

## Investigate an SFUSD Dataset

## Table of Contents

- [Introduction](#)
- [Data Quality Assessment](#)
- [Data Wrangling](#)
- [Exploratory Data Analysis](#)

# Introduction

## Data Sources

- SFUSD Lottery parents group csv file, which compiles data from these sheets: <http://www.sfusd.edu/en/enroll-in-sfUSD-schools/monitoring-student-assignment/student-assignment-system.html?fbclid=IwAR1z3WYAGrmuigkqhIS45E3ScqsnZ5cLMqEoXw9NO8cVNpaWS2XQ8xoyBHw>

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
```

In [2]:

```
df = pd.read_csv('./data/SFUSD_parents_09_2018.csv')
df2 = df.copy()
```

## Data Quality Assessment

In [3]:

```
df.head()
```

Out[3]:

[illegible]

	Elementary School (Spanish Immersion)	SE Wait Pool Jul 2018	SE Wait Pool June 2018	SE 652 SN 116 Total Requests Round 1 (2018)	SE 68, 1st choice Requests Round 1 (2018)	SE Wait Pool June 2017	758.0 Total Requests Round 1 (2017)	115.0 1st choice Requests Round 1 (2017)	44 # of K Seats	0 # of TK Seats ***	...	citywide Type	Lick Middle Feeder	K-5 Grade
1	Alvarado													
2	Elementary School, GE	9	7	995	59	8	972.0	44.0	44	18	...	attendance	Lick	K-5
3	Argonne Elementary School, GE	8	10	1275	97	6	1104.0	83.0	66 2017-18, 88 2016-17 (alternates yearly)	18	...	attendance	Presidio	K-5
4	Bryant Elementary School, GE	0	0	219	3	0	149.0	5.0	22	0	...	attendance	Everett	K-5

5 rows × 22 columns

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 22 columns):
SCHOOL Name                133 non-null object
Wait Pool Jul 2018          116 non-null object
Wait Pool June 2018         111 non-null object
Total Requests Round 1 (2018) 116 non-null object
1st choice Requests Round 1 (2018) 116 non-null object
Wait Pool June 2017         110 non-null object
Total Requests Round 1 (2017) 110 non-null float64
1st choice Requests Round 1 2017 110 non-null float64
# of K Seats                117 non-null object
# of TK Seats ***           116 non-null object
August 2016 Wait Pool #    117 non-null object
# Enrolled 2016            110 non-null float64
Type                        114 non-null object
Middle Feeder              105 non-null object
Grades                     120 non-null object
Neighborhood               120 non-null object
Address                    120 non-null object
Phone                      120 non-null object
HOURS                      119 non-null object
Uniform                    113 non-null object
WEBSITE                    31 non-null object
Before School or After School Program 119 non-null object
dtypes: float64(3), object(19)
memory usage: 23.1+ KB
```

In [5]:

```
df.tail()
```

Out[5]:

	SCHOOL Name	Wait Pool Jul 2018	Wait Pool June 2018	Total Requests Round 1 (2018)	1st choice Requests Round 1	Wait Pool June 2017	Total Requests Round 1 (2017)	1st choice Requests Round 1	# of K Seats	# of TK Seats ***	...	Type	Middle Feeder	Grades	Neigh
--	-------------	--------------------	---------------------	-------------------------------	-----------------------------	---------------------	-------------------------------	-----------------------------	--------------	-------------------	-----	------	---------------	--------	-------

		(2018)	(2018)	(2018)	(2018)	(2017)	(2017)	(2017)		# of	# of	...		Middle	Grades	Neigh
129	(NS) Spanish Newcomer Pathway	Wait Pool Jul 2018	Wait Pool June 2018	Total Requests Round 1 (2018)	1st choice Requests Round 1 (2018)	Wait Pool June 2017	Total Requests Round 1 (2017)	1st choice Requests Round 1 2017	# of K Seats	TK Seats ***	...	Type	Feeder			
130	(NX) Mixed Language Newcomer Pathway	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
131	(SB) Spanish Bilingual	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
132	(SE) Spanish Immersion Non-Target Language	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
133	(SN) Spanish Immersion Target Language	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN

5 rows x 22 columns

In [6]:

```
df.sample()
```

Out[6]:

	SCHOOL Name	Wait Pool Jul 2018	Wait Pool June 2018	Total Requests Round 1 (2018)	1st choice Requests Round 1 (2018)	Wait Pool June 2017	Total Requests Round 1 (2017)	1st choice Requests Round 1 2017	# of K Seats	# of TK Seats ***	...	Type	Middle Feeder	Grades	Ne
2	Alvarado Elementary School, GE	9	7	995	59	8	972.0	44.0	44	18	...	attendance	Lick	K-5	Nc

1 rows x 22 columns

Quality Issues to review

- column names: lowercase, spaces, descriptive
- column order
- NaN, null, outlier entries
- data types
- utility of masks, bins
- consider dropping schools with null request, waitlist data

Data Wrangling

General Properties

In [7]:

```
print("Rows: ", df.shape[0])
print("Columns: ", df.shape[1])
```

```
Rows:    134
Columns: 22
```

In [8]:

```
df.describe()
```

Out[8]:

	Total Requests Round 1 (2017)	1st choice Requests Round 1 2017	# Enrolled 2016
count	110.000000	110.000000	110.000000
mean	441.927273	41.245455	393.600000
std	424.464276	41.554913	151.803724
min	1.000000	0.000000	47.000000
25%	114.500000	12.000000	266.000000
50%	268.000000	24.000000	361.500000
75%	708.000000	51.000000	517.000000
max	1610.000000	195.000000	700.000000

## Rename columns and replace characters

In [9]:

```
df.rename(columns={'Wait Pool Jul 2018' : 'Waitpool_7_18',
                  'Wait Pool June 2018': 'Waitpool_6_18',
                  'Total Requests Round 1 (2018)': 'Total_requests_Rnd1_18',
                  '1st choice Requests Round 1 (2018)': '1st_request_Rnd1_18',
                  'Wait Pool June 2017': 'Waitpool_6_17',
                  'Total Requests Round _2017': 'total_requests_Rnd1_17',
                  'Total Requests Round 1 (2017)': 'Total_requests_Rnd1_17',
                  '1st choice Requests Round 1 2017': '1st_request_Rnd1_17',
                  'August 2016 Wait Pool #' : 'Waitpool_8_16',
                  'Before School or After School Program': 'Before_After School Program'},
          inplace=True)
```

In [10]:

```
df.columns = [x.lower() for x in df.columns]
```

In [11]:

```
df.columns = [x.strip().replace('#', 'number') for x in df.columns]
```

In [12]:

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

In [13]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 22 columns):
school_name                133 non-null object
waitpool_7_18              116 non-null object
waitpool_6_18              111 non-null object
total_requests_rnd1_18     116 non-null object
```

```

1st_request_rnd1_18      116 non-null object
waitpool_6_17           110 non-null object
total_requests_rnd1_17   110 non-null float64
1st_request_rnd1_17      110 non-null float64
number_of_k_seats        117 non-null object
number_of_tk_seats_***   116 non-null object
waitpool_8_16            117 non-null object
number_enrolled_2016     110 non-null float64
type                     114 non-null object
middle_feeder            105 non-null object
grades                   120 non-null object
neighborhood             120 non-null object
address                  120 non-null object
phone                    120 non-null object
hours                    119 non-null object
uniform                  113 non-null object
website                  31 non-null object
before_after_school_program 119 non-null object
dtypes: float64(3), object(19)
memory usage: 23.1+ KB

```

## Reorder data columns

In [14]:

```

df.reindex(sorted(df.columns), axis=1)
#https://stackoverflow.com/questions/11067027/python-pandas-re-ordering-columns-in-a-dataframe-based-on-column-name

```

Out[14]:

	1st_request_rnd1_17	1st_request_rnd1_18	address	before_after_school_program	grades	hours	middle_feeder
0	89.0	84	250 23rd Avenue	Monday - Friday: RDASC (Richmond District Afte...	K-5	8:40 a.m. - 2:40 p.m.	Presidio
1	115.0	SE 68, SN 47	625 Douglass Street	Mission Graduates	K-5	7:50 a.m.- 1:50 p.m.	Lick
2	44.0	59	625 Douglass Street	Mission Graduates	K-5	7:50 a.m.- 1:50 p.m.	Lick
3	83.0	97	680 18th Avenue	2:40-6:00 Richmond YMCA, Richmond District Aft...	K-5	8:40 am - 2:40 pm	Presidio
4	5.0	3	2641 25th Street	SFUSD Early Education Department Out of School...	K-5	7:50am-2:05pm; Thu 12:50pm	Everett
5	17.0	27	2641 25th Street	SFUSD Early Education Department Out of School...	K-5	7:50am-2:05pm; Thu 12:50pm	Everett
6	99.0	SE 34, SN 32	3351 23rd Street	BVCC runs a before-school program from 7:00 to...	K-8	9:30am-3:45pm, Thu 2:30pm	NaN
7	8.0	8	375 7th Street	(ExCEL Program), operated on-site by YMCA Emba...	K-8	8:40 a.m. - 2:40	NaN

	1st_request_rnd1_17	1st_request_rnd1_18	address	before_after_school_program	grades	p.hours	middle_feeder
8	39.0	49	375 7th Street	(ExCEL Program), operated on-site by YMCA Emba...	K-8	8:40 a.m. - 2:40 p.m.	NaN
9	6.0	15	1360 Oakdale Avenue	ExCEL after-school program sponsored by the Ba...	K-5	8:30 a.m. - 2:40 p.m.	Aptos, Brown
10	11.0	11	825 Shotwell Street	supervised recess from 8:00 - 8:35 - Playworks...	PreK-5	8:40am-2:55pm; 1:40pm Thu	Everett
11	18.0	17	825 Shotwell Street	supervised recess from 8:00 - 8:35 - Playworks...	PreK-5	8:40am-2:55pm; 1:40pm Thu	Everett
12	49.0	40	350 Broadway Street	3:35 to 5:30 PM: fee based. Run by the Chinato...	K-5	9:25AM - 3:35PM	Francisco
13	20.0	15	350 Broadway Street	3:35 to 5:30 PM: fee based. Run by the Chinato...	K-5	9:25AM - 3:35PM	Francisco
14	65.0	CE 42, CN 48	1250 Waller Street	7-8:40am, 2:40-6pm YMCA Buchanan (waitlist)	K-5	8:40 am - 2:40 pm	Roosevelt
15	144.0	122	500 Clarendon Avenue	Before care starts at 7am and after care until...	K-5	9:25AM - 3:35PM	Presidio
16	62.0	JE 75, JN 24	500 Clarendon Avenue	Before care starts at 7am and after care until...	K-5	9:25AM - 3:35PM	Presidio
17	7.0	12	455 Athens Street	ExCEL program offered at sliding scale to fami...	K-5	9:30 a.m. - 3:30 p.m	Vis Valley
18	17.0	19	455 Athens Street	ExCEL program offered at sliding scale, from e...	K-5	9:30 a.m. - 3:30 p.m	Vis Valley
19	7.0	16	2725 California Street	Sliding scale aftercare (\$0-\$500) is available...	PreK-5	8:40 am - 2:40 pm	Roosevelt
20	NaN	NaN	1601 Turk Street	Campus yard open at 8am; aftercare program ons...	K-8	MWThF K: 8:45-2:25, 1st-5th 8:45-3:35, 6th-	NaN

	1st_request_rnd1_17	1st_request_rnd1_18	address	before_after_school_program	grades	8th... hours	middle_feeder
21	19.0	21	50 Pomona Avenue	The after-school programs extend the academic ...	PreK-5	7:50am-1:50pm, Thu 1 pm	Giannini, Brown
22	8.0	13	70 Delta Street	Real Options for City Kids (ROCK) M-F 1:50 to ...	K-5	7:50 a.m. - 1:50 p.m.	Vis Valley
23	75.0	74	2550 25th Avenue	YMCA	K-5	7:50 am - 1:50 pm	Aptos
24	14.0	10	3125 Cesar Chavez Street	Mission YMCA, 7:30 AM to 6:30 PM\nExCEL progra...	K-5	8:40 a.m. - 2:40 p.m.	Lick
25	32.0	SE 20, SN 12	3125 Cesar Chavez Street	Mission YMCA, 7:30 AM to 6:30 PM\nExCEL progra...	K-5	8:40 a.m. - 2:40 p.m.	Lick
26	10.0	10	420 Filbert Street	Garfield afterschool program: 1:50 to 6:00 p.m...	K-5	7:50 pm - 1:50 pm	Francisco
27	15.0	CE 11, CE 4	420 Filbert Street	Garfield afterschool program: 1:50 to 6:00 p.m...	K-5	7:50 pm - 1:50 pm	Francisco
28	34.0	45	151 Lippard Avenue	ExCEL program is from 2:40-6:00 PM each day. T...	K-5	8:40 am - 2:40 pm	Lick
29	17.0	8	151 Lippard Avenue	ExCEL program is from 2:40-6:00 PM each day. T...	K-5	8:40 am - 2:40 pm	Lick
...	...	...	...	...	...	...	...
104	41.0	36	423 Burrows Street	ExCEL program offered at sliding scale to fami...	K-5	8:40 am - 2:40 pm	King
105	43.0	45	423 Burrows Street	ExCEL program offered at sliding scale to fami...	K-5	8:40 am - 2:40 pm	King
106	24.0	18	423 Burrows Street	ExCEL program offered at sliding scale to fami...	K-5	8:40 am - 2:40 pm	King
107	32.0	36	627 Turk Street	Up on Top, Monday through Friday, 2:40 to 6:00...	PreK-5	8:40am-2:45pm, Tue 2:15pm	Francisco
108	NaN	NaN	3531 - 22nd St	Mission Graduates	TK-8	8:20 - 3pm except Wed	NaN

	1st_request_rnd1_17	1st_request_rnd1_18	address	before_after_school_program	grades	hours	middle_feeder
						8:30am - 1pm	
109	NaN	22	2110 Greenwich Street	2:30-5:30pm OST. No before care	preK-TK	8:30 am - 2:30 pm	Sherman Elem
110	50.0	41	2650 42nd Avenue	ExCEL After School Program in partnership with...	K-5	9:30 am - 3:30 pm	Hoover
111	36.0	45	2650 42nd Avenue	ExCEL After School Program in partnership with...	K-5	9:30 am - 3:30 pm	Hoover
112	8.0	13	55 Schwerin Street	ExCEL after School Program in partnership with...	K-5	8:40 a.m. - 2:40 p.m.	Vis Valley
113	15.0	19	55 Schwerin Street	ExCEL after School Program in partnership with...	K-5	8:40 a.m. - 2:40 p.m.	Vis Valley
114	48.0	SE 65, SN 11	465 Missouri Street	After: After School Enrichment Programs (ASEP)...	K-5	8:40 am - 2:40 pm	Lick
115	21.0	35	465 Missouri Street	After: After School Enrichment Programs (ASEP)...	K-5	8:40 am - 2:40 pm	Lick
116	68.0	CE 30, CN 33	5 Lenox Way	Before: 7:00- 8:40AM. After: 2:40-6:00PM with...	K-5	8:40 am - 2:30 pm K-3. 8:40 – 2:40 4th-5th	Hoover
117	77.0	72	5 Lenox Way	Before: 7:00- 8:40AM. After: 2:40-6:00PM with...	K-5	8:40 am - 2:30 pm K-3. 8:40 – 2:40 4th-5th	Hoover
118	31.0	40	2245 Jones Street	on-site after care 3:30 - 6:30pm	K-5	9:30 AM - 3:30 PM	Francisco
119	195.0	CE 69, CT 114	1541 - 12th Avenue	Private non-profit childcare available on site.	K-8	9:30 am - 3:30 pm	NaN
120	NaN	NaN	NaN	NaN	NaN	NaN	NaN
121	NaN	NaN	NaN	NaN	NaN	NaN	NaN
122	NaN	NaN	NaN	NaN	NaN	NaN	NaN



	1st_request_rnd1_17	1st_request_rnd1_18	address	before_after_school_program	grades	hours	middle_feeder
123	NaN	NaN	NaN	NaN	NaN	NaN	NaN
124	NaN	NaN	NaN	NaN	NaN	NaN	NaN
125	NaN	NaN	NaN	NaN	NaN	NaN	NaN
126	NaN	NaN	NaN	NaN	NaN	NaN	NaN
127	NaN	NaN	NaN	NaN	NaN	NaN	NaN
128	NaN	NaN	NaN	NaN	NaN	NaN	NaN
129	NaN	NaN	NaN	NaN	NaN	NaN	NaN
130	NaN	NaN	NaN	NaN	NaN	NaN	NaN
131	NaN	NaN	NaN	NaN	NaN	NaN	NaN
132	NaN	NaN	NaN	NaN	NaN	NaN	NaN
133	NaN	NaN	NaN	NaN	NaN	NaN	NaN

134 rows × 22 columns

In [15]:

