Following are two trends from the data:

1. Top performing schools are Charter School Type and Bottom performing schools are District School Type.
2. Students across schools are doing better in reading, compared to math.

Following is the Python/Pandas Code (same as Jupyter notebook). Each section is separated by page break

# 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 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"])

#Getting all the data

total\_schools = len(school\_data\_complete["school\_name"].unique())

total\_students = school\_data\_complete["student\_name"].count()

total\_budget = school\_data["budget"].sum()

avg\_math\_score = school\_data\_complete["math\_score"].mean()

avg\_reading\_score = school\_data\_complete["reading\_score"].mean()

# Getting the passing average math percent

passing\_math = school\_data\_complete.loc[school\_data\_complete["math\_score"] >= 70]

passing\_math\_count = passing\_math["math\_score"].count()

total\_math\_count = school\_data\_complete["math\_score"].count()

passing\_math\_percent = (passing\_math\_count / total\_math\_count)\*100

# Getting the passing average reading percent

passing\_reading = school\_data\_complete.loc[school\_data\_complete["reading\_score"] >= 70]

passing\_reading\_count = passing\_reading["reading\_score"].count()

total\_reading\_count = school\_data\_complete["reading\_score"].count()

passing\_reading\_percent = (passing\_reading\_count / total\_reading\_count)\*100

overall\_passing\_rate = (school\_data\_complete["math\_score"].mean() + school\_data\_complete["reading\_score"].mean())/2

school\_summary = [{"Total Schools": total\_schools,

"Total Students": total\_students,

"Total Budget": total\_budget,

"Average Math Score": avg\_math\_score,

"Average Reading Score": avg\_reading\_score,

"% Passing Math": passing\_math\_percent,

"% Passing Reading" : passing\_reading\_percent,

"% Overall Passing Rate" : overall\_passing\_rate}]

school\_summary\_df = pd.DataFrame(school\_summary)

# Do the formating of the data

school\_summary\_df["Total Students"] = school\_summary\_df["Total Students"].map("{:,}".format)

school\_summary\_df["Total Budget"] = school\_summary\_df["Total Budget"].map("${:,}".format)

school\_summary\_df["Average Math Score"] = school\_summary\_df["Average Math Score"].map("{:.2f}".format)

school\_summary\_df["Average Reading Score"] = school\_summary\_df["Average Reading Score"].map("{:.2f}".format)

school\_summary\_df["% Passing Math"] = school\_summary\_df["% Passing Math"].map("{:.2f}%".format)

school\_summary\_df["% Passing Reading"] = school\_summary\_df["% Passing Reading"].map("{:.2f}%".format)

school\_summary\_df["% Overall Passing Rate"] = school\_summary\_df["% Overall Passing Rate"].map("{:.2f}%".format)

school\_summary\_arrange\_df = school\_summary\_df[["Total Schools", "Total Students", "Total Budget",

"Average Math Score", "Average Reading Score",

"% Passing Math", "% Passing Reading", "% Overall Passing Rate"]]

school\_summary\_arrange\_df

# Make the data set per school

# Get the first few columns of my\_school\_df dataFrame from school\_data

my\_school\_df = school\_data.iloc[:, 1:5]

my\_school\_df.rename(columns = {"school\_name":"School Name", "type":"School Type", "size":"Total Students", "budget":"Total School Budget"}, inplace = True)

# Set the index of my\_school\_df as the School Name

my\_school\_df.set\_index("School Name", inplace = True)

# group school\_data\_complete dataframe by school\_name

school\_group = school\_data\_complete.groupby(["school\_name"])

# Start building out my\_school\_df and keep formating to make it look nice

my\_school\_df["Per Student Budget"] = my\_school\_df["Total School Budget"] / my\_school\_df["Total Students"]

my\_school\_df["Total Students"] = my\_school\_df["Total Students"].map("{:,}".format)

my\_school\_df["Total School Budget"] = my\_school\_df["Total School Budget"].map("${:,}".format)

my\_school\_df["Per Student Budget"] = my\_school\_df["Per Student Budget"].map("${:.0f}".format)

my\_school\_df["Average Math Score"] = school\_group["math\_score"].mean().map("{:.3f}".format)

my\_school\_df["Average Reading Score"] = school\_group["reading\_score"].mean().map("{:.3f}".format)

# Calculate the average math passing rate per school and format

school\_math\_count = school\_group["math\_score"].count()

school\_passing\_math\_group = passing\_math.groupby("school\_name")

school\_passing\_math\_count = school\_passing\_math\_group["math\_score"].count()

school\_passing\_math\_percent = (school\_passing\_math\_count / school\_math\_count)\*100

my\_school\_df["% Passing Math"] = school\_passing\_math\_percent

my\_school\_df["% Passing Math"] = my\_school\_df["% Passing Math"].map("{:.3f}%".format)

# Calculate the average reading passing rate per school and format

school\_reading\_count = school\_group["reading\_score"].count()

school\_passing\_reading\_group = passing\_reading.groupby("school\_name")

school\_passing\_reading\_count = school\_passing\_reading\_group["reading\_score"].count()

school\_passing\_reading\_percent = (school\_passing\_reading\_count / school\_reading\_count)\*100

my\_school\_df["% Passing Reading"] = school\_passing\_reading\_percent

my\_school\_df["% Passing Reading"] = my\_school\_df["% Passing Reading"].map("{:.3f}%".format)

# Calculate the overall passing rate per school and format

school\_passing\_overall\_count = (school\_passing\_math\_percent + school\_passing\_reading\_percent) / 2

my\_school\_df["% Overall Passing Rate"] = school\_passing\_overall\_count

my\_school\_df["% Overall Passing Rate"] = my\_school\_df["% Overall Passing Rate"].map("{:.3f}%".format)

# Sort my\_school\_df dataFrame based on Overall Passing Rate with top performing schools listed first

my\_school\_sorted\_top\_df = my\_school\_df.sort\_values("% Overall Passing Rate", ascending=False)

my\_school\_sorted\_top\_df.head(5)

# Sort my\_school\_df dataFrame based on Overall Passing Rate with bottom performing schools listed first

my\_school\_sorted\_bottom\_df = my\_school\_df.sort\_values("% Overall Passing Rate", ascending=True)

my\_school\_sorted\_bottom\_df.head(5)

# 9th grade Data Per School

nineth\_grade = school\_data\_complete.loc[school\_data\_complete["grade"] == "9th"]

nineth\_grade\_school\_group = nineth\_grade.groupby(["school\_name"])

# 10th grade Data Per School

tenth\_grade = school\_data\_complete.loc[school\_data\_complete["grade"] == "10th"]

tenth\_grade\_school\_group = tenth\_grade.groupby(["school\_name"])

# 11th grade Data Per School

eleventh\_grade = school\_data\_complete.loc[school\_data\_complete["grade"] == "11th"]

eleventh\_grade\_school\_group = eleventh\_grade.groupby(["school\_name"])

# 12th grade Data Per School

twelveth\_grade = school\_data\_complete.loc[school\_data\_complete["grade"] == "12th"]

twelveth\_grade\_school\_group = twelveth\_grade.groupby(["school\_name"])

math\_grade\_summary\_df = pd.DataFrame({"9th": nineth\_grade\_school\_group["math\_score"].mean().map("{:.3f}".format),

"10th": tenth\_grade\_school\_group["math\_score"].mean().map("{:.3f}".format),

"11th": eleventh\_grade\_school\_group["math\_score"].mean().map("{:.3f}".format),

"12th": twelveth\_grade\_school\_group["math\_score"].mean().map("{:.3f}".format)})

math\_grade\_summary\_df

reading\_grade\_summary\_df = pd.DataFrame({"9th": nineth\_grade\_school\_group["reading\_score"].mean().map("{:.3f}".format),

"10th": tenth\_grade\_school\_group["reading\_score"].mean().map("{:.3f}".format),

"11th": eleventh\_grade\_school\_group["reading\_score"].mean().map("{:.3f}".format),

"12th": twelveth\_grade\_school\_group["reading\_score"].mean().map("{:.3f}".format)})

reading\_grade\_summary\_df

# Create spending bins

spending\_bins = [0, 585, 615, 645, 675]

spending\_group\_names = ["<$585", "$585-615", "$615-645", "$645-675"]

new\_school\_df = my\_school\_df.reset\_index()

new\_school\_df["Per Student Budget"] = school\_data["budget"] / school\_data["size"]

per\_student\_budegt\_df = pd.DataFrame({"Spending Ranges (Per Student)" : new\_school\_df["Per Student Budget"],

"Average Math Score" : new\_school\_df["Average Math Score"],

"Average Reading Score" : new\_school\_df["Average Reading Score"],

"% Passing Math" : new\_school\_df["% Passing Math"],

"% Passing Reading" : new\_school\_df["% Passing Reading"],

"% Overall Passing Rate" : new\_school\_df["% Overall Passing Rate"]})

# Map Bins

per\_student\_budegt\_df["Spending Ranges (Per Student)"] = pd.cut(per\_student\_budegt\_df["Spending Ranges (Per Student)"], spending\_bins, labels=spending\_group\_names)

per\_student\_budegt\_df = per\_student\_budegt\_df.groupby(["Spending Ranges (Per Student)"])

per\_student\_budegt\_df.max()

# Create Size bins

size\_bins = [0, 1000, 2000, 5000]

size\_group\_names = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]

school\_size\_df = pd.DataFrame({"School Size" : school\_data["size"],

"Average Math Score" : new\_school\_df["Average Math Score"],

"Average Reading Score" : new\_school\_df["Average Reading Score"],

"% Passing Math" : new\_school\_df["% Passing Math"],

"% Passing Reading" : new\_school\_df["% Passing Reading"],

"% Overall Passing Rate" : new\_school\_df["% Overall Passing Rate"]})

# Map Bins

school\_size\_df["School Size"] = pd.cut(school\_size\_df["School Size"], size\_bins, labels=size\_group\_names)

school\_size\_df = school\_size\_df.groupby(["School Size"])

school\_size\_df.max()

school\_type\_df = pd.DataFrame({"School Type" : school\_data["type"],

"Average Math Score" : new\_school\_df["Average Math Score"],

"Average Reading Score" : new\_school\_df["Average Reading Score"],

"% Passing Math" : new\_school\_df["% Passing Math"],

"% Passing Reading" : new\_school\_df["% Passing Reading"],

"% Overall Passing Rate" : new\_school\_df["% Overall Passing Rate"]})

school\_type\_df = school\_type\_df.groupby(["School Type"])

school\_type\_df.max()