

•

[5]

```
# Dependencies and Setup

import pandas as pd

import numpy as np


# 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 DataFrames

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"])
```

```
-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-5-efae349c28dd> in <module>
      1 # Dependencies and Setup
----> 2 import pandas as pd
      3 import numpy as np
      4
      5 # File to Load (Remember to Change These)
```

```
ModuleNotFoundError: No module named 'pandas'
```

[-]

```
## District Summary
```

```
* Calculate the total number of schools
```

```
* Calculate the total number of students
```

```
* Calculate the total budget
```

- \* Calculate the average math score
- \* Calculate the average reading score
- \* Calculate the percentage of students with a passing math score (70 or greater)
- \* Calculate the percentage of students with a passing reading score (70 or greater)
- \* Calculate the percentage of students who passed math **\*\*and\*\*** reading (% Overall Passing)
- \* Create a dataframe to hold the above results
- \* Optional: give the displayed data cleaner formatting

[5]

	Total School s	Total Studen ts	Total Budget	Average Math Score	Average Reading Score	% Passin g Math	% Passing Reading	% Overall Passing
0 15		39,170	\$24,649 ,428.00	78.985371	81.87784	74.980 853	85.80546 3	65.17232 6

[-]

## School Summary

[-]

- \* Create an overview table that summarizes key metrics about each school, including:
  - \* School Name
  - \* School Type
  - \* Total Students
  - \* Total School Budget
  - \* Per Student Budget
  - \* Average Math Score
  - \* Average Reading Score
  - \* % Passing Math

```
* % Passing Reading
* % Overall Passing (The percentage of students that passed math **and** reading.)

* Create a dataframe to hold the above results
```

```
[6]
```

```
len(school_data_complete['school_name'].unique())
```

```
school_data_complete['school_name'].unique
```

```
school_data_complete['student ID'].unique
```

```
school_data['Budget'].sum()
```

```
school_data_complete['school_name', 'budget'].drop_duplicates()
```

```
school_data_complete['math_score'].mean()
```

```
school_data_complete['reading_score'].mean()
```

```
len(school_data_complete[school_data_complete['math_score'].>=70])/num_students
```

```
(len(school_data_complete[school_data_complete['math_score'].>=70]) &
```

```
(school_data_complete['reading_score']>=70)]/
```

```
num_students
```

```
averages=school_data_complete.groupby('school_name')[['math_score',reading_score']].
```

```
mean()
```

```
averages
```

```
averages=averages.reset_index()
```

```
school_Data=school_Data.reset_index()
```

```
school_data['budget']/('school_data')
```

```
school_data['per student budget']=school_data['budget']/school_data['size']
```

```
school_Data=school_data.set_index('school_name')
```

```
school_data=pd.merge(school_data,averages)
```

```
l=[]
```

```
if school_data_complete['math_score']>=70:
```

```
    l.append(1)
```

```
ELSE:
```

```
    l.append(0)
```

```
school_data_complete['math_dummy']=l
```

```
school_data_complete['math_dummy']=np.where(school_data_complete['math_score']>=70,1,0)
```

```
math_passing=school_data_complete.groupby('school_name',as_index=False)['math_dummy'].sum()
```

```
math_passing_df=school_data_complete[school_data_complete['math_score']>=70]
```

```
math_passing_df=math_passing_df.groupby('school_name',as_index=False)['school_name'].count()
```

```
pd.merge(school_Data,math_passing_df)
```

```
school_data=pd.merge(school_Data,math_passing)
```

```
school_data['percent_passing math']=school_Data['math_dummy']/school_Data['size']
```

```
school_Data=school_Data.rename(columns={'percent passing':'percent passing math'})
```

```
school_Data=school_Data.drop('math_dummy',axis=1)
```

```
school_Data
```

```
bins = [0, 999, 1999, 9999999999]
```

```
group_name = ["Small (<1000)", "Medium (1000-2000)" , "Large (>2000)"]
```

```
all_df['size_bins'] = pd.cut(all_df['size'], bins, labels = group_name)
```

```
#group by spending
```

```
by_size = all_df.groupby('size_bins')
```

```
#calculations
```

```
avg_math = by_size['math_score'].mean()
```

```
avg_read = by_size['math_score'].mean()
```

```
pass_math = all_df[allScores by School Type_df['math_score'] >= 70].groupby('size_bins')['Student ID'].count()/by_size
```

```
['Student ID'].count()
```

```
pass_read = all_df[all_df['reading_score'] >= 70].groupby('size_bins')['Student ID'].count()/by_size['Student ID'].count()
```

```

overall = all_df[(all_df['reading_score'] >= 70) & (all_df['math_score'] >=
70)].groupby('size_bins')['Student ID'].count

()/by_size['Student ID'].count()

# df build
scores_by_size = pd.DataFrame({
    "Average Math Score": avg_math,
    "Average Reading Score": avg_read,
    '% Passing Math': pass_math,
    '% Passing Reading': pass_read,
    "Overall Passing Rate": overall

})

#reorder columns
scores_by_size = scores_by_size[[
    "Average Math Score",
    "Average Reading Score",
    '% Passing Math',
    '% Passing Reading',
    "Overall Passing Rate"
]]

Group SUMMARY

#groups by school
school = all_df.set_index('school').groupby(['school'])

types = school_df.set_index('school')['type']

# sum students by school
student_school = all_df.groupby("school")['Student ID'].count()

# school budget

```

```

school_budget = school_df.set_index('school')['budget']

#student budget
student_budget =
school_df.set_index('school')['budget']/school_df.set_index('school')['size']

#avg scores by school
avg_math = school['math_score'].mean()
avg_read = school['reading_score'].mean()

# % passing scores
pass_math = all_df[all_df['math_score'] >= 70].groupby('school')['Student
ID'].count()/student_school
pass_read = all_df[all_df['reading_score'] >= 70].groupby('school')['Student
ID'].count()/student_school
overall = all_df[(all_df['reading_score'] >= 70) & (all_df['math_score'] >=
70)].groupby('school')['Student ID'].count()/
student_school

school_summary = pd.DataFrame({
    "School Type": types,
    "Total Students": student_school,
    "Per Student Budget": student_budget,
    "Total School Budget": school_budget,
    "Average Math Score": avg_math,
    "Average Reading Score": avg_read,
    '% Passing Math': pass_math,
    '% Passing Reading': pass_read,
    "Overall Passing Rate": overall
})

```

```

#munging
school_summary = school_summary[['School Type',
                                  'Total Students',
                                  'Total School Budget',
                                  'Per Student Budget',
                                  'Average Math Score',
                                  'Average Reading Score',
                                  '% Passing Math',
                                  '% Passing Reading',
                                  'Overall Passing Rate']]

#formatting
school_summary.style.format({'Total Students': '{:,}',
                              "Total School Budget": "${:,}",
                              "Per Student Budget": "${:,.0f}",
                              'Average Math Score': "{:.1f}",
                              'Average Reading Score': "{:.1f}",
                              "% Passing Math": "{:.1%}",
                              "% Passing Reading": "{:.1%}",
                              "Overall Passing Rate": "{:.1%}"})

scores_by_size.index.name = "Total Students"
scores_by_size = scores_by_size.reindex(group_name)

#formating
scores_by_size.style.format({'Average Math Score': '{:.1f}',
                              'Average Reading Score': '{:.1f}',
                              '% Passing Math': '{:.1%}',
                              '% Passing Reading': '{:.1%}',
                              'Overall Passing Rate': '{:.1%}'})

by_type = all_df.groupby("type")

#calculations
avg_math = by_type['math_score'].mean()
avg_read = by_type['math_score'].mean()

```



[illegible]

Schools under 2000 students have much higher passing rates than those with student populations above 2000. A comparison of 95 to 75%. The same phenomenon is seen with high and low per student spending brackets and district versus charter schools.

```
File "<ipython-input-6-6bc48c1a6c89>", line 3
    school_data_complete['school_name'].nunique
                                         ^
SyntaxError: EOL while scanning string literal
```

## Top Performing Schools (By % Overall Passing)

- Sort and display the top five performing schools by % overall passing.

[10]

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
Cabrera High School	Charter	1858	\$1,081,356.00	\$582.00	83.061895	83.975780	94.13347	97.039828	91.334769

Thomas High School	Charter	1635	\$1,043,130.00	\$638.00	83.418349	83.848930	93.272171	97.308869	90.948012
Griffin High School	Charter	1468	\$917,500.00	\$625.00	83.351499	83.816757	93.392371	97.138965	90.599455
Wilson High School	Charter	2283	\$1,319,574.00	\$578.00	83.274201	83.989488	93.867718	96.539641	90.582567
Pena High School	Charter	962	\$585,858.00	\$609.00	83.839917	84.044699	94.594595	95.945946	90.540541

## Bottom Performing Schools (By % Overall Passing)

- Sort and display the five worst-performing schools by % overall passing.

```
[11]
```

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
Rodriguez High School	District	3999	\$2,547,363.00	\$637.00	76.842711	80.744686	66.366592	80.220055	52.988247
Figueroa High School	District	2949	\$1,884,411.00	\$639.00	76.711767	81.158020	65.988471	80.739234	53.204476
Huang High School	District	2917	\$1,910,635.00	\$655.00	76.629414	81.182722	65.683922	81.316421	53.513884

<b>Hernandez High School</b>	District	4635	\$3,022,020.00	\$652.00	77.289752	80.934412	66.752967	80.862999	53.527508
<b>Johnson High School</b>	District	4761	\$3,094,650.00	\$650.00	77.072464	80.966394	66.057551	81.222432	53.539172

## Math Scores by Grade

- Create a table that lists the average Reading Score for students of each grade level (9th, 10th, 11th, 12th) at each school.
  - Create a pandas series for each grade. Hint: use a conditional statement.
  - Group each series by school
  - Combine the series into a dataframe
  - Optional: give the displayed data cleaner formatting

[12]

	9th	10th	11th	12th
<b>Bailey High School</b>	77.083676	76.996772	77.515588	76.492218
<b>Cabrera High School</b>	83.094697	83.154506	82.765560	83.277487
<b>Figueroa High School</b>	76.403037	76.539974	76.884344	77.151369
<b>Ford High School</b>	77.361345	77.672316	76.918058	76.179963
<b>Griffin High School</b>	82.044010	84.229064	83.842105	83.356164
<b>Hernandez High School</b>	77.438495	77.337408	77.136029	77.186567

Holden High School	83.78740 2	83.42982 5	85.00000 0	82.85542 2
Huang High School	77.02725 1	75.90873 5	76.44660 2	77.22564 1
Johnson High School	77.18785 7	76.69111 7	77.49165 3	76.86324 8
Pena High School	83.62545 5	83.37200 0	84.32812 5	84.12154 7
Rodriguez High School	76.85996 6	76.61250 0	76.39562 6	77.69074 8
Shelton High School	83.42075 5	82.91741 1	83.38349 5	83.77897 6
Thomas High School	83.59002 2	83.08788 6	83.49879 5	83.49704 1
Wilson High School	83.08557 8	83.72442 2	83.19532 6	83.03579 4
Wright High School	83.26470 6	84.01028 8	83.83678 2	83.64498 6

## Reading Score by Grade

- Perform the same operations as above for reading scores

[13]

	9th	10th	11th	12th
Bailey High School	81.30315 5	80.90718 3	80.94564 3	80.91245 1



Wright High School	83.83333	83.81275	84.15632	84.07317
	3	7	2	1

## Scores by School Spending

- Create a table that breaks down school performances based on average Spending Ranges (Per Student). Use 4 reasonable bins to group school spending. Include in the table each of the following:
  - Average Math Score
  - Average Reading Score
  - % Passing Math
  - % Passing Reading
  - Overall Passing Rate (Average of the above two)

[18]

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
Spending Ranges (Per Student)					
<\$585	83.46	83.93	93.46	96.61	90.37
\$585-630	81.90	83.16	87.13	92.72	81.42
\$630-645	78.52	81.62	73.48	84.39	62.86
\$645-680	77.00	81.03	66.16	81.13	53.53

## Scores by School Size

- Perform the same operations as above, based on school size.

[22]

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
School Size					
Small (<1000)	83.821598	83.929843	93.550225	96.099437	89.883853
Medium (1000-2000)	83.374684	83.864438	93.599695	96.790680	90.621535
Large (2000-5000)	77.746417	81.344493	69.963361	82.766634	58.286003

## Scores by School Type

- Perform the same operations as above, based on school type

[24]

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
School Type					
Charter	83.473852	83.896421	93.620830	96.586489	90.432244
District	76.956733	80.966636	66.548453	80.799062	53.672208

[-]