In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as pp
```

In [2]:
```
data=pd.read_csv(r"C:\Users\user\Desktop\Ash\uber.csv")
```

Display top 7 and last 6 rows and print the output

In [3]:
```
data.head(7)
```

Out[3]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	p
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	-73.999512	40.723217	
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	-73.994710	40.750325	
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	-73.962565	40.772647	
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	-73.965316	40.803349	
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	-73.973082	40.761247	
5	44470845	2011-02-12 02:27:09.0000006	4.9	2011-02-12 02:27:09 UTC	-73.969019	40.755910	-73.969019	40.755910	
6	48725865	2014-10-12 07:04:00.0000002	24.5	2014-10-12 07:04:00 UTC	-73.961447	40.693965	-73.871195	40.774297	

In [4]:
```
data.tail(6)
```

Out[4]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitu
199994	3189201	2014-01-31 14:42:00.000000181	12.0	2014-01-31 14:42:00 UTC	-73.983070	40.760770	-73.972972	40.7541
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367	-73.986525	40.7402
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837	-74.006672	40.7396
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487	-73.858957	40.6925
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452	-73.983215	40.6954
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077	-73.985508	40.7687

Fill with a constant value and print the output

In [5]:
```
data.fillna(value=10)
```

Out[5]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitu
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	-73.999512	40.7232
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	-73.994710	40.7503
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	-73.962565	40.7726
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	-73.965316	40.8033
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	-73.973082	40.7612
...	...	...	...	...	...	...	...	
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367	-73.986525	40.7402
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837	-74.006672	40.7396
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487	-73.858957	40.6925
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452	-73.983215	40.6954
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077	-73.985508	40.7687

200000 rows × 9 columns

Drop the column with missing values and print the output

In [6]:
```
data.dropna(axis=1)
```

Out[6]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	passenger_count
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	1
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	1
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	1
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	3
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	5
...	...	...	...	...	...	...	...
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367	1
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837	1
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487	2
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452	1
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077	1

200000 rows × 7 columns

Drop the row with missing values and print the output

In [7]:
```
data.dropna()
```

Out[7]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitu
0	24238194	2015-05-07 19:52:06.0000003	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	-73.999512	40.7232
1	27835199	2009-07-17 20:04:56.0000002	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	-73.994710	40.7503
2	44984355	2009-08-24 21:45:00.00000061	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	-73.962565	40.7726
3	25894730	2009-06-26 08:22:21.0000001	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	-73.965316	40.8033
4	17610152	2014-08-28 17:47:00.000000188	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	-73.973082	40.7612
...	...	...	...	...	...	...	...	
199995	42598914	2012-10-28 10:49:00.00000053	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367	-73.986525	40.7402
199996	16382965	2014-03-14 01:09:00.0000008	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837	-74.006672	40.7396
199997	27804658	2009-06-29 00:42:00.00000078	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487	-73.858957	40.6925
199998	20259894	2015-05-20 14:56:25.0000004	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452	-73.983215	40.6954
199999	11951496	2010-05-15 04:08:00.00000076	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077	-73.985508	40.7687

199999 rows × 9 columns

To check the presence of missing values in your dataframe

In [8]:
```
data.isna()
```

Out[8]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	passenge
0	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	
199995	False	False	False	False	False	False	False	False	
199996	False	False	False	False	False	False	False	False	
199997	False	False	False	False	False	False	False	False	
199998	False	False	False	False	False	False	False	False	
199999	False	False	False	False	False	False	False	False	

200000 rows × 9 columns

Use operators and check the condition and print the output

In [9]:
```
data[data['fare_amount']>100]
```

Out[9]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitu
2053	31333682	2014-11-01 08:42:39.0000002	113.66	2014-11-01 08:42:39 UTC	-73.951227	40.778753	-73.949938	40.7781
4292	33491441	2014-10-24 20:20:00.000000181	350.00	2014-10-24 20:20:00 UTC	0.000000	0.000000	0.000000	0.0000
5968	28138818	2011-05-06 00:40:00.000000224	105.00	2011-05-06 00:40:00 UTC	-73.752265	40.923303	-73.752270	40.9233
6613	28579349	2013-05-03 10:05:00.000000192	137.00	2013-05-03 10:05:00 UTC	0.000000	0.000000	0.000000	0.0000
9060	33046347	2011-06-13 15:46:00.00000036	126.10	2011-06-13 15:46:00 UTC	-73.788657	40.640643	-74.001350	41.0480
...	...	...	...	...	...	...	...	
190767	19562599	2012-09-03 00:21:00.00000051	120.30	2012-09-03 00:21:00 UTC	-73.788095	40.642330	-73.976730	40.9311
194454	34209729	2013-11-22 13:07:00.000000103	130.25	2013-11-22 13:07:00 UTC	-73.982272	40.763447	-74.177182	40.6950
196616	53659256	2014-02-02 04:43:45.0000003	109.00	2014-02-02 04:43:45 UTC	-73.984697	40.749896	-74.045293	40.9731
196647	13085828	2010-08-19 16:52:45.0000003	200.00	2010-08-19 16:52:45 UTC	-73.952994	40.736298	-73.952994	40.7362
197493	54143082	2014-09-07 08:39:00.00000012	230.00	2014-09-07 08:39:00 UTC	-73.937765	40.758267	-74.382200	40.7008

84 rows × 9 columns

Display your output using loc and iloc, row and column heading

In [10]:
```
data.iloc[190767:200000]
```

Out[10]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitud
190767	19562599	2012-09-03 00:21:00.00000051	120.30	2012-09-03 00:21:00 UTC	-73.788095	40.642330	-73.976730	40.93113
190768	24624036	2011-10-08 09:25:00.00000068	5.70	2011-10-08 09:25:00 UTC	-73.978315	40.748417	-3.975457	40.75726
190769	26487358	2012-04-30 12:00:13.0000002	7.70	2012-04-30 12:00:13 UTC	-73.982882	40.771515	-73.967373	40.79265
190770	11808866	2015-06-10 15:12:05.0000005	12.00	2015-06-10 15:12:05 UTC	-73.974869	40.793243	-73.958405	40.77385
190771	20404547	2010-10-07 19:52:08.0000001	55.47	2010-10-07 19:52:08 UTC	-73.786297	40.641247	-73.863958	40.89795
...	...	...	...	...	...	...	...	.
199995	42598914	2012-10-28 10:49:00.00000053	3.00	2012-10-28 10:49:00 UTC	-73.987042	40.739367	-73.986525	40.74029
199996	16382965	2014-03-14 01:09:00.0000008	7.50	2014-03-14 01:09:00 UTC	-73.984722	40.736837	-74.006672	40.73962
199997	27804658	2009-06-29 00:42:00.00000078	30.90	2009-06-29 00:42:00 UTC	-73.986017	40.756487	-73.858957	40.69258
199998	20259894	2015-05-20 14:56:25.0000004	14.50	2015-05-20 14:56:25 UTC	-73.997124	40.725452	-73.983215	40.69541
199999	11951496	2010-05-15 04:08:00.00000076	14.10	2010-05-15 04:08:00 UTC	-73.984395	40.720077	-73.985508	40.76879

9233 rows × 9 columns

In [11]:
```
data.loc[50:60]
```

Out[11]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude
50	33899076	2009-10-03 18:25:00.000000186	7.7	2009-10-03 18:25:00 UTC	-73.983052	40.776055	-73.974093	40.763797
51	29966903	2011-08-15 13:55:53.0000001	13.3	2011-08-15 13:55:53 UTC	-73.985900	40.727700	-73.982500	40.771900
52	26382712	2012-05-25 03:19:00.00000063	14.9	2012-05-25 03:19:00 UTC	-73.997385	40.720822	-73.955628	40.782252
53	11666085	2012-12-09 02:12:29.0000004	7.5	2012-12-09 02:12:29 UTC	-73.937809	40.850558	-73.944454	40.834959
54	9716818	2009-04-09 11:27:00.000000125	16.9	2009-04-09 11:27:00 UTC	-73.986593	40.759143	-74.001767	40.720120
55	38092793	2014-11-13 20:12:25.0000005	11.0	2014-11-13 20:12:25 UTC	-73.958903	40.815239	-73.980862	40.782610
56	1239488	2010-03-05 22:41:05.0000003	6.5	2010-03-05 22:41:05 UTC	-73.990305	40.756175	-74.003678	40.759173
57	23077019	2009-01-09 14:24:24.0000003	6.5	2009-01-09 14:24:24 UTC	-73.968111	40.765625	-73.980863	40.759160
58	6922345	2009-07-10 01:01:00.00000094	5.3	2009-07-10 01:01:00 UTC	-73.989730	40.734507	-74.001847	40.740387
59	6479703	2009-02-02 16:58:00.00000011	4.1	2009-02-02 16:58:00 UTC	-73.954332	40.774060	-73.952558	40.766325
60	48399259	2015-06-17 17:52:03.0000007	9.0	2015-06-17 17:52:03 UTC	-73.972137	40.749889	-73.981857	40.732426

Display the statistical summary of data

In [12]:
```
data.describe()
```

Out[12]:

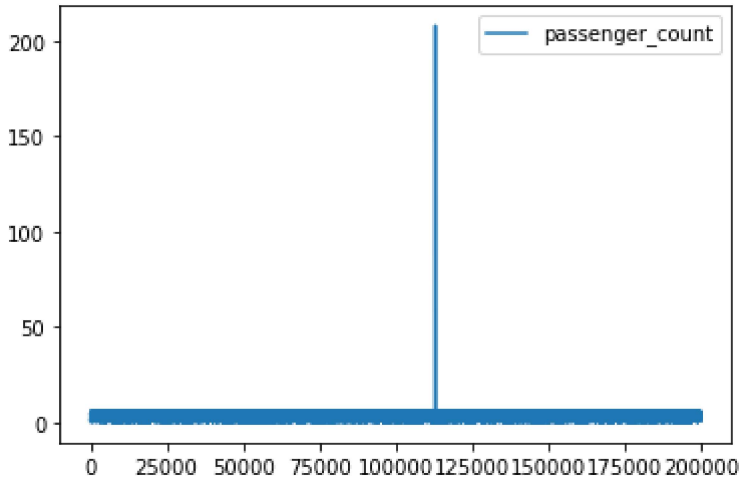
	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	passenger_count
count	2.000000e+05	200000.000000	200000.000000	200000.000000	199999.000000	199999.000000	200000.000000
mean	2.771250e+07	11.359955	-72.527638	39.935885	-72.525292	39.923890	1.684535
std	1.601382e+07	9.901776	11.437787	7.720539	13.117408	6.794829	1.385997
min	1.000000e+00	-52.000000	-1340.648410	-74.015515	-3356.666300	-881.985513	0.000000
25%	1.382535e+07	6.000000	-73.992065	40.734796	-73.991407	40.733823	1.000000
50%	2.774550e+07	8.500000	-73.981823	40.752592	-73.980093	40.753042	1.000000
75%	4.155530e+07	12.500000	-73.967154	40.767158	-73.963658	40.768001	2.000000
max	5.542357e+07	499.000000	57.418457	1644.421482	1153.572603	872.697628	208.000000

Visualization

In [13]:
```
df=data[['key','passenger_count']]
```

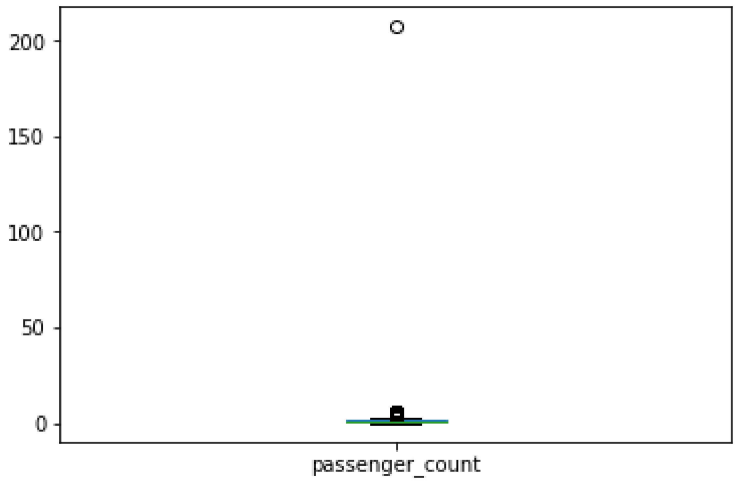
In [14]:
```
df.plot.line()
```

Out[14]: <AxesSubplot:>



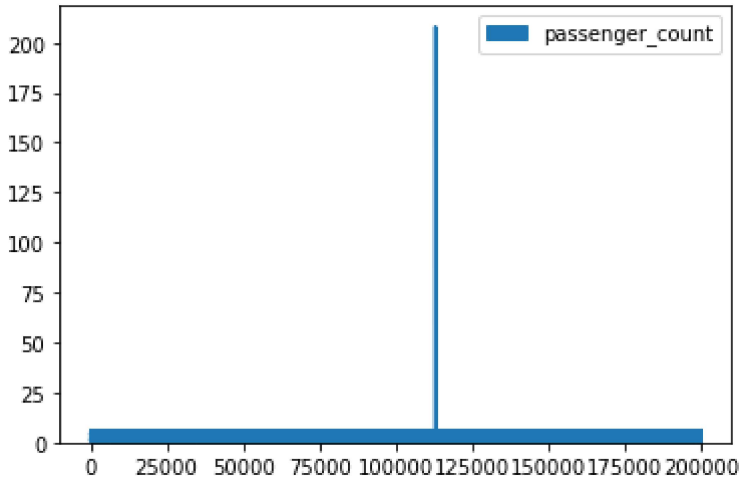
In [15]:
```
df.plot.box()
```

Out[15]: <AxesSubplot:>



In [16]:
```
df.plot.area()
```

Out[16]: <AxesSubplot:>



In [17]:
```
df.plot.hist()
```

Out[17]: <AxesSubplot:ylabel='Frequency'>