```
df.sort_index(axis=1, inplace=True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 22 columns):
1st_request_rnd1_17      110 non-null float64
1st_request_rnd1_18      116 non-null object
address                  120 non-null object
before_after_school_program  119 non-null object
grades                   120 non-null object
hours                    119 non-null object
middle_feeder            105 non-null object
neighborhood             120 non-null object
number_enrolled_2016     110 non-null float64
number_of_k_seats        117 non-null object
number_of_tk_seats_***   116 non-null object
phone                    120 non-null object
school_name              133 non-null object
total_requests_rnd1_17    110 non-null float64
total_requests_rnd1_18    116 non-null object
type                     114 non-null object
```

```
uniform                113 non-null object
waitpool_6_17          110 non-null object
waitpool_6_18          111 non-null object
waitpool_7_18          116 non-null object
waitpool_8_16          117 non-null object
website                31 non-null object
dtypes: float64(3), object(19)
memory usage: 23.1+ KB
```

In [16]:

```
cols = list(df.columns.values)
print(cols)
```

```
['1st_request_rnd1_17', '1st_request_rnd1_18', 'address', 'before_after_school_program', 'grades',
'hours', 'middle_feeder', 'neighborhood', 'number_enrolled_2016', 'number_of_k_seats',
'number_of_tk_seats_***', 'phone', 'school_name', 'total_requests_rnd1_17',
'total_requests_rnd1_18', 'type', 'uniform', 'waitpool_6_17', 'waitpool_6_18', 'waitpool_7_18', 'w
aitpool_8_16', 'website']
```

In [17]:

```
df = df[['school_name', 'type', 'grades',
'hours', 'address', 'number_of_k_seats', 'number_of_tk_seats_***', 'number_enrolled_2016', 'before_after
_school_program', 'neighborhood', '1st_request_rnd1_17', '1st_request_rnd1_18',
'total_requests_rnd1_17', 'total_requests_rnd1_18', 'waitpool_6_17', 'waitpool_6_18',
'waitpool_7_18', 'waitpool_8_16', 'website', 'phone', 'middle_feeder', 'uniform']]
#https://stackoverflow.com/questions/13148429/how-to-change-the-order-of-dataframe-columns
```

## Convert data types

In [18]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 22 columns):
school_name                133 non-null object
type                      114 non-null object
grades                    120 non-null object
hours                     119 non-null object
address                   120 non-null object
number_of_k_seats         117 non-null object
number_of_tk_seats_***     116 non-null object
number_enrolled_2016      110 non-null float64
before_after_school_program 119 non-null object
neighborhood              120 non-null object
1st_request_rnd1_17        110 non-null float64
1st_request_rnd1_18        116 non-null object
total_requests_rnd1_17     110 non-null float64
total_requests_rnd1_18     116 non-null object
waitpool_6_17             110 non-null object
waitpool_6_18             111 non-null object
waitpool_7_18             116 non-null object
waitpool_8_16             117 non-null object
website                   31 non-null object
phone                     120 non-null object
middle_feeder             105 non-null object
uniform                   113 non-null object
dtypes: float64(3), object(19)
memory usage: 23.1+ KB
```

In [19]:

```
df = df.convert_objects(convert_numeric=True)
df.dtypes
#https://stackoverflow.com/questions/28277137/how-to-convert-datatypeobject-to-float64-in-python
#https://stackoverflow.com/questions/15891038/change-data-type-of-columns-in-pandas/28648923
```

/Users/Trene/anaconda/lib/python3.6/site-packages/inukernel/ main.py:1: FutureWarning:

```
/Users/llene/anaconda/lib/python3.6/site-packages/ipykernel/_main__.py:1: 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.
if __name__ == '__main__':
```

Out[19]:

```
school_name      object
type             object
grades          object
hours           object
address         object
number_of_k_seats float64
number_of_k_seats *** float64
number_enrolled_2016 float64
before_after_school_program object
neighborhood     object
1st_request_rnd1_17 float64
1st_request_rnd1_18 float64
total_requests_rnd1_17 float64
total_requests_rnd1_18 float64
waitpool_6_17    float64
waitpool_6_18    float64
waitpool_7_18    float64
waitpool_8_16    float64
website         object
phone          object
middle_feeder   object
uniform        object
dtype: object
```

In [20]:

```
df.type = df.type.astype('category')
df.grades = df.grades.astype('category')
```

In [21]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 22 columns):
school_name      133 non-null object
type            114 non-null category
grades          120 non-null category
hours           119 non-null object
address         120 non-null object
number_of_k_seats 85 non-null float64
number_of_k_seats *** 114 non-null float64
number_enrolled_2016 110 non-null float64
before_after_school_program 119 non-null object
neighborhood     120 non-null object
1st_request_rnd1_17 110 non-null float64
1st_request_rnd1_18 98 non-null float64
total_requests_rnd1_17 110 non-null float64
total_requests_rnd1_18 98 non-null float64
waitpool_6_17    94 non-null float64
waitpool_6_18    94 non-null float64
waitpool_7_18    98 non-null float64
waitpool_8_16    86 non-null float64
website         31 non-null object
phone          120 non-null object
middle_feeder   105 non-null object
uniform        113 non-null object
dtypes: category(2), float64(11), object(9)
memory usage: 21.8+ KB
```

In [22]:

```
for feature in ['type', 'grades']:
    print("{}: {}".format(feature, df[feature].unique()))
```

```
type: [attendance, citywide, charter, NaN]
Categories (3, object): [attendance, citywide, charter]
grades: [K-5, K-8, PreK-5, preK-TK, TK-5, K-3 (goal to be a K-12), PreK-8, prek-5, TK-8, NaN]
Categories (9, object): [K-5, K-8, PreK-5, preK-TK, ..., K-3 (goal to be a K-12), PreK-8, prek-5, TK-8]
```

## Select the data that I will be using

In [23]:

```
#seperate out the columns I don't need, and keep the school_name in that df
#https://www.shanelynn.ie/select-pandas-dataframe-rows-and-columns-using-iloc-loc-and-ix/
#also, create a new table with columns I don't need in this table: https://www.youtube.com/watch?t
ime_continue=85&v=LhdGQC_vjEs
object_info_df = df.iloc[:, [0,3,4,9,18,19,20,21]]
object_info_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 8 columns):
school_name      133 non-null object
hours            119 non-null object
address          120 non-null object
neighborhood     120 non-null object
website         31 non-null object
phone            120 non-null object
middle_feeder    105 non-null object
uniform          113 non-null object
dtypes: object(8)
memory usage: 8.5+ KB
```

In [24]:

```
#seperate out the data columns that I will be working with
#https://www.shanelynn.ie/select-pandas-dataframe-rows-and-columns-using-iloc-loc-and-ix/
data = df.iloc[:, [0,1,2,5,7,8,10,11,12,13,14,15,16,17]]
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 14 columns):
school_name      133 non-null object
type             114 non-null category
grades           120 non-null category
number_of_k_seats 85 non-null float64
number_enrolled_2016 110 non-null float64
before_after_school_program 119 non-null object
1st_request_rnd1_17 110 non-null float64
1st_request_rnd1_18 98 non-null float64
total_requests_rnd1_17 110 non-null float64
total_requests_rnd1_18 98 non-null float64
waitpool_6_17     94 non-null float64
waitpool_6_18     94 non-null float64
waitpool_7_18     98 non-null float64
waitpool_8_16     86 non-null float64
dtypes: category(2), float64(10), object(2)
memory usage: 13.4+ KB
```

## Check data entries: Duplicate, Null, Unique

In [25]:

```
data.duplicated().sum()
```

Out[25]:

0

In [26]:

```
data.nunique().sum()
```

Out[26]:

701

In [27]:

```
data.isnull().any()
```

Out[27]:

```
school_name      True
type              True
grades            True
number_of_k_seats True
number_enrolled_2016 True
before_after_school_program True
1st_request_rnd1_17 True
1st_request_rnd1_18 True
total_requests_rnd1_17 True
total_requests_rnd1_18 True
waitpool_6_17    True
waitpool_6_18    True
waitpool_7_18    True
waitpool_8_16    True
dtype: bool
```

## Manage NaN entries

In [28]:

```
null_columns=data.columns[data.isnull().any()]
data>null_columns].isnull().sum()
```

Out[28]:

```
school_name      1
type              20
grades            14
number_of_k_seats 49
number_enrolled_2016 24
before_after_school_program 15
1st_request_rnd1_17 24
1st_request_rnd1_18 36
total_requests_rnd1_17 24
total_requests_rnd1_18 36
waitpool_6_17    40
waitpool_6_18    40
waitpool_7_18    36
waitpool_8_16    48
dtype: int64
```

In [29]:

```
print(data[data.isnull().any(axis=1)][null_columns].head())
```

```
      school_name      type grades \
1  Alvarado Elementary School (Spanish Immersion)  citywide  K-5
2           Alvarado Elementary School, GE  attendance  K-5
3           Argonne Elementary School, GE  attendance  K-5
6  Buena Vista Horace Mann, SPN Immersion  citywide  K-8
7 Carmichael, Bessie (K-5 Campus), FIL (Eng Lang...  citywide  K-8

      number_of_k_seats  number_enrolled_2016  \
1           44.0           535.0
2           44.0           535.0
3           NaN           422.0
6           66.0           600.0
7           NaN           653.0
```

	before_after_school_program	1st_request_rnd1_17	\
1	Mission Graduates	115.0	
2	Mission Graduates	44.0	
3	2:40-6:00 Richmond YMCA, Richmond District Aft...	83.0	
6	BVCC runs a before-school program from 7:00 to...	99.0	
7	(ExCEL Program), operated on-site by YMCA Emba...	8.0	

	1st_request_rnd1_18	total_requests_rnd1_17	total_requests_rnd1_18	\
1	NaN	758.0	NaN	
2	59.0	972.0	995.0	
3	97.0	1104.0	1275.0	
6	NaN	575.0	NaN	
7	8.0	34.0	39.0	

	waitpool_6_17	waitpool_6_18	waitpool_7_18	waitpool_8_16
1	NaN	NaN	NaN	NaN
2	8.0	7.0	9.0	NaN
3	6.0	10.0	8.0	NaN
6	NaN	NaN	NaN	NaN
7	0.0	0.0	0.0	NaN

In [30]:

```
print(data[data["1st_request_rnd1_17"].isnull()][null_columns])
```

	school_name	type	\
20	Creative Arts Charter School	charter	
55	McLaren, John Early Education School	NaN	
59	Mission Preparatory School	charter	
69	New School	charter	
71	Noriega Early Education School	NaN	
79	Presidio Early Education School	NaN	
83	Rodriguez, Zaida T. Early Education School	NaN	
99	Stockton, Commodore Early Education School	NaN	
108	Thomas Edison Charter Academy (TECA) - 2/3 Spa...	charter	
109	Tule Elk Park Early Education School	NaN	
120	NaN	NaN	
121	(CB) Chinese Biliteracy	NaN	
122	(CE) Chinese Immersion Non-Target Language	NaN	
123	(CN) Chinese Immersion Target Language	NaN	
124	(GE) General Education	NaN	
125	(JB) Japanese Bi-Cultural Program	NaN	
126	(KE) Korean Immersion Non-Target Language	NaN	
127	(KN) Korean Immersion Target Language	NaN	
128	(NC) Chinese Newcomer Pathway	NaN	
129	(NS) Spanish Newcomer Pathway	NaN	
130	(NX) Mixed Language Newcomer Pathway	NaN	
131	(SB) Spanish Biliteracy	NaN	
132	(SE) Spanish Immersion Non-Target Language	NaN	
133	(SN) Spanish Immersion Target Language	NaN	

	grades	number_of_k_seats	number_enrolled_2016	\
20	K-8	44.0	440.0	
55	preK-TK	0.0	NaN	
59	TK-5	NaN	NaN	
69	K-3 (goal to be a K-12)	48.0	NaN	
71	preK-TK	0.0	NaN	
79	preK-TK	NaN	NaN	
83	preK-TK	0.0	NaN	
99	preK-TK	0.0	NaN	
108	TK-8	66.0	700.0	
109	preK-TK	66.0	NaN	
120	NaN	NaN	NaN	
121	NaN	NaN	NaN	
122	NaN	NaN	NaN	
123	NaN	NaN	NaN	
124	NaN	NaN	NaN	
125	NaN	NaN	NaN	
126	NaN	NaN	NaN	
127	NaN	NaN	NaN	
128	NaN	NaN	NaN	
129	NaN	NaN	NaN	
130	NaN	NaN	NaN	
131	NaN	NaN	NaN	
132	NaN	NaN	NaN	

133	NaN	NaN	NaN
	before_after_school_program	1st_request_rnd1_17	\
20	Campus yard open at 8am; aftercare program ons...	NaN	
55	2:30-5:30pm. Our after-school program features...	NaN	
59	NaN	NaN	
69	Onsite before school care beginning at 7:45am ...	NaN	
71	OST 2-6pm. Our after-school program feeder sch...	NaN	
79	OST 2-6pm	NaN	
83	none	NaN	
99	The SFUSD Early Education Department administe...	NaN	
108	Mission Graduates	NaN	
109	2:30-5:30pm OST. No before care	NaN	
120	NaN	NaN	
121	NaN	NaN	
122	NaN	NaN	
123	NaN	NaN	
124	NaN	NaN	
125	NaN	NaN	
126	NaN	NaN	
127	NaN	NaN	
128	NaN	NaN	
129	NaN	NaN	
130	NaN	NaN	
131	NaN	NaN	
132	NaN	NaN	
133	NaN	NaN	
	1st_request_rnd1_18	total_requests_rnd1_17	total_requests_rnd1_18 \
20	NaN	NaN	NaN
55	13.0	NaN	90.0
59	NaN	NaN	NaN
69	NaN	NaN	NaN
71	85.0	NaN	190.0
79	36.0	NaN	179.0
83	12.0	NaN	130.0
99	28.0	NaN	84.0
108	NaN	NaN	NaN
109	22.0	NaN	149.0
120	NaN	NaN	NaN
121	NaN	NaN	NaN
122	NaN	NaN	NaN
123	NaN	NaN	NaN
124	NaN	NaN	NaN
125	NaN	NaN	NaN
126	NaN	NaN	NaN
127	NaN	NaN	NaN
128	NaN	NaN	NaN
129	NaN	NaN	NaN
130	NaN	NaN	NaN
131	NaN	NaN	NaN
132	NaN	NaN	NaN
133	NaN	NaN	NaN
	waitpool_6_17	waitpool_6_18	waitpool_7_18 waitpool_8_16
20	NaN	NaN	NaN
55	NaN	0.0	0.0 NaN
59	NaN	NaN	NaN
69	NaN	NaN	NaN
71	NaN	NaN	19.0 NaN
79	NaN	NaN	6.0 NaN
83	NaN	NaN	0.0 NaN
99	NaN	NaN	4.0 NaN
108	NaN	NaN	NaN
109	NaN	NaN	0.0 NaN
120	NaN	NaN	NaN
121	NaN	NaN	NaN
122	NaN	NaN	NaN
123	NaN	NaN	NaN
124	NaN	NaN	NaN
125	NaN	NaN	NaN
126	NaN	NaN	NaN
127	NaN	NaN	NaN
128	NaN	NaN	NaN
129	NaN	NaN	NaN
130	NaN	NaN	NaN
131	NaN	NaN	NaN

132	NaN	NaN	NaN	NaN
133	NaN	NaN	NaN	NaN

In [31]:

```
#http://www.residentmar.io/2016/06/12/null-and-missing-data-python.html
#https://stackoverflow.com/questions/40408471/select-data-when-specific-columns-have-null-value-in-pandas
data[data.filter(like='waitpool').isnull().all(1)]
```

Out[31]:

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_request
1	Alvarado Elementary School (Spanish Immersion)	citywide	K-5	44.0	535.0	Mission Graduates	115.0
6	Buena Vista Horace Mann, SPN Immersion	citywide	K-8	66.0	600.0	BVCC runs a before-school program from 7:00 to...	99.0
14	Chinese Immersion School at DeAvila (CIS), Can...	citywide	K-5	66.0	405.0	7-8:40am, 2:40-6pm YMCA Buchanan (waitlist)	65.0
20	Creative Arts Charter School	charter	K-8	44.0	440.0	Campus yard open at 8am; aftercare program ons...	NaN
25	Flynn, Leonard Elementary School, Spanish Imme...	citywide	K-5	NaN	NaN	Mission YMCA, 7:30 AM to 6:30 PM\nExCEL progra...	32.0
37	Huerta, Dolores Elementary School/ formerly Fa...	citywide	PreK-5	66.0	387.0	Before-School YMCA - Fee base 7:00 - 8:15\nYMC...	51.0
47	Lilienthal, Claire (K-2 Madison Campus), Korea...	citywide	K-8	22.0	674.0	Presidio YMCA for onsite aftercare. Presidio E...	39.0
52	Marshall Elementary School, Spanish Immersion	citywide	K-5	44.0	240.0	Mission Graduates, sliding scale for students ...	59.0
59	Mission Preparatory School	charter	TK-5	NaN	NaN	NaN	NaN
62	Monroe Elementary School, Spanish Immersion	citywide	K-5	NaN	552.0	After-school programs run from 2:25 to 6:00 p....	55.0
			K-3				



69	school name	charter	type (goal to be a K-12)	number_of_k_seats	Number_enrolled_2016	Onsite before school care beginning at 7:45am ...	request
72	Ortega, Jose Elementary School (Mandarin Immer...	citywide	PreK-5	NaN	302.0	Mon-Fri, 1:50 to 6:00 PM. Umoja After School P...	60.0
82	Revere, Paul, Spanish Immersion	citywide	PreK-8	NaN	428.0	1:50-6pm - ExCEL Extended Day program through ...	21.0
97	Starr King Elementary School, Mandarin Immersion	citywide	prek-5	44.0	NaN	YMCA Urban Services , 2:40 - 6:00pm, also have...	32.0
108	Thomas Edison Charter Academy (TECA) - 2/3 Spa...	charter	TK-8	66.0	700.0	Mission Graduates	NaN
114	Webster, Daniel Elementary School (Spanish Im...	citywide	K-5	44.0	273.0	After: After School Enrichment Programs (ASEP)...	48.0
116	West Portal Elementary School (Chinese Imm)	citywide	K-5	33.0	585.0	Before: 7:00- 8:40AM. After: 2:40-6:00PM with...	68.0
119	Yu, Alice Fong, Cantonese Immersion	citywide	K-8	66.0	587.0	Private non-profit childcare available on site.\n	195.0
120	NaN	NaN	NaN	NaN	NaN	NaN	NaN
121	(CB) Chinese Billiteracy	NaN	NaN	NaN	NaN	NaN	NaN
122	(CE) Chinese Immersion Non-Target Language	NaN	NaN	NaN	NaN	NaN	NaN
123	(CN) Chinese Immersion Target Language	NaN	NaN	NaN	NaN	NaN	NaN
124	(GE) General Education	NaN	NaN	NaN	NaN	NaN	NaN
125	(JB) Japanese Bi-Cultural Program	NaN	NaN	NaN	NaN	NaN	NaN
126	(KE) Korean Immersion Non-Target Language	NaN	NaN	NaN	NaN	NaN	NaN
127	(KN) Korean Immersion Target Language	NaN	NaN	NaN	NaN	NaN	NaN

	school_name (NC) Chinese	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_request
128	Newcomer Pathway	NaN	NaN	NaN	NaN	NaN	NaN
129	(NS) Spanish Newcomer Pathway	NaN	NaN	NaN	NaN	NaN	NaN
130	(NX) Mixed Language Newcomer Pathway	NaN	NaN	NaN	NaN	NaN	NaN
131	(SB) Spanish Biliteryacy	NaN	NaN	NaN	NaN	NaN	NaN
132	(SE) Spanish Immersion Non-Target Language	NaN	NaN	NaN	NaN	NaN	NaN
133	(SN) Spanish Immersion Target Language	NaN	NaN	NaN	NaN	NaN	NaN

In [32]:

```
#https://jakevdp.github.io/PythonDataScienceHandbook/03.04-missing-values.html
#data[data.notnull()]
data[data.filter(like='waitpool').notnull().all(1)]
```

Out[32]:

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_request
0	Alamo Elementary School, GE	attendance	K-5	88.0	530.0	Monday - Friday: RDASC (Richmond District Afte...	89.0
4	Bryant Elementary School, GE	attendance	K-5	22.0	216.0	SFUSD Early Education Department Out of School...	5.0
5	Bryant Elementary School, SPN (Eng Lang Learner)	citywide	K-5	22.0	216.0	SFUSD Early Education Department Out of School...	17.0
9	Carver, Dr. George Washington Elementary Schoo...	attendance	K-5	44.0	282.0	ExCEL after-school program sponsored by the Ba...	6.0
10	Chávez, César Elementary School, GE	attendance	PreK-5	NaN	494.0	supervised recess from 8:00 - 8:35 - Playworks...	11.0
11	Chávez, César Elementary School, SPN (Eng Lang...	citywide	PreK-5	NaN	494.0	supervised recess from 8:00 - 8:35 - Playworks...	18.0
12	Chin, John Yehall Elementary School, CHN	citywide	K-5	22.0	264.0	3:35 to 5:30 PM: fee based. Run by the Chinato...	49.0

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_requ
13	Chin, John Yehall Elementary School, GE	attendance	K-5	22.0	264.0	3:35 to 5:30 PM: fee based. Run by the Chinato...	20.0
15	Clarendon Elementary School (GE)	attendance	K-5	44.0	272.0	Before care starts at 7am and after care until...	144.0
17	Cleveland Elementary School, GE	attendance	K-5	NaN	308.0	ExCEL program offered at sliding scale to fami...	7.0
18	Cleveland Elementary School, SPN (Eng Lang Lea...	citywide	K-5	NaN	308.0	ExCEL program offered at sliding scale, from e...	17.0
19	Cobb, Dr. William L. Elementary School, GE	attendance	PreK-5	44.0	204.0	Sliding scale aftercare (\$0-\$500) is available...	7.0
21	Drew, Dr. Charles R. Elementary School, GE	attendance	PreK-5	34.0	208.0	The after-school programs extend the academic ...	19.0
22	El Dorado Elementary School, GE	attendance	K-5	44.0	274.0	Real Options for City Kids (ROCK) M-F 1:50 to ...	8.0
23	Feinstein, Dianne Elementary School, GE	attendance	K-5	88.0	518.0	YMCA	75.0
26	Garfield Elementary School, GE	attendance	K-5	22.0	250.0	Garfield afterschool program: 1:50 to 6:00 p.m...	10.0
28	Glen Park Elementary School, GE	attendance	K-5	NaN	329.0	ExCEL program is from 2:40-6:00 PM each day. T...	34.0
29	Glen Park Elementary School, SPN (Eng Lang Lea...	citywide	K-5	NaN	329.0	ExCEL program is from 2:40-6:00 PM each day. T...	17.0
30	Grattan Elementary School, GE	attendance	PreK-5	66.0	359.0	Grattan After School Program (GASP) is a slidi...	147.0
31	Guadalupe Elementary School, GE	attendance	K-5	NaN	445.0	SHPD ExCEL after-school offered at sliding sca...	20.0
32	Guadalupe Elementary School, SPN (Eng Lang Lea...	citywide	K-5	NaN	445.0	SHPD ExCEL after-school offered at sliding sca...	14.0
33	Harte, Bret Elementary School, GE	attendance	PreK-5	22.0	253.0	ExCEL: no-cost program to families of students...	7.0
	Hillcrest Elementary						

35	School Name	citywide	type	Grades	Number_of_k_seats	number_enrolled_2016	After-school program for 250 K-5th graders fro...	3st_requ
	(Eng Lang Lea...							
36	Hillcrest Elementary School, GE	attendance	K-5	NaN	493.0	After-school program for 250 K-5th graders fro...	12.0	
38	Jefferson Elementary School, GE	attendance	K-5	88.0	493.0	Before: 7:45 to 8:40 a.m.\nAfter: SFUSD Early ...	47.0	
39	Key, Francis Scott Elementary School, GE	attendance	K-5	88.0	534.0	ExCEL After School Program, Grades 2-5; 1:50 -...	48.0	
40	Lafayette Elementary School, GE	attendance	PreK-5	88.0	501.0	ExCEL after-school learning program: grades 1-...	91.0	
41	Lakeshore Elementary School, GE	attendance	K-5	88.0	553.0	ExCEL after-school program: grades K-5, 8am-9:...	51.0	
42	Lau, Gordon J. Elementary School, CHN (Eng Lan...	citywide	K-5	88.0	670.0	After-care programs offered through: Asian Wom...	49.0	
43	Lau, Gordon J. Elementary School, GE	attendance	K-5	44.0	670.0	After-care programs offered through: Asian Wom...	16.0	
...	...	...	...	...	...	...	...	...
75	Parker, Jean Elementary School, GE	attendance	K-5	22.0	263.0	2:40-5:45 PM: Daily snacks, homework support a...	8.0	
76	Parks, Rosa Elementary School, GE	attendance	K-5	44.0	390.0	"Raphael Weill Early Education School - 415-74...	21.0	
81	Revere, Paul, GE	attendance	PreK-8	NaN	428.0	1:50-6pm - ExCEL Extended Day program through ...	8.0	
84	Rooftop (PreK-4 Burnett Campus), GE	citywide	PreK-8	66.0	600.0	CASA: A nonprofit fee-based program	150.0	
85	S.F. Community, GE	citywide	K-8	44.0	286.0	City Scholars ExCEL ASP open to students K-8 f...	30.0	
86	Sanchez Elementary School, GE	attendance	PreK-5	22.0	266.0	Mission Graduates	2.0	
87	Sanchez Elementary School, SPN (Eng Lang Learn...	citywide	PreK-5	22.0	266.0	Mission Graduates	6.0	
89	Serra, Junipero Elementary School, SPN (Eng La...	citywide	K-5	22.0	269.0	After School Success Club (ExCEL After School ...	22.0	
91	Sherman Elementary School, GE	attendance	K-5	66.0	428.0	Sherman afterschool program for all students f...	94.0	

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_requ
92	Commodore Elementary School, GE	attendance	K-5	66.0	396.0	Anchor Program: Mon. to Fri., 2:40 to 6:30 PM\n	63.0
93	Spring Valley Elementary School, CHN (Eng Lang...	citywide	K-5	22.0	343.0	SFUSD Early Education Department Out of School...	11.0
94	Spring Valley Elementary School, GE	attendance	K-5	22.0	343.0	SFUSD Early Education Department Out of School...	12.0
95	Spring Valley Elementary School, SPN (Eng Lang...	citywide	K-5	22.0	343.0	SFUSD Early Education Department Out of School...	13.0
96	Starr King Elementary School, GE	attendance	PreK-5	22.0	329.0	YMCA Urban Services , 2:40 - 6:00pm, also have...	4.0
98	Stevenson, Robert Louis Elementary School, GE	attendance	K-5	88.0	463.0	KEEP, 2:40 to 5:40 p.m.: fee-based afterschool...	48.0
100	Sunnyside Elementary School, GE	attendance	K-5	66.0	385.0	YMCA	103.0
101	Sunset Elementary School, GE	attendance	K-5	66.0	377.0	ExCEL After School Program in partnership with...	135.0
102	Sutro Elementary School, CHN (Eng Lang Learners)	citywide	K-5	22.0	239.0	SFUSD Early Education Department Out of School...	21.0
103	Sutro Elementary School, GE	attendance	K-5	22.0	239.0	SFUSD Early Education Department Out of School...	13.0
104	Taylor, E.R. Elementary School, CHN (Eng Lang...	citywide	K-5	NaN	638.0	ExCEL program offered at sliding scale to fami...	41.0
105	Taylor, E.R. Elementary School, GE	attendance	K-5	NaN	638.0	ExCEL program offered at sliding scale to fami...	43.0
106	Taylor, E.R. Elementary School, SPN (Eng Lang ...	citywide	K-5	NaN	638.0	ExCEL program offered at sliding scale to fami...	24.0
107	Tenderloin Elementary School, GE	attendance	PreK-5	54.0	354.0	Up on Top, Monday through Friday, 2:40 to 6:00...	32.0
110	Ulloa Elementary School, CHN (Eng Lang Learners)	citywide	K-5	NaN	514.0	ExCEL After School Program in partnership with...	50.0
111	Ulloa Elementary School, GE	attendance	K-5	NaN	514.0	ExCEL After School Program in partnership with...	36.0

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_request
112	Valley Visitacion Elementary School, CHN (Eng ...	citywide	K-5	NaN	400.0	ExCEL after School Program in partnership with...	8.0
113	Valley Visitacion Elementary School, GE	attendance	K-5	NaN	400.0	ExCEL after School Program in partnership with...	15.0
115	Webster, Daniel Elementary School, GE	attendance	K-5	22.0	273.0	After: After School Enrichment Programs (ASEP)...	21.0
117	West Portal Elementary School, GE	attendance	K-5	66.0	585.0	Before: 7:00- 8:40AM. After: 2:40-6:00PM with...	77.0
118	Yick Wo Elementary School, GE	attendance	K-5	44.0	263.0	on-site after care 3:30 - 6:30pm	31.0

82 rows × 14 columns

In [33]:

```
#http://www.residentmar.io/2016/06/12/null-and-missing-data-python.html
#https://stackoverflow.com/questions/40408471/select-data-when-specific-columns-have-null-value-in
-pandas
data[data.filter(like='request').isnull().all(1)]
```

Out[33]:

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_request
20	Creative Arts Charter School	charter	K-8	44.0	440.0	Campus yard open at 8am; aftercare program ons...	NaN
59	Mission Preparatory School	charter	TK-5	NaN	NaN	NaN	NaN
69	New School	charter	K-3 (goal to be a K- 12)	48.0	NaN	Onsite before school care beginning at 7:45am ...	NaN
108	Thomas Edison Charter Academy (TECA) - 2/3 Spa...	charter	TK-8	66.0	700.0	Mission Graduates	NaN
120	NaN	NaN	NaN	NaN	NaN	NaN	NaN
121	(CB) Chinese Biliteracy	NaN	NaN	NaN	NaN	NaN	NaN
122	(CE) Chinese Immersion Non-Target Language	NaN	NaN	NaN	NaN	NaN	NaN
123	(CN) Chinese Immersion Target Language	NaN	NaN	NaN	NaN	NaN	NaN

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_request
124	Education	NaN	NaN	NaN	NaN	NaN	NaN
125	(JB) Japanese Bi-Cultural Program	NaN	NaN	NaN	NaN	NaN	NaN
126	(KE) Korean Immersion Non-Target Language	NaN	NaN	NaN	NaN	NaN	NaN
127	(KN) Korean Immersion Target Language	NaN	NaN	NaN	NaN	NaN	NaN
128	(NC) Chinese Newcomer Pathway	NaN	NaN	NaN	NaN	NaN	NaN
129	(NS) Spanish Newcomer Pathway	NaN	NaN	NaN	NaN	NaN	NaN
130	(NX) Mixed Language Newcomer Pathway	NaN	NaN	NaN	NaN	NaN	NaN
131	(SB) Spanish Biliteracy	NaN	NaN	NaN	NaN	NaN	NaN
132	(SE) Spanish Immersion Non-Target Language	NaN	NaN	NaN	NaN	NaN	NaN
133	(SN) Spanish Immersion Target Language	NaN	NaN	NaN	NaN	NaN	NaN

In [34]:

```
data[data.filter(like='request').notnull().all(1)]
```

Out[34]:

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_request
0	Alamo Elementary School, GE	attendance	K-5	88.0	530.0	Monday - Friday: RDASC (Richmond District Afte...	89.0
2	Alvarado Elementary School, GE	attendance	K-5	44.0	535.0	Mission Graduates	44.0
3	Argonne Elementary School, GE	attendance	K-5	NaN	422.0	2:40-6:00 Richmond YMCA, Richmond District Aft...	83.0
4	Bryant Elementary School, GE	attendance	K-5	22.0	216.0	SFUSD Early Education Department Out of School...	5.0
5	Bryant Elementary School, SPN (Eng Lang Learner)	citywide	K-5	22.0	216.0	SFUSD Early Education Department Out of School...	17.0

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_requ
7	Carmichael, Bessie (K-5 Campus), FIL (Eng Lang...	citywide	K-8	NaN	653.0	(ExCEL Program), operated on-site by YMCA Emba...	8.0
8	Carmichael, Bessie (K-5 Campus), GE	citywide	K-8	NaN	653.0	(ExCEL Program), operated on-site by YMCA Emba...	39.0
9	Carver, Dr. George Washington Elementary Schoo...	attendance	K-5	44.0	282.0	ExCEL after-school program sponsored by the Ba...	6.0
10	Chávez, César Elementary School, GE	attendance	PreK-5	NaN	494.0	supervised recess from 8:00 - 8:35 - Playworks...	11.0
11	Chávez, César Elementary School, SPN (Eng Lang...	citywide	PreK-5	NaN	494.0	supervised recess from 8:00 - 8:35 - Playworks...	18.0
12	Chin, John Yehall Elementary School, CHN (Eng ...	citywide	K-5	22.0	264.0	3:35 to 5:30 PM: fee based. Run by the Chinato...	49.0
13	Chin, John Yehall Elementary School, GE	attendance	K-5	22.0	264.0	3:35 to 5:30 PM: fee based. Run by the Chinato...	20.0
15	Clarendon Elementary School (GE)	attendance	K-5	44.0	272.0	Before care starts at 7am and after care until...	144.0
17	Cleveland Elementary School, GE	attendance	K-5	NaN	308.0	ExCEL program offered at sliding scale to fami...	7.0
18	Cleveland Elementary School, SPN (Eng Lang Lea...	citywide	K-5	NaN	308.0	ExCEL program offered at sliding scale, from e...	17.0
19	Cobb, Dr. William L. Elementary School, GE	attendance	PreK-5	44.0	204.0	Sliding scale aftercare (\$0-\$500) is available...	7.0
21	Drew, Dr. Charles R. Elementary School, GE	attendance	PreK-5	34.0	208.0	The after-school programs extend the academic ...	19.0
22	El Dorado Elementary School, GE	attendance	K-5	44.0	274.0	Real Options for City Kids (ROCK) M-F 1:50 to ...	8.0
23	Feinstein, Dianne Elementary School, GE	attendance	K-5	88.0	518.0	YMCA	75.0
24	Flynn, Leonard	attendance	K-5	NaN	478.0	Mission YMCA, 7:30 AM to 6:30 PM ExCEL program	14.0



	Elementary School, GE	type	grades	number_of_k_seats	number_enrolled_2016	6:30 PM ExCEL program... before_after_school_program	1st_requ
26	Garfield Elementary School, GE	attendance	K-5	22.0	250.0	Garfield afterschool program: 1:50 to 6:00 p.m...	10.0
28	Glen Park Elementary School, GE	attendance	K-5	NaN	329.0	ExCEL program is from 2:40- 6:00 PM each day. T...	34.0
29	Glen Park Elementary School, SPN (Eng Lang Lea...	citywide	K-5	NaN	329.0	ExCEL program is from 2:40- 6:00 PM each day. T...	17.0
30	Grattan Elementary School, GE	attendance	PreK-5	66.0	359.0	Grattan After School Program (GASP) is a slidi...	147.0
31	Guadalupe Elementary School, GE	attendance	K-5	NaN	445.0	SHPD ExCEL after-school offered at sliding sca...	20.0
32	Guadalupe Elementary School, SPN (Eng Lang Lea...	citywide	K-5	NaN	445.0	SHPD ExCEL after-school offered at sliding sca...	14.0
33	Harte, Bret Elementary School, GE	attendance	PreK-5	22.0	253.0	ExCEL: no-cost program to families of students...	7.0
35	Hillcrest Elementary School, CHN (Eng Lang Lea...	citywide	K-5	NaN	493.0	After-school program for 250 K-5th graders fro...	3.0
36	Hillcrest Elementary School, GE	attendance	K-5	NaN	493.0	After-school program for 250 K-5th graders fro...	12.0
38	Jefferson Elementary School, GE	attendance	K-5	88.0	493.0	Before: 7:45 to 8:40 a.m.\nAfter: SFUSD Early ...	47.0
...	...	...	...	...	...	...	...
81	Revere, Paul, GE	attendance	PreK-8	NaN	428.0	1:50-6pm - ExCEL Extended Day program through ...	8.0
84	Rooftop (PreK-4 Burnett Campus), GE	citywide	PreK-8	66.0	600.0	CASA: A nonprofit fee-based program	150.0
85	S.F. Community, GE	citywide	K-8	44.0	286.0	City Scholars ExCEL ASP open to students K-8 f...	30.0
86	Sanchez Elementary School, GE	attendance	PreK-5	22.0	266.0	Mission Graduates	2.0
87	Sanchez Elementary School, SPN (Eng Lang Learn...	citywide	PreK-5	22.0	266.0	Mission Graduates	6.0
88	Serra, Junipero	attendance	K-5	22.0	269.0	After School Success Club	9.0

	School name	type	grades	number_of_k_seats	number_enrolled_2016	(ExCEL After School before_after_school program	1st_requ
89	Serra, Junipero Elementary School, SPN (Eng La...	citywide	K-5	22.0	269.0	After School Success Club (ExCEL After School ...	22.0
90	Sheridan Elementary School, GE	attendance	K-5	44.0	215.0	On-site YMCA program: 2:00 to 5:30 p.m.\nExCEL...	12.0
91	Sherman Elementary School, GE	attendance	K-5	66.0	428.0	Sherman afterschool program for all students f...	94.0
92	Sloat, Commodore Elementary School, GE	attendance	K-5	66.0	396.0	Anchor Program: Mon. to Fri., 2:40 to 6:30 PM\n	63.0
93	Spring Valley Elementary School, CHN (Eng Lang...	citywide	K-5	22.0	343.0	SFUSD Early Education Department Out of School...	11.0
94	Spring Valley Elementary School, GE	attendance	K-5	22.0	343.0	SFUSD Early Education Department Out of School...	12.0
95	Spring Valley Elementary School, SPN (Eng Lang...	citywide	K-5	22.0	343.0	SFUSD Early Education Department Out of School...	13.0
96	Starr King Elementary School, GE	attendance	PreK-5	22.0	329.0	YMCA Urban Services , 2:40 - 6:00pm, also have...	4.0
98	Stevenson, Robert Louis Elementary School, GE	attendance	K-5	88.0	463.0	KEEP, 2:40 to 5:40 p.m.: fee-based afterschool...	48.0
100	Sunnyside Elementary School, GE	attendance	K-5	66.0	385.0	YMCA	103.0
101	Sunset Elementary School, GE	attendance	K-5	66.0	377.0	ExCEL After School Program in partnership with...	135.0
102	Sutro Elementary School, CHN (Eng Lang Learners)	citywide	K-5	22.0	239.0	SFUSD Early Education Department Out of School...	21.0
103	Sutro Elementary School, GE	attendance	K-5	22.0	239.0	SFUSD Early Education Department Out of School...	13.0
104	Taylor, E.R. Elementary School, CHN (Eng Lang...	citywide	K-5	NaN	638.0	ExCEL program offered at sliding scale to fami...	41.0
105	Taylor, E.R. Elementary School, GE	attendance	K-5	NaN	638.0	ExCEL program offered at sliding scale to fami...	43.0
106	Taylor, E.R. Elementary School, SPN	citywide	K-5	NaN	638.0	ExCEL program offered at ...	24.0

	School, SPN school_name (Eng Lang ...	type	grades	number_of_k_seats	number_enrolled_2016	sidling scare to rami... before_after_school_program	1st_requ
107	Tenderloin Elementary School, GE	attendance	PreK-5	54.0	354.0	Up on Top, Monday through Friday, 2:40 to 6:00...	32.0
110	Ulloa Elementary School, CHN (Eng Lang Learners)	citywide	K-5	NaN	514.0	ExCEL After School Program in partnership with...	50.0
111	Ulloa Elementary School, GE	attendance	K-5	NaN	514.0	ExCEL After School Program in partnership with...	36.0
112	Visitation Valley Elementary School, CHN (Eng ...	citywide	K-5	NaN	400.0	ExCEL after School Program in partnership with...	8.0
113	Visitation Valley Elementary School, GE	attendance	K-5	NaN	400.0	ExCEL after School Program in partnership with...	15.0
115	Webster, Daniel Elementary School, GE	attendance	K-5	22.0	273.0	After: After School Enrichment Programs (ASEP)...	21.0
117	West Portal Elementary School, GE	attendance	K-5	66.0	585.0	Before: 7:00- 8:40AM. After: 2:40-6:00PM with...	77.0
118	Yick Wo Elementary School, GE	attendance	K-5	44.0	263.0	on-site after care 3:30 - 6:30pm	31.0

92 rows × 14 columns

```
data[(data['1st_request_rnd1_17'].isnull())& data['1st_request_rnd1_18'].isnull())& data['total_requests_rnd1_17'].isnull())&
data['total_requests_rnd1_18'].isnull())]
```

In [35]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 14 columns):
school_name      133 non-null object
type             114 non-null category
grades           120 non-null category
number_of_k_seats 85 non-null float64
number_enrolled_2016 110 non-null float64
before_after_school_program 119 non-null object
1st_request_rnd1_17 110 non-null float64
1st_request_rnd1_18 98 non-null float64
total_requests_rnd1_17 110 non-null float64
total_requests_rnd1_18 98 non-null float64
waitpool_6_17    94 non-null float64
waitpool_6_18    94 non-null float64
waitpool_7_18    98 non-null float64
waitpool_8_16    86 non-null float64
dtypes: category(2), float64(10), object(2)
memory usage: 13.4+ KB
```

Drop rows

In [36]:

```
#https://stackoverflow.com/questions/36370839/better-way-to-drop-nan-rows-in-pandas
#clean_data = data.dropna(how='all')
#cdf // clean data frame
data_simplified = data.dropna(subset=[
    '1st_request_rnd1_17',
    '1st_request_rnd1_18',
    'total_requests_rnd1_17',
    'total_requests_rnd1_18',
    'waitpool_6_17',
    'waitpool_6_18',
    'waitpool_7_18',
    'waitpool_8_16'])
```

In [37]:

```
data_simplified.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 82 entries, 0 to 118
Data columns (total 14 columns):
school_name      82 non-null object
type             82 non-null category
grades           82 non-null category
number_of_k_seats 58 non-null float64
number_enrolled_2016 82 non-null float64
before_after_school_program 82 non-null object
1st_request_rnd1_17 82 non-null float64
1st_request_rnd1_18 82 non-null float64
total_requests_rnd1_17 82 non-null float64
total_requests_rnd1_18 82 non-null float64
waitpool_6_17    82 non-null float64
waitpool_6_18    82 non-null float64
waitpool_7_18    82 non-null float64
waitpool_8_16    82 non-null float64
dtypes: category(2), float64(10), object(2)
memory usage: 9.0+ KB
```

In [38]:

```
#https://medium.com/dunder-data/selecting-subsets-of-data-in-pandas-6fcd0170be9c
data_simplified[['school_name']]
```

Out[38]:

	school_name
0	Alamo Elementary School, GE
4	Bryant Elementary School, GE
5	Bryant Elementary School, SPN (Eng Lang Learner)
9	Carver, Dr. George Washington Elementary Schoo...
10	Chávez, César Elementary School, GE
11	Chávez, César Elementary School, SPN (Eng Lang...
12	Chin, John Yehall Elementary School, CHN (Eng ...
13	Chin, John Yehall Elementary School, GE
15	Clarendon Elementary School (GE)
17	Cleveland Elementary School, GE
18	Cleveland Elementary School, SPN (Eng Lang Lea...
19	Cobb, Dr. William L. Elementary School, GE
21	Drew, Dr. Charles R. Elementary School, GE
22	El Dorado Elementary School, GE
23	El Dorado Elementary School, GE

23	Feinstein, Dianne Elementary School, GE
	<b>school_name</b>
26	Garfield Elementary School, GE
28	Glen Park Elementary School, GE
29	Glen Park Elementary School, SPN (Eng Lang Lea...
30	Grattan Elementary School, GE
31	Guadalupe Elementary School, GE
32	Guadalupe Elementary School, SPN (Eng Lang Lea...
33	Harte, Bret Elementary School, GE
35	Hillcrest Elementary School, CHN (Eng Lang Lea...
36	Hillcrest Elementary School, GE
38	Jefferson Elementary School, GE
39	Key, Francis Scott Elementary School, GE
40	Lafayette Elementary School, GE
41	Lakeshore Elementary School, GE
42	Lau, Gordon J. Elementary School, CHN (Eng Lan...
43	Lau, Gordon J. Elementary School, GE
...	...
75	Parker, Jean Elementary School, GE
76	Parks, Rosa Elementary School, GE
81	Revere, Paul, GE
84	Rooftop (PreK-4 Burnett Campus), GE
85	S.F. Community, GE
86	Sanchez Elementary School, GE
87	Sanchez Elementary School, SPN (Eng Lang Learn...
89	Serra, Junipero Elementary School, SPN (Eng La...
91	Sherman Elementary School, GE
92	Sloat, Commodore Elementary School, GE
93	Spring Valley Elementary School, CHN (Eng Lang...
94	Spring Valley Elementary School, GE
95	Spring Valley Elementary School, SPN (Eng Lang...
96	Starr King Elementary School, GE
98	Stevenson, Robert Louis Elementary School, GE
100	Sunnyside Elementary School, GE
101	Sunset Elementary School, GE
102	Sutro Elementary School, CHN (Eng Lang Learners)
103	Sutro Elementary School, GE
104	Taylor, E.R. Elementary School, CHN (Eng Lang...
105	Taylor, E.R. Elementary School, GE
106	Taylor, E.R. Elementary School, SPN (Eng Lang ...
107	Tenderloin Elementary School, GE
110	Ulloa Elementary School, CHN (Eng Lang Learners)
111	Ulloa Elementary School, GE
112	Visitacion Valley Elementary School, CHN (Eng ...
113	Visitacion Valley Elementary School, GE
115	Webster, Daniel Elementary School, GE
117	West Portal Elementary School, GE

117	West Portal Elementary School, GE	school_name
118	Yick Wo Elementary School, GE	

82 rows x 1 columns

In [39]:

```
data_simplified.head()
```

Out[39]:

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_reque
0	Alamo Elementary School, GE	attendance	K-5	88.0	530.0	Monday - Friday: RDASC (Richmond District Afte...	89.0
4	Bryant Elementary School, GE	attendance	K-5	22.0	216.0	SFUSD Early Education Department Out of School...	5.0
5	Bryant Elementary School, SPN (Eng Lang Learner)	citywide	K-5	22.0	216.0	SFUSD Early Education Department Out of School...	17.0
9	Carver, Dr. George Washington Elementary Schoo...	attendance	K-5	44.0	282.0	ExCEL after-school program sponsored by the Ba...	6.0
10	Chávez, César Elementary School, GE	attendance	PreK-5	NaN	494.0	supervised recess from 8:00 - 8:35 - Playworks...	11.0

## Check for outliers

In [40]:

```
print ("range - 1st_request_rnd1_17:", sorted(data_simplified['1st_request_rnd1_17'].unique()))
```

```
range - 1st_request_rnd1_17: [0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, 16.0, 17.0, 18.0, 19.0, 20.0, 21.0, 22.0, 24.0, 27.0, 30.0, 31.0, 32.0, 34.0, 36.0, 40.0, 41.0, 42.0, 43.0, 47.0, 48.0, 49.0, 50.0, 51.0, 63.0, 75.0, 77.0, 78.0, 86.0, 89.0, 91.0, 93.0, 94.0, 103.0, 128.0, 135.0, 144.0, 147.0, 150.0, 176.0]
```

In [41]:

```
print ("range - 1st_request_rnd1_18:", sorted(data_simplified['1st_request_rnd1_18'].unique()))
```

```
range - 1st_request_rnd1_18: [0.0, 2.0, 3.0, 4.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 15.0, 16.0, 17.0, 18.0, 19.0, 20.0, 21.0, 23.0, 27.0, 35.0, 36.0, 38.0, 40.0, 41.0, 45.0, 47.0, 56.0, 57.0, 58.0, 62.0, 71.0, 72.0, 74.0, 76.0, 78.0, 82.0, 84.0, 96.0, 113.0, 120.0, 122.0, 124.0, 155.0, 169.0, 176.0]
```

In [42]:

```
#Checking 'Total Requests Round 1 2018' entries for outliers
ds_min_num = data_simplified['1st_request_rnd1_18'].min()
ds_max_num = data_simplified['1st_request_rnd1_18'].max()
print ("cdf range for 1st_request_rnd1_18: ", ds_min_num, "-", ds_max_num)
```

```
cdf range for 1st_request_rnd1_18: 0.0 - 176.0
```

In [43]:

```
data_simplified.number_of_k_seats.sort_values()
```

Out[43]:

```
45      9.0
93     22.0
66     22.0
67     22.0
68     22.0
51     22.0
74     22.0
75     22.0
86     22.0
87     22.0
89     22.0
65     22.0
33     22.0
64     22.0
26     22.0
95     22.0
96     22.0
102    22.0
103    22.0
13     22.0
12     22.0
115    22.0
5      22.0
4      22.0
94     22.0
21     34.0
85     44.0
76     44.0
70     44.0
63     44.0
...
23     88.0
40     88.0
98     88.0
41     88.0
42     88.0
0      88.0
10     NaN
11     NaN
17     NaN
18     NaN
28     NaN
29     NaN
31     NaN
32     NaN
35     NaN
36     NaN
48     NaN
49     NaN
50     NaN
58     NaN
60     NaN
61     NaN
81     NaN
104    NaN
105    NaN
106    NaN
110    NaN
111    NaN
112    NaN
113    NaN
```

Name: number\_of\_k\_seats, Length: 82, dtype: float64

In [44]:

```
wjune17 = data_simplified.waitpool_6_17.value_counts().sort_index()
wjune18 = data_simplified.waitpool_6_18.value_counts().sort_index()
wjuly18 = data_simplified.waitpool_7_18.value_counts().sort_index()
waug16 = data_simplified.waitpool_8_16.value_counts().sort_index()
```

```
print(wjuly18)
```

```
0.0    32
1.0     7
2.0     3
3.0    10
4.0     2
5.0     2
6.0     4
8.0     2
9.0     2
10.0    3
12.0    2
13.0    2
14.0    2
15.0    1
17.0    2
18.0    2
19.0    1
20.0    2
23.0    1
Name: waitpool_7_18, dtype: int64
```

## Exploratory analysis

From there you can do things like remove outliers and create new and more descriptive features from existing data, also known as feature engineering. Or detect and remove outliers so your model's fit is better.

```
In [45]:
```

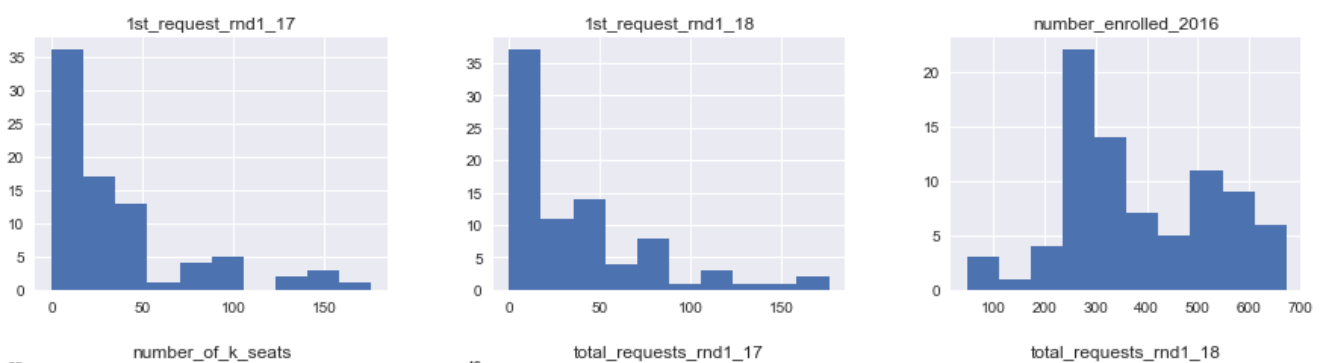
```
data_simplified.describe()
```

```
Out[45]:
```

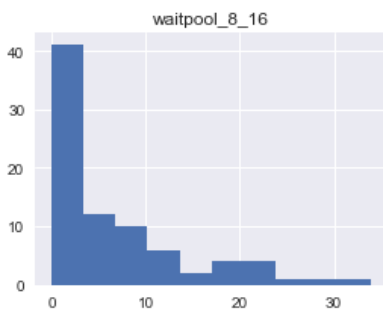
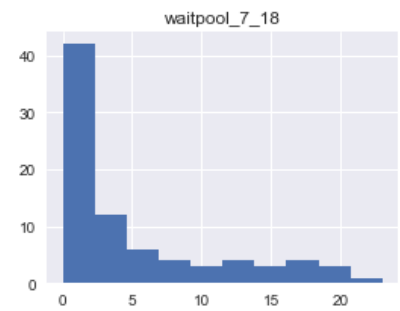
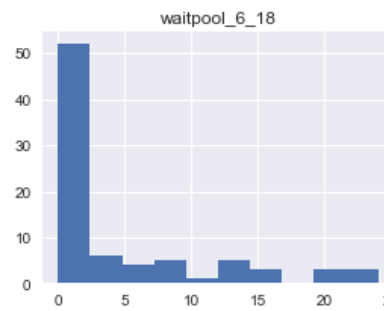
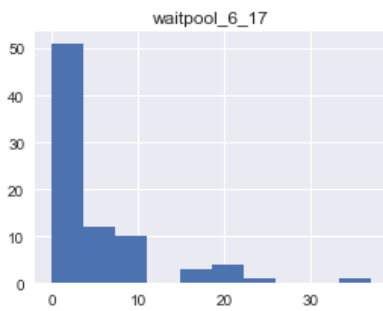
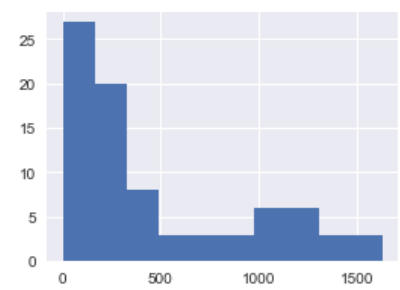
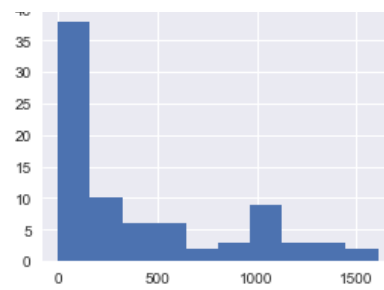
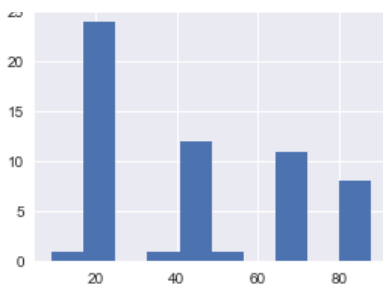
	number_of_k_seats	number_enrolled_2016	1st_request_rnd1_17	1st_request_rnd1_18	total_requests_rnd1_17	total_requests_rnd1_18
count	58.000000	82.000000	82.000000	82.000000	82.000000	82.000000
mean	44.534483	381.878049	37.329268	37.890244	435.109756	477.000000
std	24.364351	149.841677	40.166030	39.790503	454.332804	470.000000
min	9.000000	47.000000	0.000000	0.000000	1.000000	0.000000
25%	22.000000	264.500000	11.000000	11.000000	97.500000	89.500000
50%	44.000000	343.000000	21.000000	19.000000	199.000000	255.000000
75%	66.000000	499.250000	48.000000	46.500000	762.000000	814.000000
max	88.000000	674.000000	176.000000	176.000000	1610.000000	163.000000

```
In [46]:
```

```
data_simplified.hist(figsize=(15,15));
```

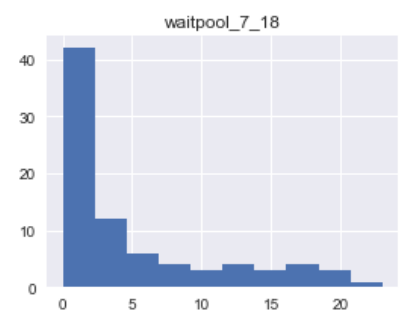
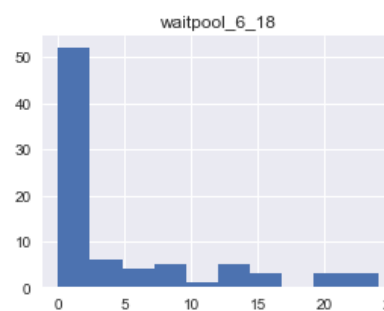
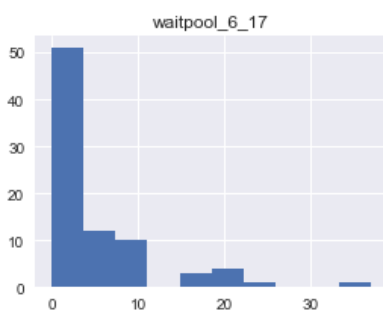
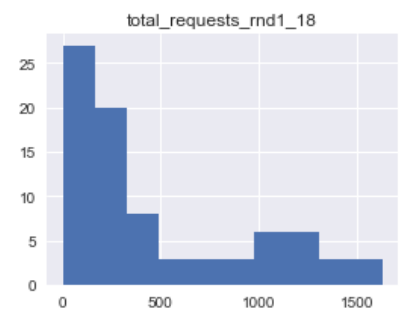
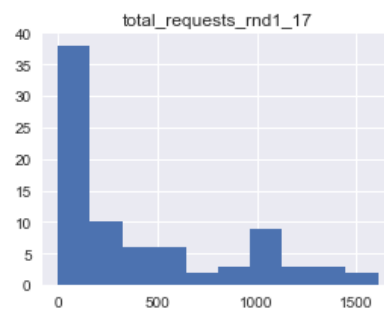
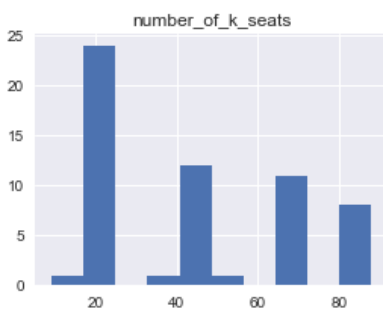
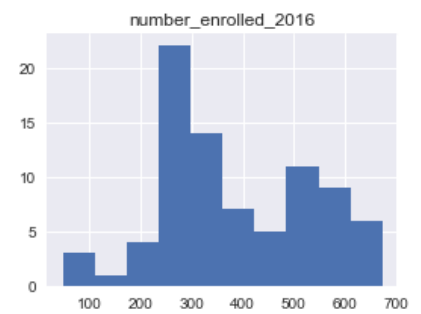
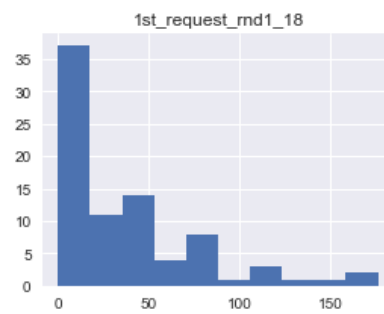
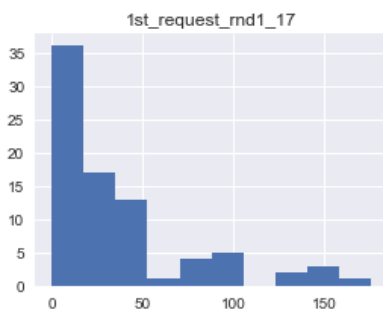


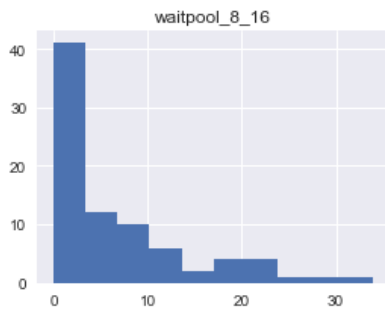




In [47]:

```
data_simplified.hist(figsize=(15,15));
```





## Questions:

In [48]:

```
data_simplified = data_simplified.sort_values(['1st_request_rnd1_18',
'1st_request_rnd1_17',
'total_requests_rnd1_18',
'total_requests_rnd1_17',
'waitpool_6_18',
'waitpool_7_18',
'waitpool_6_17',
'waitpool_8_16'], ascending=False)
```

*#also: loop through <https://stackoverflow.com/questions/45927905/loop-though-pandas-dataframe-columns-sort-and-return-the-top-n-results>*

In [49]:

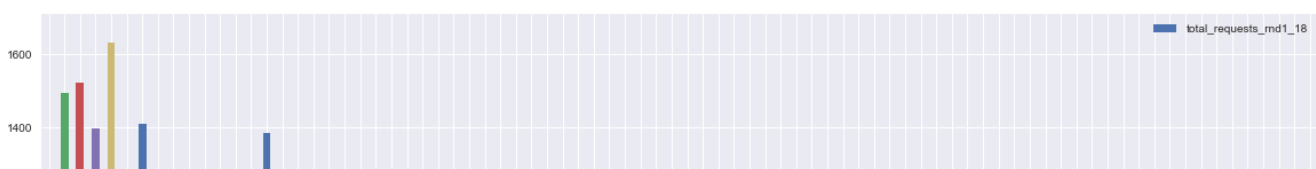
```
data_simplified.head()
```

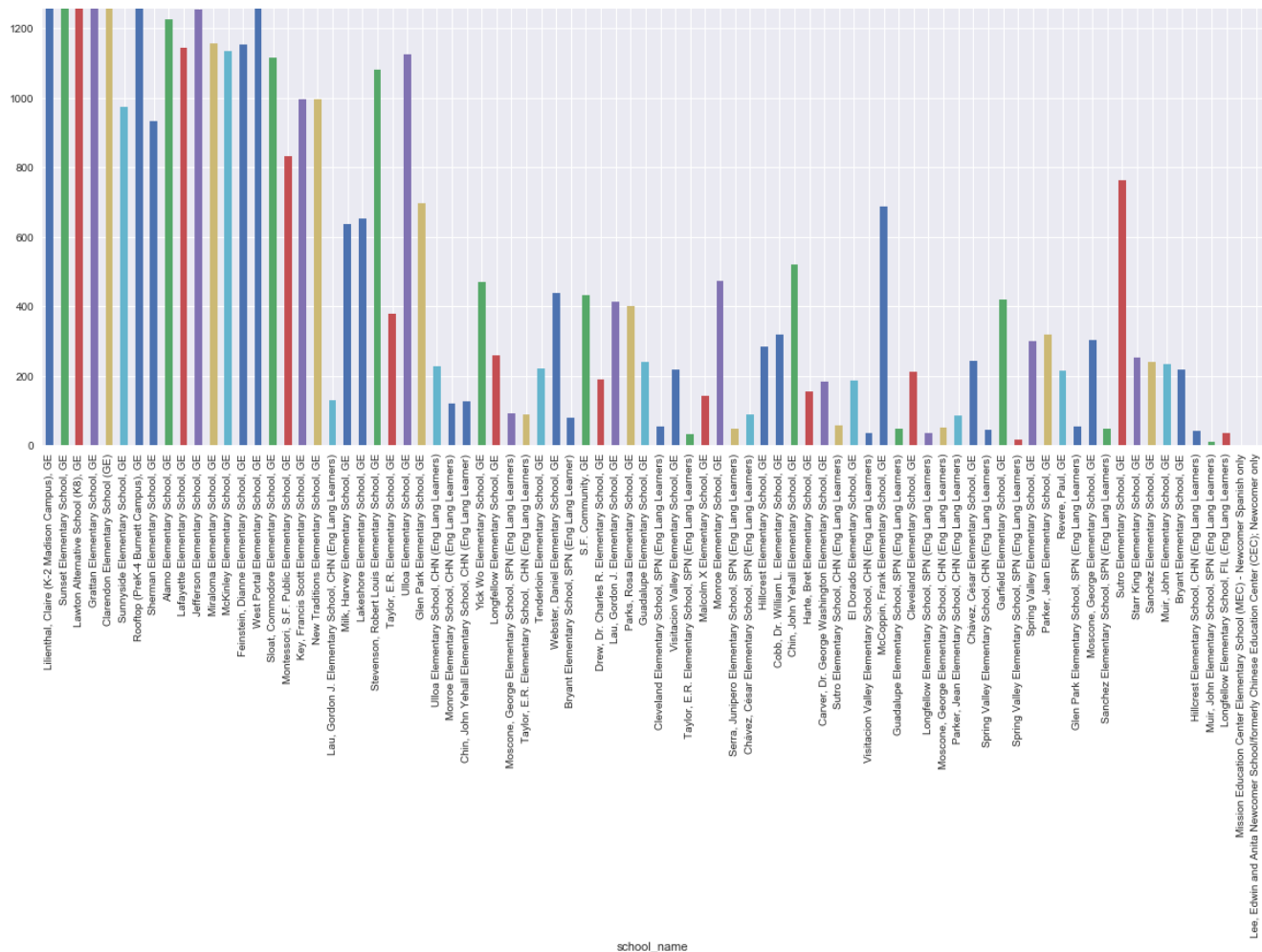
Out[49]:

	school_name	type	grades	number_of_k_seats	number_enrolled_2016	before_after_school_program	1st_requ
46	Lilienthal, Claire (K-2 Madison Campus), GE	citywide	K-8	66.0	674.0	CLASP (independent nonprofit Claire Lilienthal...	176.0
101	Sunset Elementary School, GE	attendance	K-5	66.0	377.0	ExCEL After School Program in partnership with...	135.0
44	Lawton Alternative School (K8), GE	citywide	K-8	66.0	610.0	K.E.E.P - Fee-based AM and PM childcare. Call ...	128.0
30	Grattan Elementary School, GE	attendance	PreK-5	66.0	359.0	Grattan After School Program (GASP) is a slidi...	147.0
15	Clarendon Elementary School (GE)	attendance	K-5	44.0	272.0	Before care starts at 7am and after care until...	144.0

In [50]:

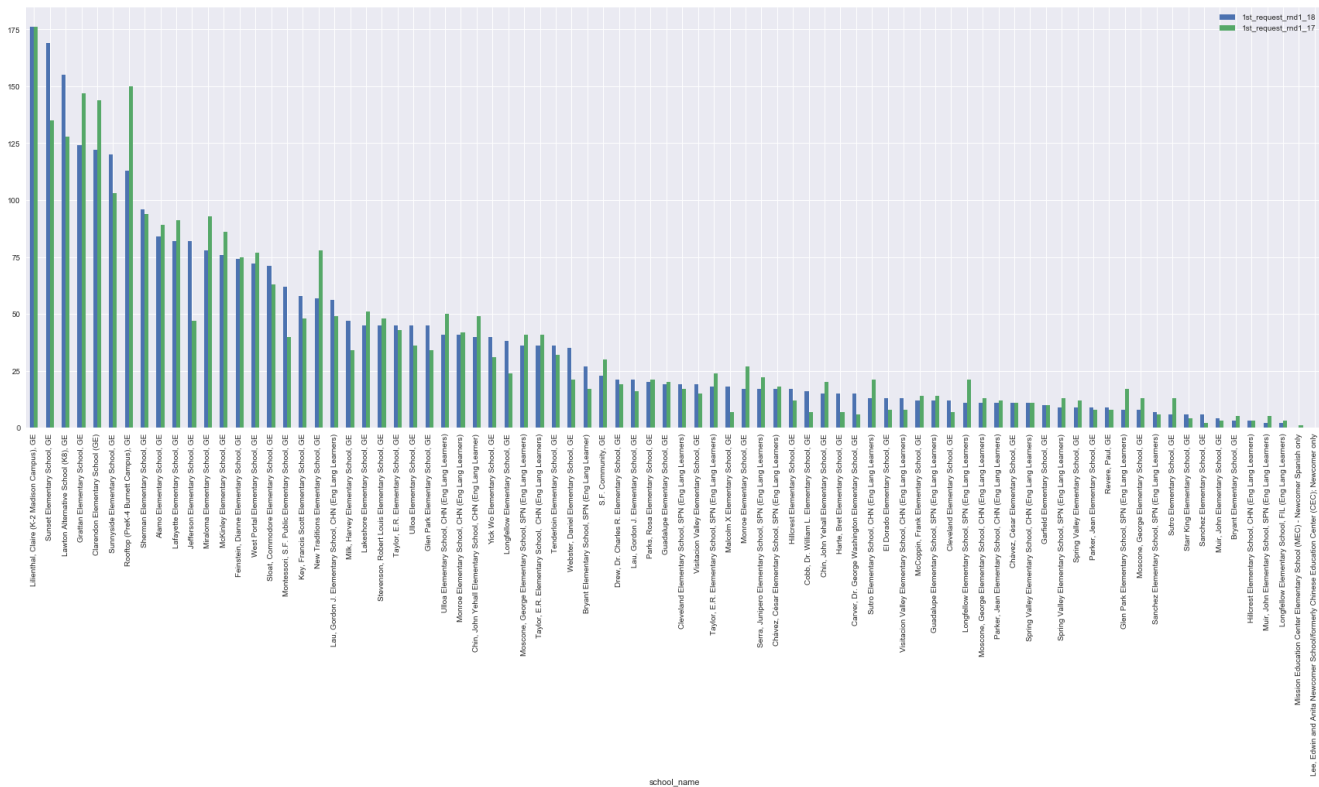
```
data_simplified.plot(x='school_name', y='total_requests_rnd1_18', kind='bar', figsize=(20,10))
plt.show();
```





In [51]:

```
#plot 1st_request_rnd1 columns to see a comparison over time
data_simplified.plot(x='school_name', y=["1st_request_rnd1_18", "1st_request_rnd1_17"], kind="bar",
figsize=(30,10));
```



In [52]:

In [53]:

School	total_requests_md1_18
Lilienthal, Claire (K-2 Madison Campus), GE	1280
Sunset Elementary School, GE	1500
Lawton Alternative School (K8), GE	1520
Grattan Elementary School, GE	1400
Clarendon Elementary School (GE)	980
Sunnyside Elementary School, GE	1000
Rooftop (PreK-4 Burnett Campus), GE	1420
Sherman Elementary School, GE	950
Alamo Elementary School, GE	1230
Lafayette Elementary School, GE	1150
Jefferson Elementary School, GE	1260
Miraloma Elementary School, GE	1160
McKinley Elementary School, GE	1140
Feinstein, Dianne Elementary School, GE	1380
West Portal Elementary School, GE	1120
Sloat, Commodore Elementary School, GE	1120
Montessori, S.F. Public Elementary School, GE	840
Key, Francis Scott Elementary School, GE	980
New Traditions Elementary School, GE	1000
Lau, Gordon J. Elementary School, CHN (Eng Lang Learners)	140
Milk, Harvey Elementary School, GE	650
Lakeshore Elementary School, GE	660
Robert Louis Elementary School, GE	1090
Taylor, E.R. Elementary School, GE	380
Ulloa Elementary School, GE	1130
Glen Park Elementary School, GE	700
Ulloa Elementary School, CHN (Eng Lang Learners)	230
Moscone Elementary School, CHN (Eng Lang Learners)	120
Chin, John Yehall Elementary School, CHN (Eng Lang Learners)	130
Yick Wo Elementary School, GE	480
Longfellow Elementary School, GE	270
Moscone, George Elementary School, SPN (Eng Lang Learners)	80
Taylor, E.R. Elementary School, CHN (Eng Lang Learners)	100
Taylor, E.R. Elementary School, GE	230
Webster, Daniel Elementary School, GE	450
Bryant Elementary School, SPN (Eng Lang Learner)	80
S.F. Community, GE	440
Drew, Dr. Charles R. Elementary School, GE	440
Lau, Gordon J. Elementary School, GE	190
Parks, Rosa Elementary School, GE	420
Guadalupe Elementary School, GE	400
Cleveland Elementary School, SPN (Eng Lang Learners)	250
Visitation Valley Elementary School, GE	50
Visitation Valley Elementary School, SPN (Eng Lang Learners)	30
Taylor, E.R. Elementary School, SPN (Eng Lang Learners)	150
Malcolm X Elementary School, GE	480
Monroe Elementary School, GE	40
Serra, Junipero Elementary School, SPN (Eng Lang Learners)	100
Chavez, César Elementary School, GE	100
Hillcrest Elementary School, GE	300
Cobb, Dr. William L. Elementary School, GE	320
Chin, John Yehall Elementary School, GE	540
Harte, Bret Elementary School, GE	160
Carver, Dr. George Washington Elementary School, GE	180
Sutro Elementary School, CHN (Eng Lang Learners)	50
El Dorado Elementary School, GE	190
Visitation Valley Elementary School, CHN (Eng Lang Learners)	30
McCoppin, Frank Elementary School, GE	30
Guadalupe Elementary School, SPN (Eng Lang Learners)	40
Cleveland Elementary School, GE	220
Longfellow Elementary School, SPN (Eng Lang Learners)	30
Moscone, George Elementary School, CHN (Eng Lang Learners)	50
Parker, Jean Elementary School, CHN (Eng Lang Learners)	100
Chavez, Cesar Elementary School, GE	250
Spring Valley Elementary School, CHN (Eng Lang Learners)	40
Garfield Elementary School, GE	430
Spring Valley Elementary School, SPN (Eng Lang Learners)	10
Spring Valley Elementary School, GE	300
Parker, Jean Elementary School, GE	320
Revere, Paul, GE	50
Glen Park Elementary School, SPN (Eng Lang Learners)	300
Moscone, George Elementary School, GE	300
Sanchez Elementary School, SPN (Eng Lang Learners)	40
Sutro Elementary School, GE	760
Starr King Elementary School, GE	260
Sanchez Elementary School, GE	240
Muir, John Elementary School, GE	240
Bryant Elementary School, GE	220
Hillcrest Elementary School, CHN (Eng Lang Learners)	40
Muir, John Elementary School, SPN (Eng Lang Learners)	30
Longfellow Elementary School, FL (Eng Lang Learners)	20
Mission Education Center/formerly Chinese Education Center (CEC), Newcomer only	20

In [54]:

```
#https://stackoverflow.com/questions/42128467/matplotlib-plot-multiple-columns-of-pandas-data-frame-on-the-bar-chart/42131286
requests_total_sorted.plot(x='school_name', y=["total_requests_rnd1_l8", "total_requests_rnd1_l7"],
kind='bar', figsize=(20,10))
plt.show();
```

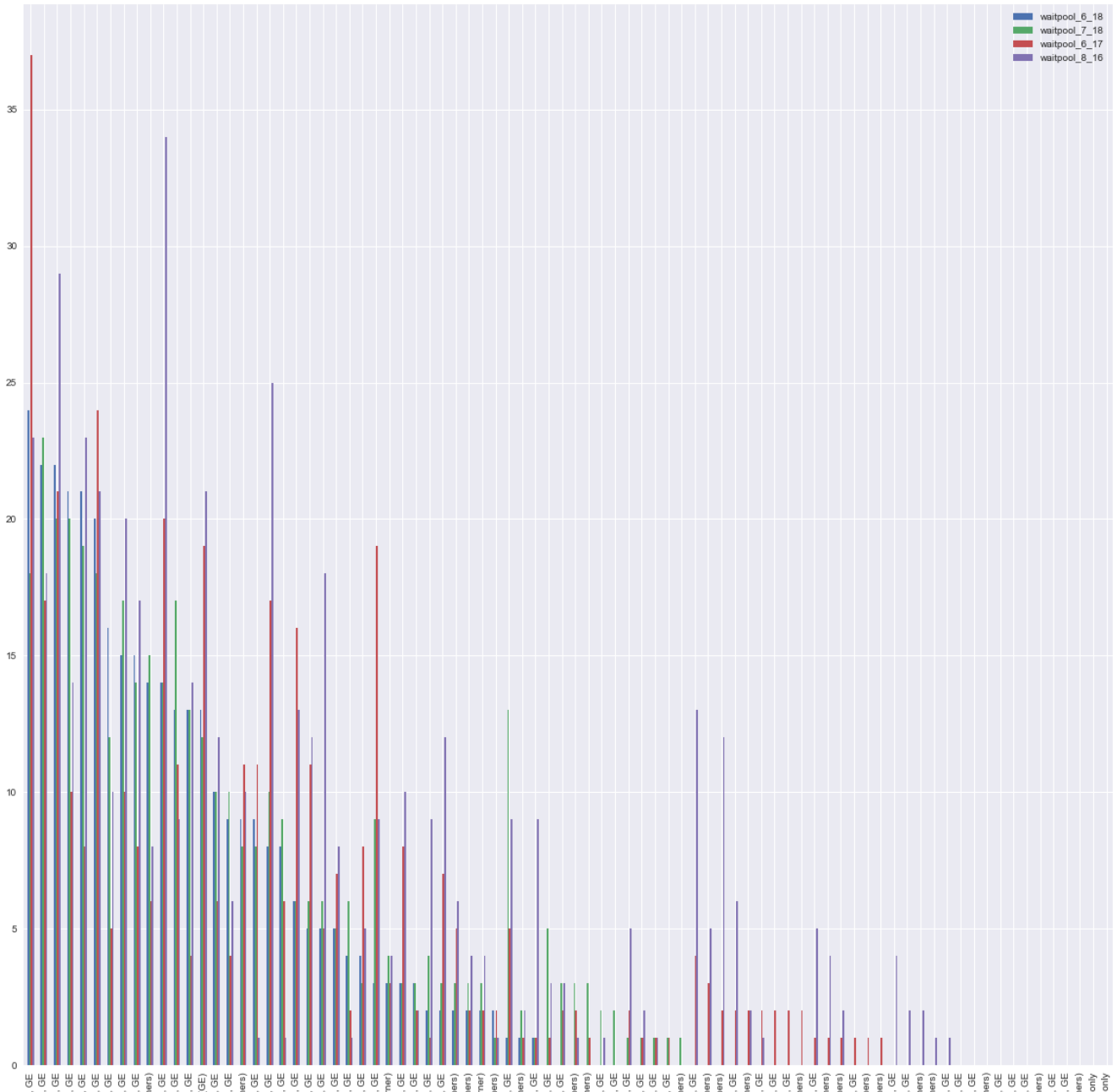


Lilienthal, Claire (K-2 Madison Campus), (Sunset Elementary School, (Lawton Alternative School (K8), (Grattan Elementary School, (Clarendon Elementary School, (Summyside Elementary School, (Roodtop (PreK-4 Burnett Campus), (Sherman Elementary School, (Alamo Elementary School, (Lafayette Elementary School, (Jefferson Elementary School, (McDonnell Elementary School, (McDonnell Elementary School, (Feinstein, Daniel (Young Elementary School, (West Portland Elementary School, (Stoat Commodore Elementary School, (Montessori, S.F. Public Elementary School, (Key Francis Scott Elementary School, (New Traditions Elementary School, (Lau, Gordon J. Elementary School, CHN (Eng Lang Learne Milk, Harvey Elementary School, (Lakeshore Elementary School, (Stevenson, Robert Louis Elementary School, (Taylor, E.R. Elementary School, (Ulioa Elementary School, (Glen Park Elementary School, (Ulioa Elementary School, CHN (Eng Lang Learne Monroe Elementary School, CHN (Eng Lang Learne Chin, John Yehall Elementary School, CHN (Eng Lang Learne Yick Wo Elementary School, (Longfellow Elementary School, (Moscone, George Elementary School, SPN (Eng Lang Learne Taylor, E.R. Elementary School, CHN (Eng Lang Learne Tenderloin Elementary School, (Webster, Daniel Elementary School, (Bryant Elementary School, SPN (Eng Lang Learne S.F. Community, (Drew, Dr. Charles R. Elementary School, (Lau, Gordon J. Elementary School, (Parks, Rosa Elementary School, (Guadalupe Elementary School, (Cleveland Elementary School, SPN (Eng Lang Learne Visitation Valley Elementary School, (Taylor, E.R. Elementary School, SPN (Eng Lang Learne Malcolm X Elementary School, (Monroe Elementary School, (Serre, Junipero Elementary School, SPN (Eng Lang Learne Chavez, César Elementary School, SPN (Eng Lang Learne Hillcrest Elementary School, (Cobb, Dr. William L. Elementary School, (Chin, John Yehall Elementary School, (Carver, Dr. George Washington Elementary School, (Sutro Elementary School, CHN (Eng Lang Learne Visitation Valley Elementary School, CHN (Eng Lang Learne McCoppin, Frank Elementary School, (Guadalupe Elementary School, SPN (Eng Lang Learne Cleveland Elementary School, SPN (Eng Lang Learne Longfellow Elementary School, SPN (Eng Lang Learne Moscone, George Elementary School, CHN (Eng Lang Learne Parker, Jean Elementary School, CHN (Eng Lang Learne Chavez, César Elementary School, CHN (Eng Lang Learne Spring Valley Elementary School, CHN (Eng Lang Learne Spring Valley Elementary School, SPN (Eng Lang Learne Spring Valley Elementary School, (Parker, Jean Elementary School, (Rosen, Paul, (Glen Park Elementary School, SPN (Eng Lang Learne Moscone, George Elementary School, (Sanchez Elementary School, SPN (Eng Lang Learne Starr King Elementary School, (Sanchez Elementary School, (Muir, John Elementary School, (Bryant Elementary School, (Hillcrest Elementary School, CHN (Eng Lang Learne Muir, John Elementary School, SPN (Eng Lang Learne Longfellow Elementary School, FIL (Eng Lang Learne Mission Education Center Elementary School (MEC) - Newcomer Spanish or Lee, Edwin and Anita Newcomer School/formerly Chinese Education Center (CEC); Newcomer or

school\_name

In [56]:

```
#plot multiple waitlist columns to see a comparison over time
waitlist_sorted.plot(x='school_name', y=[ 'waitpool_6_18',
'waitpool_7_18',
'waitpool_6_17',
'waitpool_8_16'], kind="bar", figsize=(20,20));
```



Sunset Elementary School  
Sunnyside Elementary School  
Lawton Alternative School (K8)  
Stevenson, Robert Louis Elementary School  
Jefferson Elementary School  
Llinterthal, Claire (K-2 Madison Campus)  
Key, Francis Scott Elementary School  
Sherman Elementary School  
Sloat, Commodore Elementary School  
Monroe Elementary School, CHN (Eng Lang Lear  
Grattan Elementary School  
Ulloa Elementary School  
Feinstein, Dianne Elementary School  
Clarendon Elementary School  
Miraloma Elementary School  
Monroe Elementary School  
Ulloa Elementary School, CHN (Eng Lang Lear  
Amano Elementary School  
Rooftop (Pak-4 Burnett Campus)  
Webster, Daniel Elementary School  
New Traditions Elementary School  
Lafayette Elementary School  
McKinley Elementary School  
Montessori, S.F. Public Elementary School  
Longfellow Elementary School  
S.F. Community  
West Portal Elementary School  
Chin, John Yehall Elementary School, CHN (Eng Lang Lear  
Mik, Harvey Elementary School  
Chin, John Yehall Elementary School  
Yick Wo Elementary School  
Glen Park Elementary School  
Taylor, E.R. Elementary School, CHN (Eng Lang Lear  
Moscone, George Elementary School, SPN (Eng Lang Lear  
Bryant Elementary School, SPN (Eng Lang Lear  
Serra, Junipero Elementary School, SPN (Eng Lang Lear  
Lakeshore Elementary School  
Sutro Elementary School, CHN (Eng Lang Lear  
McCoppin, Frank Elementary School  
Parks, Rosa Elementary School  
Cleveland Elementary School, SPN (Eng Lang Lear  
Lau, Gordon J. Elementary School, CHN (Eng Lang Lear  
Burr King Elementary School  
Lau, Gordon J. Elementary School  
Sutro Elementary School  
Tendler Elementary School  
Cobb, Dr. William L. Elementary School  
Harris, Bret Elementary School  
Moscone, George Elementary School, CHN (Eng Lang Lear  
Taylor, E.R. Elementary School, SPN (Eng Lang Lear  
Taylor, E.R. Elementary School, SPN (Eng Lang Lear  
Glen Park Elementary School, SPN (Eng Lang Lear  
Sanchez Elementary School  
Longfellow Elementary School, SPN (Eng Lang Lear  
Guadalupe Elementary School  
Drew, Dr. Charles R. Elementary School  
Chavez, Cesar Elementary School  
Spring Valley Elementary School, CHN (Eng Lang Lear  
Visitation Valley Elementary School  
Chavez, Cesar Elementary School, SPN (Eng Lang Lear  
Hillcrest Elementary School, CHN (Eng Lang Lear  
Parker, Jean Elementary School, CHN (Eng Lang Lear  
Mur, John Elementary School, SPN (Eng Lang Lear  
Garfield Elementary School  
Hillcrest Elementary School  
Guadalupe Elementary School, SPN (Eng Lang Lear  
Spring Valley Elementary School, SPN (Eng Lang Lear  
Spring Valley Elementary School  
Malcolm X Elementary School  
Carver, Dr. George Washington Elementary School  
Visitation Valley Elementary School, CHN (Eng Lang Lear  
Parker, Jean Elementary School  
Revere, Paul  
Moscone, George Elementary School  
Sanchez Elementary School, SPN (Eng Lang Lear  
Mur, John Elementary School  
Bryant Elementary School  
Longfellow Elementary School, FL (Eng Lang Lear  
Mission Education Center Elementary School (MEC) - Newcomer Spanish  
Lee, Edwin and Anita Newcomer School/formerly Chinese Education Center (CEC); Newcomer

school\_name

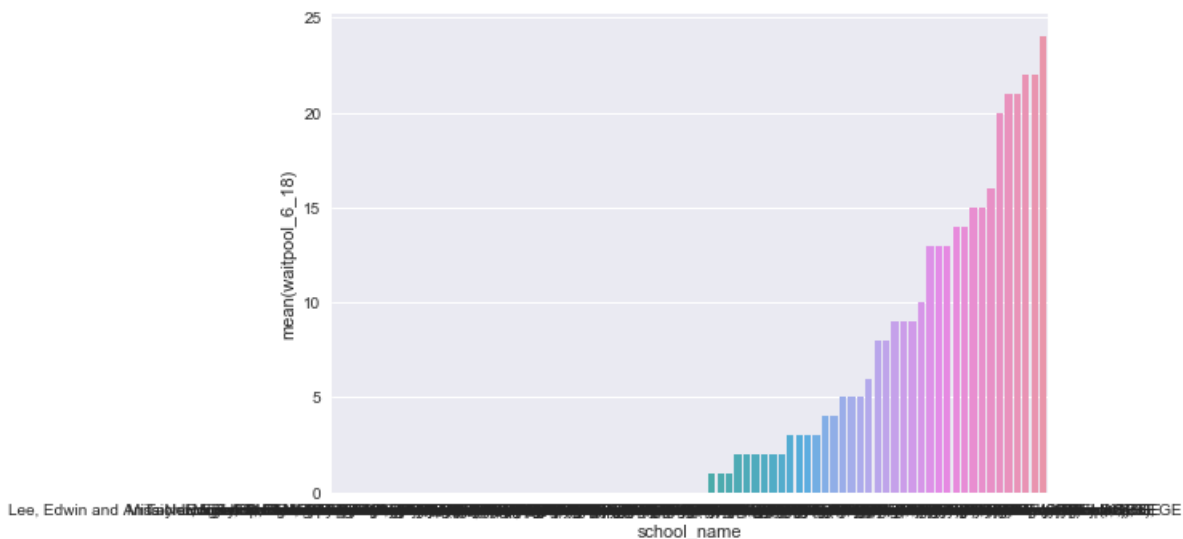
In [57]:

```
#https://nbviewer.jupyter.org/gist/fomightez/bb5a9c727d93d1508187677b4d74d7c1
waitpool_6_18_by_school = data_simplified.groupby(["school_name"]['waitpool_6_18']).aggregate(np.m
edian).reset_index().sort_values('waitpool_6_18')
```

In [58]:

```
sns.barplot(x='school_name', y="waitpool_6_18", data=data_simplified, order=waitpool_6_18_by_school
['school_name'])
plt.show();
```

/Users/Irene/anaconda/lib/python3.6/site-packages/seaborn/categorical.py:1428: FutureWarning:  
remove\_na is deprecated and is a private function. Do not use.  
stat\_data = remove\_na(group\_data)



## Export data files

In [59]:

```
#export to csv
data_simplified.to_csv('data_simplified.csv')
#from pandas import DataFrame
#cdf.to_excel('xcdf.xlsx', sheet_name='sheet1', index=False)
```

## Wrapping Up

Conclusion: round 1 request data did not differ very much year over year. The major differences are seen in the comparison of waitpool

lists for each school.

In [ ]:

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

In [ ]: