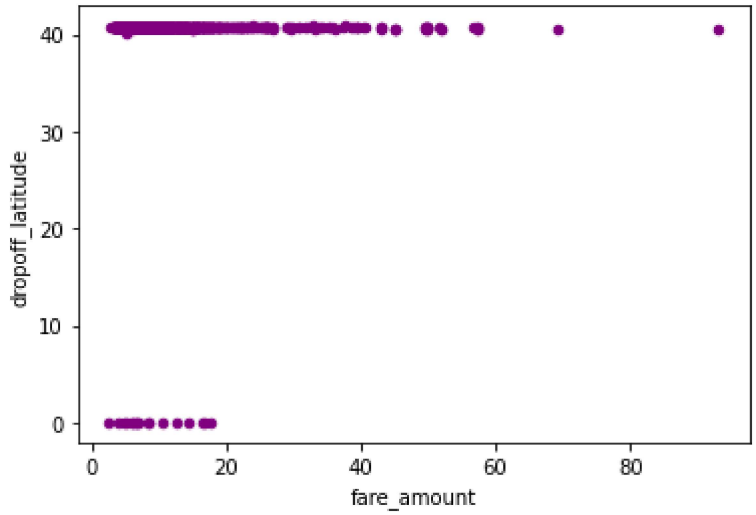
In [16]:
```
da.plot.box()
```

Out[16]: <AxesSubplot:>



In [29]:
```
da.plot.scatter(x='fare_amount',y='dropoff_latitude',color='purple')
```

Out[29]:
```
<AxesSubplot:xlabel='fare_amount', ylabel='dropoff_latitude'>
```



In [18]:
```
da.mean()
```

Out[18]:
```
Unnamed: 0           2.810554e+07
fare_amount          1.106570e+01
pickup_longitude    -7.101976e+01
pickup_latitude      3.912362e+01
dropoff_longitude   -7.101548e+01
dropoff_latitude     3.912629e+01
passenger_count      1.640000e+00
dtype: float64
```

In [19]:
```
da.median()
```

Out[19]:
```
Unnamed: 0           2.710896e+07
fare_amount          8.100000e+00
pickup_longitude    -7.398200e+01
pickup_latitude      4.075276e+01
dropoff_longitude   -7.397940e+01
dropoff_latitude     4.075708e+01
passenger_count      1.000000e+00
dtype: float64
```

In [20]:
```
da.mode()
```

Out[20]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitud
0	226870	2009-01-09 14:24:24.0000003	4.9	2009-01-09 14:24:24 UTC	0.0	0.
1	781560	2009-01-10 22:43:36.0000007	7.7	2009-01-10 22:43:36 UTC	NaN	Na
2	1239488	2009-02-02 16:58:00.00000011	NaN	2009-02-02 16:58:00 UTC	NaN	Na
3	1347588	2009-02-12 17:52:18.0000001	NaN	2009-02-12 17:52:18 UTC	NaN	Na
4	1454546	2009-02-19 08:28:42.0000001	NaN	2009-02-19 08:28:42 UTC	NaN	Na
...	...	...	...	...	...	.
95	51671648	2015-05-17 14:56:39.0000002	NaN	2015-05-17 14:56:39 UTC	NaN	Na
96	53475676	2015-05-21 01:35:16.0000006	NaN	2015-05-21 01:35:16 UTC	NaN	Na
97	53705359	2015-05-22 17:32:27.0000004	NaN	2015-05-22 17:32:27 UTC	NaN	Na
98	54858310	2015-05-28 13:31:03.0000006	NaN	2015-05-28 13:31:03 UTC	NaN	Na
99	55085966	2015-06-17 17:52:03.0000007	NaN	2015-06-17 17:52:03 UTC	NaN	Na

100 rows × 9 columns

In [21]:
```
da.sum()
```

Out[21]:
```
Unnamed: 0                                                2810553617
key                 2015-05-07 19:52:06.00000032009-07-17 20:04:56...
fare_amount                                                  1106.57
pickup_datetime     2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...
pickup_longitude                                         -7101.975887
pickup_latitude                                           3912.362079
dropoff_longitude                                        -7101.547911
dropoff_latitude                                          3912.629481
passenger_count                                                   164
dtype: object
```

In [22]:
```
da.cumsum()
```

Out[22]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_l
0	24238194	2015-05-07 19:52:06.0000003	7.50	2015-05-07 19:52:06 UTC	-73.999817	40.
1	52073393	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	15.20	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC	-147.994172	81.
2	97057748	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	28.10	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-221.999215	122.
3	122952478	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	33.40	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-295.975339	162.
4	140562630	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	49.40	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-369.900362	203.
...	...	...	...	...	...	
95	2680223709	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	1040.97	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-6806.017481	3749.
96	2725015721	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	1045.47	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-6880.007536	3790.
97	2743586741	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	1048.77	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-6953.989862	3830.
98	2781529145	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	1079.67	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-7027.985750	3871.
99	2810553617	2015-05-07 19:52:06.00000032009-07-17 20:04:56...	1106.57	2015-05-07 19:52:06 UTC2009-07-17 20:04:56 UTC...	-7101.975887	3912.

100 rows × 9 columns

In [23]:
```
da.count()
```

Out[23]:
```
Unnamed: 0           100
key                  100
fare_amount          100
pickup_datetime      100
pickup_longitude     100
pickup_latitude      100
dropoff_longitude    100
dropoff_latitude     100
passenger_count      100
dtype: int64
```

In [24]:
```
da.max()
```

Out[24]:
```
Unnamed: 0                            55085966
key                  2015-06-17 17:52:03.0000007
fare_amount                                 56.8
pickup_datetime         2015-06-17 17:52:03 UTC
pickup_longitude                            0.0
pickup_latitude                       40.850558
dropoff_longitude                           0.0
dropoff_latitude                      40.876687
passenger_count                               5
dtype: object
```

In [25]:
```
da.min()
```

Out[25]:
```
Unnamed: 0                              226870
key                  2009-01-09 14:24:24.0000003
fare_amount                                  2.5
pickup_datetime         2009-01-09 14:24:24 UTC
pickup_longitude                      -74.013173
pickup_latitude                              0.0
dropoff_longitude                     -74.016152
dropoff_latitude                             0.0
passenger_count                              1
dtype: object
```

In [26]:
```
da.cov()
```

Out[26]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longit
Unnamed: 0	2.673334e+14	6.570581e+06	-2.872125e+07	1.584530e+07	-2.866167e
fare_amount	6.570581e+06	8.153650e+01	-1.576000e+01	8.704432e+00	-1.571441e
pickup_longitude	-2.872125e+07	-1.576000e+01	2.122821e+02	-1.169423e+02	2.122689e
pickup_latitude	1.584530e+07	8.704432e+00	-1.169423e+02	6.442242e+01	-1.169353e
dropoff_longitude	-2.866167e+07	-1.571441e+01	2.122689e+02	-1.169353e+02	2.122566e
dropoff_latitude	1.581474e+07	8.756788e+00	-1.169502e+02	6.442633e+01	-1.169429e
passenger_count	1.853260e+05	8.575273e-01	-1.902809e+00	1.054606e+00	-1.905182e

In [27]:
```
da.corr()
```

Out[27]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude
Unnamed: 0	1.000000	0.044504	-0.120565	0.120741	-0.120322
fare_amount	0.044504	1.000000	-0.119791	0.120101	-0.119451
pickup_longitude	-0.120565	-0.119791	1.000000	-0.999992	0.999998
pickup_latitude	0.120741	0.120101	-0.999992	1.000000	-0.999992
dropoff_longitude	-0.120322	-0.119451	0.999998	-0.999992	1.000000
dropoff_latitude	0.120500	0.120815	-0.999992	0.999993	-0.999989
passenger_count	0.009564	0.080135	-0.110201	0.110871	-0.110345

In [ ]:
```

```