**import** numpy **as** np

In [2]:

**import** pandas **as** pd

In [3]:

**import** matplotlib.pyplot **as** pp

**Import dataset**

In [4]:

data **=** pd**.**read\_csv(r"C:\Users\user\Desktop\7\_uber.csv")

**Display top 7 and last 6 rows and print the output**

In [5]:

data**.**head(7)

Out[5]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_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 | -73.999512 | 40.723217 | 1 |
| **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 | 1 |
| **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 | 1 |
| **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 | 3 |
| **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 |
| **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 | 1 |
| **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 | 5 |

In [6]:

data**.**tail(6)

Out[6]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_latitude** | **passenger\_count** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **199994** | 3189201 | 2014-01-31 14:42:00.000000181 | 12.0 | 2014-01-31 14:42:00 UTC | -73.983070 | 40.760770 | -73.972972 | 40.754177 | 1 |
| **199995** | 42598914 | 2012-10-28 10:49:00.00000053 | 3.0 | 2012-10-28 10:49:00 UTC | -73.987042 | 40.739367 | -73.986525 | 40.740297 | 1 |
| **199996** | 16382965 | 2014-03-14 01:09:00.0000008 | 7.5 | 2014-03-14 01:09:00 UTC | -73.984722 | 40.736837 | -74.006672 | 40.739620 | 1 |
| **199997** | 27804658 | 2009-06-29 00:42:00.00000078 | 30.9 | 2009-06-29 00:42:00 UTC | -73.986017 | 40.756487 | -73.858957 | 40.692588 | 2 |
| **199998** | 20259894 | 2015-05-20 14:56:25.0000004 | 14.5 | 2015-05-20 14:56:25 UTC | -73.997124 | 40.725452 | -73.983215 | 40.695415 | 1 |
| **199999** | 11951496 | 2010-05-15 04:08:00.00000076 | 14.1 | 2010-05-15 04:08:00 UTC | -73.984395 | 40.720077 | -73.985508 | 40.768793 | 1 |

**Fill with a constant value and print the output**

In [7]:

data**.**fillna("2010")

Out[7]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_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 | -73.999512 | 40.723217 | 1 |
| **1** | 27835199 | 2009-07-17 20:04:56.0000002 | 7.7 | 2009-07-17 20:04:56 UTC | -73.994355 | 40.728225 | -73.99471 | 40.750325 | 1 |
| **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 | 1 |
| **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 | 3 |
| **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 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **199995** | 42598914 | 2012-10-28 10:49:00.00000053 | 3.0 | 2012-10-28 10:49:00 UTC | -73.987042 | 40.739367 | -73.986525 | 40.740297 | 1 |
| **199996** | 16382965 | 2014-03-14 01:09:00.0000008 | 7.5 | 2014-03-14 01:09:00 UTC | -73.984722 | 40.736837 | -74.006672 | 40.73962 | 1 |
| **199997** | 27804658 | 2009-06-29 00:42:00.00000078 | 30.9 | 2009-06-29 00:42:00 UTC | -73.986017 | 40.756487 | -73.858957 | 40.692588 | 2 |
| **199998** | 20259894 | 2015-05-20 14:56:25.0000004 | 14.5 | 2015-05-20 14:56:25 UTC | -73.997124 | 40.725452 | -73.983215 | 40.695415 | 1 |
| **199999** | 11951496 | 2010-05-15 04:08:00.00000076 | 14.1 | 2010-05-15 04:08:00 UTC | -73.984395 | 40.720077 | -73.985508 | 40.768793 | 1 |

200000 rows × 9 columns

**Drop the column with missing values and print the output**

In [8]:

data**.**dropna(axis **=** 1)

Out[8]:

|  | **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 [9]:

data**.**dropna()

Out[9]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_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 | -73.999512 | 40.723217 | 1 |
| **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 | 1 |
| **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 | 1 |
| **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 | 3 |
| **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 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **199995** | 42598914 | 2012-10-28 10:49:00.00000053 | 3.0 | 2012-10-28 10:49:00 UTC | -73.987042 | 40.739367 | -73.986525 | 40.740297 | 1 |
| **199996** | 16382965 | 2014-03-14 01:09:00.0000008 | 7.5 | 2014-03-14 01:09:00 UTC | -73.984722 | 40.736837 | -74.006672 | 40.739620 | 1 |
| **199997** | 27804658 | 2009-06-29 00:42:00.00000078 | 30.9 | 2009-06-29 00:42:00 UTC | -73.986017 | 40.756487 | -73.858957 | 40.692588 | 2 |
| **199998** | 20259894 | 2015-05-20 14:56:25.0000004 | 14.5 | 2015-05-20 14:56:25 UTC | -73.997124 | 40.725452 | -73.983215 | 40.695415 | 1 |
| **199999** | 11951496 | 2010-05-15 04:08:00.00000076 | 14.1 | 2010-05-15 04:08:00 UTC | -73.984395 | 40.720077 | -73.985508 | 40.768793 | 1 |

