

# PyCitySchools-report\_pdf

By: Aswathy Mohan

December 8, 2017

## 0.1 PYCITYSCHOOLS

```
In [1]: #import the neccessary packages
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [4]: school_df.head()
```

```
Out [4]:
```

	School ID	name	type	size	budget
0	0	Huang High School	District	2917	1910635
1	1	Figueroa High School	District	2949	1884411
2	2	Shelton High School	Charter	1761	1056600
3	3	Hernandez High School	District	4635	3022020
4	4	Griffin High School	Charter	1468	917500

```
In [5]: student_df.head()
```

```
Out [5]:
```

	Student ID	name	gender	grade	school
0	0	Paul Bradley	M	9th	Huang High School
1	1	Victor Smith	M	12th	Huang High School
2	2	Kevin Rodriguez	M	12th	Huang High School
3	3	Dr. Richard Scott	M	12th	Huang High School
4	4	Bonnie Ray	F	9th	Huang High School

  

	reading_score	math_score
0	66	79
1	94	61
2	90	60
3	67	58
4	97	84

```
In [6]: # rename the column
school_df=school_df.rename(columns={"name":"school"})
school_df.head()
```

```
Out [6]:
```

	School ID	school	type	size	budget
0	0	Huang High School	District	2917	1910635
1	1	Figueroa High School	District	2949	1884411

2	2	Shelton High School	Charter	1761	1056600
3	3	Hernandez High School	District	4635	3022020
4	4	Griffin High School	Charter	1468	917500

Now the column names are same, perform merge, on school

```
In [7]: student_school_df=pd.merge(student_df,school_df,on="school")
```

```
In [8]: student_school_df.head()
```

```
Out[8]:
```

	Student ID	name	gender	grade	school	\
0	0	Paul Bradley	M	9th	Huang High School	
1	1	Victor Smith	M	12th	Huang High School	
2	2	Kevin Rodriguez	M	12th	Huang High School	
3	3	Dr. Richard Scott	M	12th	Huang High School	
4	4	Bonnie Ray	F	9th	Huang High School	

  

	reading_score	math_score	School ID	type	size	budget
0	66	79	0	District	2917	1910635
1	94	61	0	District	2917	1910635
2	90	60	0	District	2917	1910635
3	67	58	0	District	2917	1910635
4	97	84	0	District	2917	1910635

## 0.2 District

```
In [11]: summary_of_District_Schools=district_summary()
summary_of_District_Schools
```

```
Out[11]:
```

	School Count	student Count	Budget	Avg Maths Score	Avg Reading Score	\
0	7	26976	17347923	76.99	80.96	

  

	Maths Pass %	Reading Pass %	Overall Pass %
0	64.31	78.37	71.34

From above summary , it can be seen that, 64% of students passed Maths and 78.37% passed reading and Overall % is 71.34

## 0.3 Pass % by School

```
In [13]: #use the function to get summary
school_tot=passSchool()
school_tot.head()
```

```
Out[13]:
```

	School ID	School Name	Type	Budget	Budget Per Student	\
0	0	Huang High School	District	1910635	655.0	
1	1	Figueroa High School	District	1884411	639.0	
2	2	Shelton High School	Charter	1056600	600.0	
3	3	Hernandez High School	District	3022020	652.0	

4	4	Griffin High School	Charter	917500	625.0	
	Student Count	Avg Math Score	Avg Read Score	Math %	Read %	Overall %
0	2917	76.63	81.18	63.32	78.81	71.06
1	2949	76.71	81.16	63.75	78.43	71.09
2	1761	83.36	83.73	89.89	92.62	91.25
3	4635	77.29	80.93	64.75	78.19	71.47
4	1468	83.35	83.82	89.71	93.39	91.55

## Top 5 Schools

```
In [14]: #sort the schools and display 5 top schools based on Overall %
sort_school_top=school_tot.sort_values(["Overall %"],ascending=False)
sort_school_top.head()
```

```
Out[14]:
```

	School ID	School Name	Type	Budget	Budget Per Student	\
5	5	Wilson High School	Charter	1319574	578.0	
9	9	Pena High School	Charter	585858	609.0	
10	10	Wright High School	Charter	1049400	583.0	
6	6	Cabrera High School	Charter	1081356	582.0	
8	8	Holden High School	Charter	248087	581.0	
	Student Count	Avg Math Score	Avg Read Score	Math %	Read %	Overall %
5	2283	83.27	83.99	90.93	93.25	92.09
9	962	83.84	84.04	91.68	92.20	91.94
10	1800	83.68	83.95	90.28	93.44	91.86
6	1858	83.06	83.98	89.56	93.86	91.71
8	427	83.80	83.81	90.63	92.74	91.69

The top school is Wilson High School with a Overall pass % of 92.09. It can be seen that , all schools in the top 5 are type: Charter. Also pass % for Maths is lower than reading % for schools.

## Bottom 5 Schools

```
In [15]: #sort the schools and display 5 bottom schools based on Overall %
sort_school_bottom=school_tot.sort_values(["Overall %"],ascending=True)
sort_school_bottom.head()
```

```
Out[15]:
```

	School ID	School Name	Type	Budget	Budget Per Student	\
11	11	Rodriguez High School	District	2547363	637.0	
0	0	Huang High School	District	1910635	655.0	
12	12	Johnson High School	District	3094650	650.0	
1	1	Figueroa High School	District	1884411	639.0	
3	3	Hernandez High School	District	3022020	652.0	
	Student Count	Avg Math Score	Avg Read Score	Math %	Read %	Overall %
11	3999	76.84	80.74	64.07	77.74	70.91

0	2917	76.63	81.18	63.32	78.81	71.06
12	4761	77.07	80.97	63.85	78.28	71.06
1	2949	76.71	81.16	63.75	78.43	71.09
3	4635	77.29	80.93	64.75	78.19	71.47

The Rodriguez High School has the lowest Overall pass % 70.91%. It can be seen that , all schools in the bottom 5 are type: District . Also pass % for Maths is lower than reading % for schools.

## 0.4 Score by Grade

```
In [17]: grade_math_df=grade_sort("math_score")
         grade_read_df=grade_sort("reading_score")
```

### AVERAGE MATH SCORE BY GRADE

```
In [18]: grade_math_df.head()
```

```
Out[18]:
```

	School Name	9th	10th	11th	12th
0	Huang High School	77.03	75.91	76.45	77.23
1	Figueroa High School	76.40	76.54	76.88	77.15
2	Shelton High School	83.42	82.92	83.38	83.78
3	Hernandez High School	77.44	77.34	77.14	77.19
4	Griffin High School	82.04	84.23	83.84	83.36

The 9th grade students have scored maths average score higher than other grades and the 10th grade students have lowest maths average scores

### AVERAGE READING SCORE BY GRADE

```
In [19]: grade_read_df.head()
```

```
Out[19]:
```

	School Name	9th	10th	11th	12th
0	Huang High School	81.29	81.51	81.42	80.31
1	Figueroa High School	81.20	81.41	80.64	81.38
2	Shelton High School	84.12	83.44	84.37	82.78
3	Hernandez High School	80.87	80.66	81.40	80.86
4	Griffin High School	83.37	83.71	84.29	84.01

The 10th grade students have scored reading average score higher than other grades for most schools and the 12th grade students have lowest maths average scores for most schools.

## 0.4.1 SCORE BY SCHOOL SPENDING

```
In [21]: # call the function to retrieve df
         spend_df=spending_student()
         #create bins
         bins=[0,585, 615, 645, 675]
```

```
# Create the names for the four bins
spend_labels = ['<585', '585-615', '616-645', '646-675']
# cut bins according to values
spend_df["Spend Group"] = pd.cut(spend_df["Budget per Student"],bins,labels=spend_labels)
spend_df.head()
```

```
Out[21]:
```

	School ID	Type	Student Count	School Name	School Budget \
0	0	District	2917	Huang High School	1910635
1	1	District	2949	Figueroa High School	1884411
2	2	Charter	1761	Shelton High School	1056600
3	3	District	4635	Hernandez High School	3022020
4	4	Charter	1468	Griffin High School	917500

  

	Budget per Student	Maths	Reading	Maths Pass Count	Read Pass Count \
0	655.0	223528	236810	1847	2299
1	639.0	226223	239335	1880	2313
2	600.0	146796	147441	1583	1631
3	652.0	358238	375131	3001	3624
4	625.0	122360	123043	1317	1371

  

	Spend Group
0	646-675
1	616-645
2	585-615
3	646-675
4	616-645

```
In [23]: grade_school_spend=spending_group(spend_df,"Spend Group")
grade_school_spend
```

```
Out[23]:
```

	Spend Group Range	Avg Maths Score	Avg Reading Score	Maths %	Read % \
0	646-675	77.05	81.01	64.06	78.37
1	616-645	78.06	81.43	68.96	80.95
2	585-615	83.53	83.84	90.53	92.47
3	<585	83.36	83.96	90.33	93.45

  

	Overall %
0	71.22
1	74.95
2	91.50
3	91.89

The overall % is higher for group less than 585 spend range and lowest for 646-675 spend range . Here also it can be seen that Maths % is greater than read % for spend range above 616 and lesser for other two groups.

## 0.4.2 Score By School Size

Use the `spending_student()` create df for grouping by size

```
In [24]: size_df=spending_student()
```

```
size_df.head()
```

```
Out [24]:
```

	School ID	Type	Student Count	School Name	School Budget \
0	0	District	2917	Huang High School	1910635
1	1	District	2949	Figueroa High School	1884411
2	2	Charter	1761	Shelton High School	1056600
3	3	District	4635	Hernandez High School	3022020
4	4	Charter	1468	Griffin High School	917500

  

	Budget per Student	Maths	Reading	Maths Pass Count	Read Pass Count
0	655.0	223528	236810	1847	2299
1	639.0	226223	239335	1880	2313
2	600.0	146796	147441	1583	1631
3	652.0	358238	375131	3001	3624
4	625.0	122360	123043	1317	1371

```
In [25]: #create bins and cut based on labels
```

```
bins_size=[0,1700, 3400, 5100]
```

```
size_labels = ['<1700', '1700-3400', '3400-5100']
```

```
size_df["Size Group"] = pd.cut(spend_df["Student Count"],bins_size,labels=size_labels
```

```
size_df.head()
```

```
Out [25]:
```

	School ID	Type	Student Count	School Name	School Budget \
0	0	District	2917	Huang High School	1910635
1	1	District	2949	Figueroa High School	1884411
2	2	Charter	1761	Shelton High School	1056600
3	3	District	4635	Hernandez High School	3022020
4	4	Charter	1468	Griffin High School	917500

  

	Budget per Student	Maths	Reading	Maths Pass Count	Read Pass Count \
0	655.0	223528	236810	1847	2299
1	639.0	226223	239335	1880	2313
2	600.0	146796	147441	1583	1631
3	652.0	358238	375131	3001	3624
4	625.0	122360	123043	1317	1371

  

	Size Group
0	1700-3400
1	1700-3400
2	1700-3400
3	3400-5100
4	<1700

```
In [26]: #call the spending_group function to calculate the group based %
```

```
grade_school_size=spending_group(size_df,"Size Group")
```

```
#grade_school_size.rest_index()
```

```
grade_school_size.set_index('Size Group Range')
grade_school_size.head()
```

```
Out [26]:
```

	Size Group Range	Avg Maths Score	Avg Reading Score	Maths %	Read %	\
0	1700-3400	79.89	82.40	76.51	85.37	
1	3400-5100	77.07	80.93	64.34	78.42	
2	<1700	83.52	83.88	90.41	92.90	

  

	Overall %
0	80.94
1	71.38
2	91.66

The overall % is higher for group less than 1700 size range and lowest for 3400-5100 Size range . Here also it can be seen that Maths % is greater than read % for size range all groups.

### 0.4.3 Scores by School Type

```
In [27]: #Use the spending_student() create df for grouping by type
type_df=spending_student()
```

```
type_df.head()
```

```
Out [27]:
```

	School ID	Type	Student Count	School Name	School Budget	\
0	0	District	2917	Huang High School	1910635	
1	1	District	2949	Figueroa High School	1884411	
2	2	Charter	1761	Shelton High School	1056600	
3	3	District	4635	Hernandez High School	3022020	
4	4	Charter	1468	Griffin High School	917500	

  

	Budget per Student	Maths	Reading	Maths Pass Count	Read Pass Count
0	655.0	223528	236810	1847	2299
1	639.0	226223	239335	1880	2313
2	600.0	146796	147441	1583	1631
3	652.0	358238	375131	3001	3624
4	625.0	122360	123043	1317	1371

```
In [28]: #call the spending_group function to calculate the group based %
grade_school_district=spending_group(size_df,"Type")
grade_school_district.set_index("Type Range")
grade_school_district
```

```
Out [28]:
```

	Type Range	Avg Maths Score	Avg Reading Score	Maths %	Read %	Overall %
0	District	76.99	80.96	64.31	78.37	71.34
1	Charter	83.41	83.90	90.28	93.15	91.72

The overall % is higher for Charter type by more than 20% .

## **0.5 Final Results:**

1. It can be analyzed from above data that Charter schools score higher pass % than District type schools.
2. The schools with higher students counts has the higher pass %
3. For schools with highest Budget Per Student has lower Pass % rate , which means the schools which have low budget per students scored more marks.
4. Also schools in top 5 are all of type Charter.