

# Analysis of the 2017-2018 school year kindergarten lottery

## Table of Contents

- [Introduction](#)
- [Data Wrangling](#)
- [Questions about the data set](#)

## 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 reviews data analyzed by KQED for this news article: <https://www.kqed.org/news/11641019/s-f-s-kindergarten-lottery-do-parents-tricks-work>, using this base workbook: <https://github.com/pickoffwhite/San-Francisco-Kindergarten-Lottery>

In addition to the baseline review, already set forth by the KQED team, this notebook looks at

- those applications that listed three schools
- the impact of zipcode and CTIP status on choices, assignments, and enrollments

Data limitations: Additional data would help ascertain how selections connect with language requirements for the bilingual and immersion programs. And, additional data can help assess how AA and sibling preference impact assignments and enrollment.

In [1]:

```
# Importing packages and functions
import pandas as pd
import numpy as np
import datetime # operations to parse dates
import time
import csv
from pprint import pprint #for pretty printing dictionaries
import ggplot as gplot
import matplotlib as mpl # Still required to change the sizes of plots
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
%matplotlib inline
import seaborn as sns
```

## Data Wrangling

### General Properties

In [2]:

```
lotto =
pd.read_csv('./data/kqed_data/20171103_KQED_KinderAssignmentData_201718_k_placement_2017_2018.csv',
, dtype = object)
df=lotto.copy()
```

In [3]:

```
lotto.info();
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4611 entries, 0 to 4610
Data columns (total 98 columns):
StudentNo      4611 non-null object
1              4611 non-null object
2              3856 non-null object
3              3441 non-null object
4              3028 non-null object
```

4	3020 non-null object
5	2657 non-null object
6	2361 non-null object
7	2114 non-null object
8	1938 non-null object
9	1804 non-null object
10	1668 non-null object
11	1147 non-null object
12	1027 non-null object
13	992 non-null object
14	949 non-null object
15	912 non-null object
16	881 non-null object
17	844 non-null object
18	796 non-null object
19	769 non-null object
20	738 non-null object
21	685 non-null object
22	649 non-null object
23	627 non-null object
24	605 non-null object
25	577 non-null object
26	544 non-null object
27	521 non-null object
28	503 non-null object
29	477 non-null object
30	444 non-null object
31	397 non-null object
32	374 non-null object
33	353 non-null object
34	338 non-null object
35	322 non-null object
36	300 non-null object
37	283 non-null object
38	271 non-null object
39	260 non-null object
40	245 non-null object
41	226 non-null object
42	212 non-null object
43	201 non-null object
44	188 non-null object
45	174 non-null object
46	166 non-null object
47	161 non-null object
48	154 non-null object
49	143 non-null object
50	137 non-null object
51	118 non-null object
52	115 non-null object
53	109 non-null object
54	106 non-null object
55	106 non-null object
56	99 non-null object
57	95 non-null object
58	91 non-null object
59	91 non-null object
60	88 non-null object
61	84 non-null object
62	83 non-null object
63	79 non-null object
64	77 non-null object
65	73 non-null object
66	69 non-null object
67	62 non-null object
68	59 non-null object
69	56 non-null object
70	53 non-null object
71	49 non-null object
72	44 non-null object
73	41 non-null object
74	40 non-null object
75	39 non-null object
76	38 non-null object
77	37 non-null object
78	35 non-null object
79	33 non-null object
80	29 non-null object
81	23 non-null object

```

81 20 non-null object
82 21 non-null object
83 13 non-null object
84 9 non-null object
85 7 non-null object
86 2 non-null object
87 2 non-null object
88 2 non-null object
89 2 non-null object
90 2 non-null object
91 1 non-null object
92 1 non-null object
Round 1 Assignment 4470 non-null object
School Enrolled In As Of 11/03/2017 4043 non-null object
Student's Ethnicity 4389 non-null object
Does Student Live In CTIP1 Zone? (y/n) 4611 non-null object
Student's Residential Zip Code 4611 non-null object
dtypes: object(98)
memory usage: 3.4+ MB

```

In [4]:

```
lotto.duplicated().sum()
```

Out[4]:

0

In [5]:

```
lotto.nunique().sum()
```

Out[5]:

8892

In [6]:

```
lotto.head()
```

Out[6]:

	StudentNo	1	2	3	4	5	6	7	8	9	...	88	89	90	91	92	Round 1 Assignment	School Enrolled In As Of 11/03/2017	Student's Ethnicity
0	1	848	569	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	848	848	White Not c Hispa Origin
1	2	420	420	537	729	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	729	729	Hispa
2	3	478	589	722	664	539	796	420	413	569	...	NaN	NaN	NaN	NaN	NaN	722	722	White Not c Hispa Origin
3	4	782	676	478	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	782	782	Asian Pacif Islanc
4	5	420	618	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	420	420	Hispa

5 rows × 98 columns

## Data Cleaning

## Data Cleaning

```
lotto.head.columns = [x.lower() for x in lotto.head.columns] lotto.head.columns = [x.strip().replace(' ', '_') for x in lotto.head.columns]
```

### Manage CTIP column

In [7]:

```
#seperate out the data columns that I will be working with
zipdata = df.iloc[:, [0,93,94,96,97]]
zipdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4611 entries, 0 to 4610
Data columns (total 5 columns):
StudentNo                4611 non-null object
Round 1 Assignment       4470 non-null object
School Enrolled In As Of 11/03/2017    4043 non-null object
Does Student Live In CTIP1 Zone? (y/n) 4611 non-null object
Student's Residential Zip Code          4611 non-null object
dtypes: object(5)
memory usage: 180.2+ KB
```

In [8]:

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

```
/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
```

Out[8]:

```
StudentNo                int64
Round 1 Assignment       float64
School Enrolled In As Of 11/03/2017    float64
Does Student Live In CTIP1 Zone? (y/n)  object
Student's Residential Zip Code          int64
dtype: object
```

In [9]:

```
#Replace 'Y' and 'N' with 1 and 0 for 'Not CTIP'
#1 = yes CTIP #0 = is not a CTIP applicant
zipdata['Does Student Live In CTIP1 Zone? (y/n)'].replace({'N':0, 'Y': 1}, inplace=True)
```

In [10]:

```
zipdata.dtypes
```

Out[10]:

```
StudentNo                int64
Round 1 Assignment       float64
School Enrolled In As Of 11/03/2017    float64
Does Student Live In CTIP1 Zone? (y/n)  int64
Student's Residential Zip Code          int64
dtype: object
```

In [11]:

```
pd.options.display.float_format = '{:,.0f}'.format
```

In [12]:

```
zipdata.head()
```

Out[12]:

	StudentNo	Round 1 Assignment	School Enrolled In As Of 11/03/2017	Does Student Live In CTIP1 Zone? (y/n)	Student's Residential Zip Code
0	1	848	848	0	94118
1	2	729	729	0	94112
2	3	722	722	0	94131
3	4	782	782	0	94127
4	5	420	420	1	94110

## Questions about the data

Which are the top most requested schools?

How many students enrolled in their Round 1 top choice school?

2810 students... 60.94% of students

There are students that entered the lottery late. The school that those students enroll in becomes their first choice in the data. We are going to exclude them from the dataset. We can tell who they are because they did not received a round one assignment, so their data is null, according to the San Francisco School District. This year there were 141 students that entered the lotter late.

In [13]:

```
late = lotto['Round 1 Assignment']
late.isna().sum()
```

Out[13]:

141

In [14]:

```
lotto.dropna(subset=['Round 1 Assignment'])
```

Out[14]:

	StudentNo	1	2	3	4	5	6	7	8	9	...	88	89	90	91	92	Round 1 Assignment	School Enrolled In As Of 11/03/2017
0	1	848	569	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	848	848
1	2	420	420	537	729	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	729	729
2	3	478	589	722	664	539	796	420	413	569	...	NaN	NaN	NaN	NaN	NaN	722	722

3	4	782	676	478	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	782	782	School
4	StudentNo	1	2	3	4	5	6	7	8	9	...	88	89	90	91	92	Round 1	Enrolled In	As Of
5	5	420	618	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	Assignment	420	11/03/2017
6	6	838	746	478	876	488	848	729	569	796	...	NaN	NaN	NaN	NaN	NaN	838	838	
7	7	485	876	513	513	478	862	862	782	760	...	NaN	NaN	NaN	NaN	NaN	760	614	
8	8	859	449	790	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	859	859	
9	9	420	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	420	420	
10	10	782	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	782	782	
11	11	485	676	862	750	746	539	435	782	644	...	NaN	NaN	NaN	NaN	NaN	862	862	
12	12	676	796	485	478	478	420	420	497	589	...	NaN	NaN	NaN	NaN	NaN	676	676	
13	13	842	722	575	478	488	876	420	796	539	...	NaN	NaN	NaN	NaN	NaN	842	842	
14	14	862	876	509	862	876	488	490	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	490	490	
15	15	746	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	746	746	
16	16	676	782	750	485	644	876	862	876	544	...	NaN	NaN	NaN	NaN	NaN	782	782	
17	17	435	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	435	435	
18	18	796	420	478	722	876	718	589	479	413	...	NaN	NaN	NaN	NaN	NaN	420	NaN	
19	19	676	782	644	750	862	876	539	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	676	676	
20	20	478	735	420	435	485	676	750	544	876	...	NaN	NaN	NaN	NaN	NaN	478	478	
21	21	867	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	867	867	
22	22	488	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	488	488	
23	23	569	479	782	750	644	413	876	435	485	...	NaN	NaN	NaN	NaN	NaN	656	860	
24	24	862	867	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	867	525	
25	25	562	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	562	562	
26	26	782	676	485	509	544	750	644	569	549	...	NaN	NaN	NaN	NaN	NaN	782	782	

27	28	664	796	479	625	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	625	875
28	29	562	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	562	562 School
	<b>StudentNo</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	...	<b>88</b>	<b>89</b>	<b>90</b>	<b>91</b>	<b>92</b>	<b>Round 1 Assignment</b>	<b>Enrolled In As Of</b>
29	30	718	589	796	735	786	420	505	478	569	...	NaN	NaN	NaN	NaN	NaN	718	11/03/2017
30	31	509	485	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	509	509
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4581	4582	420	420	876	722	796	862	488	478	842	...	NaN	NaN	NaN	NaN	NaN	656	656
4582	4583	485	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	485	485
4583	4584	449	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	449	859
4584	4585	814	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	814	814
4585	4586	589	718	478	796	722	814	735	569	420	...	NaN	NaN	NaN	NaN	NaN	589	938
4586	4587	507	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	507	507
4587	4588	644	676	782	750	589	664	862	569	544	...	NaN	NaN	NaN	NaN	NaN	644	644
4588	4589	537	420	456	680	618	760	723	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	456	456
4589	4590	796	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	796	796
4590	4591	603	656	722	575	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	603	614
4591	4592	435	413	664	569	676	848	485	750	544	...	NaN	NaN	NaN	NaN	NaN	859	544
4592	4593	823	525	801	790	718	638	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	525	525
4593	4594	796	478	589	722	876	842	718	513	NaN	...	NaN	NaN	NaN	NaN	NaN	796	796
4594	4595	521	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	521	521
4595	4596	760	449	796	420	435	488	478	722	479	...	NaN	NaN	NaN	NaN	NaN	760	760
4596	4597	750	876	876	478	478	485	676	644	NaN	...	NaN	NaN	NaN	NaN	NaN	750	750
4597	4598	796	420	420	876	478	589	718	485	722	...	NaN	NaN	NaN	NaN	NaN	420	420
4598	4599	816	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	816	816
4599	4600	670	750	644	876	796	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	670	670
4600	4601	872	490	638	834	801	562	562	823	790	...	NaN	NaN	NaN	NaN	NaN	490	490

4601	4602	449	796	537	814	735	723	478	664	676	...	NaN	NaN	NaN	NaN	NaN	449	449	School
4602	StudentNo 4603	1 478	2 478	3 420	4 796	5 644	6 842	7 722	8 676	9 479	...	88 NaN	89 NaN	90 NaN	91 NaN	92 NaN	Round 1 478 Assignment	Enrolled In 478 As Of 11/03/2017	
4603	4604	618	714	537	729	723	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	618	618	
4604	4605	478	478	876	876	796	589	420	420	575	...	NaN	NaN	NaN	NaN	NaN	816	201	
4605	4606	644	413	664	435	670	497	569	848	750	...	NaN	NaN	NaN	NaN	NaN	670	670	
4606	4607	867	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	867	867	
4607	4608	618	796	760	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	618	618	
4608	4609	820	488	691	691	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	820	820	
4609	4610	834	490	638	562	525	650	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	834	834	
4610	4611	718	722	814	823	413	420	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	718	718	

4470 rows × 98 columns

In [15]:

```
lotto['got1stchoice'] = np.where(lotto['1'] == lotto['School Enrolled In As Of 11/03/2017'], 'yes', 'no')
```

In [16]:

```
lotto.got1stchoice.value_counts()
```

Out[16]:

```
yes      2810
no       1801
Name: got1stchoice, dtype: int64
```

In [17]:

```
lotto.got1stchoice.value_counts('yes')
```

Out[17]:

```
yes      1
no       0
Name: got1stchoice, dtype: float64
```

## How many students enrolled in one of the schools in their first three, five, and ten choices.

We're filtering down the dataframe to the first five schools, and then seeing how many match.

Top 3: 73%

Top 5: 76%



Top 5: 78%

Top 10: 79%

In [18]:

```
lotto2 = lotto[['School Enrolled In As Of 11/03/2017', '1', '2']]
```

In [19]:

```
lotto['gottop2'] = lotto2.drop("School Enrolled In As Of 11/03/2017", 1).isin(lotto2["School Enrolled In As Of 11/03/2017"]).any(1)
```

In [20]:

```
lotto.gottop2.value_counts(True)
```

Out[20]:

```
True    1
False   0
Name: gottop2, dtype: float64
```

In [21]:

```
lotto3 = lotto[['School Enrolled In As Of 11/03/2017', '1', '2', '3']]
```

In [22]:

```
lotto['gottop3'] = lotto3.drop("School Enrolled In As Of 11/03/2017", 1).isin(lotto3["School Enrolled In As Of 11/03/2017"]).any(1)
```

In [23]:

```
lotto.gottop3.value_counts(True)
```

Out[23]:

```
True    1
False   0
Name: gottop3, dtype: float64
```

How many students enrolled in one of the schools in their top 5 schools. We're filtering down the dataframe to the first ten schools, and then seeing how many match.

In [24]:

```
lotto5 = lotto[['School Enrolled In As Of 11/03/2017', '1', '2', '3', '4', '5']]
```

In [25]:

```
lotto['gottop5'] = lotto5.drop("School Enrolled In As Of 11/03/2017", 1).isin(lotto5["School Enrolled In As Of 11/03/2017"]).any(1)
```

In [26]:

```
lotto.gottop5.value_counts(True)
```

Out[26]:

```
True    1
False   0
Name: gottop5, dtype: float64
```

How many students enrolled in one of the schools in their top 10 choices. We're filtering down the dataframe to the first ten schools, and then seeing how many match.

In [27]:

```
lotto10 = lotto[['School Enrolled In As Of 11/03/2017', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10']]
```

In [28]:

```
lotto['gottop10'] = lotto10.drop("School Enrolled In As Of 11/03/2017", 1).isin(lotto10["School Enrolled In As Of 11/03/2017"]).any(1)
```

In [29]:

```
lotto.gottop10.value_counts(True)
```

Out[29]:

```
True      1
False     0
Name: gottop10, dtype: float64
```

In [30]:

```
lotto.reset_index()
```

Out[30]:

	index	StudentNo	1	2	3	4	5	6	7	8	...	Round 1 Assignment	School Enrolled In As Of 11/03/2017	Student's Ethnicity	Does Student Live In CTIP1 Zone? (y/n)	Stude Reside Zip C
0	0	1	848	569	NaN	NaN	NaN	NaN	NaN	NaN	...	848	848	White, Not of Hispanic Origin	N	94118
1	1	2	420	420	537	729	NaN	NaN	NaN	NaN	...	729	729	Hispanic	N	94112
2	2	3	478	589	722	664	539	796	420	413	...	722	722	White, Not of Hispanic Origin	N	94131
3	3	4	782	676	478	NaN	NaN	NaN	NaN	NaN	...	782	782	Asian or Pacific Islander	N	94127
4	4	5	420	618	NaN	NaN	NaN	NaN	NaN	NaN	...	420	420	Hispanic	Y	94110
5	5	6	838	746	478	876	488	848	729	569	...	838	838	Asian or Pacific Islander	N	94134
6	6	7	485	876	513	513	478	862	862	782	...	760	614	Declines To State	N	94134
7	7	8	859	449	790	NaN	NaN	NaN	NaN	NaN	...	859	859	Hispanic	N	94102
8	8	9	420	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	420	420	Declines To State	Y	94107
9	9	10	782	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	782	782	Declines To State	N	94112
10	10	11	485	676	862	750	746	539	435	782	...	862	862	Asian or Pacific	N	94116

															Islander	Does	
11	11	12	676	796	485	478	478	420	420	497	...	676	Round 1	School	Declines	Student	Stude
index	StudentNo	1	2	3	4	5	6	7	8	...	Assignment	Enrolled In	As Of	Student's	Live In	Reside	
													11/03/2017	Ethnicity	CTIP1	Zip C	
12	12	13	842	722	575	478	488	876	420	796	...	842		842	White, Not of	Zone? (y/n)	94127
															Hispanic Origin		
13	13	14	862	876	509	862	876	488	490	NaN	...	490		490	Asian or Pacific Islander	N	94112
14	14	15	676	485	750	644	782	876	746	876	...	NaN		NaN	Asian or Pacific Islander	N	94132
15	15	16	746	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	746		746	Declines To State	N	94112
16	16	17	676	782	750	485	644	876	862	876	...	782		782	Asian or Pacific Islander	N	94132
17	17	18	435	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	435		435	Declines To State	N	94121
18	18	19	796	420	478	722	876	718	589	479	...	420		NaN	Declines To State	N	94114
19	19	20	676	782	644	750	862	876	539	NaN	...	676		676	Asian or Pacific Islander	N	94112
20	20	21	478	735	420	435	485	676	750	544	...	478		478	White, Not of Hispanic Origin	N	94122
21	21	22	867	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	867		867	Asian or Pacific Islander	N	94134
22	22	23	488	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	488		488	Declines To State	N	94112
23	23	24	569	479	782	750	644	413	876	435	...	656		860	Declines To State	N	94131
24	24	25	862	867	NaN	NaN	NaN	NaN	NaN	NaN	...	867		525	Declines To State	N	94116
25	25	26	562	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	562		562	Hispanic	N	94117
26	26	27	782	676	485	509	544	750	644	569	...	782		782	Asian or Pacific Islander	N	94103
27	27	28	664	796	479	625	NaN	NaN	NaN	NaN	...	625		875	Hispanic	N	94129
28	28	29	562	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	562		562	Hispanic	N	94107
29	29	30	718	589	796	735	786	420	505	478	...	718		718	White, Not of Hispanic Origin	N	94114
...	...	...	...	...	...	...	...	...	...	...	...	...		...	...	...	...
4581	4581	4582	420	420	876	722	796	862	488	478	...	656		656	White, Not of Hispanic Origin	N	94112
4582	4582	4583	485	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	485		485	Asian or Pacific	N	94112

															Islander	Does	
4583	4583	4584	449	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	449	859	School	NaN	Student	94115
4584	4584	4585	814	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	814	Round 1	Enrolled In	Student's	Live In	Reside
													Assignment	As Of	Ethnicity	CTIP1	Zip C
4585	4585	4586	589	718	478	796	722	814	735	569	...	589	11/03/2017	White, Not of	Zone? (y/n)		94117
														Hispanic Origin			
4586	4586	4587	507	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	507	507	NaN	Y		94124
4587	4587	4588	644	676	782	750	589	664	862	569	...	644	644	Asian or Pacific Islander	N		94122
4588	4588	4589	537	420	456	680	618	760	723	NaN	...	456	456	Hispanic	N		94124
4589	4589	4590	796	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	796	796	Hispanic	N		94134
4590	4590	4591	603	656	722	575	NaN	NaN	NaN	NaN	...	603	614	Hispanic	N		94112
4591	4591	4592	435	413	664	569	676	848	485	750	...	859	544	Asian or Pacific Islander	N		94121
4592	4592	4593	823	525	801	790	718	638	NaN	NaN	...	525	525	Black, Not of Hispanic Origin	N		94115
4593	4593	4594	796	478	589	722	876	842	718	513	...	796	796	Asian or Pacific Islander	N		94134
4594	4594	4595	521	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	521	521	Hispanic	N		94134
4595	4595	4596	760	449	796	420	435	488	478	722	...	760	760	Asian or Pacific Islander	N		94124
4596	4596	4597	750	876	876	478	478	485	676	644	...	750	750	Asian or Pacific Islander	N		94122
4597	4597	4598	796	420	420	876	478	589	718	485	...	420	420	NaN	N		94114
4598	4598	4599	816	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	816	816	Hispanic	Y		94110
4599	4599	4600	670	750	644	876	796	NaN	NaN	NaN	...	670	670	White, Not of Hispanic Origin	N		94127
4600	4600	4601	872	490	638	834	801	562	562	823	...	490	490	Asian or Pacific Islander	N		94108
4601	4601	4602	449	796	537	814	735	723	478	664	...	449	449	NaN	Y		94124
4602	4602	4603	478	478	420	796	644	842	722	676	...	478	478	White, Not of Hispanic Origin	Y		94109
4603	4603	4604	618	714	537	729	723	NaN	NaN	NaN	...	618	618	Hispanic	Y		94110
4604	4604	4605	478	478	876	876	796	589	420	420	...	816	201	NaN	N		94131
4605	4605	4606	644	413	664	435	670	497	569	848	...	670	670	White, Not of Hispanic Origin	N		94118
4606	4606	4607	867	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	867	867	Asian or Pacific Islander	N		94134

4607	4607	4608	618	796	760	NaN	NaN	NaN	NaN	NaN	...	618	618	NaN	N	94110	
4608	4608	4609	820	488	691	691	NaN	NaN	NaN	NaN	...	820	820	School	Hispanic	Does Student	94112
4609	4609	4610	834	490	638	562	525	650	NaN	NaN	...	834	834	Round 1 Enrolled In Assignment	Student's NaN As Of	Live In Y CTIP1 Zone? (y/n)	94113
														11/03/2017	White, Not of Ethnicity		Zip C
4610	4610	4611	718	722	814	823	413	420	NaN	NaN	...	718	718		Hispanic Origin		94115

4611 rows x 104 columns

## How many students are assigned a school not in their list in round one?

3% of students were assigned a school not listed in their list of selected schools

In [31]:

```
randomAssignment = lotto.drop(['StudentNo', 'School Enrolled In As Of 11/03/2017'], axis=1)
```

In [32]:

```
randomAssignment['notrandom'] = randomAssignment.drop('Round 1 Assignment', 1).isin(randomAssignment['Round 1 Assignment']).any(1)
```

In [33]:

```
randomAssignment.notrandom.value_counts(False)
```

Out[33]:

```
True      4470
False      141
Name: notrandom, dtype: int64
```

## How many people put down more than 5, 10 school choices?

57% listed more than 5 choices

36% listed more than 10 choices

In [34]:

```
amountchoices = lotto.drop(['StudentNo', "Does Student Live In CTIP1 Zone? (y/n)", 'School Enrolled In As Of 11/03/2017', 'Round 1 Assignment', "Student's Ethnicity", "Student's Residential Zip Code", 'got1stchoice', 'gottop5', 'gottop10'], axis=1)
```

In [35]:

```
lotto['choices'] = amountchoices.notnull().sum(axis=1)
```

In [36]:

```
lotto['morethan3'] = np.where(lotto['choices'] > 3, 'yes', 'no')
```

In [37]:

```
lotto.morethan3.value_counts('yes')
```

Out[37]:

```
yes    1
no      0
Name: morethan3, dtype: float64
```

In [38]:

```
lotto['morethan5'] = np.where(lotto['choices'] > 5, 'yes', 'no')
```

In [39]:

```
lotto.morethan5.value_counts()
```

Out[39]:

```
yes    3028
no     1583
Name: morethan5, dtype: int64
```

In [40]:

```
lotto.morethan5.value_counts('yes')
```

Out[40]:

```
yes    1
no      0
Name: morethan5, dtype: float64
```

In [41]:

```
lotto['morethan10'] = np.where(lotto['choices'] > 10, 'yes', 'no')
```

In [42]:

```
lotto.morethan10.value_counts()
```

Out[42]:

```
no      2807
yes     1804
Name: morethan10, dtype: int64
```

In [43]:

```
lotto.morethan10.value_counts('yes')
```

Out[43]:

```
no      1
yes      0
Name: morethan10, dtype: float64
```

In [44]:

```
lotto['lessthan10'] = np.where(lotto['choices'] < 10, 'yes', 'no')
```

In [45]:

```
lotto.lessthan10.value_counts('yes')
```

Out[45]:

```
yes    1
no      0
```

Name: lessthan10, dtype: float64

In [46]:

```
lotto['lessthan5'] = np.where(lotto['choices'] < 5, 'yes', 'no')
```

In [47]:

```
lotto.lessthan5.value_counts('yes')
```

Out[47]:

```
no    1
yes    0
Name: lessthan5, dtype: float64
```

## Does putting down more than ten schools impact the chance of enrolling in your first choice?

In [48]:

```
lotto.groupby(['got1stchoice', 'morethan10']).size()
```

Out[48]:

```
got1stchoice  morethan10
no            no          769
              yes        1032
yes           no        2038
              yes         772
dtype: int64
```

In [49]:

```
lotto.groupby(['got1stchoice', 'morethan5']).size()
```

Out[49]:

```
got1stchoice  morethan5
no            no         243
              yes       1558
yes           no       1340
              yes       1470
dtype: int64
```

In [50]:

```
lotto.groupby(['got1stchoice', 'lessthan5']).size()
```

Out[50]:

```
got1stchoice  lessthan5
no            no       1683
              yes        118
yes           no       1758
              yes       1052
dtype: int64
```

In [51]:

```
lotto.head()
```

Out[51]:

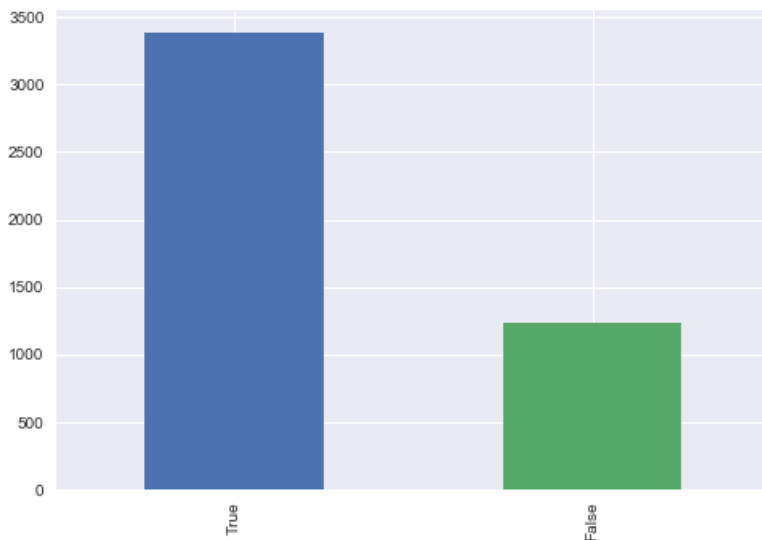
	StudentNo	1	2	3	4	5	6	7	8	9	...	gottop2	gottop3	gottop5	gottop10	choices	morethan3
0	1	848	569	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	True	True	True	True	4	yes
1	2	420	420	537	729	NaN	NaN	NaN	NaN	NaN	...	False	False	True	True	6	yes

	StudentNo	1	2	3	4	5	6	7	8	9	...	gottop2	gottop3	gottop5	gottop10	choices	morethan3
2	3	478	589	722	664	539	796	420	413	569	...	False	True	True	True	17	yes
3	4	782	676	478	NaN	NaN	NaN	NaN	NaN	NaN	...	True	True	True	True	5	yes
4	5	420	618	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	True	True	True	True	4	yes

5 rows × 109 columns

In [52]:

```
lotto['gottop3'].value_counts().plot(kind='bar');
```



## Export data

In [53]:

```
lotto.to_csv('./data/exports/lotto_edited.csv')
```

In [54]:

```
zipdata.to_csv('./data/exports/zipdata.csv')
```

## CTIP

The Census Tract Integration Preference (CTIP) operates as a preference/tie-breaking factor in the choice student assignment system. – Designed to meet the school board's goals to reverse the trend of racial isolation and the concentration of underserved students in the same school, and provide equitable access to the range of opportunities offered to students – Based on average test scores – Children living in areas of the city with the lowest average test scores get the CTIP1 tiebreaker

## How many CTIP applicants are there?

705

In [55]:

```
zipdata['Does Student Live In CTIP1 Zone? (y/n)'].value_counts()
```

Out[55]:

```
0    3906
1     705
```

Name: Does Student Live In CTIP1 Zone? (y/n), dtype: int64



## Did any of the CTIP applicants apply late?

no

In [56]:

```
ctip_1 = zipdata[(zipdata['Does Student Live In CTIP1 Zone? (y/n)']==1)]  
print(ctip_1)
```

	StudentNo	Round 1 Assignment	School Enrolled In As Of 11/03/2017 \
4	5	420	420
8	9	420	420
11	12	676	676
35	36	485	485
36	37	760	760
49	50	723	723
55	56	420	420
66	67	603	603
69	70	714	517
78	79	786	786
92	93	680	680
97	98	420	420
98	99	420	844
99	100	859	859
104	105	722	722
141	142	735	735
151	152	680	680
168	169	420	420
169	170	676	676
172	173	638	638
204	205	420	420
205	206	513	453
213	214	838	838
225	226	664	845
238	239	507	507
250	251	490	490
252	253	714	714
255	256	618	618
271	272	618	618
273	274	680	680
...	...	...	...
4506	4507	525	525
4510	4511	867	521
4511	4512	723	723
4522	4523	603	603
4529	4530	760	680
4532	4533	859	859
4534	4535	420	420
4537	4538	830	830
4538	4539	449	704
4541	4542	614	614
4542	4543	449	449
4546	4547	625	625
4551	4552	625	625
4555	4556	723	723
4562	4563	449	449
4563	4564	830	951
4566	4567	834	809
4570	4571	589	589
4573	4574	680	680
4574	4575	867	867
4579	4580	614	614
4580	4581	796	575
4583	4584	449	859
4586	4587	507	507
4598	4599	816	816
4601	4602	449	449
4602	4603	478	478
4603	4604	618	618
4609	4610	834	834

4610	4611	718	718
	Does Student Live In CTIP1 Zone? (y/n)	Student's Residential	Zip Code
4	1		94110
8	1		94107
11	1		94107
35	1		94124
36	1		94124
49	1		94110
55	1		94110
66	1		94110
69	1		94124
78	1		94115
92	1		94110
97	1		94110
98	1		94107
99	1		94124
104	1		94131
141	1		94117
151	1		94110
168	1		94107
169	1		94102
172	1		94124
204	1		94110
205	1		94107
213	1		94107
225	1		94102
238	1		94124
250	1		94103
252	1		94103
255	1		94110
271	1		94110
273	1		94110
...	...		...
4506	1		94124
4510	1		94134
4511	1		94110
4522	1		94110
4529	1		94134
4532	1		94109
4534	1		94103
4537	1		94124
4538	1		94124
4541	1		94124
4542	1		94130
4546	1		94124
4551	1		94124
4555	1		94103
4562	1		94124
4563	1		94124
4566	1		94115
4570	1		94115
4573	1		94110
4574	1		94112
4579	1		94130
4580	1		94124
4583	1		94115
4586	1		94124
4598	1		94110
4601	1		94124
4602	1		94109
4603	1		94110
4609	1		94109
4610	1		94115

[705 rows x 5 columns]

In [57]:

```
ctip_late = ctip_1['Round 1 Assignment']
ctip_late.isna().sum()
```

Out[57]:

0

## How many CTIP applicants were assigned their top choice school? compared to non CTIP applicants?

**14% or 88 CTIP applicants out of 705 received their top choice school in the first round...compared to 46% of non CTIP applicants** Note: non CTIP applicants includes AA designation students and students with an older sibling already at their top choice school...both factors would increase their likelihood of receiving their top choice according to SF Parents For Public Schools.

In [58]:

```
zipdata['got1stchoice'] = np.where(zipdata['Round 1 Assignment'] == zipdata['School Enrolled In As Of 11/03/2017'], 0, 1)
```

In [59]:

```
zipdata.head()
```

Out[59]:

	StudentNo	Round 1 Assignment	School Enrolled In As Of 11/03/2017	Does Student Live In CTIP1 Zone? (y/n)	Student's Residential Zip Code	got1stchoice
0	1	848	848	0	94118	0
1	2	729	729	0	94112	0
2	3	722	722	0	94131	0
3	4	782	782	0	94127	0
4	5	420	420	1	94110	0

In [60]:

```
zipdata.groupby(['Does Student Live In CTIP1 Zone? (y/n)', 'got1stchoice']).size()
```

Out[60]:

```
Does Student Live In CTIP1 Zone? (y/n)  got1stchoice
0                                         0          2667
                                         1          1239
1                                         0           617
                                         1           88
dtype: int64
```

## Where did CTIP applicants end up?

In [61]:

```
codes = pd.read_csv('./data/codes/master_schools_list_complete.csv')
df_codes = codes.copy()
```

In [62]:

```
codes.head()
```

Out[62]:

	Unnamed: 0	school_name_x	district_code	cds_code	school_name_y	campus_address	zipcode
0	0	Alamo ES	413	6040695	Alamo Elementary School	250 23RD AVE	94121
		Alamo ES					



2	Unnamed: 0	Alvarado ES school_name_x	420 district_code	6040703 cds_code	Alvarado Elementary School school_name_y	625 DOUGLASS ST campus_address	94114 campus_zipcode
3	2	Argonne ES	435	6040737	Argonne Elementary School	680 18TH AVE	94121
4	3	Bryant ES	456	6040778	Bryant Early Education / Bryant Elementary	2641 25TH ST	94110

In [69]:

```
ctip_1.head()
```

Out[69]:

	StudentNo	Round 1 Assignment	School Enrolled In As Of 11/03/2017	Does Student Live In CTIP1 Zone? (y/n)	Student's Residential Zip Code
4	5	420	420	1	94110
8	9	420	420	1	94107
11	12	676	676	1	94107
35	36	485	485	1	94124
36	37	760	760	1	94124

In [70]:

```
ctip_enrolled = ctip_1.drop(['Round 1 Assignment'], axis=1)
```

In [71]:

```
ctip_enrolled.head()
```

Out[71]:

	StudentNo	School Enrolled In As Of 11/03/2017	Does Student Live In CTIP1 Zone? (y/n)	Student's Residential Zip Code
4	5	420	1	94110
8	9	420	1	94107
11	12	676	1	94107
35	36	485	1	94124
36	37	760	1	94124

In [72]:

```
ctip_enrolled.rename(columns={'School Enrolled In As Of 11/03/2017': 'district_code'}, inplace=True)
```

In [73]:

```
ctip_enrolled_w_school_zip = pd.merge(ctip_enrolled, codes, on='district_code', how='outer')
```

In [74]:

```
ctip_enrolled_w_school_zip.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 754 entries, 0 to 753
Data columns (total 10 columns):
StudentNo                739 non-null float64
district_code            727 non-null float64
Does Student Live In CTIP1 Zone? (y/n)  739 non-null float64
Student's Residential Zip Code  739 non-null float64
Unnamed: 0               754 non-null float64
```

```

unnamed: 0
school_name_x      694 non-null object
cds_code           694 non-null float64
school_name_y      690 non-null object
campus_address     690 non-null object
campus_zipcode     694 non-null float64
dtypes: float64(7), object(3)
memory usage: 64.8+ KB

```

In [75]:

```

ctip_enrolled_w_school_zip['campus_zipcode'] = ctip_enrolled_w_school_zip['campus_zipcode'].fillna(
0).astype(int)
ctip_enrolled_w_school_zip['campus_zipcode'] = ctip_enrolled_w_school_zip['campus_zipcode'].astype(
np.int64)
ctip_enrolled_w_school_zip['Student\'s Residential Zip Code'] =
ctip_enrolled_w_school_zip['Student\'s Residential Zip Code'].fillna(0).astype(int)
ctip_enrolled_w_school_zip['Student\'s Residential Zip Code'] =
ctip_enrolled_w_school_zip['Student\'s Residential Zip Code'].astype(np.int64)
ctip_enrolled_w_school_zip['cds_code'] = ctip_enrolled_w_school_zip['cds_code'].fillna(0).astype(in
t)
ctip_enrolled_w_school_zip['cds_code'] = ctip_enrolled_w_school_zip['cds_code'].astype(np.int64)

```

In [76]:

```
ctip_enrolled_w_school_zip.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 754 entries, 0 to 753
Data columns (total 10 columns):
StudentNo      739 non-null float64
district_code  727 non-null float64
Does Student Live In CTIP1 Zone? (y/n)  739 non-null float64
Student's Residential Zip Code  754 non-null int64
Unnamed: 0     694 non-null float64
school_name_x  694 non-null object
cds_code       754 non-null int64
school_name_y  690 non-null object
campus_address 690 non-null object
campus_zipcode 754 non-null int64
dtypes: float64(4), int64(3), object(3)
memory usage: 64.8+ KB

```

In [77]:

```
ctip_enrolled_w_school_zip.head()
```

Out[77]:

	StudentNo	district_code	Does Student Live In CTIP1 Zone? (y/n)	Student's Residential Zip Code	Unnamed: 0	school_name_x	cds_code	school_name_y	campus_address
0	5	420	1	94110	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
1	9	420	1	94107	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
2	56	420	1	94110	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
3	98	420	1	94110	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST

4	169	420	Does Student Live In	94107 Student's Residential	1 Unnamed:	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
	StudentNo	district_code				school_name_x	cds_code	school_name_y	campus_address

In [78]:

```
ctip_enrolled_w_school_zip['enrolled_in_home_zipcode'] =
np.where(ctip_enrolled_w_school_zip['Student\'s Residential Zip Code'] ==
ctip_enrolled_w_school_zip['campus_zipcode'], 0, 1)
```

In [79]:

```
ctip_enrolled_w_school_zip.head()
```

Out[79]:

	StudentNo	district_code	Does Student Live In CTIP1 Zone? (y/n)	Student's Residential Zip Code	Unnamed: 0	school_name_x	cds_code	school_name_y	campus_address
0	5	420	1	94110	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
1	9	420	1	94107	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
2	56	420	1	94110	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
3	98	420	1	94110	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST
4	169	420	1	94107	1	Alvarado ES	6040703	Alvarado Elementary School	625 DOUGLASS ST

In [80]:

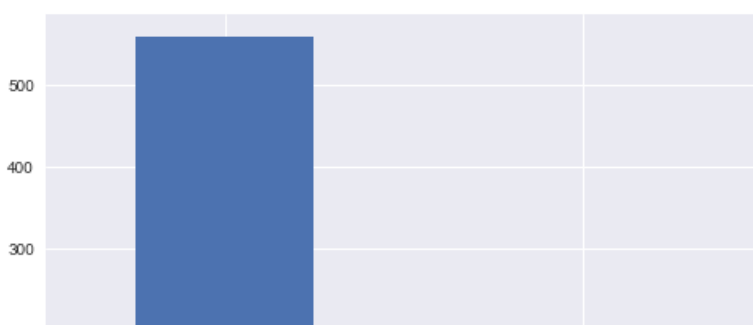
```
ctip_enrolled_w_school_zip.enrolled_in_home_zipcode.value_counts()
```

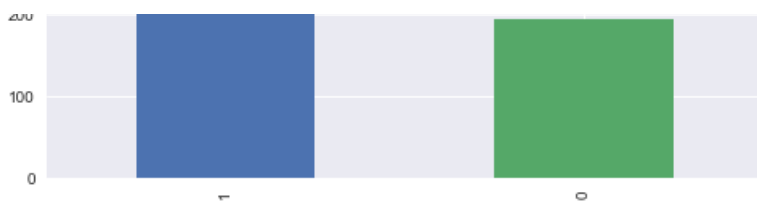
Out[80]:

```
1    559
0    195
Name: enrolled_in_home_zipcode, dtype: int64
```

In [82]:

```
ctip_enrolled_w_school_zip['enrolled_in_home_zipcode'].value_counts().plot(kind='bar');
```





In [ ]:

```
import pdfkit
pdfkit.from_file('online-job-postings.html', 'online-job-postings.pdf')
```