199999 rows × 9 columns

**To check the presence of missing values in your dataframe**

In [10]:

data**.**isna()

Out[10]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_latitude** | **passenger\_count** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | False | False | False | False | False | False | False | False | False |
| **1** | False | False | False | False | False | False | False | False | False |
| **2** | False | False | False | False | False | False | False | False | False |
| **3** | False | False | False | False | False | False | False | False | False |
| **4** | False | False | False | False | False | False | False | False | False |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **199995** | False | False | False | False | False | False | False | False | False |
| **199996** | False | False | False | False | False | False | False | False | False |
| **199997** | False | False | False | False | False | False | False | False | False |
| **199998** | False | False | False | False | False | False | False | False | False |
| **199999** | False | False | False | False | False | False | False | False | False |

200000 rows × 9 columns

**Use operators and check the condition and print the output**

In [11]:

data[data['fare\_amount']**>**15]

Out[11]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_latitude** | **passenger\_count** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **4** | 17610152 | 2014-08-28 17:47:00.000000188 | 16.00 | 2014-08-28 17:47:00 UTC | -73.925023 | 40.744085 | -73.973082 | 40.761247 | 5 |
| **6** | 48725865 | 2014-10-12 07:04:00.0000002 | 24.50 | 2014-10-12 07:04:00 UTC | -73.961447 | 40.693965 | -73.871195 | 40.774297 | 5 |
| **30** | 31945670 | 2011-05-21 09:00:00.00000031 | 25.70 | 2011-05-21 09:00:00 UTC | -73.944815 | 40.834367 | -73.989332 | 40.721920 | 4 |
| **34** | 19277743 | 2014-06-04 06:49:00.000000102 | 39.50 | 2014-06-04 06:49:00 UTC | -73.788080 | 40.642187 | -73.865042 | 40.725997 | 4 |
| **39** | 38703737 | 2014-02-13 17:57:00.000000102 | 29.00 | 2014-02-13 17:57:00 UTC | -73.992600 | 40.753172 | -73.908508 | 40.816192 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **199977** | 21117828 | 2012-11-20 21:04:30.0000001 | 43.50 | 2012-11-20 21:04:30 UTC | -73.996671 | 40.737483 | -73.867758 | 40.897563 | 1 |
| **199982** | 13096190 | 2014-08-06 11:06:06.0000001 | 57.33 | 2014-08-06 11:06:06 UTC | -73.969204 | 40.754771 | -73.790351 | 40.643802 | 1 |
| **199985** | 25830754 | 2015-04-18 15:16:06.0000005 | 24.00 | 2015-04-18 15:16:06 UTC | -74.005089 | 40.737301 | -73.945290 | 40.774162 | 5 |
| **199991** | 13512837 | 2015-06-08 10:49:14.0000001 | 17.50 | 2015-06-08 10:49:14 UTC | -73.981453 | 40.743919 | -74.013908 | 40.712635 | 1 |
| **199997** | 27804658 | 2009-06-29 00:42:00.00000078 | 30.90 | 2009-06-29 00:42:00 UTC | -73.986017 | 40.756487 | -73.858957 | 40.692588 | 2 |

35251 rows × 9 columns

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

In [12]:

data**.**iloc[300:500]

Out[12]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_latitude** | **passenger\_count** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **300** | 29569353 | 2009-08-14 17:30:14.0000002 | 9.3 | 2009-08-14 17:30:14 UTC | -73.987011 | 40.776157 | -73.994649 | 40.750440 | 1 |
| **301** | 20159924 | 2012-12-31 23:57:56.0000001 | 15.0 | 2012-12-31 23:57:56 UTC | -73.976850 | 40.764368 | -73.911379 | 40.747178 | 1 |
| **302** | 15291924 | 2012-04-05 13:25:36.0000002 | 6.5 | 2012-04-05 13:25:36 UTC | -73.970823 | 40.798273 | -73.955982 | 40.787948 | 1 |
| **303** | 1443240 | 2013-05-21 00:17:00.00000045 | 5.0 | 2013-05-21 00:17:00 UTC | -73.973030 | 40.752042 | -73.977175 | 40.745387 | 5 |
| **304** | 3286192 | 2013-06-18 16:19:00.00000018 | 13.5 | 2013-06-18 16:19:00 UTC | -73.973317 | 40.750867 | -73.991122 | 40.747067 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **495** | 1204312 | 2012-06-03 12:18:02.0000001 | 25.7 | 2012-06-03 12:18:02 UTC | -73.862765 | 40.770908 | -73.989013 | 40.688776 | 1 |
| **496** | 2511529 | 2014-12-24 05:54:45.0000001 | 8.0 | 2014-12-24 05:54:45 UTC | -73.918530 | 40.743330 | -73.946696 | 40.749438 | 1 |
| **497** | 24116460 | 2010-01-18 02:18:16.0000001 | 10.5 | 2010-01-18 02:18:16 UTC | -74.005734 | 40.743641 | -74.006287 | 40.708330 | 2 |
| **498** | 42607669 | 2015-03-30 10:58:37.0000001 | 5.5 | 2015-03-30 10:58:37 UTC | -74.001648 | 40.740940 | -74.005730 | 40.750175 | 1 |
| **499** | 36533403 | 2015-03-09 16:16:21.0000006 | 10.0 | 2015-03-09 16:16:21 UTC | -73.960037 | 40.780624 | -73.971756 | 40.765934 | 1 |

200 rows × 9 columns

In [13]:

data**.**loc[100:200]

Out[13]:

|  | **Unnamed: 0** | **key** | **fare\_amount** | **pickup\_datetime** | **pickup\_longitude** | **pickup\_latitude** | **dropoff\_longitude** | **dropoff\_latitude** | **passenger\_count** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **100** | 29350780 | 2014-02-19 18:02:00.00000011 | 9.0 | 2014-02-19 18:02:00 UTC | -73.958280 | 40.768900 | -73.973510 | 40.782907 | 5 |
| **101** | 30977645 | 2011-02-20 21:01:58.0000003 | 4.1 | 2011-02-20 21:01:58 UTC | -73.977633 | 40.753166 | -73.981589 | 40.757359 | 1 |
| **102** | 7290820 | 2011-02-04 11:48:00.000000216 | 14.9 | 2011-02-04 11:48:00 UTC | -73.962218 | 40.640218 | -73.968043 | 40.640505 | 1 |
| **103** | 47729464 | 2009-07-31 09:29:00.000000167 | 6.5 | 2009-07-31 09:29:00 UTC | -74.000545 | 40.737462 | -74.009545 | 40.713217 | 5 |
| **104** | 46435788 | 2015-05-15 18:58:16.0000001 | 43.0 | 2015-05-15 18:58:16 UTC | -73.862701 | 40.768959 | -73.999092 | 40.741829 | 2 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **196** | 51452192 | 2009-05-12 10:32:00.000000154 | 24.0 | 2009-05-12 10:32:00 UTC | -73.981558 | 40.783752 | -73.900931 | 40.893366 | 1 |
| **197** | 45317989 | 2012-08-07 20:53:18.0000001 | 10.5 | 2012-08-07 20:53:18 UTC | -73.965930 | 40.805358 | -73.949923 | 40.778868 | 1 |
| **198** | 41858701 | 2009-09-24 16:21:42.0000001 | 8.9 | 2009-09-24 16:21:42 UTC | -73.952080 | 40.790119 | -73.963637 | 40.798090 | 1 |
| **199** | 13472186 | 2011-04-03 00:01:40.0000002 | 14.1 | 2011-04-03 00:01:40 UTC | -74.000190 | 40.718336 | -73.956801 | 40.764745 | 1 |
| **200** | 12508194 | 2012-03-22 07:14:24.0000002 | 6.1 | 2012-03-22 07:14:24 UTC | -73.965017 | 40.764308 | -73.981057 | 40.760887 | 1 |

101 rows × 9 columns

**Display the statistical summary of data**

In [14]:

data**.**describe()

Out[14]:

|  | **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**

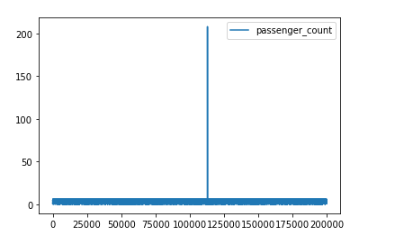
In [19]:

df**=**data[['key','passenger\_count']]

df**.**plot**.**line()

Out[19]:

<AxesSubplot:>

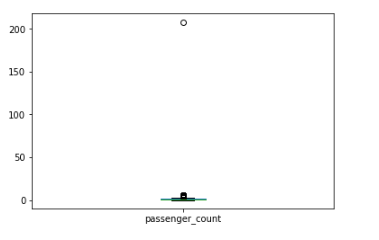


In [16]:

df**.**plot**.**box()

Out[16]:

<AxesSubplot:>

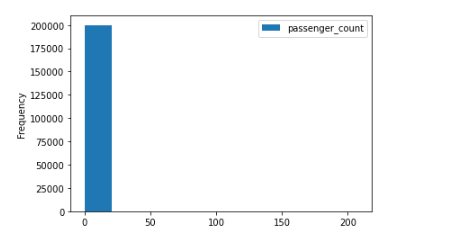


In [20]:

df**.**plot**.**hist()

Out[20]:

<AxesSubplot:ylabel='Frequency'>



In [21]:

df**.**plot**.**area()

Out[21]:

<AxesSubplot:>

In [ ]:

