fetch_city_data

November 23, 2021

[14]:
```
import pandas as pd
from requests.auth import HTTPBasicAuth
from uszipcode import SearchEngine
```

[47]:
```
df = pd.read_csv('../ref/unique_zips.csv', index_col=0)
print(df.columns)
```

```
Index(['zipcode'], dtype='object')
```

[12]:
```
def search_by_zip(zip_code):
    search = SearchEngine(simple_zipcode=True)  # set simple_zipcode=False to␣
 ↪use rich info database
    zipcode = search.by_zipcode(zip_code)
    return zipcode
```

[17]:
```
test_zip_search = search_by_zip('75234').to_dict()
```

[50]:
```
for index, row in df.iterrows():
    zip_code = row['zipcode']
    zip_search = search_by_zip(zip_code).to_dict()
    for key in zip_search.keys():
        val = zip_search[key]
        if isinstance(val, list):
            for list_index, list_val in enumerate(val):
                if list_index == 0:
                    df.loc[index,key] = list_val
                else:
                    list_key = key + str(list_index)
                    df.loc[index,list_key] = list_val
            print(key, "Is List")
        else:
            df.loc[index,key] = val
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

```
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
common_city_list Is List
area_code_list Is List
```

[57]:
```
print(df.columns)
```

```
Index(['zipcode', 'zipcode_type', 'major_city', 'post_office_city',
       'common_city_list', 'county', 'state', 'lat', 'lng', 'timezone',
       'radius_in_miles', 'area_code_list', 'population', 'population_density',
       'land_area_in_sqmi', 'water_area_in_sqmi', 'housing_units',
       'occupied_housing_units', 'median_home_value',
       'median_household_income', 'bounds_west', 'bounds_east', 'bounds_north',
       'bounds_south', 'area_code_list1', 'area_code_list2',
       'common_city_list1', 'area_code_list3', 'area_code_list4',
       'area_code_list5', 'common_city_list2', 'common_city_list3',
       'area_code_list6', 'area_code_list7', 'area_code_list8',
       'common_city_list4'],
      dtype='object')
```

[56]:
```
#print(df[~df['area_code_list8'].isnull()])
print()
```

[58]:
```
df.to_csv('../ref/unique_zips_and_population.csv', index=False)
```

[12]:
```
import pandas as pd

df = pd.read_csv('../ref/train_with_zips.csv')
df_pop = pd.read_csv('../ref/unique_zips_and_population.csv')
```

[23]:
```
print(df.dtypes)
df['zip'] = df.zip.astype('int64')
print(df.dtypes)
print("df ^^^")
print(df_pop.dtypes)
```

```
RowId                       int64
IntersectionId              int64
Latitude                  float64
Longitude                 float64
EntryStreetName            object
ExitStreetName             object
EntryHeading               object
ExitHeading                object
Hour                        int64
Weekend                     int64
Month                       int64
Path                       object
TotalTimeStopped_p20      float64
TotalTimeStopped_p40      float64
TotalTimeStopped_p50      float64
TotalTimeStopped_p60      float64
TotalTimeStopped_p80      float64
TimeFromFirstStop_p20     float64
TimeFromFirstStop_p40     float64
TimeFromFirstStop_p50     float64
TimeFromFirstStop_p60     float64
TimeFromFirstStop_p80     float64
DistanceToFirstStop_p20   float64
DistanceToFirstStop_p40   float64
DistanceToFirstStop_p50   float64
DistanceToFirstStop_p60   float64
DistanceToFirstStop_p80   float64
City                       object
zip                         int64
dtype: object
RowId                       int64
IntersectionId              int64
Latitude                  float64
Longitude                 float64
EntryStreetName            object
ExitStreetName             object
EntryHeading               object
ExitHeading                object
Hour                        int64
Weekend                     int64
Month                       int64
```

```
Path                         object
TotalTimeStopped_p20        float64
TotalTimeStopped_p40        float64
TotalTimeStopped_p50        float64
TotalTimeStopped_p60        float64
TotalTimeStopped_p80        float64
TimeFromFirstStop_p20       float64
TimeFromFirstStop_p40       float64
TimeFromFirstStop_p50       float64
TimeFromFirstStop_p60       float64
TimeFromFirstStop_p80       float64
DistanceToFirstStop_p20     float64
DistanceToFirstStop_p40     float64
DistanceToFirstStop_p50     float64
DistanceToFirstStop_p60     float64
DistanceToFirstStop_p80     float64
City                         object
zip                           int64
dtype: object
df ^^^
zip                           int64
zipcode_type                 object
major_city                   object
post_office_city             object
common_city_list             object
county                       object
state                        object
lat                         float64
lng                         float64
timezone                     object
radius_in_miles             float64
area_code_list                int64
population                  float64
population_density          float64
land_area_in_sqmi           float64
water_area_in_sqmi          float64
housing_units               float64
occupied_housing_units      float64
median_home_value           float64
median_household_income     float64
bounds_west                 float64
bounds_east                 float64
bounds_north                float64
bounds_south                float64
area_code_list1             float64
area_code_list2             float64
common_city_list1            object
area_code_list3             float64
```

```
area_code_list4            float64
area_code_list5            float64
common_city_list2           object
common_city_list3           object
area_code_list6            float64
area_code_list7            float64
area_code_list8            float64
common_city_list4           object
dtype: object
```

[ ]:
```
result = pd.merge(df, df_pop, on="zip")
```

[24]:
```
# df.head()
# df_pop.head()
result.head()
```

[24]:

	RowId	IntersectionId	Latitude	Longitude	\
0	1921357	0	33.791659	-84.430032	
1	1921358	0	33.791659	-84.430032	
2	1921359	0	33.791659	-84.430032	
3	1921360	0	33.791659	-84.430032	
4	1921361	0	33.791659	-84.430032	

	EntryStreetName	ExitStreetName	EntryHeading	\
0	Marietta Boulevard Northwest	Marietta Boulevard Northwest	NW	
1	Marietta Boulevard Northwest	Marietta Boulevard Northwest	SE	
2	Marietta Boulevard Northwest	Marietta Boulevard Northwest	NW	
3	Marietta Boulevard Northwest	Marietta Boulevard Northwest	SE	
4	Marietta Boulevard Northwest	Marietta Boulevard Northwest	NW	

	ExitHeading	Hour	Weekend	...	common_city_list1	area_code_list3	\
0	NW	0	0	...	NaN	NaN	
1	SE	0	0	...	NaN	NaN	
2	NW	1	0	...	NaN	NaN	
3	SE	1	0	...	NaN	NaN	
4	NW	2	0	...	NaN	NaN	

	area_code_list4	area_code_list5	common_city_list2	common_city_list3	\
0	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	

	area_code_list6	area_code_list7	area_code_list8	common_city_list4
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN

```
3              NaN              NaN              NaN              NaN
4              NaN              NaN              NaN              NaN

[5 rows x 64 columns]
```

[27]:
```
print(len(result))
print(len(df))
```

```
856387
856387
```

[28]:
```
result.to_csv('../ref/train_with_zip_and_pop.csv', index=False)
```

[3]:
```
import pandas as pd
zip_df = pd.read_csv('../ref/unique_zips.csv', dtype={'zipcode':str})
zip_string = ""
for index, row in zip_df.iterrows():
    zip_string = zip_string + str(row['zipcode']) + ", "
print(zip_string)
```

```
30318, 02130, 60641, 60637, 30306, 02142, 60601, 60644, 30126, 19134, 60652,
60638, 02110, 02129, 60636, 19109, 60171, 30307, 30303, 19103, 02126, 60649,
02143, 60608, 02203, 60630, 02135, 19121, 60623, 02115, 02134, 30349, 60634,
30316, 60609, 60654, 30312, 60617, 60805, 19112, 60615, 19143, 60706, 60661,
60304, 60653, 19139, 19137, 60604, 60625, 30305, 60632, 60651, 60603, 02111,
19147, 60613, 02109, 19131, 60624, 19130, 60620, 30324, 02127, 30030, 02125,
60611, 30344, 02210, 60827, 19106, 30311, 60643, 19129, 02467, 02150, 19146,
60602, 19107, 60605, 30308, 19120, 02113, 19123, 30315, 60606, 02122, 30314,
30332, 02446, 02120, 60657, 60655, 30331, 30313, 60610, 30310, 60406, 19111,
30363, 19104, 19145, 30317, 02199, 19125, 02163, 60633, 19140, 02215, 30354,
19135, 60639, 02136, 30334, 19151, 19148, 02118, 60628, 02108, 46320, 02141,
02138, 60707, 60629, 60607, 60619, 02132, 02121, 19149, 60616, 60614, 19124,
19132, 02119, 60622, 60647, 30309, 19122, 02124, 19102, 60621, 02131, 02128,
60804, 02445, 19142, 02116, 60642, 60640, 60456, 19133, 60302, 60612, 60618,
02139, 02114,
```