

Process to create a master list of SFUSD elementary schools including district codes, state codes, and addresses

Table of Contents

- [Introduction](#)
- [Data Wrangling](#)
- [Master List](#)

Introduction

4,611 children entered the 2017-2018 San Francisco Unified School District kindergarten lottery. Parents can select multiple schools, and they made more than 45,000 unique choices. This Jupyter Notebook prepares the master schools list to be used in the analysis of the SFUSD Lottery data.

In [1]:

```
# Importing packages and functions
import pandas as pd
import numpy as np
import csv
```

In [2]:

```
#file 1
sfusd_cds= pd.read_csv('./data/codes/sfusd_cds_codes.csv')
#http://web.sfusd.edu/Services/research_public/rpadc_lib/SFUSD%20CDS%20Codes%20SchYr2012-13_(08-20-12).pdf

#file2
sfdata_addresses_cds= pd.read_csv('./data/codes/schools_concise.csv')
#https://data.sfgov.org/Economy-and-Community/Schools/tpp3-epx2/data

#file3
caaspp_cds_district= pd.read_csv('./data/codes/sfusd_district_cds_codes_clean.csv')
#http://caaspp.edsource.org/
#https://www2.cde.ca.gov/dataresourceguide/
#https://caaspp.cde.ca.gov/sb2018/ResearchFileListCAA?ps=true&lstCounty=38&lstDistrict=68478-000&lstSchool=&lstCntyNam=San%20Francisco&lstDistNam=San%20Francisco%20Unified&lstTestYear=2018&lstTestA&rft=true
```

Data Wrangling

In [3]:

```
#clean up column names
sfusd_cds.columns = [x.strip().replace('-', '_') for x in sfusd_cds.columns]
sfusd_cds.rename(columns={'cds_cd' : 'cds_code'}, inplace=True)

sfdata_addresses_cds.columns = [x.lower() for x in sfdata_addresses_cds.columns]
sfdata_addresses_cds.columns = [x.strip().replace(' ', '_') for x in sfdata_addresses_cds.columns]
sfdata_addresses_cds.rename(columns={'campus_name' : 'school_name'}, inplace=True)

caaspp_cds_district.rename(columns={'state_code' : 'cds_code'}, inplace=True)
```

In [4]:

```
#convert data types
sfusd_cds = sfusd_cds.convert_objects(convert numeric=True)
```

```
sfdata_addresses_cds = sfdata_addresses_cds.convert_objects(convert_numeric=True)
caaspp_cds_district = caaspp_cds_district.convert_objects(convert_numeric=True)
```

```
/Users/Irene/anaconda/lib/python3.6/site-packages/ipykernel/__main__.py:2: FutureWarning:
convert_objects is deprecated. To re-infer data dtypes for object columns, use
DataFrame.infer_objects()
For all other conversions use the data-type specific converters pd.to_datetime, pd.to_timedelta and
pd.to_numeric.
from ipykernel import kernelapp as app
/Users/Irene/anaconda/lib/python3.6/site-packages/ipykernel/__main__.py:3: FutureWarning:
convert_objects is deprecated. To re-infer data dtypes for object columns, use
DataFrame.infer_objects()
For all other conversions use the data-type specific converters pd.to_datetime, pd.to_timedelta and
pd.to_numeric.
app.launch_new_instance()
/Users/Irene/anaconda/lib/python3.6/site-packages/ipykernel/__main__.py:4: FutureWarning:
convert_objects is deprecated. To re-infer data dtypes for object columns, use
DataFrame.infer_objects()
For all other conversions use the data-type specific converters pd.to_datetime, pd.to_timedelta and
pd.to_numeric.
```

In [5]:

```
sfusd_cds['cds_code'] = sfusd_cds['cds_code'].fillna(0).astype(int)
sfusd_cds['cds_code'] = sfusd_cds['cds_code'].astype(np.int64)
```

In [6]:

```
sfusd_cds.dtypes
```

Out[6]:

```
school_name    object
cds_code        int64
dtype: object
```

In [7]:

```
sfdata_addresses_cds.dtypes
```

Out[7]:

```
school_name    object
campus_address object
zipcode        int64
cds_code        int64
dtype: object
```

```
sfdata_addresses_cds['cds_code'] = sfdata_addresses_cds['cds_code'].fillna(0).astype(int) sfdata_addresses_cds['cds_code'] =
sfdata_addresses_cds['cds_code'].astype(np.int64) #school_addresses_cds.options.display.float_format = '{:,.0f}'.format
#sfusd_cds_district_codes.options.display.float_format = '{:,.0f}'.format #pd.astype(int)
#https://stackoverflow.com/questions/11548005/numpy-or-pandas-keeping-array-type-as-integer-while-having-a-nan-value
```

In [8]:

```
sfdata_addresses_cds.dtypes
```

Out[8]:

```
school_name    object
campus_address object
zipcode        int64
cds_code        int64
dtype: object
```

In [9]:

```
caaspp_cds_district.dtypes
```

Out[9]:

```
school_name    object
```

```
school_name      object
district_code    int64
cds_code         int64
dtype: object
```

In [10]:

```
cols_file1 = list(sfusd_cds.columns.values)
cols_file2 = list(sfdata_addresses_cds.columns.values)
cols_file3 = list(caaspp_cds_district.columns.values)
print(cols_file1)
print(cols_file2)
print(cols_file3)
```

```
['school_name', 'cds_code']
['school_name', 'campus_address', 'zipcode', 'cds_code']
['school_name', 'district_code', 'cds_code']
```

In [11]:

```
sfusd_cds.head()
```

Out[11]:

	school_name	cds_code
0	Alamo Elementary School, GE	6040695
1	Alvarado Elementary School (Spanish Immersion)	6040703
2	Alvarado Elementary School, GE	6040703
3	Argonne Elementary School, GE	6040737
4	Bryant Elementary School, GE	6040778

In [12]:

```
sfdata_addresses_cds.head()
```

Out[12]:

	school_name	campus_address	zipcode	cds_code
0	Carmichael, Bessie Carmichael K-5 Campus / Ear...	375 07TH ST	94103	384001456
1	Feinstein, Dianne Feinstein Elementary School	2550 25TH AVE	94116	111427
2	Stockton, Commodore Stockton Early Education S...	1 TRENTON ST	94108	117465
3	Noriega Early Education School	1775 44TH AVE	94122	117473
4	San Miguel Early Education School	300 SENECA AVE	94112	117481

In [13]:

```
sfdata_addresses_cds.head()
```

Out[13]:

	school_name	campus_address	zipcode	cds_code
0	Carmichael, Bessie Carmichael K-5 Campus / Ear...	375 07TH ST	94103	384001456
1	Feinstein, Dianne Feinstein Elementary School	2550 25TH AVE	94116	111427
2	Stockton, Commodore Stockton Early Education S...	1 TRENTON ST	94108	117465
3	Noriega Early Education School	1775 44TH AVE	94122	117473
4	San Miguel Early Education School	300 SENECA AVE	94112	117481

In [14]:

```
caaspp_cds_district.head()
```

Out[14]:

	school_name	district_code	cds_code
0	Alamo ES	413	6040695
1	Alvarado ES	420	6040703
2	Argonne ES	435	6040737
3	Bryant ES	456	6040778
4	Carver ES	625	6093496

Getting the Master List

In [15]:

```
df = pd.merge(sfusd_cds, caaspp_cds_district, on='cds_code', how='outer')
```

In [16]:

```
df.head()
```

Out[16]:

	school_name_x	cds_code	school_name_y	district_code
0	Alamo Elementary School, GE	6040695	Alamo ES	413.0
1	Alvarado Elementary School (Spanish Immersion)	6040703	Alvarado ES	420.0
2	Alvarado Elementary School, GE	6040703	Alvarado ES	420.0
3	Argonne Elementary School, GE	6040737	Argonne ES	435.0
4	Bryant Elementary School, GE	6040778	Bryant ES	456.0

In [17]:

```
df.rename(columns={'school_name_x' : 'school_name_sfusd_cds_codes', 'school_name_y' : 'school_name_caaspp_cds_district_codes'}, inplace=True)
```

In [18]:

```
df_big = pd.merge(caaspp_cds_district, sfdata_addresses_cds, on='cds_code', how='outer')
```

In [19]:

```
df_big.head()
```

Out[19]:

	school_name_x	district_code	cds_code	school_name_y	campus_address	zipcode
0	Alamo ES	413.0	6040695	Alamo Elementary School	250 23RD AVE	94121.0
1	Alvarado ES	420.0	6040703	Alvarado Elementary School	625 DOUGLASS ST	94114.0
2	Argonne ES	435.0	6040737	Argonne Elementary School	680 18TH AVE	94121.0
3	Bryant ES	456.0	6040778	Bryant Early Education / Bryant Elementary	2641 25TH ST	94110.0

4	Carver ES	625.0	6093496	Carver, Dr. George Washington Carver Elementar...	1360 OAKDALE AVE	94124.0
	school_name_x	district_code	cds_code	school_name_y	campus_address	zipcode

In [20]:

```
#convert data types
df_big = df_big.convert_objects(convert_numeric=True)
```

/Users/Irene/anaconda/lib/python3.6/site-packages/ipykernel/__main__.py:2: FutureWarning: convert_objects is deprecated. To re-infer data dtypes for object columns, use DataFrame.infer_objects()
For all other conversions use the data-type specific converters pd.to_datetime, pd.to_timedelta and pd.to_numeric.
from ipykernel import kernelapp as app

In [21]:

```
df_big.dtypes
```

Out[21]:

```
school_name_x      object
district_code      float64
cds_code            int64
school_name_y      object
campus_address     object
zipcode            float64
dtype: object
```

In [22]:

```
df_big['zipcode'] = df_big['zipcode'].fillna(0).astype(int)
df_big['zipcode'] = df_big['zipcode'].astype(np.int64)
```

In [23]:

```
df_big['district_code'] = df_big['district_code'].fillna(0).astype(int)
df_big['district_code'] = df_big['district_code'].astype(np.int64)
```

In [24]:

```
df_big.dtypes
```

Out[24]:

```
school_name_x      object
district_code      int64
cds_code            int64
school_name_y      object
campus_address     object
zipcode            int64
dtype: object
```

In [25]:

```
#export to csv
df_big.to_csv('./data/exports/master_schools_list.csv')
```

In []:

```
import pdfkit
pdfkit.from_file('concat_schools.html', 'concat_schools.pdf')
```

In []: