Crime_outlook

August 20, 2018

```
In [4]: # Import all libraries needed
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        from sklearn import linear_model


        # Enable inline plotting
        %matplotlib inline
```

```
In [12]: crime = pd.read_excel('table_8_offenses_known_to_law_enforcement_new_york_by_city_2013.
```

```
In [13]: crime
```

```
Out[13]:                          City  Population  Violent\ncrime  \
         0                Adams Village        1861               0
         1      Addison Town and Village        2577               3
         2                Akron Village        2846               3
         3                       Albany       97956             791
         4               Albion Village        6388              23
         5               Alfred Village        4089               5
         6             Allegany Village        1781               3
         7                 Amherst Town      118296             107
         8           Amityville Village        9519               9
         9                    Amsterdam       18182              30
         10              Arcade Village        2053               0
         11             Ardsley Village        4523               5
         12           Asharoken Village         658               0
         13              Attica Village        2522               2
         14                      Auburn       27270              96
         15                Avon Village        3333               1
         16        Baldwinsville Village        7473               5
         17        Ballston Spa Village        5418               8
         18                     Batavia       15374              57
         19                Bath Village        5733              15
         20                 Bedford Town       17627               5
         21               Bethlehem Town       34243              13
```

1

22	Binghamton	46304	284
23	Black River	1410	0
24	Blooming Grove Town	12001	8
25	Bolivar Village	1022	1
26	Bolton Town	2325	2
27	Boonville Village	2054	1
28	Brant Town	2065	4
29	Brewster	2358	0
..	...	...	...
318	Tupper Lake Village	3642	3
319	Tuxedo Park Village	615	0
320	Ulster Town	12195	14
321	Utica	61686	361
322	Vestal Town	27980	16
323	Walden Village	6887	18
324	Wallkill Town	27940	40
325	Walton Village	3012	6
326	Warsaw Village	3453	6
327	Washingtonville Village	5828	2
328	Waterford Town and Village	8392	4
329	Waterloo Village	5175	11
330	Watertown	28179	109
331	Watervliet	10245	26
332	Watkins Glen Village	1862	6
333	Waverly Village	4337	5
334	Webster Town and Village	43777	21
335	Weedsport Village	1797	0
336	Wellsville Village	4574	11
337	Westhampton Beach Village	1732	2
338	West Seneca Town	44821	35
339	Whitehall Village	2604	6
340	White Plains	57559	78
341	Whitesboro Village	3733	4
342	Whitestown Town	9141	0
343	Woodbury Town	10685	3
344	Woodridge Village	829	7
345	Woodstock Town	5931	2
346	Yonkers	199134	1036
347	Yorktown Town	36643	15

	Murder and\nnonnegligent\nmanslaughter	Rape\n(revised\ndefinition)1 \
0	0	NaN
1	0	NaN
2	0	NaN
3	8	NaN
4	0	NaN
5	0	NaN
6	0	NaN

7	1	NaN
8	0	NaN
9	0	NaN
10	0	NaN
11	0	NaN
12	0	NaN
13	0	NaN
14	1	NaN
15	0	NaN
16	0	NaN
17	0	NaN
18	0	NaN
19	0	NaN
20	0	NaN
21	0	NaN
22	3	NaN
23	0	NaN
24	0	NaN
25	0	NaN
26	0	NaN
27	0	NaN
28	0	NaN
29	0	NaN
..	...	...
318	0	NaN
319	0	NaN
320	0	NaN
321	7	NaN
322	0	NaN
323	1	NaN
324	0	NaN
325	0	NaN
326	0	NaN
327	0	NaN
328	0	NaN
329	0	NaN
330	0	NaN
331	1	NaN
332	0	NaN
333	0	NaN
334	0	NaN
335	0	NaN
336	0	NaN
337	0	NaN
338	1	NaN
339	0	NaN
340	1	NaN
341	0	NaN

342	0	NaN
343	0	NaN
344	0	NaN
345	0	NaN
346	6	NaN
347	0	NaN

	Rape\n(legacy\ndefinition)2	Robbery	Aggravated\nassault \
0	0	0	0
1	0	0	3
2	0	0	3
3	30	227	526
4	3	4	16
5	0	3	2
6	0	0	3
7	7	31	68
8	2	4	3
9	0	12	18
10	0	0	0
11	0	3	2
12	0	0	0
13	0	0	2
14	20	22	53
15	0	1	0
16	0	1	4
17	0	3	5
18	7	13	37
19	1	3	11
20	0	1	4
21	0	3	10
22	19	101	161
23	0	0	0
24	1	0	7
25	0	0	1
26	0	0	2
27	0	0	1
28	0	1	3
29	0	0	0
..	...	...	...
318	0	0	3
319	0	0	0
320	3	2	9
321	27	102	225
322	2	6	8
323	2	1	14
324	3	16	21
325	1	0	5
326	1	0	5

327	0	1	1
328	0	0	4
329	0	3	8
330	12	16	81
331	3	11	11
332	0	0	6
333	0	1	4
334	2	3	16
335	0	0	0
336	1	0	10
337	0	0	2
338	1	9	24
339	0	0	6
340	3	28	46
341	0	0	4
342	0	0	0
343	0	2	1
344	0	0	7
345	0	0	2
346	25	390	615
347	0	2	13

	Property\ncrime	Burglary	Larceny-\ntheft	Motor\nvehicle\ntheft	Arson3
0	12	2	10	0	0.0
1	24	3	20	1	0.0
2	16	1	15	0	0.0
3	4090	705	3243	142	NaN
4	223	53	165	5	NaN
5	46	10	36	0	NaN
6	10	0	10	0	0.0
7	2118	204	1882	32	3.0
8	210	16	188	6	1.0
9	405	99	291	15	0.0
10	39	3	35	1	0.0
11	33	4	28	1	0.0
12	2	0	2	0	0.0
13	5	0	4	1	0.0
14	859	132	721	6	NaN
15	25	3	20	2	NaN
16	88	10	77	1	0.0
17	94	16	78	0	1.0
18	559	103	454	2	NaN
19	191	32	157	2	2.0
20	156	26	127	3	NaN
21	442	50	388	4	NaN
22	2349	525	1767	57	NaN
23	14	2	12	0	0.0
24	99	20	72	7	NaN

25	4	2	2	0	0.0
26	24	4	20	0	0.0
27	20	6	14	0	0.0
28	31	9	20	2	0.0
29	13	3	10	0	0.0
..	...	...	...	...	...
318	83	22	59	2	NaN
319	0	0	0	0	0.0
320	437	25	410	2	NaN
321	2528	449	1997	82	NaN
322	571	39	523	9	NaN
323	142	13	124	5	NaN
324	767	91	650	26	1.0
325	68	18	50	0	NaN
326	78	5	73	0	0.0
327	58	5	51	2	0.0
328	59	6	52	1	0.0
329	194	23	169	2	NaN
330	1165	169	953	43	NaN
331	242	58	163	21	NaN
332	92	10	80	2	NaN
333	54	9	41	4	NaN
334	469	101	361	7	NaN
335	25	5	20	0	0.0
336	164	32	129	3	1.0
337	44	4	37	3	0.0
338	896	166	710	20	0.0
339	62	12	49	1	NaN
340	1232	77	1134	21	NaN
341	36	14	21	1	0.0
342	77	13	61	3	0.0
343	541	9	529	3	NaN
344	17	8	9	0	0.0
345	58	13	45	0	NaN
346	2368	470	1662	236	10.0
347	334	45	287	2	NaN

[348 rows x 13 columns]

In [17]: crime.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 348 entries, 0 to 347
Data columns (total 13 columns):
City   348 non-null object
Population   348 non-null int64
Violent
crime   348 non-null int64

6

```
Murder and
nonnegligent
manslaughter     348 non-null int64
Rape
(revised
definition)1               0 non-null float64
Rape
(legacy
definition)2            348 non-null int64
Robbery                            348 non-null int64
Aggravated
assault                   348 non-null int64
Property
crime                      348 non-null int64
Burglary                          348 non-null int64
Larceny-
theft                      348 non-null int64
Motor
vehicle
theft                   348 non-null int64
Arson3                             187 non-null float64
dtypes: float64(2), int64(10), object(1)
memory usage: 35.4+ KB
```

```
In [25]: crime['Arson3'].fillna(value=0, inplace=True)
         #df[['a', 'b']] = df[['a','b']].fillna(value=0)
         crime.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 348 entries, 0 to 347
Data columns (total 13 columns):
City                               348 non-null object
Population                         348 non-null int64
Violent
crime                      348 non-null int64
Murder and
nonnegligent
manslaughter     348 non-null int64
Rape
(revised
definition)1               0 non-null float64
Rape
(legacy
definition)2            348 non-null int64
Robbery                            348 non-null int64
Aggravated
assault                   348 non-null int64
```

7

```
Property
crime                            348 non-null int64
Burglary                         348 non-null int64
Larceny-
theft                            348 non-null int64
Motor
vehicle
theft                            348 non-null int64
Arson3                           348 non-null float64
dtypes: float64(2), int64(10), object(1)
memory usage: 35.4+ KB
```

In [26]: crime.head(10)

Out[26]:   City Population Violent\ncrime \
        0   Adams Village   1861   0
        1   Addison Town and Village   2577   3
        2   Akron Village   2846   3
        3   Albany   97956   791
        4   Albion Village   6388   23
        5   Alfred Village   4089   5
        6   Allegany Village   1781   3
        7   Amherst Town   118296   107
        8   Amityville Village   9519   9
        9   Amsterdam   18182   30

          Murder and\nnonnegligent\nmanslaughter Rape\n(revised\ndefinition)1 \
        0   0   NaN
        1   0   NaN
        2   0   NaN
        3   8   NaN
        4   0   NaN
        5   0   NaN
        6   0   NaN
        7   1   NaN
        8   0   NaN
        9   0   NaN

          Rape\n(legacy\ndefinition)2 Robbery Aggravated\nassault Property\ncrime \
        0   0   0   0   12
        1   0   0   3   24
        2   0   0   3   16
        3   30   227   526   4090
        4   3   4   16   223
        5   0   3   2   46
        6   0   0   3   10
        7   7   31   68   2118
```
                                    8
```

```
8                               2      4           3      210
9                               0     12          18      405

   Burglary  Larceny-\ntheft  Motor\nvehicle\ntheft  Arson3
0         2              10                       0     0.0
1         3              20                       1     0.0
2         1              15                       0     0.0
3       705            3243                     142     0.0
4        53             165                       5     0.0
5        10              36                       0     0.0
6         0              10                       0     0.0
7       204            1882                      32     3.0
8        16             188                       6     1.0
9        99             291                      15     0.0
```

In [27]: `crime['Population_sq'] = crime['Population'] ** 2`

In [30]: `crime.head(10)`

```
Out[30]:                        City  Population  Violent\ncrime  \
0                 Adams Village        1861               0
1       Addison Town and Village      2577               3
2                 Akron Village        2846               3
3                        Albany       97956             791
4                 Albion Village       6388              23
5                 Alfred Village       4089               5
6               Allegany Village      1781               3
7                   Amherst Town     118296             107
8             Amityville Village      9519               9
9                     Amsterdam      18182              30

   Murder and\nnonnegligent\nmanslaughter  Rape\n(revised\ndefinition)1  \
0                                        0                          NaN
1                                        0                          NaN
2                                        0                          NaN
3                                        8                          NaN
4                                        0                          NaN
5                                        0                          NaN
6                                        0                          NaN
7                                        1                          NaN
8                                        0                          NaN
9                                        0                          NaN

   Rape\n(legacy\ndefinition)2  Robbery  Aggravated\nassault  Property\ncrime  \
0                            0        0                    0              12
1                            0        0                    3              24
2                            0        0                    3              16
3                           30      227                  526            4090
```

```
4                            3         4                    16          223
5                            0         3                     2           46
6                            0         0                     3           10
7                            7        31                    68         2118
8                            2         4                     3          210
9                            0        12                    18          405

      Burglary  Larceny-\ntheft  Motor\nvehicle\ntheft  Arson3  Population_sq
0            2               10                       0     0.0        3463321
1            3               20                       1     0.0        6640929
2            1               15                       0     0.0        8099716
3          705             3243                     142     0.0     9595377936
4           53              165                       5     0.0       40806544
5           10               36                       0     0.0       16719921
6            0               10                       0     0.0        3171961
7          204             1882                      32     3.0    13993943616
8           16              188                       6     1.0       90611361
9           99              291                      15     0.0      330585124
```
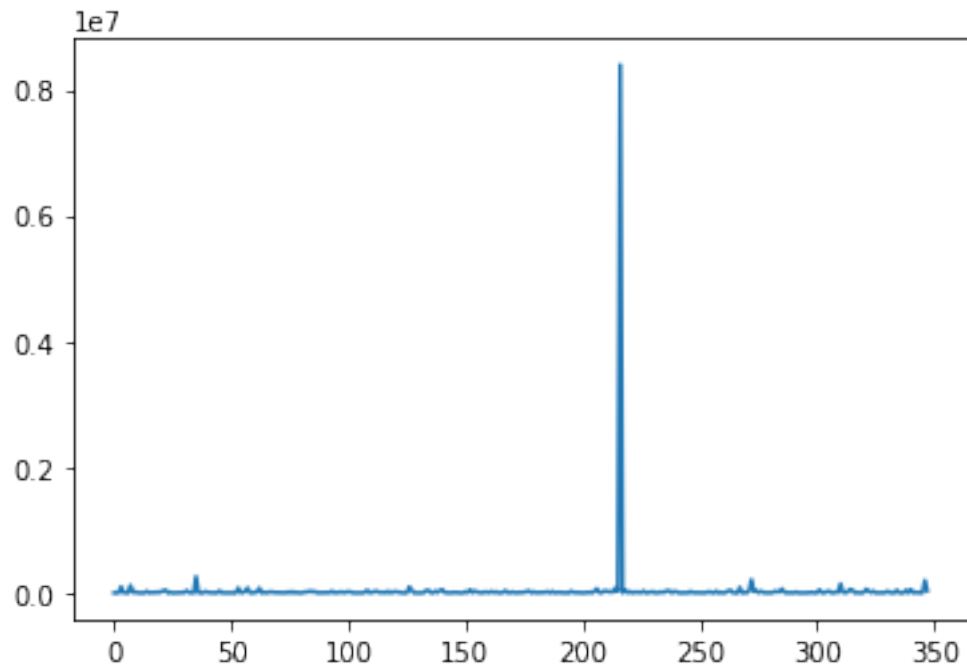
In [37]: `crime.columns`

Out[37]: Index(['City', 'Population', 'Violent\ncrime',
       'Murder and\nnonnegligent\nmanslaughter',
       'Rape\n(revised\ndefinition)1', 'Rape\n(legacy\ndefinition)2',
       'Robbery', 'Aggravated\nassault', 'Property\ncrime', 'Burglary',
       'Larceny-\ntheft', 'Motor\nvehicle\ntheft', 'Arson3', 'Population_sq'],
      dtype='object')

In [44]: ```
for row in crime:
if crime['Murder and\nnonnegligent\nmanslaughter'] == 0:
crime['Murder_cat'] = 0
else:
crime['Murder_cat'] = 1

crime['Murder_cat'] = crime['Murder and\nnonnegligent\nmanslaughter'].where(crime['Murd
crime['Robbery_cat'] = crime['Robbery'].where(crime['Robbery']==0,other=1)
```

In [45]: `crime.head(10)`

Out[45]:
```
                          City Population Violent\ncrime \
0   Adams Village   1861   0
1   Addison Town and Village   2577   3
2   Akron Village   2846   3
3   Albany   97956   791
4   Albion Village   6388   23
5   Alfred Village   4089   5
6   Allegany Village   1781   3
7   Amherst Town   118296   107
8   Amityville Village   9519   9
```

10

| | | | |
|---|---|---|---|
| 9 | Amsterdam | 18182 | 30 |

| | Murder and\nnonnegligent\nmanslaughter | Rape\n(revised\ndefinition)1 |
|---|---|---|
| 0 | 0 | NaN |
| 1 | 0 | NaN |
| 2 | 0 | NaN |
| 3 | 8 | NaN |
| 4 | 0 | NaN |
| 5 | 0 | NaN |
| 6 | 0 | NaN |
| 7 | 1 | NaN |
| 8 | 0 | NaN |
| 9 | 0 | NaN |

| | Rape\n(legacy\ndefinition)2 | Robbery | Aggravated\nassault | Property\ncrime |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 12 |
| 1 | 0 | 0 | 3 | 24 |
| 2 | 0 | 0 | 3 | 16 |
| 3 | 30 | 227 | 526 | 4090 |
| 4 | 3 | 4 | 16 | 223 |
| 5 | 0 | 3 | 2 | 46 |
| 6 | 0 | 0 | 3 | 10 |
| 7 | 7 | 31 | 68 | 2118 |
| 8 | 2 | 4 | 3 | 210 |
| 9 | 0 | 12 | 18 | 405 |

| | Burglary | Larceny-\ntheft | Motor\nvehicle\ntheft | Arson3 | Population_sq |
|---|---|---|---|---|---|
| 0 | 2 | 10 | 0 | 0.0 | 3463321 |
| 1 | 3 | 20 | 1 | 0.0 | 6640929 |
| 2 | 1 | 15 | 0 | 0.0 | 8099716 |
| 3 | 705 | 3243 | 142 | 0.0 | 9595377936 |
| 4 | 53 | 165 | 5 | 0.0 | 40806544 |
| 5 | 10 | 36 | 0 | 0.0 | 16719921 |
| 6 | 0 | 10 | 0 | 0.0 | 3171961 |
| 7 | 204 | 1882 | 32 | 3.0 | 13993943616 |
| 8 | 16 | 188 | 6 | 1.0 | 90611361 |
| 9 | 99 | 291 | 15 | 0.0 | 330585124 |

| | Murder_cat | Robbery_cat |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 2 | 0 | 0 |
| 3 | 1 | 1 |
| 4 | 0 | 1 |
| 5 | 0 | 1 |
| 6 | 0 | 0 |
| 7 | 1 | 1 |
| 8 | 0 | 1 |

```
9                0                1
```

In [51]: plt.plot(crime['Population'])

Out[51]: [<matplotlib.lines.Line2D at 0x10e802d30>]

In [50]: crime[210:220]

Out[50]:

	City	Population	Violent\ncrime \
210	Newburgh Town	30984	38
211	New Castle Town	17864	2
212	New Hartford Town and Village	20355	13
213	New Paltz Town and Village	14248	34
214	New Rochelle	78800	175
215	New Windsor Town	25767	31
216	New York	8396126	52384
217	New York Mills Village	3324	2
218	Niagara Falls	49574	584
219	Niagara Town	8241	13

	Murder and\nnonnegligent\nmanslaughter	Rape\n(revised\ndefinition)1 \
210	0	NaN
211	0	NaN
212	0	NaN
213	0	NaN
214	0	NaN

215		0	NaN
216		335	NaN
217		0	NaN
218		3	NaN
219		0	NaN

	Rape\n(legacy\ndefinition)2	Robbery	Aggravated\nassault	\
210	5	15	18	
211	0	0	2	
212	1	3	9	
213	4	4	26	
214	5	81	89	
215	1	8	22	
216	1112	19170	31767	
217	0	0	2	
218	12	166	403	
219	0	2	11	

	Property\ncrime	Burglary	Larceny-\ntheft	Motor\nvehicle\ntheft	\
210	1353	143	1182	28	
211	77	13	62	2	
212	806	72	730	4	
213	200	16	180	4	
214	1391	150	1172	69	
215	471	65	394	12	
216	141971	16606	117931	7434	
217	64	16	47	1	
218	2807	746	1949	112	
219	489	48	427	14	

	Arson3	Population_sq	Murder_cat	Robbery_cat
210	2.0	960008256	0	1
211	0.0	319122496	0	0
212	0.0	414326025	0	1
213	0.0	203005504	0	1
214	0.0	6209440000	0	1
215	4.0	663938289	0	1
216	0.0	70494931807876	1	1
217	0.0	11048976	0	0
218	22.0	2457581476	1	1
219	0.0	67914081	0	1

In [52]: crime.columns

Out[52]: Index(['City', 'Population', 'Violent\ncrime',
        'Murder and\nnonnegligent\nmanslaughter',
        'Rape\n(revised\ndefinition)1', 'Rape\n(legacy\ndefinition)2',
        'Robbery', 'Aggravated\nassault', 'Property\ncrime', 'Burglary',

```
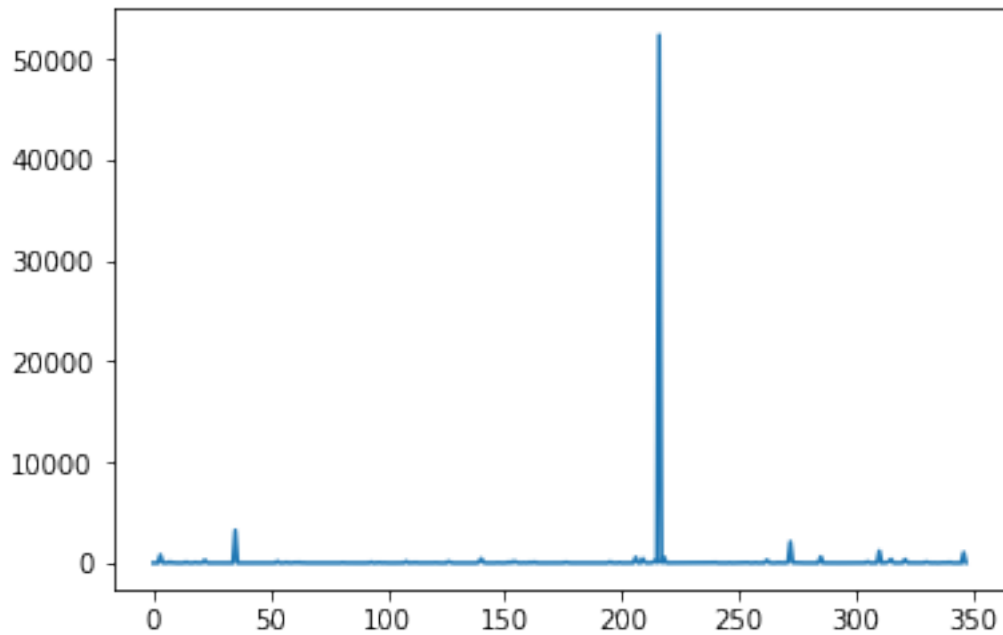                'Larceny-\ntheft', 'Motor\nvehicle\ntheft', 'Arson3', 'Population_sq',
                'Murder_cat', 'Robbery_cat'],
            dtype='object')
```

In [58]: `plt.plot(crime['Violent\ncrime'])`

Out[58]: `[<matplotlib.lines.Line2D at 0x10eb65d30>]`

Clearly NYC is a problem. Most likely we'll have to drop it.

In [59]: `crime_no_NY = crime.drop(216)`

In [60]: `crime_no_NY[210:220]`

Out[60]:

	City	Population	Violent\ncrime \
210	Newburgh Town	30984	38
211	New Castle Town	17864	2
212	New Hartford Town and Village	20355	13
213	New Paltz Town and Village	14248	34
214	New Rochelle	78800	175
215	New Windsor Town	25767	31
217	New York Mills Village	3324	2
218	Niagara Falls	49574	584
219	Niagara Town	8241	13
220	Niskayuna Town	22097	13

```
        Murder and\nnonnegligent\nmanslaughter  Rape\n(revised\ndefinition)1  \
```

		NaN
210	0	NaN
211	0	NaN
212	0	NaN
213	0	NaN
214	0	NaN
215	0	NaN
217	0	NaN
218	3	NaN
219	0	NaN
220	0	NaN

	Rape\n(legacy\ndefinition)2	Robbery	Aggravated\nassault \
210	5	15	18
211	0	0	2
212	1	3	9
213	4	4	26
214	5	81	89
215	1	8	22
217	0	0	2
218	12	166	403
219	0	2	11
220	1	5	7

	Property\ncrime	Burglary	Larceny-\ntheft	Motor\nvehicle\ntheft \
210	1353	143	1182	28
211	77	13	62	2
212	806	72	730	4
213	200	16	180	4
214	1391	150	1172	69
215	471	65	394	12
217	64	16	47	1
218	2807	746	1949	112
219	489	48	427	14
220	341	51	282	8

	Arson3	Population_sq	Murder_cat	Robbery_cat
210	2.0	960008256	0	1
211	0.0	319122496	0	0
212	0.0	414326025	0	1
213	0.0	203005504	0	1
214	0.0	6209440000	0	1
215	4.0	663938289	0	1
217	0.0	11048976	0	0
218	22.0	2457581476	1	1
219	0.0	67914081	0	1
220	0.0	488277409	0	1

```
In [62]: plt.plot(crime_no_NY['Population'])
```