

1. IMPORTING LIBRARIES

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

2. Importing dataset

In [2]:

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

Out[2]:

		Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_l
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 08:22:21		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 01:09:00		7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837	-
199997	27804658	2009-06-29 00:42:00		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 04:08:00		14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077	-

200000 rows × 9 columns

3.head

In [3]:

data.head(8)

Out[3]:

		Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude
0	24238194	2015-05-07 19:52:06		7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	-73.9991
1	27835199	2009-07-17 20:04:56		7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	-73.9941
2	44984355	2009-08-24 21:45:00		12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	-73.9621
3	25894730	2009-06-26 08:22:21		5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	-73.9651
4	17610152	2014-08-28 17:47:00		16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	-73.9731
5	44470845	2011-02-12 02:27:09		4.9	2011-02-12 02:27:09 UTC	-73.969019	40.755910	-73.9691
6	48725865	2014-10-12 07:04:00		24.5	2014-10-12 07:04:00 UTC	-73.961447	40.693965	-73.8711
7	44195482	2012-12-11 13:52:00		2.5	2012-12-11 13:52:00 UTC	0.000000	0.000000	0.0000

4.tail

In [4]:

data.tail(7)

Out[4]:

		Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude
199993	28359558	2012-09-29 19:51:27		9.5	2012-09-29 19:51:27 UTC	-73.987798	40.721210	-73.9991
199994	3189201	2014-01-31		12.0	2014-01-31 14:42:00 UTC	-73.983070	40.760770	-73.9941

		Unnamed: 0	14:42:00 key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude
			2012-						
199995	42598914	10-28 10:49:00		3.0	2012-10-28 10:49:00 UTC		-73.987042	40.739367	-
199996	16382965	2014- 03-14 01:09:00		7.5	2014-03-14 01:09:00 UTC		-73.984722	40.736837	-
199997	27804658	2009- 06-29 00:42:00		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 04:08:00		14.1	2010-05-15 04:08:00 UTC		-73.984395	40.720077	-

5.describe()

In [5]:

```
data.describe()
```

Out[5]:

	Unnamed: 0	fare_amount	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude
count	2.000000e+05	200000.000000	200000.000000	200000.000000	199999.000000	199999.000000
mean	2.771250e+07	11.359955	-72.527638	39.935885	-72.525292	39.92381
std	1.601382e+07	9.901776	11.437787	7.720539	13.117408	6.79481
min	1.000000e+00	-52.000000	-1340.648410	-74.015515	-3356.666300	-881.9855
25%	1.382535e+07	6.000000	-73.992065	40.734796	-73.991407	40.73381
50%	2.774550e+07	8.500000	-73.981823	40.752592	-73.980093	40.75301
75%	4.155530e+07	12.500000	-73.967153	40.767158	-73.963659	40.76801
max	5.542357e+07	499.000000	57.418457	1644.421482	1153.572603	872.69761

6.shape()

In [6]:

```
np.shape(data)
```

Out[6]: (200000, 9)

7.size()

In [7]: `np.size(data)`

Out[7]: 1800000

8.isna()

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

Out[8]:

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

200000 rows × 9 columns

9.dropna

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

Out[9]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude
0	24238194	2015-05-07 19:52:06	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	-74.005043	40.740770
1	27835199	2009-07-17 20:04:56	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	-74.005043	40.740770
2	44984355	2009-08-24 21:45:00	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	-74.005043	40.740770

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude
3	25894730	2009-06-26 08:22:21	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	-73.987042	40.739367
4	17610152	2014-08-28 17:47:00	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	-73.984722	40.736837
...
199995	42598914	2012-10-28 10:49:00	3.0	2012-10-28 10:49:00 UTC	-73.987042	40.739367	-73.997124	40.725452
199996	16382965	2014-03-14 01:09:00	7.5	2014-03-14 01:09:00 UTC	-73.984722	40.736837	-73.984395	40.720077
199997	27804658	2009-06-29 00:42:00	30.9	2009-06-29 00:42:00 UTC	-73.986017	40.756487	-73.984395	40.720077
199998	20259894	2015-05-20 14:56:25	14.5	2015-05-20 14:56:25 UTC	-73.997124	40.725452	-73.984395	40.720077
199999	11951496	2010-05-15 04:08:00	14.1	2010-05-15 04:08:00 UTC	-73.984395	40.720077	-73.984395	40.720077

199999 rows × 9 columns

10.selecting specific column

In [10]:

```
da=data[["fare_amount","passenger_count"]]
da
```

Out[10]:

	fare_amount	passenger_count
0	7.5	1
1	7.7	1
2	12.9	1
3	5.3	3
4	16.0	5
...
199995	3.0	1
199996	7.5	1
199997	30.9	2

	fare_amount	passenger_count
199998	14.5	1
199999	14.1	1

200000 rows × 2 columns

In [11]:

```
dat=data.head(100)
dat
```

Out[11]:

	Unnamed: 0	key	fare_amount	pickup_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude
0	24238194	2015-05-07 19:52:06	7.5	2015-05-07 19:52:06 UTC	-73.999817	40.738354	-73.991000	40.738354
1	27835199	2009-07-17 20:04:56	7.7	2009-07-17 20:04:56 UTC	-73.994355	40.728225	-73.991000	40.728225
2	44984355	2009-08-24 21:45:00	12.9	2009-08-24 21:45:00 UTC	-74.005043	40.740770	-73.961000	40.740770
3	25894730	2009-06-26 08:22:21	5.3	2009-06-26 08:22:21 UTC	-73.976124	40.790844	-73.961000	40.790844
4	17610152	2014-08-28 17:47:00	16.0	2014-08-28 17:47:00 UTC	-73.925023	40.744085	-73.971000	40.744085
...
95	25431833	2015-04-11 08:47:47	9.5	2015-04-11 08:47:47 UTC	-73.978432	40.752399	-74.001000	40.752399
96	44792012	2011-10-03 20:29:00	4.5	2011-10-03 20:29:00 UTC	-73.990055	40.756413	-73.981000	40.756413
97	18571020	2010-04-26 03:12:44	3.3	2010-04-26 03:12:44 UTC	-73.982326	40.731314	-73.981000	40.731314
98	37942404	2011-11-18 09:51:00	30.9	2011-11-18 09:51:00 UTC	-73.995888	40.759078	-73.861000	40.759078
99	29024472	2009-08-30 14:03:55	26.9	2009-08-30 14:03:55 UTC	-73.990137	40.756007	-73.921000	40.756007

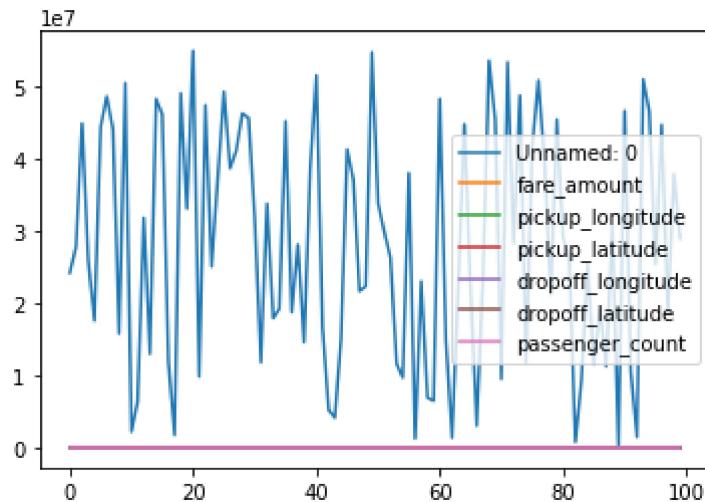
100 rows × 9 columns



11.line plot

```
In [12]: dat.plot.line()
```

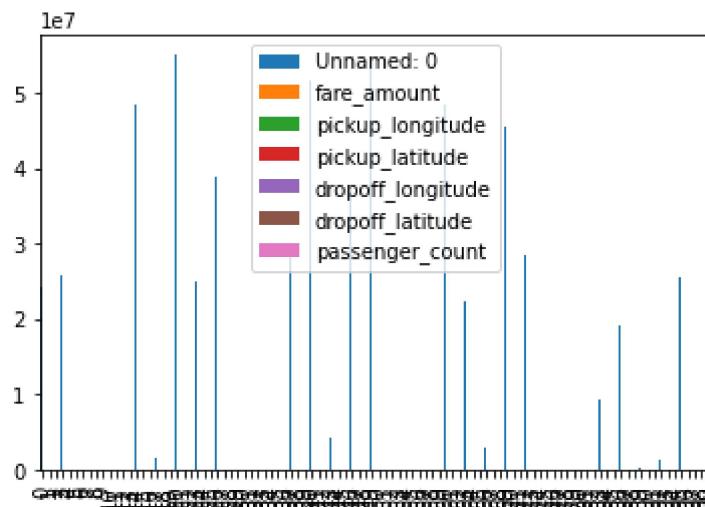
```
Out[12]: <AxesSubplot:>
```



12.bar plot

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

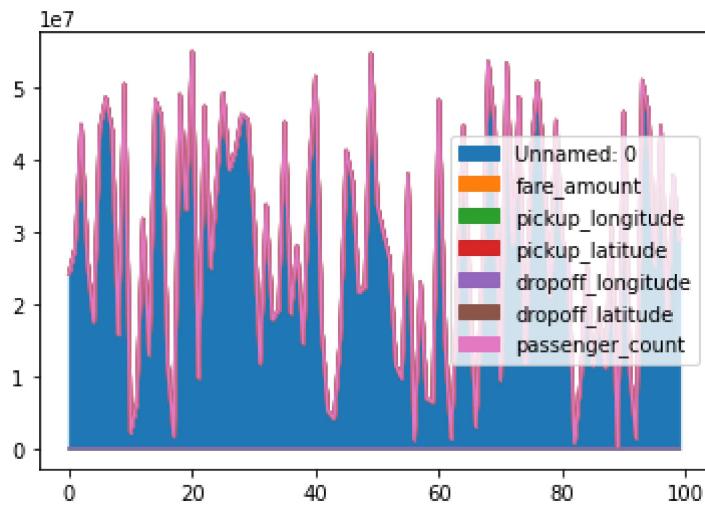
```
Out[13]: <AxesSubplot:>
```



13.area plot

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

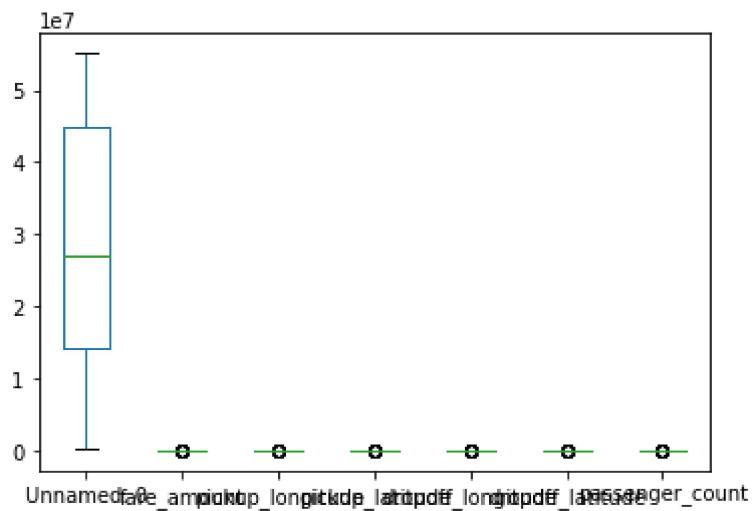
```
Out[14]: <AxesSubplot:>
```



14.box plot

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

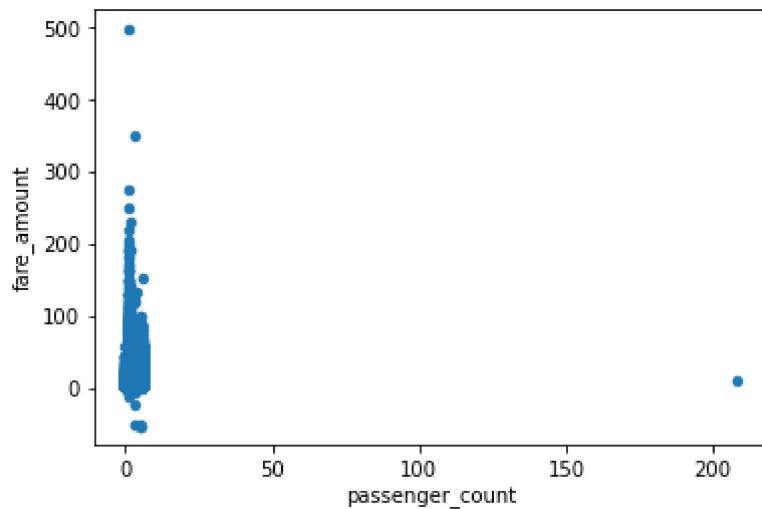
```
Out[15]: <AxesSubplot:>
```



15.scatter plot

```
In [16]: data.plot.scatter("passenger_count", "fare_amount")
```

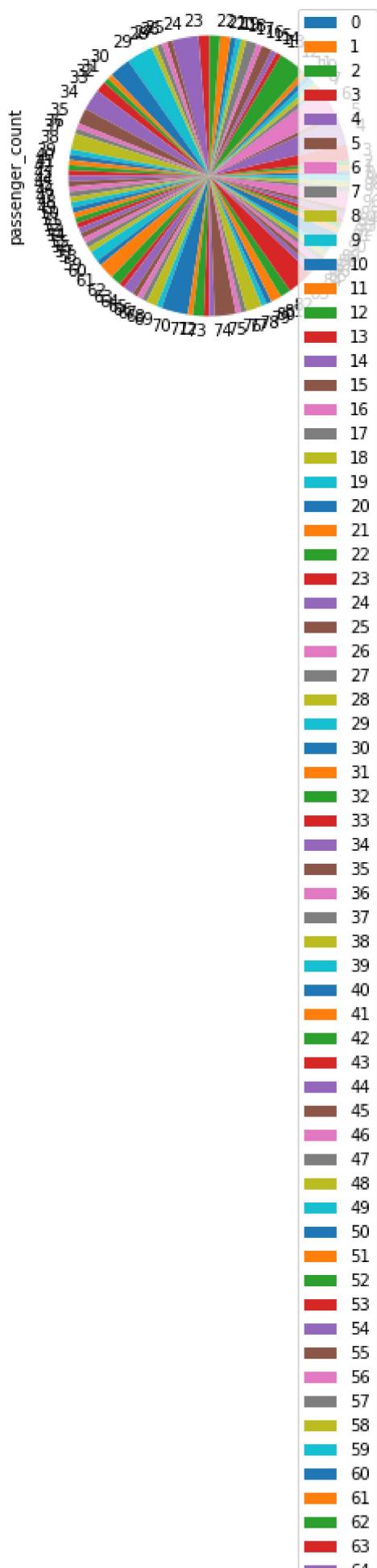
```
Out[16]: <AxesSubplot:xlabel='passenger_count', ylabel='fare_amount'>
```

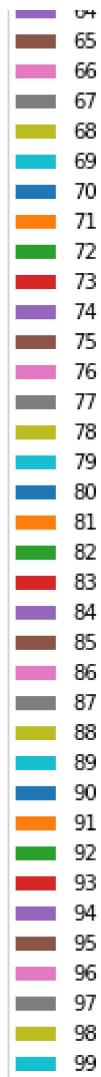


16.pie plot

```
In [17]: dat.plot.pie(y="passenger_count")
```

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



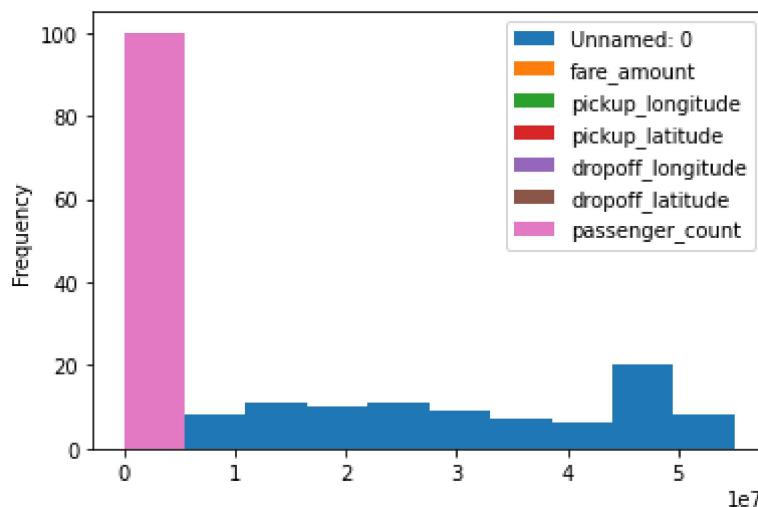


17.histogram

In [18]:

```
dat.plot.hist()
```

Out[18]: <AxesSubplot:ylabel='Frequency'>



In []: