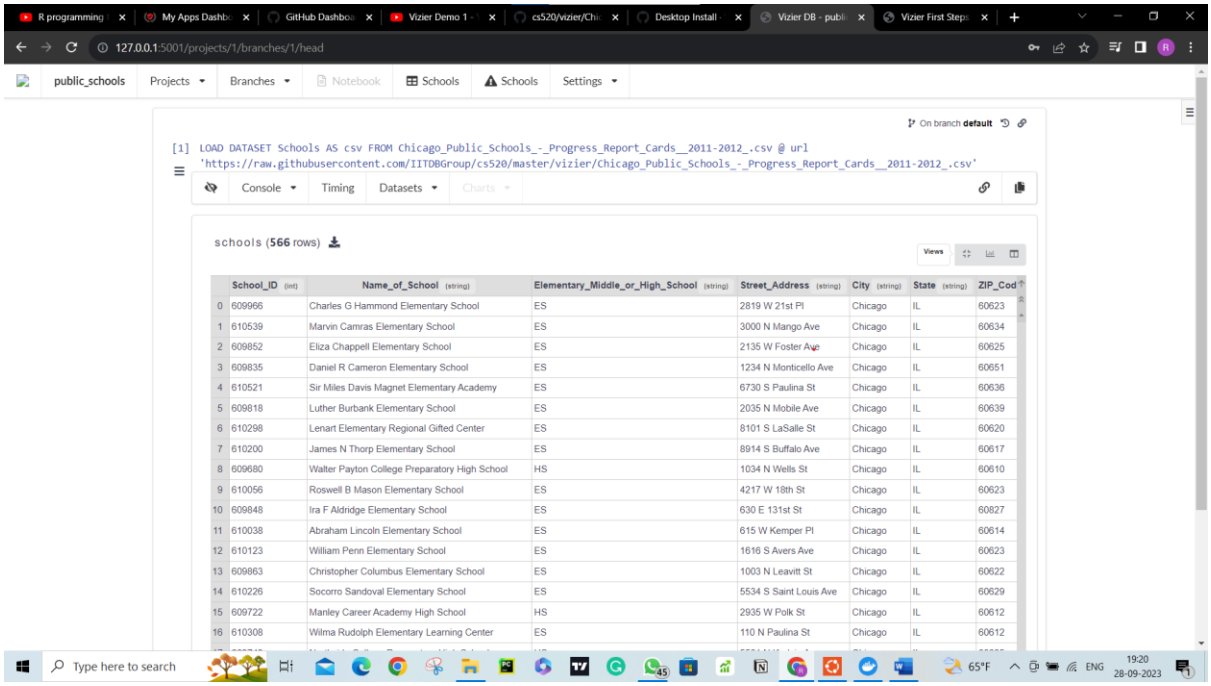
VIZIER FIRST STEPS ASSIGNMENT

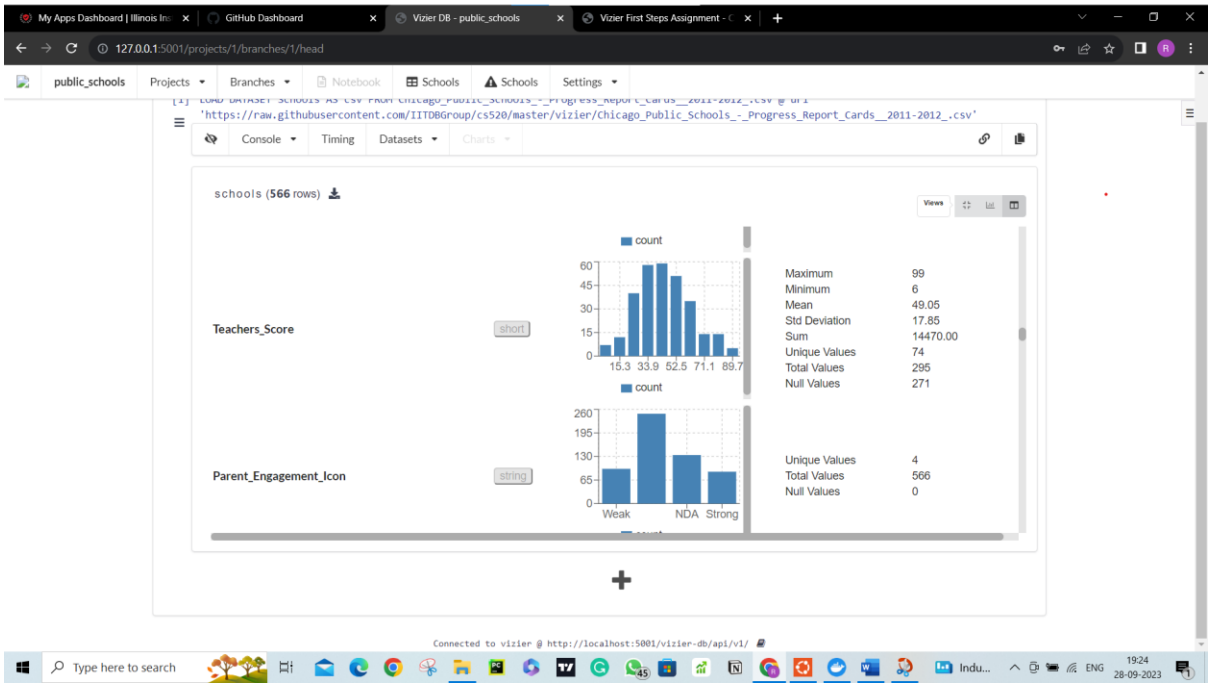
DATA INTEGRATION, WAREHOUSING, AND PROVENANCE

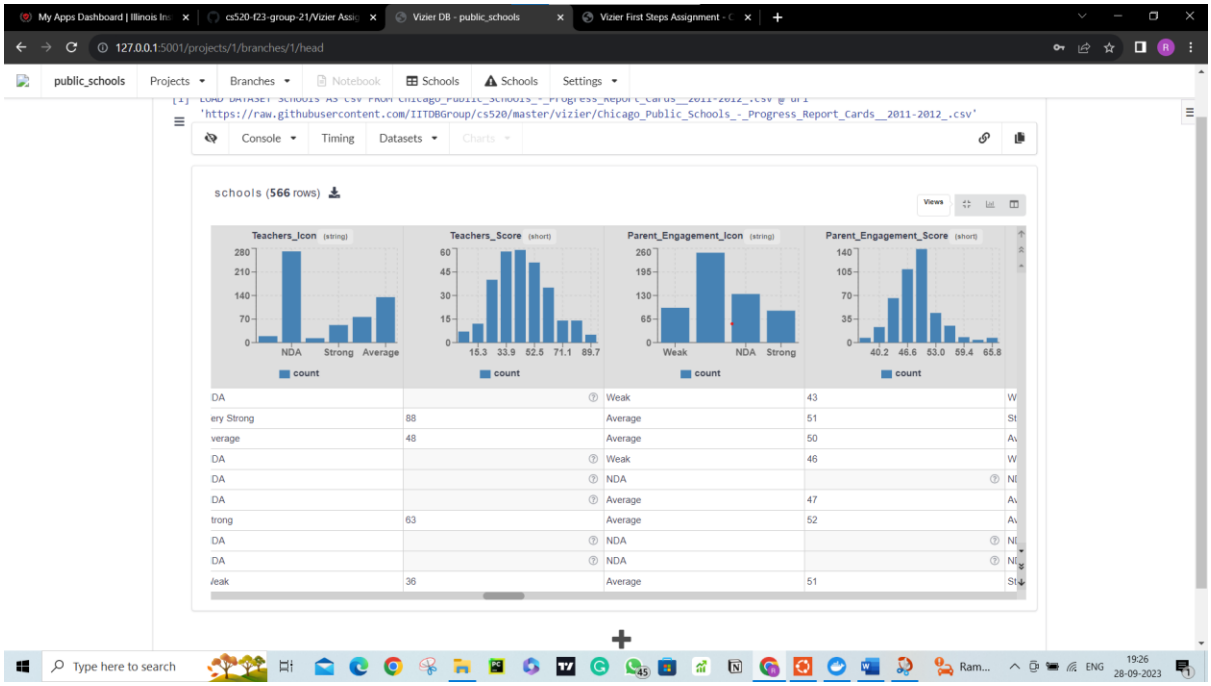
CS-520-01 (FALL 2023) TEAM-16

TASK1: Load a dataset and take a screenshot of the result

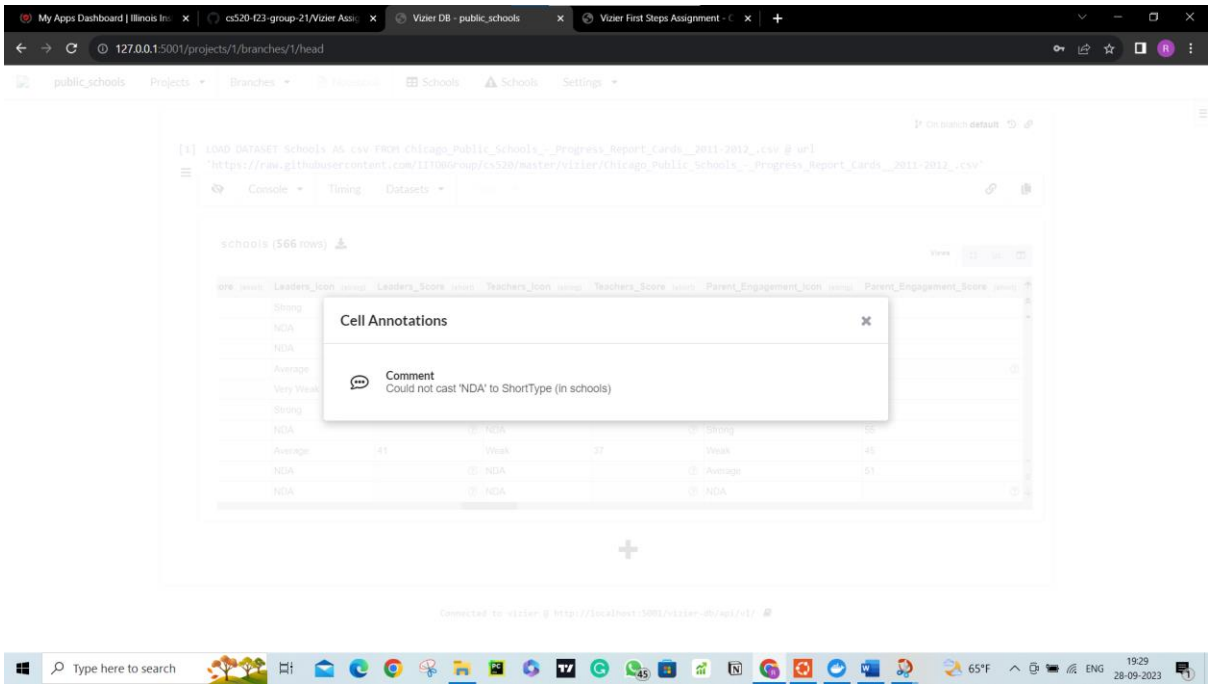


TASK2: Select the detail view and look at the distributions of some columns. Then look at the column view and take a screenshot of the distribution for column Teachers_Score.

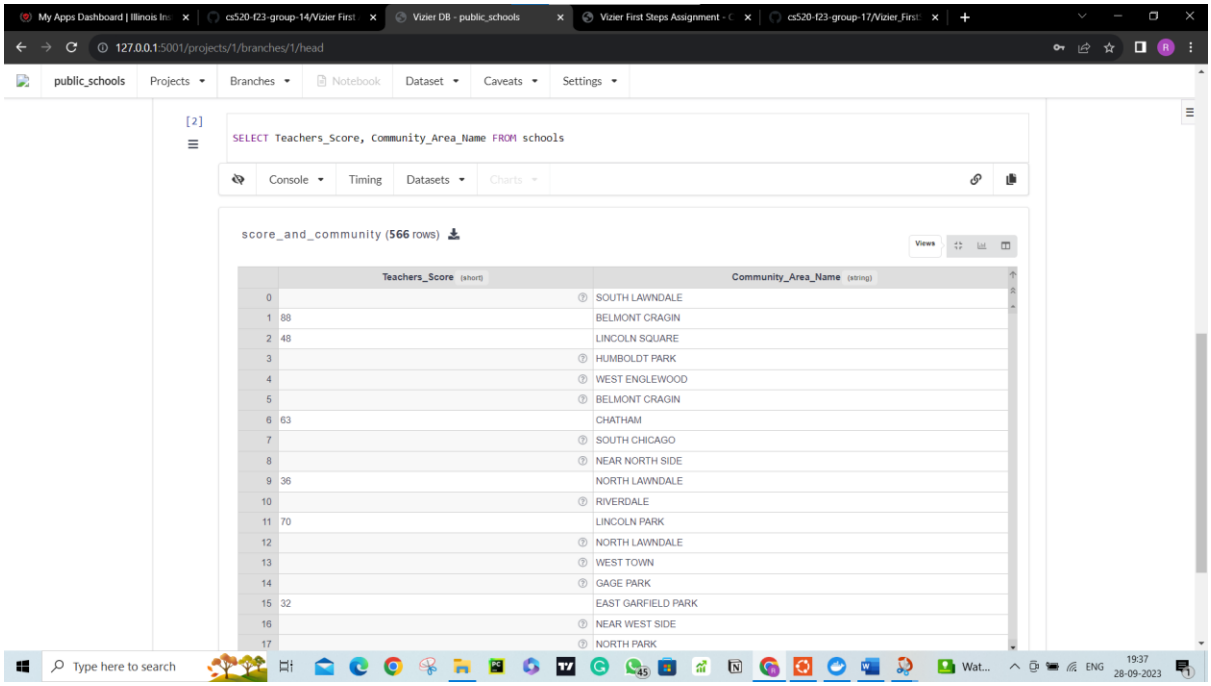




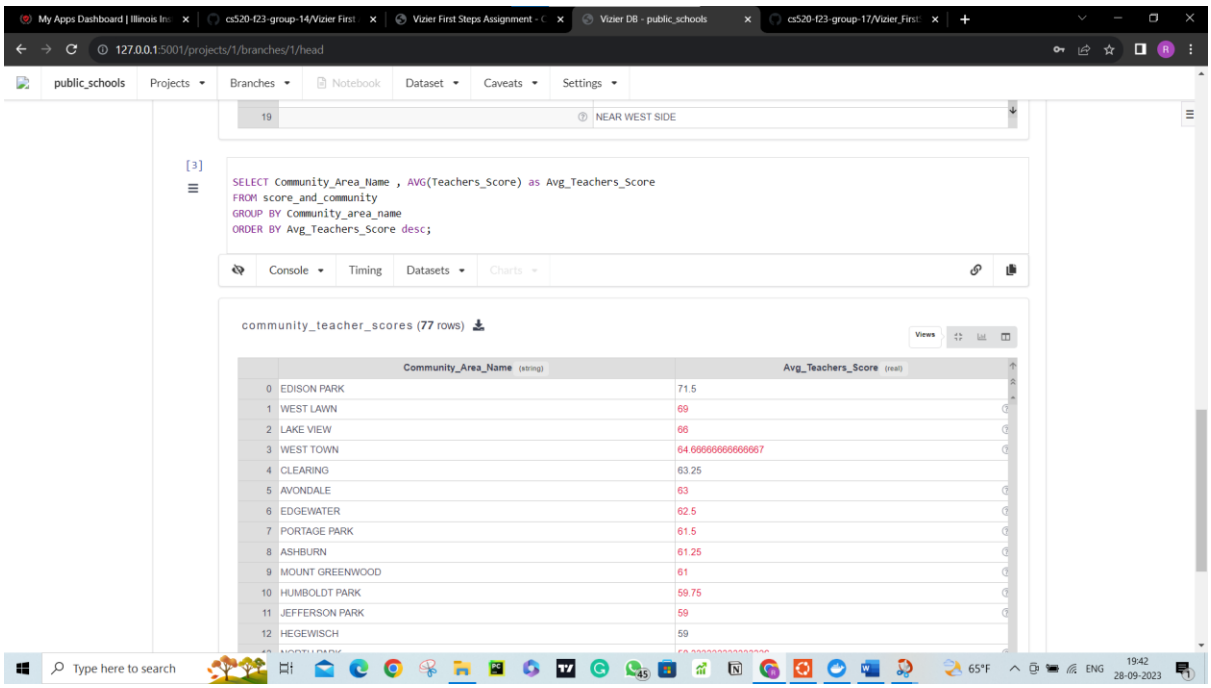
TASK3: Click on one of the question marks for values in the teachers column and take a screenshot.



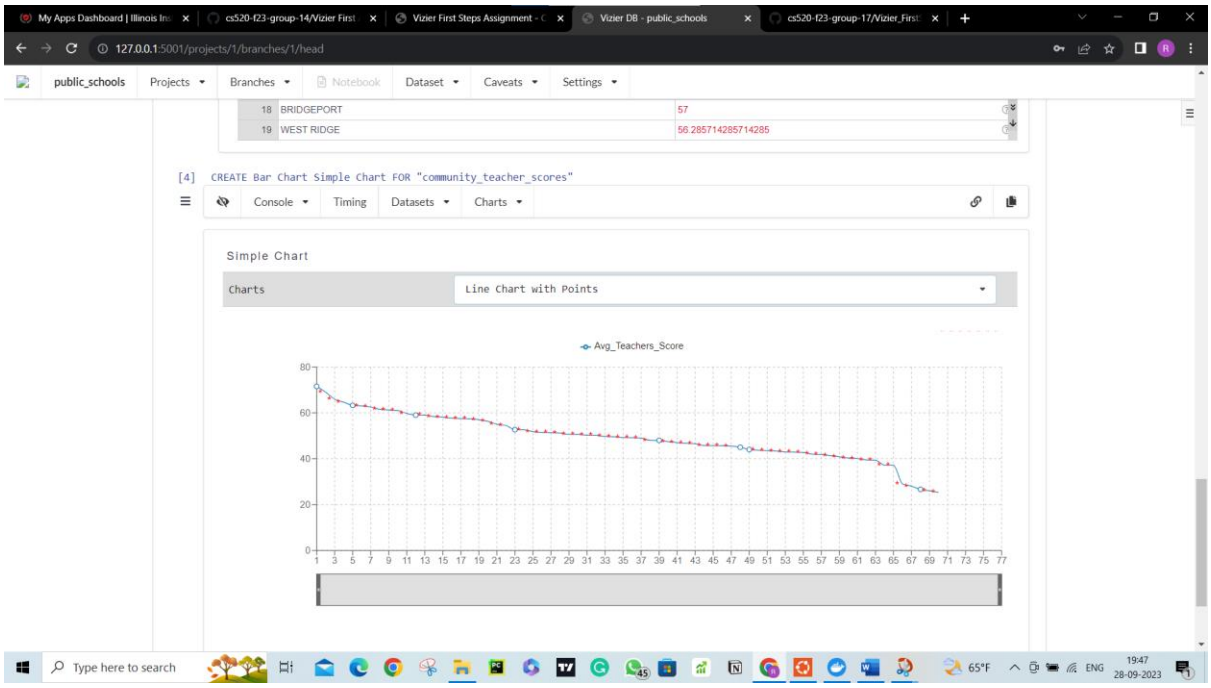
TASK4: Create a SQL cell and write a query that returns columns Teachers_Score and Community_Area_Name. SQL results can be stored as new datasets in Vizier. Call the result dataset score_and_community. And take a screenshot of the result.



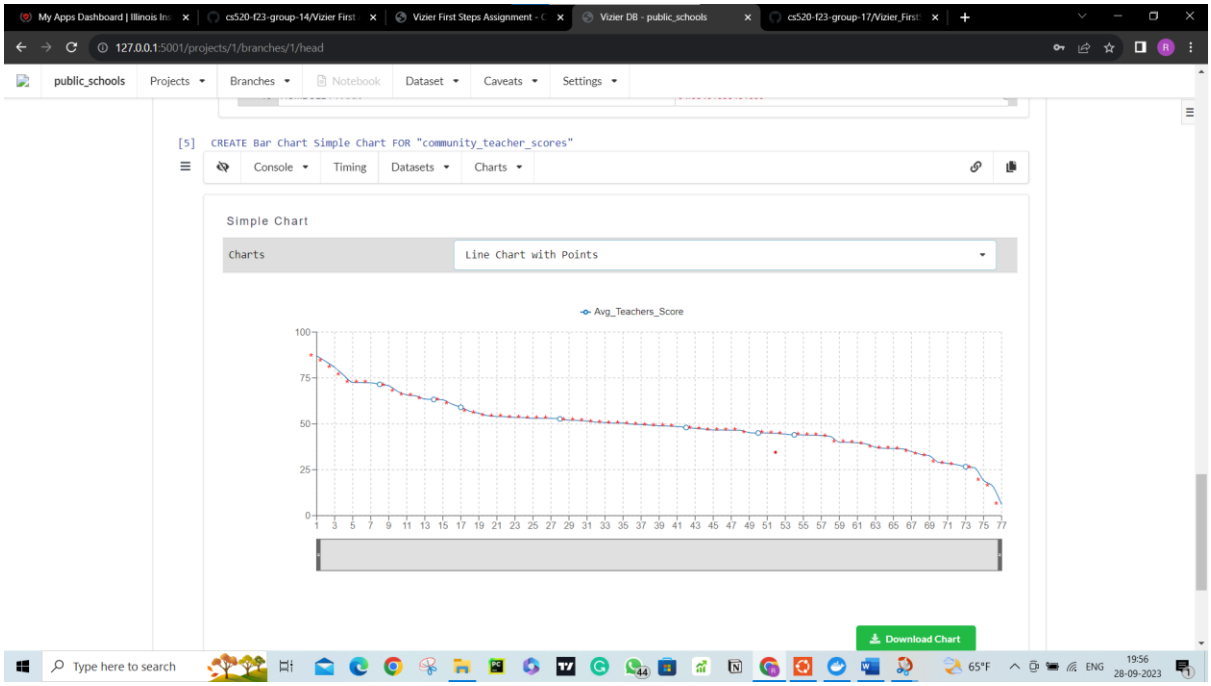
TASK5: Create a SQL cell and write a query over the over the score_and_community dataset that computes the result as described above. Call the result dataset community_teacher_scores. And take a screenshot of the result.



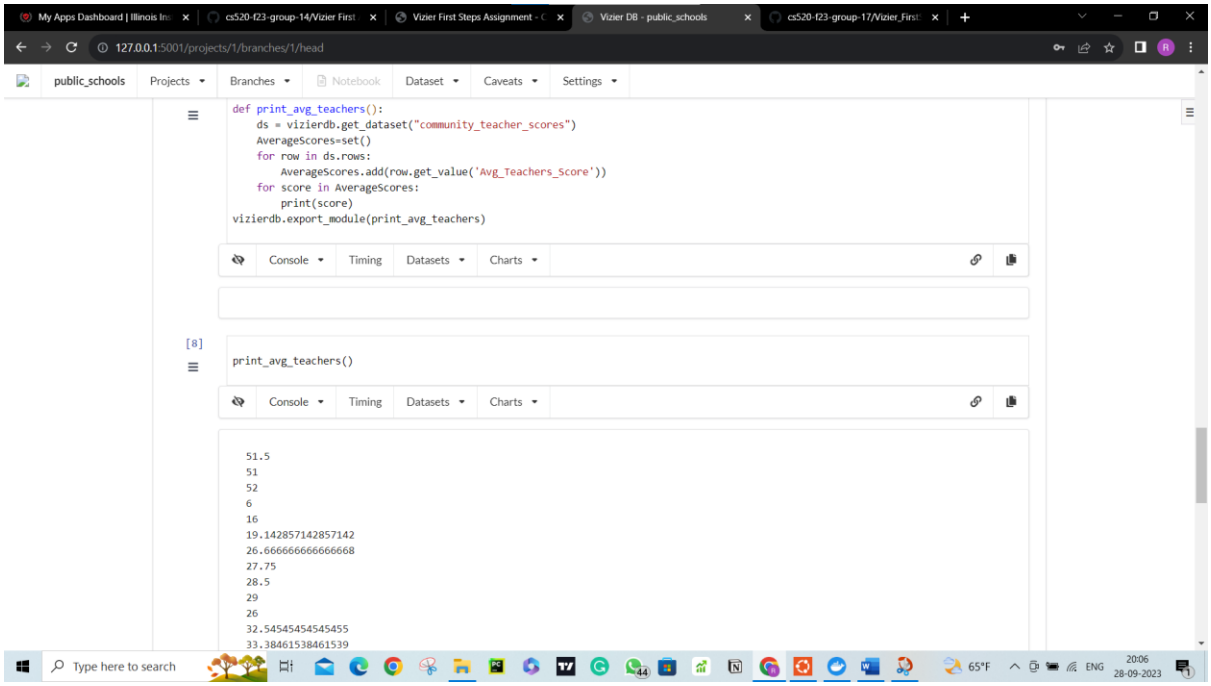
TASK6: Create a line chart of the aggregation result by creating a plot cell and take a screenshot of the result.



TASK7: Insert a new cell above the SQL cell that computes the average teacher scores (notebooks in Vizier are executed top down) by pressing the three bars below the cell number. Select *"Impute Missing Values"*, select the score_and_community dataset and Teachers_Score as the column to be imputed, and select mean as the imputation method and take a screenshot of the updated line chart.



TASK8:. Then use vizierdb.export_module to export the function. Then create a second Python cell and use vizierdb.get_model("print_avg_teachers") for importing the function and then call it. Take a screenshot of the result.



TASK 9: Create another Python cell and use Vizier's API to access the dataset community_teacher_scores as a DataFrame, then filter out rows where the avg_teacher_score is larger than or equal to 30.0 and then print the remaining rows and take a screenshot.