Data set 3

In [1]:

```python
import numpy as np
import pandas as pd
```

In [2]:

```python
a=pd.read_csv(r"C:\Users\user\Downloads\uber - uber.csv")
```

To print top rows:

In [3]:

```
c=a.head(100)
c
```

Out[3]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dro
0	24238194	2015-05-07 19:52:06	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	
1	27835199	2009-07-17 20:04:56	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	
2	44984355	2009-08-24 21:45:00	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	
3	25894730	2009-06-26 8:22:21	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	
4	17610152	2014-08-28 17:47:00	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	
...	...	...	...	...	...	...	
95	25431833	2015-04-11 8:47:47	9.5	2015-04-11 08:47:47 UTC	-73.978432	40.752399	
96	44792012	2011-10-03 20:29:00	4.5	2011-10-03 20:29:00 UTC	-73.990055	40.756413	
97	18571020	2010-04-26 3:12:44	3.3	2010-04-26 03:12:44 UTC	-73.982326	40.731314	
98	37942404	2011-11-18 9:51:00	30.9	2011-11-18 09:51:00 UTC	-73.995888	40.759078	
99	29024472	2009-08-30 14:03:55	26.9	2009-08-30 14:03:55 UTC	-73.990137	40.756007	

100 rows × 9 columns
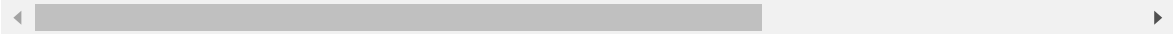
To print Last rows:

In [4]:

```
a.tail()
```

Out[4]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude
199995	42598914	2012-10-28 10:49:00	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367
199996	16382965	2014-03-14 1:09:00	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837
199997	27804658	2009-06-29 0:42:00	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487
199998	20259894	2015-05-20 14:56:25	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452
199999	11951496	2010-05-15 4:08:00	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077

Statistical Summary:

In [5]:

```
a.describe()
```

Out[5]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dro
count	2.000000e+05	200000.000000	200000.000000	200000.000000	199999.000000	19
mean	2.771250e+07	11.359955	-72.527638	39.935885	-72.525292	
std	1.601382e+07	9.901776	11.437787	7.720539	13.117408	
min	1.000000e+00	-52.000000	-1340.648410	-74.015515	-3356.666300	
25%	1.382535e+07	6.000000	-73.992065	40.734796	-73.991407	
50%	2.774550e+07	8.500000	-73.981823	40.752592	-73.980093	
75%	4.155530e+07	12.500000	-73.967153	40.767158	-73.963659	
max	5.542357e+07	499.000000	57.418457	1644.421482	1153.572603	

To print no of rows and columns

In [6]:

```
a.shape
```

Out[6]:

```
(200000, 9)
```

To print no of elements

In [7]:

```
a.size
```

Out[7]:

```
1800000
```

In [8]:

```
a.isna()
```

Out[8]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dr
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 [9]:

```
a.dropna()
```

Out[9]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude
0	24238194	2015-05-07 19:52:06	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354
1	27835199	2009-07-17 20:04:56	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225
2	44984355	2009-08-24 21:45:00	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770
3	25894730	2009-06-26 8:22:21	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844
4	17610152	2014-08-28 17:47:00	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085
...	...	...	...	...	...	...
199995	42598914	2012-10-28 10:49:00	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367
199996	16382965	2014-03-14 1:09:00	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837
199997	27804658	2009-06-29 0:42:00	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487
199998	20259894	2015-05-20 14:56:25	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452
199999	11951496	2010-05-15 4:08:00	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077

199999 rows × 9 columns

In [10]:

```
conda install matplotlib
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.


Note: you may need to restart the kernel to use updated packages.


==> WARNING: A newer version of conda exists. <==
  current version: 4.10.1
  latest version: 23.5.2

Please update conda by running

    $ conda update -n base -c defaults conda
```

In [11]:

```python
import matplotlib.pyplot as pp
```

In [12]:

```python
b=a[['fare_amount','pickup_longitude']]
b
```

Out[12]:

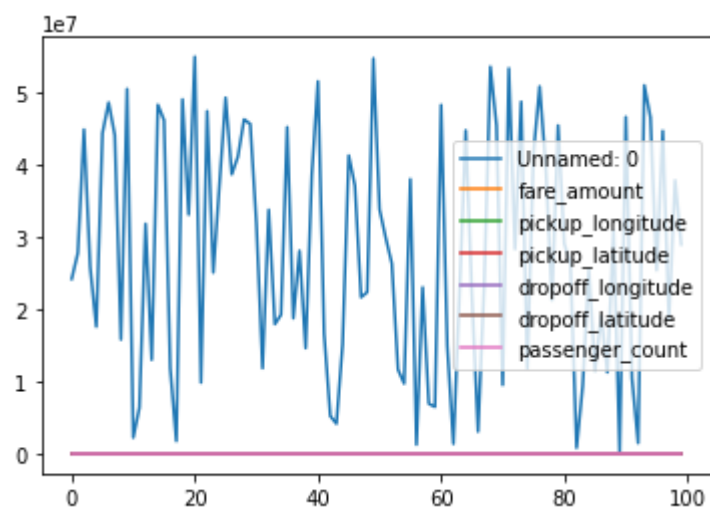|	fare_amount	pickup_longitude
0	7.5	-73.999817
1	7.7	-73.994355
2	12.9	-74.005043
3	5.3	-73.976124
4	16.0	-73.925023
...	...	...
199995	3.0	-73.987042
199996	7.5	-73.984722
199997	30.9	-73.986017
199998	14.5	-73.997124
199999	14.1	-73.984395

200000 rows × 2 columns

In [20]:

```
c.plot.line()
```

Out[20]:

<AxesSubplot:>

In [16]:

```
c.plot.bar()
```
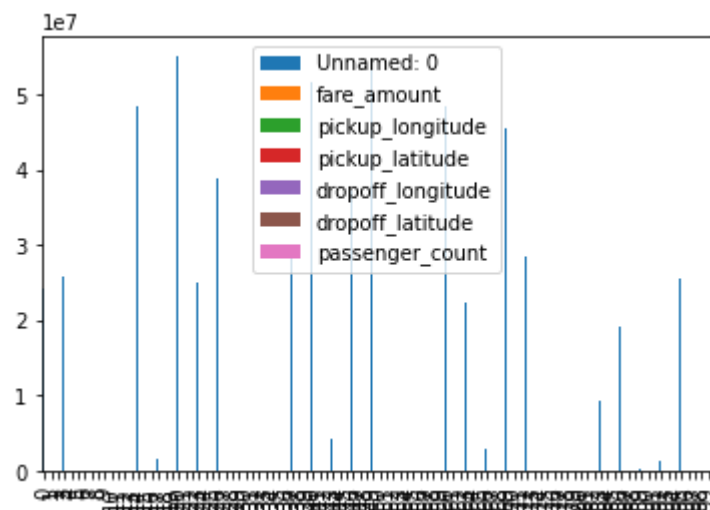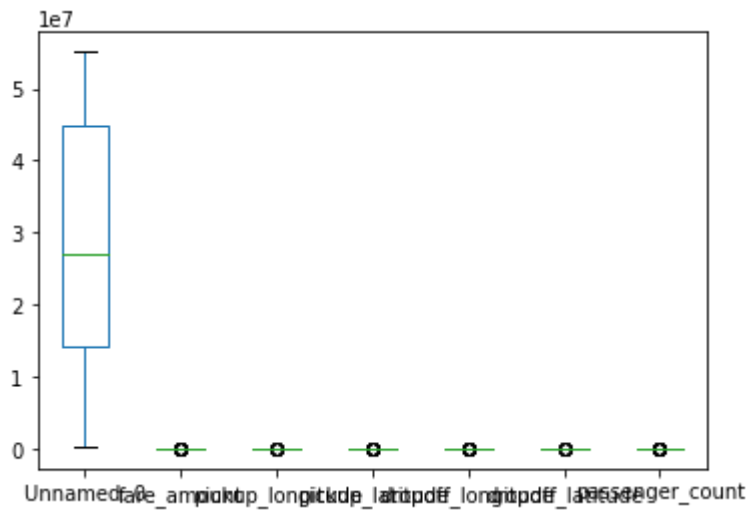
Out[16]:

<AxesSubplot:>

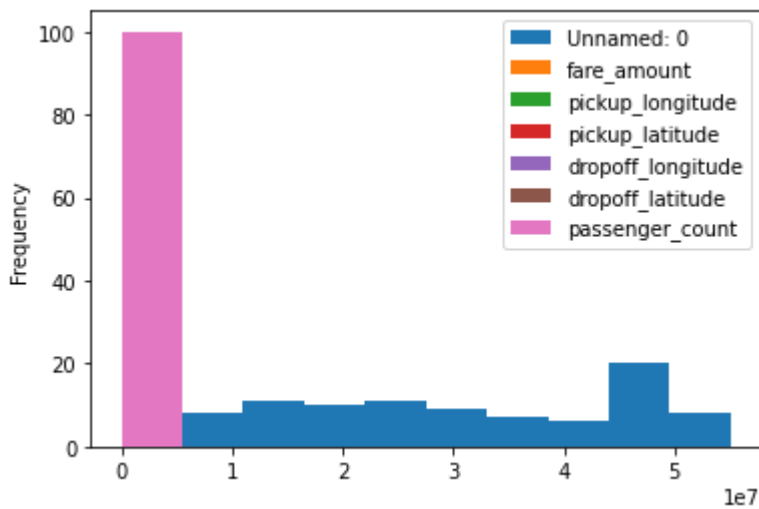In [17]:

```
c.plot.box()
```

Out[17]:

```
<AxesSubplot:>
```


In [18]:

```
c.plot.hist()
```

Out[18]:

```
<AxesSubplot:ylabel='Frequency'>
```
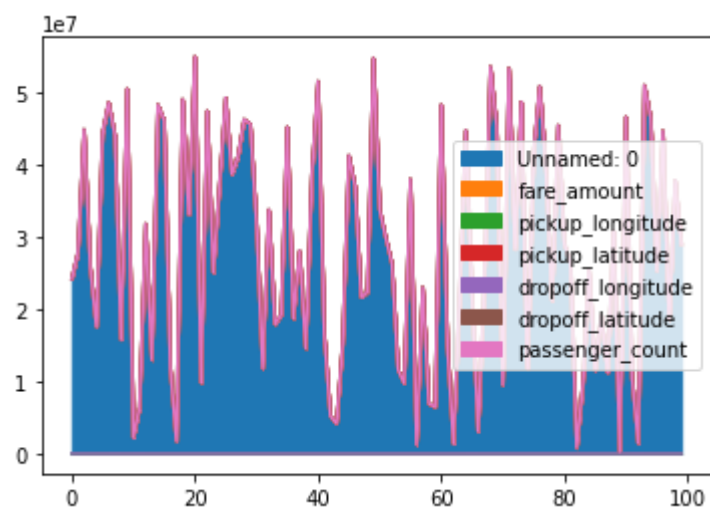
In [19]:

```
c.plot.area()
```

Out[19]:

`<AxesSubplot:>`

In [14]:

```python
c.plot.pie(y='pickup_latitude',figsize=(3,3))
```

Out[14]:

```
<AxesSubplot:ylabel='pickup_latitude'>
```

In [15]:

```python
c.plot.scatter(x='pickup_latitude',y='dropoff_latitude')
```

Out[15]:

```
<AxesSubplot:xlabel='pickup_latitude', ylabel='dropoff_latitude'>
```

In [ ]: