Schools

November 9, 2019

```
In [1]: import pandas as pd
        from sqlalchemy import create_engine

In [2]: schools_csv = "Schools.csv"
        schools_df = pd.read_csv(schools_csv)
        schools_df.head()
```

Out[2]:

	School_ID	Legacy_Unit_ID	Finance_ID	Short_Name	Long_Name	School_Type	Primary_Category	Is_High_School	Is_Middle_School	Is_Elementary_School	...	Third_Contact_Name	Fourth_Contact_Title	Fourth_Contact_Name	Fifth_Contact_Title	Fifth_Contact_Name	Sixth_Contact_Title
0	610163	5770	30081	STOCK	Frederick Stock Elementary School	Neighborhood	ES	N	N	N	...	NaN	NaN	NaN	NaN	NaN	NaN
1	610558	9598	46611	GOODE HS	Sarah E. Goode STEM Academy	Citywide-Option	HS	Y	N	N	...	NaN	NaN	NaN	NaN	NaN	NaN
2	609750	1750	49051	SIMPSON HS	Simpson Academy HS for Young Women	Citywide-Option	HS	Y	Y	Y	...	Rita Somen	NaN	NaN	NaN	NaN	NaN
3	610571	9636	65015	OMBUDSMAN - WEST HS	Ombudsman Chicago- West	Citywide-Option	HS	Y	N	N	...	NaN	NaN	NaN	NaN	NaN	NaN
4	610123	5370	24911	PENN	William Penn Elementary School	Neighborhood	ES	N	Y	Y	...	NaN	NaN	NaN	NaN	NaN	NaN

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

  Sixth_Contact_Name Seventh_Contact_Title Seventh_Contact_Name  \
0                NaN                   NaN                  NaN
1                NaN                   NaN                  NaN
2                NaN                   NaN                  NaN
3                NaN                   NaN                  NaN
4                NaN                   NaN                  NaN

                                          Location
0  7507 W BIRCHWOOD AVE\nChicago, Illinois 60631\...
1  7651 S HOMAN AVE\nChicago, Illinois 60652\n(41...
2  1321 S PAULINA ST\nChicago, Illinois 60608\n(4...
3  2401 W CONGRESS PKWY\nChicago, Illinois 60612\...
4  1616 S AVERS AVE\nChicago, Illinois 60623\n(41...

[5 rows x 91 columns]
```

In [3]:
```
new_schools_df = schools_df[['Long_Name', 'School_Type', 'Primary_Category','Zip','Stu
new_schools_df.head()
```

Out[3]:
```
                            Long_Name      School_Type Primary_Category  \
0   Frederick Stock Elementary School     Neighborhood               ES
1         Sarah E. Goode STEM Academy  Citywide-Option               HS
2  Simpson Academy HS for Young Women  Citywide-Option               HS
3            Ombudsman Chicago- West  Citywide-Option               HS
4      William Penn Elementary School     Neighborhood               ES

     Zip  Student_Count_Total  Student_Count_Low_Income  \
0  60631                 232                        37
1  60652                 900                       788
2  60608                  38                        37
3  60612                 341                       320
4  60623                 311                       279

   Student_Count_Special_Ed  Student_Count_English_Learners  \
0                        90                              27
1                       153                              57
2                         6                               2
3                        57                              31
4                        78                              13

   Student_Count_Black  Student_Count_Hispanic  Student_Count_White  \
0                    1                      39                  175
1                  459                     420                    7
2                   28                       8                    2
```

```
3                  187                     148                    4
4                  283                      26                    1

   Student_Count_Asian  Student_Count_Native_American  Student_Count_Multi  \
0                   16                              0                    0
1                    2                              5                    6
2                    0                              0                    0
3                    0                              1                    1
4                    0                              1                    0

      Overall_Rating
0  Inability to Rate
1           Level 1+
2            Level 2
3            Level 2
4           Level 1+
```

In [4]:
```
housing_csv = "Housing.csv"
housing_df = pd.read_csv(housing_csv)
housing_df.head()
```

Out[4]:
```
  Community_Area_Name  Community_Area_Number Property_Description  \
0        Portage Park                     15                  ARO
1      West Englewood                     67          Multifamily
2           Englewood                     68          Multifamily
3     Washington Park                     40       Senior HUD 202
4       Humboldt Park                     23          Multifamily

                Property_Name                   Address  ZIP_Code  \
0  4812-15 W. Montrose Apts.  4812-15 W. Montrose Ave.     60641
1   New West Englewood Homes          2109 W. 63rd St.     60636
2           Antioch Homes II     301 W. Marquette Road     60621
3       St. Edmund's Corners     5556 S. Michigan Ave.     60637
4       Nelson Mandela Apts.           526 N. Troy St.     60624

   Phone_Number                 Management_Company  Units   Latitude  \
0  630-694-6968                       @properties      2        NaN
1  773-434-4929           Interfaith Housing Corp.     12        NaN
2  773-994-4546  Universal Management Service, Inc.     69  41.772564
3  773-667-7583    St. Edmund's Redevelopment Corp.     53  41.792975
4  773-227-6332                  Bickerdike Apts.      6  41.891173

   Longitude
0        NaN
1        NaN
2 -87.632419
3 -87.622569
4 -87.705338
```

```
In [13]: housing_zip = housing_df.groupby(['ZIP_Code'])['Units']
         housing_zip_count = housing_zip.sum()
         housing_zip_df = pd.DataFrame(housing_zip_count).reset_index()
         housing_zip_df = housing_zip_df.rename(columns={"ZIP_Code":"zip",'Units':'units'})
         housing_zip_df.head()
         # housing_zip_count.dtypes
```

Out[13]:

	zip	units
0	60601	16
1	60605	276
2	60607	233
3	60608	1022
4	60609	1207

```
In [6]: students_zip = new_schools_df.groupby(['Zip'])['Student_Count_Total','Student_Count_Lo
        students_zip_count = students_zip.sum()
        students_zip_count.head()
```

Out[6]:

Zip	Student_Count_Total	Student_Count_Low_Income
60602	1326	1142
60605	2645	975
60607	5477	2358
60608	11009	9798
60609	12972	11467

```
In [7]: students_zip_count['low_inc_percent'] = 100*students_zip_count['Student_Count_Low_Inco
        students_zip_count.head()
```

Out[7]:

Zip	Student_Count_Total	Student_Count_Low_Income	low_inc_percent
60602	1326	1142	86.123680
60605	2645	975	36.862004
60607	5477	2358	43.052766
60608	11009	9798	88.999909
60609	12972	11467	88.398088

```
In [8]: rds_connection_string = "postgres:postgres@localhost:5432/chicago"
        engine = create_engine(f'postgresql://{rds_connection_string}')

In [9]: engine.table_names()
```

Out[9]: ['housing', 'schools']

```
In [10]: students_zip_count.reset_index(level=0, inplace=True)
         students_zip_count = students_zip_count.rename(columns={'Zip':'zip','Student_Count_To
         students_zip_count.head()
```

Out[10]:

	zip	student_count_total	student_count_low_income	low_inc_perc
0	60602	1326	1142	86.123680
1	60605	2645	975	36.862004
2	60607	5477	2358	43.052766
3	60608	11009	9798	88.999909
4	60609	12972	11467	88.398088

In [11]:
```
students_zip_count.to_sql(name='schools', con=engine, if_exists='append', index=False)
```

In [14]:
```
housing_zip_df.to_sql(name='housing', con=engine, if_exists='append', index=False)
```