

```
In [1]: import pandas as pd
import numpy as np
import csv
import os
```

```
In [2]: #Join path for data and create a data frame for the student data
students = os.path.join('students_complete.csv')
students_df = pd.read_csv(students)
students_df['Passed Reading'] = np.where(students_df['reading_score'] >= 70, 1, 0)
students_df['Passed Math'] = np.where(students_df['math_score'] >= 70, 1, 0)
```

```
In [3]: #Import schools data, sort values for the school, create a data frame for data
schools = os.path.join('schools_complete.csv')
schools_df = pd.read_csv(schools)
schools_df = schools_df.sort_values('name').reset_index()
schools_df = schools_df.drop(columns = ['index', 'School ID'])
schools_df = pd.DataFrame(schools_df)
```

```
In [4]: school_district = schools_df.loc[schools_df['type'] == 'District', :]
school_district = pd.DataFrame(school_district)
```

```
In [5]: student_district = pd.DataFrame()
for school in school_district['name']:
    student_district = student_district.append(students_df.loc[students_df['School ID'] == school])
```

```
In [6]: Create a high level snapshot (in table form) of the district's key metrics
average_reading_score = student_district['reading_score'].mean()
average_math_score = student_district['math_score'].mean()
total_school = school_district['name'].count()
total_student = school_district['size'].sum()
total_budget = school_district['budget'].sum()
reading_score = student_district.loc[student_district['reading_score'] >= 70, 'reading_score'].mean()
passing_reading = (reading_score/total_student)*100
math_score = student_district.loc[student_district['math_score'] >= 70, 'math_score'].mean()
passing_math = (math_score/total_student)*100
average_passing = (passing_math + passing_reading)/2
district_summary = pd.DataFrame({"Total Schools": [total_school], "Total Students": [total_student], "Total Budget": [total_budget], "Average Reading Score": [average_reading_score], "Average Math Score": [average_math_score], "% Passing Reading": [passing_reading], "% Passing Math": [passing_math], "% Overall Passing Rate": [average_passing]})
district_summary
```

Out[6]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
0	7	26976	17347923	76.987026	80.962485	66.518387	80.905249	73.711818

```

In [7]: #Part 2 for schools summary
summary = pd.DataFrame()
summary[''] = schools_df['name']
summary['School Type'] = schools_df['type']
summary['Total Students'] = schools_df['size']
summary['Total Budget'] = schools_df['budget']
summary['Per Student Budget'] = summary['Total Budget']/summary['Total Students']
summary['Average Reading Score'] = students_df.groupby(students_df['school'])['Reading Score'].mean()
summary['Average Math Score'] = students_df.groupby(students_df['school'])['Math Score'].mean()
reading_passed = students_df.groupby(students_df['school'])['Passed Reading'].sum()
summary['% Passing Reading'] = (reading_passed/schools_df['size']).round(2)
math_passed = students_df.groupby(students_df['school'])['Passed Math'].sum()
summary['% Passing Math'] = (math_passed/schools_df['size']).round(2)
summary['% Overall Passing Rate'] = (summary['% Passing Reading'] + summary['% Passing Math'])/2
summary.set_index('', inplace = True)
summary

```

Out[7]:

	School Type	Total Students	Total Budget	Per Student Budget	Average Reading Score	Average Math Score	% Passing Reading	% Passing Math	% Overall Passing Rate
Bailey High School	District	4976	3124928	628.0	81.033963	77.048432	81.933280	66.680064	74.306672
Cabrera High School	Charter	1858	1081356	582.0	83.975780	83.061895	97.039828	94.133477	95.586653
Figueroa High School	District	2949	1884411	639.0	81.158020	76.711767	80.739234	65.988471	73.363853
Ford High School	District	2739	1763916	644.0	80.746258	77.102592	79.299014	68.309602	73.804308
Griffin High School	Charter	1468	917500	625.0	83.816757	83.351499	97.138965	93.392371	95.265168
Hernandez High School	District	4635	3022020	652.0	80.934412	77.289752	80.862999	66.752967	73.807483
Holden High School	Charter	427	248087	581.0	83.814988	83.803279	96.252927	92.505855	94.379391
Huang High School	District	2917	1910635	655.0	81.182722	76.629414	81.316421	65.683922	73.500172
Johnson High School	District	4761	3094650	650.0	80.966394	77.072464	81.222432	66.057551	73.639992
Pena High School	Charter	962	585858	609.0	84.044699	83.839917	95.945946	94.594595	95.270271
Rodriguez High School	District	3999	2547363	637.0	80.744686	76.842711	80.220055	66.366592	73.293324

	School Type	Total Students	Total Budget	Per Student Budget	Average Reading Score	Average Math Score	% Passing Reading	% Passing Math	% Overall Passing Rate
Shelton High School	Charter	1761	1056600	600.0	83.725724	83.359455	95.854628	93.867121	94.860
Thomas High School	Charter	1635	1043130	638.0	83.848930	83.418349	97.308869	93.272171	95.290
Wilson High School	Charter	2283	1319574	578.0	83.989488	83.274201	96.539641	93.867718	95.203
Wright High School	Charter	1800	1049400	583.0	83.955000	83.682222	96.611111	93.333333	94.972

```
In [8]: #Top 5 schools have the highest overall passing rate
top_5 = summary.sort_values('% Overall Passing Rate').tail()
top_5
```

Out[8]:

	School Type	Total Students	Total Budget	Per Student Budget	Average Reading Score	Average Math Score	% Passing Reading	% Passing Math	% Overall Passing Rate
Wilson High School	Charter	2283	1319574	578.0	83.989488	83.274201	96.539641	93.867718	95.20367
Griffin High School	Charter	1468	917500	625.0	83.816757	83.351499	97.138965	93.392371	95.26566
Pena High School	Charter	962	585858	609.0	84.044699	83.839917	95.945946	94.594595	95.27027
Thomas High School	Charter	1635	1043130	638.0	83.848930	83.418349	97.308869	93.272171	95.29052
Cabrera High School	Charter	1858	1081356	582.0	83.975780	83.061895	97.039828	94.133477	95.58665

```
In [9]: #Bottom 5 schools have the lowest overall passing rate
bottom_5 = summary.sort_values('% Overall Passing Rate').head()
bottom_5
```

Out[9]:

	School Type	Total Students	Total Budget	Per Student Budget	Average Reading Score	Average Math Score	% Passing Reading	% Passing Math	% Overall Passing Rate
Rodriguez High School	District	3999	2547363	637.0	80.744686	76.842711	80.220055	66.366592	73.2931
Figueroa High School	District	2949	1884411	639.0	81.158020	76.711767	80.739234	65.988471	73.3631
Huang High School	District	2917	1910635	655.0	81.182722	76.629414	81.316421	65.683922	73.5001
Johnson High School	District	4761	3094650	650.0	80.966394	77.072464	81.222432	66.057551	73.6391
Ford High School	District	2739	1763916	644.0	80.746258	77.102592	79.299014	68.309602	73.8041

```
In [10]: #Create a dataframe of math score by grade for each school
math_score_by_grade = students_df.groupby(['school', 'grade'])['math_score']
math_score_by_grade = math_score_by_grade.rename(columns = {'school': '', 'grade': ''})
math_score_by_grade = math_score_by_grade.pivot(index = '', columns=' ', values='math_score')
math_score_by_grade
```

Out[10]:

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

```
In [11]: #Create a dataframe of reading score by grade for each school
reading_score_by_grade = students_df.groupby(['school', 'grade'])['reading_score'].mean()
reading_score_by_grade = reading_score_by_grade.rename(columns = {'school': 'School'})
reading_score_by_grade = reading_score_by_grade.pivot(index = 'School', columns='grade', values='reading_score_by_grade')
```

Out[11]:

	10th	11th	12th	9th
Bailey High School	80.907183	80.945643	80.912451	81.303155
Cabrera High School	84.253219	83.788382	84.287958	83.676136
Figueroa High School	81.408912	80.640339	81.384863	81.198598
Ford High School	81.262712	80.403642	80.662338	80.632653
Griffin High School	83.706897	84.288089	84.013699	83.369193
Hernandez High School	80.660147	81.396140	80.857143	80.866860
Holden High School	83.324561	83.815534	84.698795	83.677165
Huang High School	81.512386	81.417476	80.305983	81.290284
Johnson High School	80.773431	80.616027	81.227564	81.260714
Pena High School	83.612000	84.335938	84.591160	83.807273
Rodriguez High School	80.629808	80.864811	80.376426	80.993127
Shelton High School	83.441964	84.373786	82.781671	84.122642
Thomas High School	84.254157	83.585542	83.831361	83.728850
Wilson High School	84.021452	83.764608	84.317673	83.939778
Wright High School	83.812757	84.156322	84.073171	83.833333

```
In [12]: #Scores by School Spending
school_spending = summary.drop(columns = ['School Type', 'Total Budget'])
school_spending['Reading'] = (school_spending['Average Reading Score'] * school_spending['Total Budget'])
school_spending['Math'] = (school_spending['Average Math Score'] * school_spending['Total Budget'])
school_spending['Number Passed Reading'] = (school_spending['% Passing Reading'] * school_spending['Total Budget'])
school_spending['Number Passed Math'] = (school_spending['% Passing Math'] * school_spending['Total Budget'])
school_spending = school_spending.drop(columns = ['Average Reading Score', 'Average Math Score', '% Passing Reading', '% Passing Math'])
```

```

In [13]: #Create a new dataframe for the school spending result
average_reading = []
average_math = []
percent_reading = []
percent_math = []
less_585 = school_spending.loc[school_spending['Per Student Budget'] < float(585)]
average_reading.append(less_585['Reading'].sum()/less_585['Total Students'].sum())
average_math.append(less_585['Math'].sum()/less_585['Total Students'].sum())
percent_reading.append((less_585['Number Passed Reading'].sum()/less_585['Total Students'].sum())*100)
percent_math.append((less_585['Number Passed Math'].sum()/less_585['Total Students'].sum())*100)

more_585 = school_spending.loc[(float(585) <= school_spending['Per Student Budget'] < float(615))]
average_reading.append(more_585['Reading'].sum()/more_585['Total Students'].sum())
average_math.append(more_585['Math'].sum()/more_585['Total Students'].sum())
percent_reading.append((more_585['Number Passed Reading'].sum()/more_585['Total Students'].sum())*100)
percent_math.append((more_585['Number Passed Math'].sum()/more_585['Total Students'].sum())*100)

more_615 = school_spending.loc[(float(615) <= school_spending['Per Student Budget'] < float(645))]
average_reading.append(more_615['Reading'].sum()/more_615['Total Students'].sum())
average_math.append(more_615['Math'].sum()/more_615['Total Students'].sum())
percent_reading.append((more_615['Number Passed Reading'].sum()/more_615['Total Students'].sum())*100)
percent_math.append((more_615['Number Passed Math'].sum()/more_615['Total Students'].sum())*100)

more_645 = school_spending.loc[(float(645) <= school_spending['Per Student Budget'] < float(675))]
average_reading.append(more_645['Reading'].sum()/more_645['Total Students'].sum())
average_math.append(more_645['Math'].sum()/more_645['Total Students'].sum())
percent_reading.append((more_645['Number Passed Reading'].sum()/more_645['Total Students'].sum())*100)
percent_math.append((more_645['Number Passed Math'].sum()/more_645['Total Students'].sum())*100)

result = pd.DataFrame({"Spending Ranges (Per Student)": ["<$585", "$585-615", "$615-645", "$645-675"],
                        "% Overall Passing Rate": [95.194724, 94.968784, 77.507599, 73.670105],
                        "% Passing Math": [93.702889, 94.087404, 71.400428, 66.230813],
                        "% Passing Reading": [96.686558, 95.850165, 83.614770, 81.109397],
                        "Average Math Score": [83.363065, 83.838414, 81.434031, 81.005604],
                        "Average Reading Score": [83.964039, 83.838414, 81.434031, 81.005604]})
result['% Overall Passing Rate'] = (result['% Passing Math'] + result['% Passing Reading'])/2
result = result[['Spending Ranges (Per Student)', 'Average Math Score', 'Average Reading Score', '% Passing Math', '% Passing Reading', '% Overall Passing Rate']]
result.set_index('Spending Ranges (Per Student)', inplace = True)
result

```

Out[13]:

Spending Ranges (Per Student)	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
<\$585	83.363065	83.964039	93.702889	96.686558	95.194724
\$585-615	83.529196	83.838414	94.087404	95.850165	94.968784
\$615-645	78.061578	81.434031	71.400428	83.614770	77.507599
\$645-675	77.049297	81.005604	66.230813	81.109397	73.670105

```

In [14]: reading = []
math = []
percentage_math = []
percentage_reading = []
small = school_spending.loc[school_spending['Total Students'] < 1000, :]
reading.append(small['Reading'].sum()/small['Total Students'].sum())
math.append(small['Math'].sum()/small['Total Students'].sum())
percentage_reading.append((small['Number Passed Reading'].sum()/small['Total Students'].sum()))
percentage_math.append((small['Number Passed Math'].sum()/small['Total Students'].sum()))

medium = school_spending.loc[(1000 <= school_spending['Total Students']) & (2000 > school_spending['Total Students']), :]
reading.append(medium['Reading'].sum()/medium['Total Students'].sum())
math.append(medium['Math'].sum()/medium['Total Students'].sum())
percentage_reading.append((medium['Number Passed Reading'].sum()/medium['Total Students'].sum()))
percentage_math.append((medium['Number Passed Math'].sum()/medium['Total Students'].sum()))

large = school_spending.loc[(2000 <= school_spending['Total Students']) & (5000 > school_spending['Total Students']), :]
reading.append(large['Reading'].sum()/large['Total Students'].sum())
math.append(large['Math'].sum()/large['Total Students'].sum())
percentage_reading.append((large['Number Passed Reading'].sum()/large['Total Students'].sum()))
percentage_math.append((large['Number Passed Math'].sum()/large['Total Students'].sum()))

result_2 = pd.DataFrame({"School Size": ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"],
"% Overall Passing Rate"] = (result_2["% Passing Math"] + result_2["% Passing Reading"])/2
result_2 = result_2[["School Size", "Average Math Score", "Average Reading Score", "% Overall Passing Rate"]]
result_2.set_index('School Size', inplace = True)
result_2

```

Out[14]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
School Size					
Small (<1000)	83.828654	83.974082	93.952484	95.968323	94.960403
Medium (1000-2000)	83.372682	83.867989	93.604788	96.773058	95.188923
Large (2000-5000)	77.477562	81.198640	68.652380	82.125158	75.388769

```

In [15]: #Create a for loop function for charter school type
school_charter = schools_df.loc[schools_df['type'] == 'Charter', :]
student_charter = pd.DataFrame()
for school in school_charter['name']:
    student_charter = student_charter.append(students_df.loc[students_df['school'] == school])

```

```

In [16]: # Create a high level snapshot (in table form) of the charter's key metrics
average_reading_score_1 = student_charter['reading_score'].mean()
average_math_score_1 = student_charter['math_score'].mean()
total_student_1 = school_charter['size'].sum()
reading_passed_1 = ((student_charter['Passed Reading'].sum())/total_student_1)*100
math_passed_1 = ((student_charter['Passed Math'].sum())/total_student_1)*100
average_passing_1 = (math_passed_1 + reading_passed_1)/2
charter_summary = pd.DataFrame({"School Type": "Charter", "Average Math Score": average_math_score_1, "Average Reading Score": average_reading_score_1, "% Overall Passing Rate": average_passing_1})
charter_summary = charter_summary[["School Type", "Average Math Score", "Average Reading Score", "% Overall Passing Rate"]]

```



```
In [17]: district_summary_1 = district_summary.drop(columns = ['Total Schools', 'Total  
district_summary_1['School Type'] = "District"  
district_summary_1 = district_summary_1[['School Type', 'Average Math Score',
```

```
In [18]: by_school_type = pd.concat([charter_summary, district_summary_1])  
by_school_type.set_index('School Type', inplace = True)  
by_school_type
```

Out[18]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
School Type					
Charter	83.406183	83.902821	93.701821	96.645891	95.173856
District	76.987026	80.962485	66.518387	80.905249	73.711818



Present



Slides



Themes



Help