**PYCITY SCHOOL WRITTEN REPORT**

* The First we did was to import dependencies which is pandas.
* Next, we load the csv files and then merge them into one dataset
* After that next step was to get District Summary. For that I have find total number of schools and students using nunique, for total budget I used sum(), for averages I used mean(),to get the percentages I used “.loc” with condition then calculating percentages. After that , I made a new dataframe and put everything in a new dataframe.
* Next step was School Summary, I collected the school names, type by grouping it. I have collected total students by grouping school name then counting the student ids. For total school budget I first created group for school name then perform mean() on budget. For average math score and average reading score I have grouped it first then used sum() and then divide it by total students. For percent passing math and reading first calculated their sums respectively for condition where score >= 70 and then find their lengths and after that divide them and multiplied with 100 to get the percent value. For overall percentage firstly I used loc to check the conditions both math and reading scores greater than or equals to 70 , then grouped it together and then find the percentage. Finally, I stored all the values in a “grouped\_data” dataframe which gives us desired results.
* Now for next part of the activity, Top Performing School, I used the dataset we created above “grouped\_data” and used “sort\_values” on the “% Overall Passsing” and set ascending to false , which will give us the top performers.
* Now for Bottom Performing School, I used the dataset we created above “grouped\_data” and used “sort\_values” on the “% Overall Passsing” and set ascending to true , which will give us the bottom performers.
* Next part was Math Scores by Grades(9,10,11, 12). For that I have created a new dataset which contains “school name”,”grade”,”math\_score”,”reading\_score” and used this dataset to first set the “grade” to 9th,10th, 11th,12th respectively and after that grouped the math score by school name and getting the mean values individually. After that I put everything in a new dataframe and print it, it will show the average math scores of the classes grouped by the school name.
* For next part, Reading Score by Grade, I just did the same thing I did above but this time for “reading\_score”.
* **CONCLUSIONS:**
* As we can see from the very first dataset, which is ‘school\_data\_complete’ it just gives us the general outlook of the data, but as we go on the next part it shows us data in more informative way, as it is in ‘new\_data\_df’, which shows us how number of schools, students, total budget, average math score, average reading score, percent passing math, percent passing reading, overall passing. Which is quite useful in general analysis of the data and give clear image about dataset.
* Same in the case of School Summary, it separates the data based on the school name which gives us more clearer image about the dataset. As all the entries here are based on specific school name.
* Second one we can say as from the first dataset ‘school\_data\_complete’ it it very very difficult to see who is the top scorers and who are the bottom scorers. But when we made the new data sets ‘top\_schools’ and ‘bottom\_schools’ it gives us very clear image of who are top and bottom scorers in the data.