# 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)

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

student\_df = pd.DataFrame(student\_data)

# Combine the data into a single dataset

school\_data\_complete = pd.merge(student\_data, school\_data, how="left", on=["school\_name", "school\_name"])

school\_data.head()

school\_df.head()

student\_df.head()

school\_name = school\_df["school\_name"]

school\_types = school\_df["type"]

total\_school\_budget = school\_df["budget"]

total\_students = student\_df.groupby(["school\_name"])["student\_name"].count().rename("Student Counts")

total\_student\_df = pd.DataFrame(total\_students)

pd\_new = pd.DataFrame({"school\_name": school\_name,

"school type": school\_types,

"Total School Budget": total\_school\_budget,

})

total\_students

#new2=pd\_new.join(total\_student\_df('school\_name'), on ='school\_name')

new3\_df = pd.merge(pd\_new, total\_student\_df, on= 'school\_name')

new3\_df['Per Student Budget'] = new3\_df["Total School Budget"]/ new3\_df["Student Counts"]

math = student\_df.groupby("school\_name")["math\_score"].sum().rename("Math Score")

avg\_math = math/total\_students

read=student\_df.groupby("school\_name")["reading\_score"].sum().rename("Reading Score")

avg\_read = read/total\_students

avg\_math\_df = pd.DataFrame(avg\_math)

avg\_read\_df =pd.DataFrame(avg\_read)

new4\_df = pd.merge(new3\_df, avg\_math\_df, on= 'school\_name')

new4\_df.rename(columns ={0:"Average Math Score"}, inplace = True)

new5\_df =pd.merge(new4\_df, avg\_read\_df, on= 'school\_name')

new5\_df.rename(columns ={0:"Average Reading Score"}, inplace = True)

pass\_math = student\_df.loc[student\_df['math\_score'] >= 70]['math\_score']

pass\_math = student\_df[student\_df['math\_score'] >= 70].groupby('school\_name')['Student ID'].count()/total\_students

pass\_read = student\_df[student\_df['reading\_score'] >= 70].groupby('school\_name')['Student ID'].count()/total\_students

pass\_perc\_math = pass\_math\*100

pass\_perc\_read = pass\_read\*100

pass\_perc\_math\_df =pd.DataFrame(pass\_perc\_math)

new6\_df = pd.merge(new5\_df, pass\_perc\_math\_df, on= 'school\_name')

pass\_perc\_read\_df =pd.DataFrame(pass\_perc\_read)

new7\_df = pd.merge(new6\_df, pass\_perc\_read\_df, on= 'school\_name')

new7\_df.rename(columns={"0\_x": "% Passing Math",

"0\_y": "% Passing Read"}, inplace= True)

new7\_df["Overall Passing rate"] = (new7\_df["% Passing Math"]+new7\_df["% Passing Read"])/2

#total\_student\_df

#new3\_df

#math

#new5\_df

list(new6\_df.columns.values)

pass\_math

students

pass\_perc\_math

new7\_df

list(new7\_df.columns.values)

new7\_df.set\_index("school\_name")

# Sorting for top five passing rate

new8\_df= new7\_df.sort\_values("Overall Passing rate",ascending = False)

# Top five schools in overall pass rate

new8\_df.head(5)

# Bottom Performing School

new9\_df=new7\_df.sort\_values("Overall Passing rate")

new9\_df.tail(5)

ninth = student\_df.loc[student\_df['grade'] == "9th"]

ninth\_math=ninth.groupby('school\_name')["math\_score"].mean()

tenth = student\_df.loc[student\_df['grade'] == "10th"]

tenth\_math=tenth.groupby('school\_name')["math\_score"].mean()

eleven = student\_df.loc[student\_df['grade'] == "11th"]

eleven\_math=eleven.groupby('school\_name')["math\_score"].mean()

twelve = student\_df.loc[student\_df['grade'] == "12th"]

twelve\_math=twelve.groupby('school\_name')["math\_score"].mean()

df3 = pd.DataFrame({"9th":ninth\_math,

"10th": tenth\_math,

"11th": eleven\_math,

"12th": twelve\_math})

df3.round(2)

ninth = student\_df.loc[student\_df['grade'] == "9th"]

ninth\_read=ninth.groupby('school\_name')["reading\_score"].mean()

tenth = student\_df.loc[student\_df['grade'] == "10th"]

tenth\_read=tenth.groupby('school\_name')["reading\_score"].mean()

eleven = student\_df.loc[student\_df['grade'] == "11th"]

eleven\_read=eleven.groupby('school\_name')["reading\_score"].mean()

twelve = student\_df.loc[student\_df['grade'] == "12th"]

twelve\_read=twelve.groupby('school\_name')["reading\_score"].mean()

df4 = pd.DataFrame({"9th":ninth\_read,

"10th": tenth\_read,

"11th": eleven\_read,

"12th": twelve\_read})

df4.round(2)