main

September 28, 2019

[1]:
```
# Dependencies and Setup
import pandas as pd
```

[2]:
```
# File to Load (Remember to Change These)
school_data_to_load = "Resources/schools_complete.csv"
student_data_to_load = "Resources/students_complete.csv"

# Read School and Student Data File and store into Pandas Data Frames
school_data = pd.read_csv(school_data_to_load)
student_data = pd.read_csv(student_data_to_load)

# Combine the data into a single dataset
school_data_complete = pd.merge(student_data, school_data, how="left",␣
 ↪on=["school_name", "school_name"])
```

[3]:
```
school_data_complete.columns
```

[3]:
```
Index(['Student ID', 'student_name', 'gender', 'grade', 'school_name',
       'reading_score', 'math_score', 'School ID', 'type', 'size', 'budget'],
      dtype='object')
```

[5]:
```
Total_Schools = len(school_data_complete["school_name"].unique())
Total_Students = school_data_complete["Student ID"].count()
Total_Budget = school_data_complete["budget"].sum()
Average_Math_Score = school_data_complete["math_score"].mean()
Average_Reading_Score = school_data_complete["reading_score"].mean()
math_score_pass =␣
 ↪school_data_complete[(school_data_complete["math_score"]>=70)].
 ↪count()["math_score"]
reading_score_pass =␣
 ↪school_data_complete[(school_data_complete["reading_score"]>=70)].
 ↪count()["reading_score"]
math_pass_percent = (math_score_pass/Total_Students)*100
reading_pass_percent = (reading_score_pass/Total_Students)*100
overall_passing_rate = (Average_Math_Score + Average_Reading_Score)/2
```

[6]:
```
district_summary = pd.DataFrame({"Total Schools":[Total_Schools],
                                 "Total Students":[Total_Students],
                                 "Total Budget":[Total_Budget],
```

```
                                "Average Math Score":[Average_Math_Score],
                                "Average Reading Score":
 ↪[Average_Reading_Score],
                                "% Passing Math":[math_pass_percent],
                                "% Passing Reading":[reading_pass_percent],
                                "% Overall Passing Rate":
 ↪[overall_passing_rate]})
district_summary
```

[6]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
0	15	39170	82932329558	78.985371	81.87784	74.980853	85.805463	80.431606

[7]:
```
school_data_complete.groupby(["school_name"]).mean()
```

[7]:

school_name	Student ID	reading_score	math_score	School ID	size	budget
Bailey High School	20358.5	81.033963	77.048432	7.0	4976.0	3124928.0
Cabrera High School	16941.5	83.975780	83.061895	6.0	1858.0	1081356.0
Figueroa High School	4391.0	81.158020	76.711767	1.0	2949.0	1884411.0
Ford High School	36165.0	80.746258	77.102592	13.0	2739.0	1763916.0
Griffin High School	12995.5	83.816757	83.351499	4.0	1468.0	917500.0
Hernandez High School	9944.0	80.934412	77.289752	3.0	4635.0	3022020.0
Holden High School	23060.0	83.814988	83.803279	8.0	427.0	248087.0
Huang High School	1458.0	81.182722	76.629414	0.0	2917.0	1910635.0
Johnson High School	32415.0	80.966394	77.072464	12.0		
Pena High School	23754.5	84.044699	83.839917	9.0		
Rodriguez High School	28035.0	80.744686	76.842711	11.0		
Shelton High School	6746.0	83.725724	83.359455	2.0		
Thomas High School	38352.0	83.848930	83.418349	14.0		
Wilson High School	14871.0	83.989488	83.274201	5.0		
Wright High School	25135.5	83.955000	83.682222	10.0		

```
Johnson High School    4761.0  3094650.0
Pena High School        962.0   585858.0
Rodriguez High School  3999.0  2547363.0
Shelton High School    1761.0  1056600.0
Thomas High School     1635.0  1043130.0
Wilson High School     2283.0  1319574.0
Wright High School     1800.0  1049400.0
```

[8]:
```
school_summary = school_data_complete.groupby(["school_name"]).mean()
school_summary["school_math_pass"] =␣
 ↪school_data_complete[(school_data_complete["math_score"]>=70)].
 ↪groupby(["school_name"])["size"].count() / school_data_complete.
 ↪groupby(["school_name"])["size"].count()
school_summary["school_reading_pass"] =␣
 ↪school_data_complete[(school_data_complete["reading_score"]>=70)].
 ↪groupby(["school_name"])["size"].count() / school_data_complete.
 ↪groupby(["school_name"])["size"].count()
school_summary["per_student_budget"] = school_data_complete.
 ↪groupby(["school_name"]).mean()["budget"] / school_data_complete.
 ↪groupby(["school_name"]).mean()["size"]
school_summary
```

[8]:
```
                      Student ID  reading_score  math_score  School ID  \
school_name
Bailey High School       20358.5      81.033963   77.048432        7.0
Cabrera High School      16941.5      83.975780   83.061895        6.0
Figueroa High School      4391.0      81.158020   76.711767        1.0
Ford High School         36165.0      80.746258   77.102592       13.0
Griffin High School      12995.5      83.816757   83.351499        4.0
Hernandez High School     9944.0      80.934412   77.289752        3.0
Holden High School       23060.0      83.814988   83.803279        8.0
Huang High School         1458.0      81.182722   76.629414        0.0
Johnson High School      32415.0      80.966394   77.072464       12.0
Pena High School         23754.5      84.044699   83.839917        9.0
Rodriguez High School    28035.0      80.744686   76.842711       11.0
Shelton High School       6746.0      83.725724   83.359455        2.0
Thomas High School       38352.0      83.848930   83.418349       14.0
Wilson High School       14871.0      83.989488   83.274201        5.0
Wright High School       25135.5      83.955000   83.682222       10.0

                        size     budget  school_math_pass  \
school_name
Bailey High School     4976.0  3124928.0          0.666801
Cabrera High School    1858.0  1081356.0          0.941335
Figueroa High School   2949.0  1884411.0          0.659885
Ford High School       2739.0  1763916.0          0.683096
Griffin High School    1468.0   917500.0          0.933924
Hernandez High School  4635.0  3022020.0          0.667530
```

```
Holden High School      427.0   248087.0          0.925059
Huang High School      2917.0  1910635.0          0.656839
Johnson High School    4761.0  3094650.0          0.660576
Pena High School        962.0   585858.0          0.945946
Rodriguez High School  3999.0  2547363.0          0.663666
Shelton High School    1761.0  1056600.0          0.938671
Thomas High School     1635.0  1043130.0          0.932722
Wilson High School     2283.0  1319574.0          0.938677
Wright High School     1800.0  1049400.0          0.933333
```

school_name	school_reading_pass	per_student_budget
Bailey High School	0.819333	628.0
Cabrera High School	0.970398	582.0
Figueroa High School	0.807392	639.0
Ford High School	0.792990	644.0
Griffin High School	0.971390	625.0
Hernandez High School	0.808630	652.0
Holden High School	0.962529	581.0
Huang High School	0.813164	655.0
Johnson High School	0.812224	650.0
Pena High School	0.959459	609.0
Rodriguez High School	0.802201	637.0
Shelton High School	0.958546	600.0
Thomas High School	0.973089	638.0
Wilson High School	0.965396	578.0
Wright High School	0.966111	583.0

[13]:
```
school_summary_df = pd.DataFrame({  "School Type":school_data_complete.
 ↪groupby(["school_name"])["type"].unique(),
                                    "Total Students":school_data_complete.
 ↪groupby(["school_name"])["size"].unique(),
                                    "Total School Budget":school_data_complete.
 ↪groupby(["school_name"])["budget"].unique(),
                                    "Per Student Budget":
 ↪school_summary["per_student_budget"],
                                    "Average Math Score":school_data_complete.
 ↪groupby(["school_name"]).mean()["math_score"],
                                    "Average Reading Score":school_data_complete.
 ↪groupby(["school_name"]).mean()["reading_score"],
                                    "% Passing Math":
 ↪school_summary["school_math_pass"]*100,
                                    "% Passing Reading":
 ↪school_summary["school_reading_pass"]*100,
                                    "% Overall Passing Rate":
 ↪((school_summary["school_math_pass"] +␣
 ↪school_summary["school_reading_pass"])*50)})
```

```
#school_summary_df["Total School Budget"] = school_summary_df["Total School␣
 ↪Budget"].map("${:.2f}".format)
#school_summary_df["Per Student Budget"] = school_summary_df["Per Student␣
 ↪Budget"].map("${:.2f}".format)
#school_summary_df["Total Students"] = school_summary_df["Total Students"].
 ↪map("{:,}".format)
```

[14]:
```
schools_passing_rate = school_summary_df.sort_values(["% Overall Passing␣
 ↪Rate"],ascending=False)
schools_passing_rate.head()
```

[14]:

school_name	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Cabrera High School	[Charter]	[1858]	[1081356]	582.0	83.061895	83.975780	94.133477	97.039828	95.586652
Thomas High School	[Charter]	[1635]	[1043130]	638.0	83.418349	83.848930	93.272171	97.308869	95.290520
Pena High School	[Charter]	[962]	[585858]	609.0	83.839917	84.044699	94.594595	95.945946	95.270270
Griffin High School	[Charter]	[1468]	[917500]	625.0	83.351499	83.816757	93.392371	97.138965	95.265668
Wilson High School	[Charter]	[2283]	[1319574]	578.0	83.274201	83.989488	93.867718	96.539641	95.203679

[39]:
```
bins = [0, 585, 615, 645, 675]
group_names = ["<585", "585-615","615-645","645-675"]
school_data_complete["bps"] = school_data_complete['budget']/
 ↪school_data_complete['size']
```

```
school_data_complete["school_bins"] = pd.
 ↪cut(school_data_complete["bps"],bins,labels=group_names)
school_data_complete.groupby('school_bins').mean()

#school_data_complete["per_student_budget"] = school_data_complete["budget"]/
 ↪school_data_complete["size"]
#school_data_complete["bins"] = pd.
 ↪cut(school_data_complete["per_student_budget"], bins, labels=group_names)
#school_data_complete["% Passing Math"] =␣
 ↪school_data_complete[(school_data_complete["math_score"]>=70)].
 ↪groupby(["bins"])["size"].count() / school_data_complete.
 ↪groupby(["bins"])["size"].count()
#school_data_complete["% Passing Reading"] =␣
 ↪school_data_complete[(school_data_complete["reading_score"]>=70)].
 ↪groupby(["bins"])["size"].count() / school_data_complete.
 ↪groupby(["bins"])["size"].count()
#school_data_complete.groupby(["bins"]).mean()
```

[39]:

school_bins	Student ID	reading_score	math_score	School ID	size	budget	bps
<585	18925.615578	83.964039	83.363065	6.906250	1898.018530	1.101853e+06	580.781564
585-615	12754.878810	83.838414	83.529196	4.473008	1478.723834	8.902931e+05	603.179581
615-645	22920.396769	81.434088	78.061635	8.225768	3477.403355	2.205192e+06	634.990881
645-675	16622.369934	81.005604	77.049297	5.769268	4276.718509	2.786812e+06	651.937383

[85]:
```
#school_grade = school_data_complete.
 ↪groupby(["school_name","grade"])["math_score"].mean()
#pd.pivot_table(school_grade,values =␣
 ↪"grade",index="school_name",columns=["9th","10th","11th","12th"]).
 ↪reset_index()

tjw = school_data_complete[['school_name','grade','math_score']]
hello = tjw.pivot_table(values =␣
 ↪"math_score",index="school_name",columns='grade').reset_index()
#ninth_grade = school_data_complete[school_data_complete["grade"]="9th"]).
 ↪groupby(["school_name"])["math_score"].mean()
```

[94]:
```
cols = ['school_name','9th','10th','11th','12th']
hello = hello[cols]
```

[95]:
```
hello
```

[95]:

grade	school_name	9th	10th	11th	12th
0	Bailey High School	77.083676	76.996772	77.515588	76.492218
1	Cabrera High School	83.094697	83.154506	82.765560	83.277487
2	Figueroa High School	76.403037	76.539974	76.884344	77.151369
3	Ford High School	77.361345	77.672316	76.918058	76.179963
4	Griffin High School	82.044010	84.229064	83.842105	83.356164
5	Hernandez High School	77.438495	77.337408	77.136029	77.186567
6	Holden High School	83.787402	83.429825	85.000000	82.855422
7	Huang High School	77.027251	75.908735	76.446602	77.225641
8	Johnson High School	77.187857	76.691117	77.491653	76.863248
9	Pena High School	83.625455	83.372000	84.328125	84.121547
10	Rodriguez High School	76.859966	76.612500	76.395626	77.690748
11	Shelton High School	83.420755	82.917411	83.383495	83.778976
12	Thomas High School	83.590022	83.087886	83.498795	83.497041
13	Wilson High School	83.085578	83.724422	83.195326	83.035794
14	Wright High School	83.264706	84.010288	83.836782	83.644986

[ ]: