In [94]:

```
import zipcode
import pandas as pd
import numpy as np

df_movielens = pd.read_csv('sampleTableMin.csv',  encoding = 'ISO-8859-1')
```

In [95]:

```
df_movielens.columns
```

Out[95]:

```
Index(['Unnamed: 0', 'Unnamed: 0.1', 'userID', 'gender', 'age', 'occ
upation',
       'zip_code', 'movieID', 'ratings', 'timestamp', 'year', 'title
',
       'Animation', 'Children's', 'Comedy', 'Adventure', 'Fantasy',
'Romance',
       'Drama', 'Action', 'Crime', 'Thriller', 'Horror', 'Sci-Fi',
       'Documentary', 'War', 'Musical', 'Mystery', 'Film-Noir', 'Wes
tern'],
      dtype='object')
```

In [96]:

```
from uszipcode import ZipcodeSearchEngine

search = ZipcodeSearchEngine()
zipcode = search.by_zipcode("04901")
print(zipcode)
```


```
{
    "City": "Waterville",
    "Density": 333.0541809415048,
    "HouseOfUnits": 11921,
    "LandArea": 78.81,
    "Latitude": 44.55432769999999,
    "Longitude": -69.54199840000001,
    "NEBoundLatitude": 44.6456249,
    "NEBoundLongitude": -69.393123,
    "Population": 26248,
    "SWBoundLatitude": 44.480602000000005,
    "SWBoungLongitude": -69.72410500000001,
    "State": "ME",
    "TotalWages": 343517918.0,
    "WaterArea": 2.98,
    "Wealthy": 13087.394010972263,
    "Zipcode": "04901",
    "ZipcodeType": "Standard"
}
```


In [57]:

```
print(zipcode['State'])
```

```
None
```

In [64]:

```
df_movielens.iloc[:,6]
```

Out[64]:

```
0    04901
1    92688
2    95405
3    44333
4    15701
5    37843
6    19801
7    94110
8    55337
9    28134
Name: zip_code, dtype: object
```

In [66]:

```
ziplist = []
i = 0
for item_zipcode in df_movielens.iloc[:,6]:
    myzip = search.by_zipcode(item_zipcode)
    ziplist.append(myzip['State'])
```

In [67]:

```
print(df_movielens.shape)
print(len(ziplist))
print(ziplist[:200])
```

```
(100000, 30)
100000
['ME', 'CA', 'CA', 'OH', 'PA', 'TN', 'DE', 'CA', 'MN', 'NC', 'IL', '
CO', 'GA', 'MN', 'NY', 'WA', 'CA', 'OH', 'CA', 'CT', None, 'RI', 'CT
', 'MN', None, 'KS', 'IA', 'TN', 'CA', 'MD', 'CA', 'WA', 'CA', 'TX',
'NE', 'IL', 'MA', 'VA', 'CA', 'NY', 'CA', 'CA', 'MN', 'IL', 'IL', 'W
A', None, 'VA', 'MA', 'TX', 'CA', 'CA', 'GA', 'OK', 'MD', 'WI', 'MN'
, 'VA', 'NY', 'FL', 'FL', 'CA', None, 'NY', 'DE', 'CA', 'AZ', 'NC',
'MI', 'MA', 'TN', 'NJ', 'MA', 'GA', 'CA', 'RI', 'MA', 'MN', 'AL', 'N
Y', None, 'IL', 'TX', 'MN', 'WA', 'NC', 'MN', 'MD', 'MI', 'FL', 'CA'
, 'NC', 'NC', 'GA', 'NY', 'OH', 'WA', 'MN', 'MA', 'CA', 'MN', None,
'CA', None, 'WA', 'DC', 'TN', 'WI', 'PA', 'IA', 'FL', 'AZ', 'MD', 'W
V', 'MN', 'CA', 'OR', 'KY', None, 'MA', 'NY', 'TX', 'IL', 'CO', 'IL'
, 'FL', 'CA', 'CA', 'NY', 'CA', 'NE', 'CA', 'PR', 'CA', 'CA', 'OR',
'NJ', 'CA', 'MA', 'IL', 'CA', 'OH', 'CA', 'VT', 'CA', 'TN', 'VA', 'N
C', 'AL', 'MN', 'FL', 'NM', 'WA', 'OR', 'IA', 'CA', 'IL', 'MD', 'MO'
, 'CA', 'MI', 'CA', 'TN', 'TX', 'AR', 'OR', 'CA', 'CA', 'CO', 'MA',
'DE', 'OR', 'CA', 'MI', 'MN', 'WI', 'MN', 'TX', 'ID', 'IN', 'IL', 'M
N', 'RI', 'PA', 'VA', 'WA', 'CA', 'SC', 'CA', 'NJ', 'MI', 'IL', 'NY'
, 'CA', 'TX', 'CA', 'PR', 'IL', None, 'MN']
```

In [68]:

```
print(df_movielens.columns)
```

```
Index(['Unnamed: 0', 'Unnamed: 0.1', 'userID', 'gender', 'age', 'occ
upation',
       'zip_code', 'movieID', 'ratings', 'timestamp', 'year', 'title
',
       'Animation', 'Children's', 'Comedy', 'Adventure', 'Fantasy',
'Romance',
       'Drama', 'Action', 'Crime', 'Thriller', 'Horror', 'Sci-Fi',
       'Documentary', 'War', 'Musical', 'Mystery', 'Film-Noir', 'Wes
tern'],
      dtype='object')
```

In [71]:

```
ziplist = np.asarray(ziplist)
df_movielens['state'] = ziplist
df_movielens = df_movielens[['userID', 'movieID', 'gender', 'age', 'occupation',
       'zip_code', 'state', 'ratings', 'timestamp', 'year', 'title',
       'Animation', 'Children\'s', 'Comedy', 'Adventure', 'Fantasy', 'Romance',
       'Drama', 'Action', 'Crime', 'Thriller', 'Horror', 'Sci-Fi',
       'Documentary', 'War', 'Musical', 'Mystery', 'Film-Noir', 'Western']]
```

In [72]:

```
df_movielens.to_csv('movielens100Nones.csv', index = False)
```

In [91]:

```
df_movielens_with_states = df_movielens.loc[df_movielens['state'].notnull()]
```

In [92]:

```
df_movielens_with_states.shape
```

Out[92]:

```
(95315, 29)
```

In [97]:

```
df_movielens_with_states.to_csv('movielens100States.csv', index = False)
```