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

In [2]:
```
data=pd.read_csv(r"C:\Users\user\Desktop\Vicky\7_uber.csv")
```

In [3]:
```
data.shape
```

Out[3]:
```
(200000, 9)
```

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

Out[4]:
```
<bound method NDFrame.head of         Unnamed: 0                                 k
ey  fare_amount  \
0         24238194    2015-05-07 19:52:06.0000003          7.5
1         27835199    2009-07-17 20:04:56.0000002          7.7
2         44984355   2009-08-24 21:45:00.00000061         12.9
3         25894730    2009-06-26 08:22:21.0000001          5.3
4         17610152  2014-08-28 17:47:00.000000188         16.0
...            ...                           ...          ...
199995    42598914   2012-10-28 10:49:00.00000053          3.0
199996    16382965    2014-03-14 01:09:00.0000008          7.5
199997    27804658   2009-06-29 00:42:00.00000078         30.9
199998    20259894    2015-05-20 14:56:25.0000004         14.5
199999    11951496   2010-05-15 04:08:00.00000076         14.1

                pickup_datetime  pickup_longitude  pickup_latitude  \
0       2015-05-07 19:52:06 UTC        -73.999817        40.738354
1       2009-07-17 20:04:56 UTC        -73.994355        40.728225
2       2009-08-24 21:45:00 UTC        -74.005043        40.740770
3       2009-06-26 08:22:21 UTC        -73.976124        40.790844
4       2014-08-28 17:47:00 UTC        -73.925023        40.744085
...                         ...               ...              ...
199995  2012-10-28 10:49:00 UTC        -73.987042        40.739367
199996  2014-03-14 01:09:00 UTC        -73.984722        40.736837
199997  2009-06-29 00:42:00 UTC        -73.986017        40.756487
199998  2015-05-20 14:56:25 UTC        -73.997124        40.725452
199999  2010-05-15 04:08:00 UTC        -73.984395        40.720077

        dropoff_longitude  dropoff_latitude  passenger_count
0              -73.999512         40.723217                1
1              -73.994710         40.750325                1
2              -73.962565         40.772647                1
3              -73.965316         40.803349                3
4              -73.973082         40.761247                5
...                   ...               ...              ...
199995         -73.986525         40.740297                1
199996         -74.006672         40.739620                1
199997         -73.858957         40.692588                2
199998         -73.983215         40.695415                1
199999         -73.985508         40.768793                1

[200000 rows x 9 columns]>
```

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 | dropoff |
|---|---|---|---|---|---|---|
| **count** | 2.000000e+05 | 200000.000000 | 200000.000000 | 200000.000000 | 199999.000000 | 19999 |
| **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 | -88 |
| **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 | 87 |

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

Out[7]:

| | Unnamed: 0 | key | fare_amount | pickup_datetime | pickup_longitude | pickup_latitude | drop |
|---|---|---|---|---|---|---|---|
| **0** | False | False | False | False | False | False | |
| **1** | False | False | False | False | False | False | |
| **2** | False | False | False | False | False | False | |
| **3** | False | False | False | False | False | False | |
| **4** | False | False | False | False | False | False | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **199995** | False | False | False | False | False | False | |
| **199996** | False | False | False | False | False | False | |
| **199997** | False | False | False | False | False | False | |
| **199998** | False | False | False | False | False | False | |
| **199999** | False | False | False | False | False | False | |

200000 rows × 9 columns

In [8]:
```
data.size
```

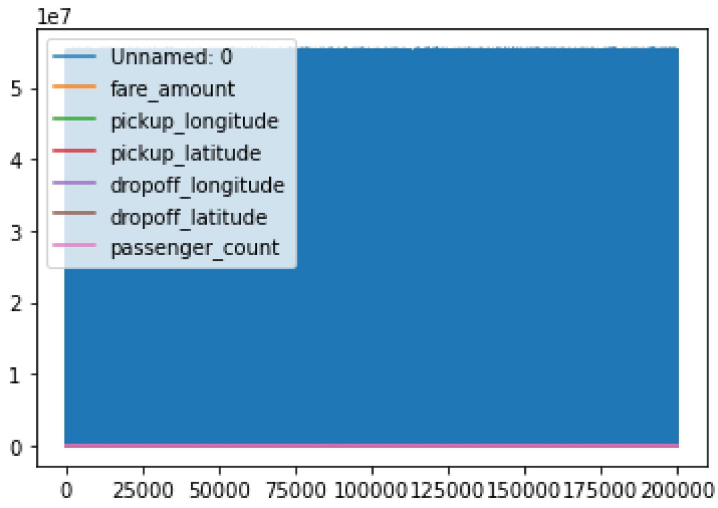
Out[8]: 1800000

In [9]:
```
data.isna
```

Out[9]:
```
<bound method DataFrame.isna of         Unnamed: 0
key  fare_amount  \
0         24238194    2015-05-07 19:52:06.0000003          7.5
1         27835199    2009-07-17 20:04:56.0000002          7.7
2         44984355   2009-08-24 21:45:00.00000061         12.9
3         25894730    2009-06-26 08:22:21.0000001          5.3
4         17610152  2014-08-28 17:47:00.000000188         16.0
...            ...                           ...          ...
199995    42598914   2012-10-28 10:49:00.00000053          3.0
199996    16382965    2014-03-14 01:09:00.0000008          7.5
199997    27804658   2009-06-29 00:42:00.00000078         30.9
199998    20259894    2015-05-20 14:56:25.0000004         14.5
199999    11951496   2010-05-15 04:08:00.00000076         14.1

                pickup_datetime  pickup_longitude  pickup_latitude  \
0       2015-05-07 19:52:06 UTC        -73.999817        40.738354
1       2009-07-17 20:04:56 UTC        -73.994355        40.728225
2       2009-08-24 21:45:00 UTC        -74.005043        40.740770
3       2009-06-26 08:22:21 UTC        -73.976124        40.790844
4       2014-08-28 17:47:00 UTC        -73.925023        40.744085
...                         ...               ...              ...
199995  2012-10-28 10:49:00 UTC        -73.987042        40.739367
199996  2014-03-14 01:09:00 UTC        -73.984722        40.736837
199997  2009-06-29 00:42:00 UTC        -73.986017        40.756487
199998  2015-05-20 14:56:25 UTC        -73.997124        40.725452
199999  2010-05-15 04:08:00 UTC        -73.984395        40.720077

        dropoff_longitude  dropoff_latitude  passenger_count
0              -73.999512         40.723217                1
1              -73.994710         40.750325                1
2              -73.962565         40.772647                1
3              -73.965316         40.803349                3
4              -73.973082         40.761247                5
...                   ...               ...              ...
199995         -73.986525         40.740297                1
199996         -74.006672         40.739620                1
199997         -73.858957         40.692588                2
199998         -73.983215         40.695415                1
199999         -73.985508         40.768793                1

[200000 rows x 9 columns]>
```

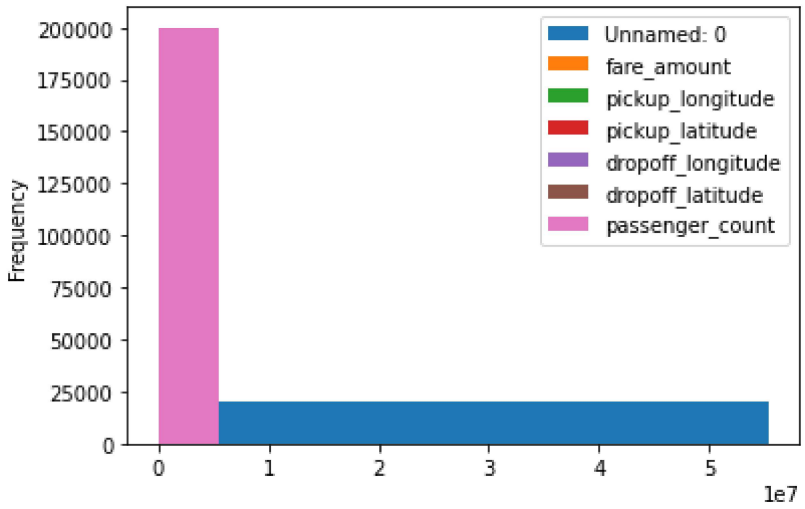
In [10]:
```
data.plot.line()
```

Out[10]: <AxesSubplot:>



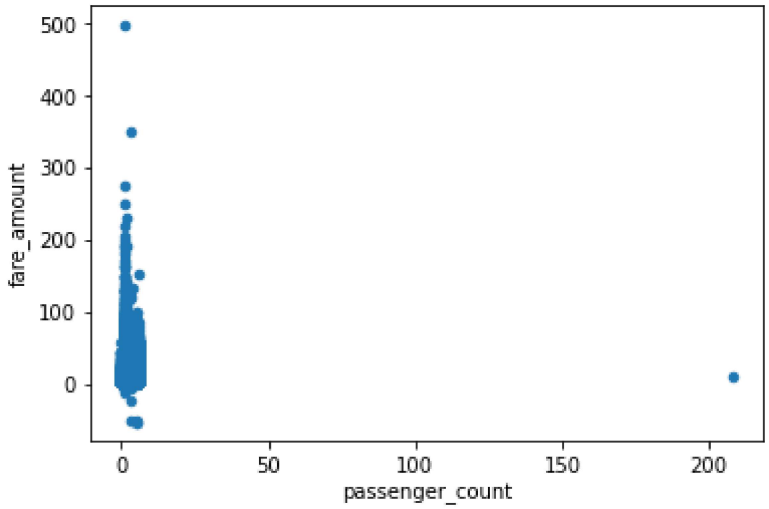

In [11]:
```
data.plot.hist()
```

Out[11]: <AxesSubplot:ylabel='Frequency'>



In [12]:
```
data.plot.scatter(x="passenger_count",y='fare_amount')
```

Out[12]: <AxesSubplot:xlabel='passenger_count', ylabel='fare_amount'>




In [13]:
```
data.fillna(value=5)
```

Out[13]:

| | Unnamed: 0 | key | fare_amount | pickup_datetime | pickup_longitude | pickup_l |
|---|---|---|---|---|---|---|
| **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 [14]:
```
pd.isna(data)
```

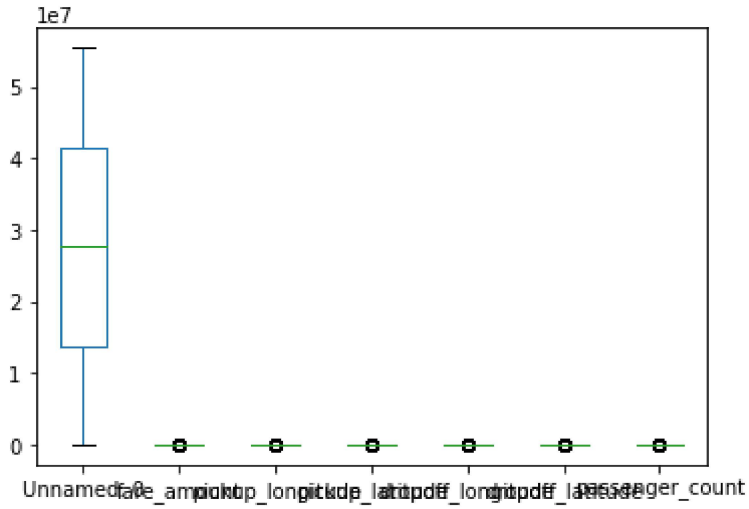
Out[14]:

| | Unnamed: 0 | key | fare_amount | pickup_datetime | pickup_longitude | pickup_latitude | dropo |
|---|---|---|---|---|---|---|---|
| **0** | False | False | False | False | False | False | |
| **1** | False | False | False | False | False | False | |
| **2** | False | False | False | False | False | False | |
| **3** | False | False | False | False | False | False | |
| **4** | False | False | False | False | False | False | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **199995** | False | False | False | False | False | False | |
| **199996** | False | False | False | False | False | False | |
| **199997** | False | False | False | False | False | False | |
| **199998** | False | False | False | False | False | False | |
| **199999** | False | False | False | False | False | False | |

200000 rows × 9 columns

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

Out[16]: <AxesSubplot:>




In [*]:
```
data.plot.pie(y="passenger_count")
```

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

In [ ]:
```

```