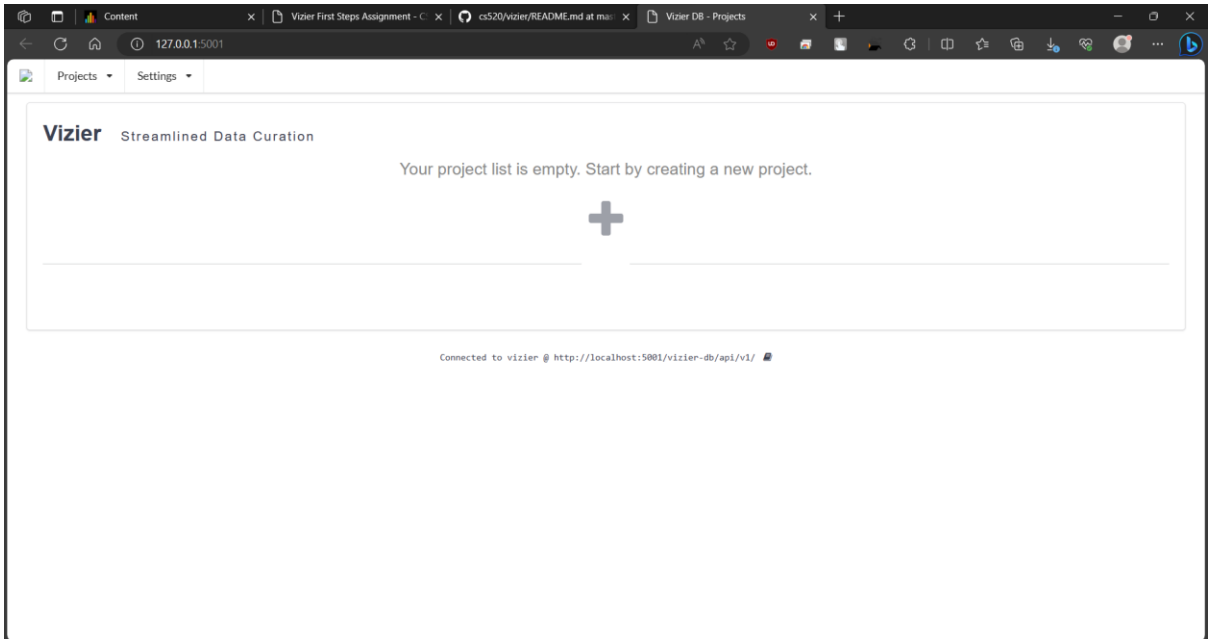
Dharmik Patel
A20526771

CS 520

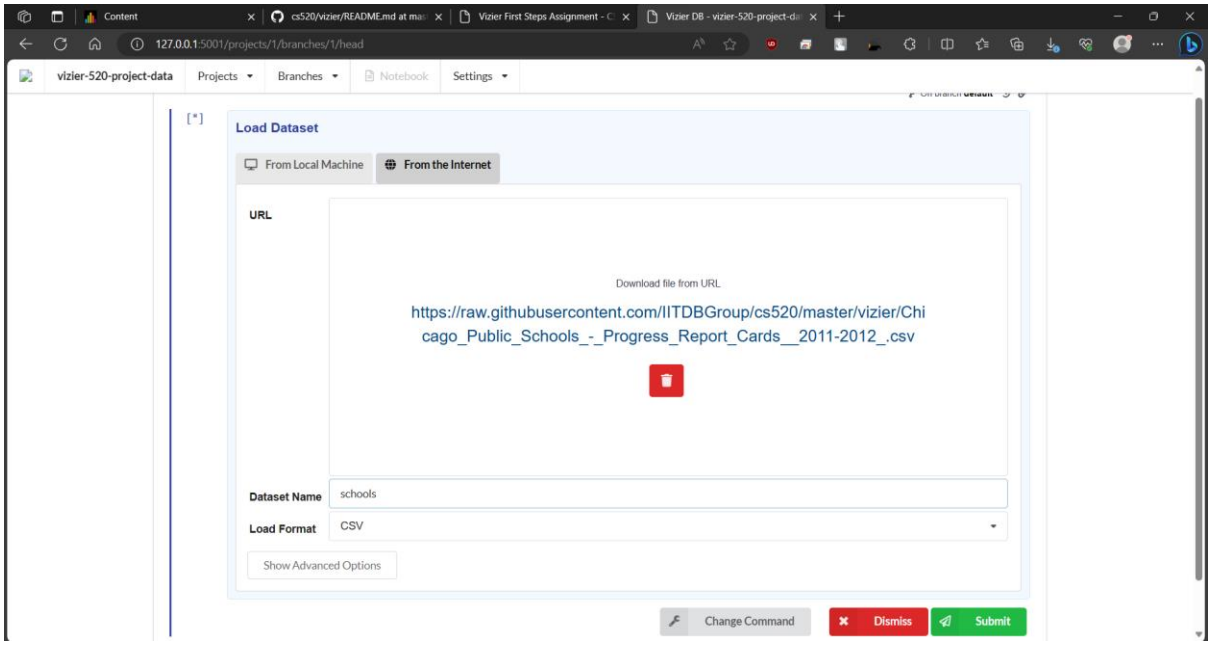
Data Integration, Warehousing, and Provenance

Vizier Project CS 520

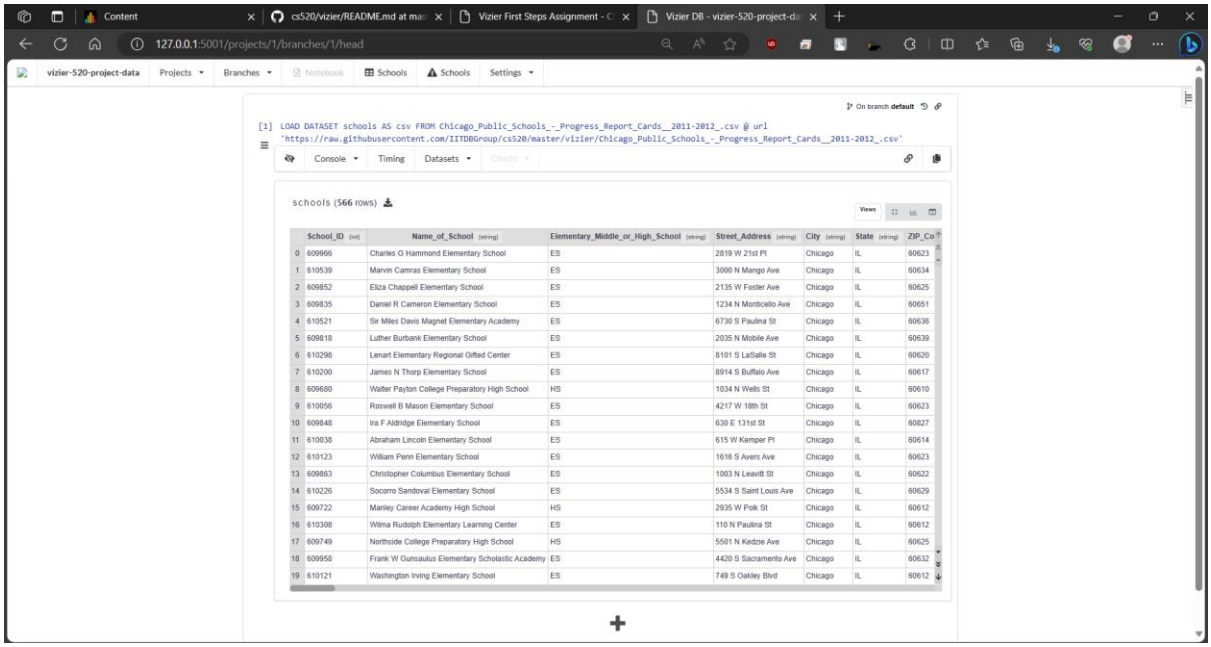
```
C:\Windows\System32>docker pull iitdbgroup/vizier_iit_cs520_fall23:intel
intel: Pulling from iitdbgroup/vizier_iit_cs520_fall23
f84cab65f19f: Pull complete
ad1d7f4beb56: Pull complete
9392614fb6cf: Pull complete
fbbd90443348: Pull complete
5c33c86c1b15: Pull complete
38aebece3a62: Pull complete
9295e6117d6d: Pull complete
139432960069: Pull complete
e10f1f4cc279: Pull complete
2475152a3569: Pull complete
ca4fa56d4de4: Pull complete
dd06735a2d33: Pull complete
Digest: sha256:5e6278701dcde2e69510e7469ed2feb732f3c637b756e51f047f57d78dd31afe
Status: Downloaded newer image for iitdbgroup/vizier_iit_cs520_fall23:intel
docker.io/iitdbgroup/vizier_iit_cs520_fall23:intel

What's Next?
  View a summary of image vulnerabilities and recommendations → docker scout quickview iitdbgroup/vizier_iit_cs520_fall23:intel

C:\Windows\System32>
```



Dharmik Patel
A20526771

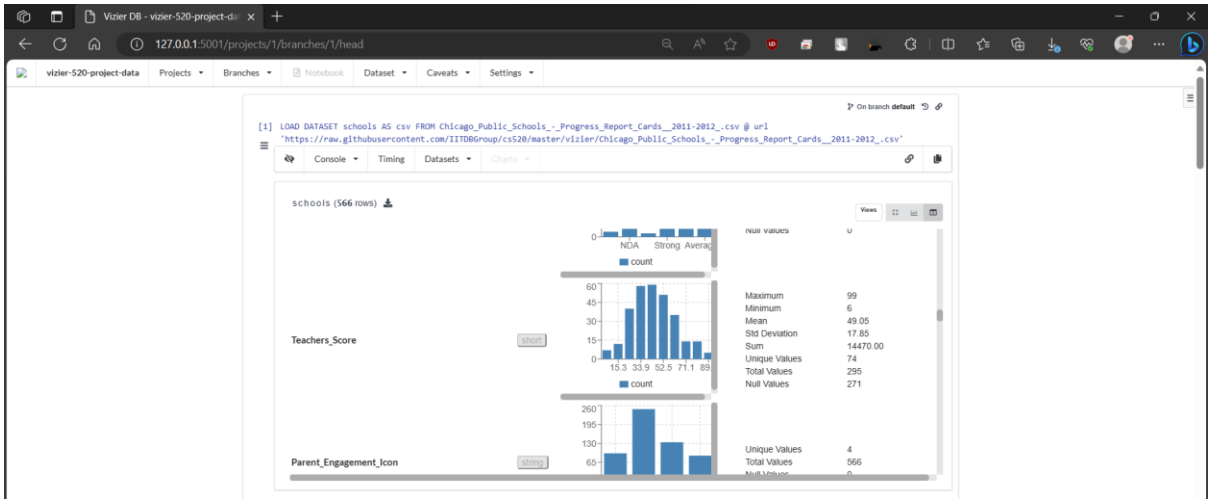


Task 1: load a dataset and take a screenshot of the result

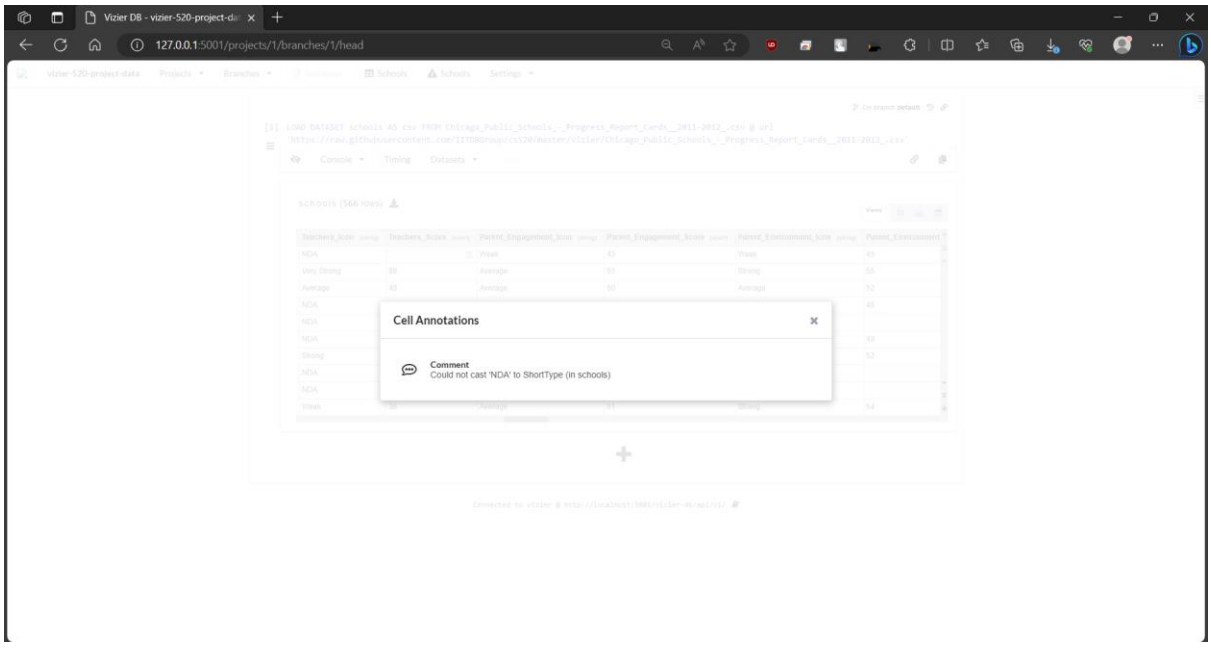


Dharmik Patel
A20526771

Task 2: 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.

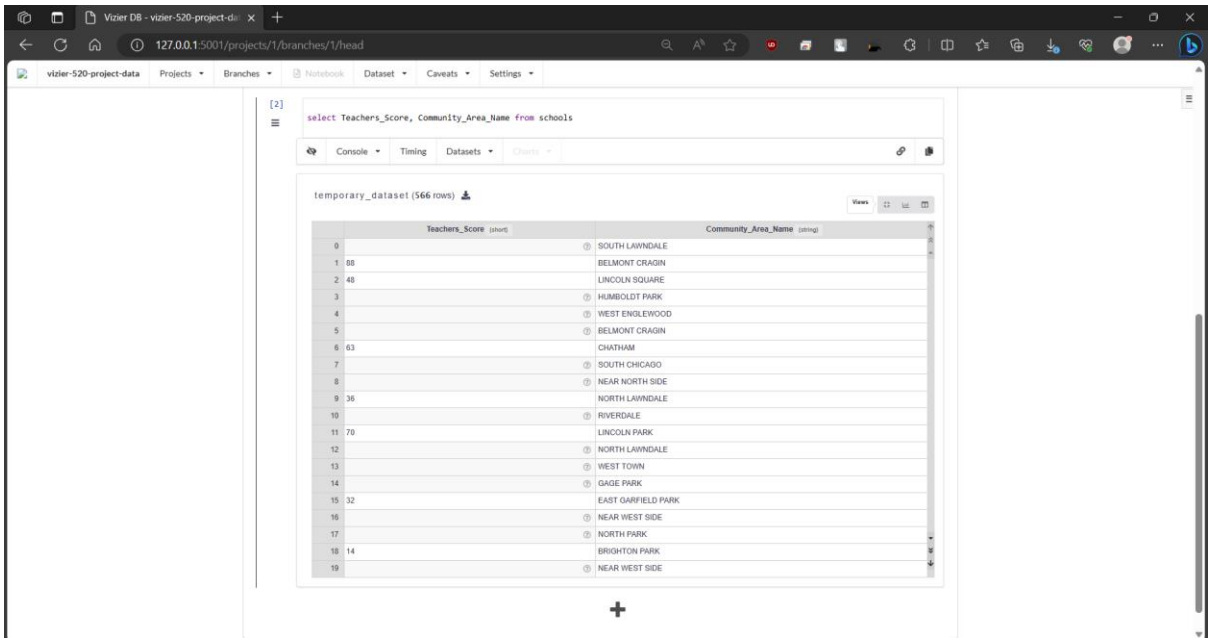


Task 3: Click on one of the question marks for values in the teachers column and take a screenshot.

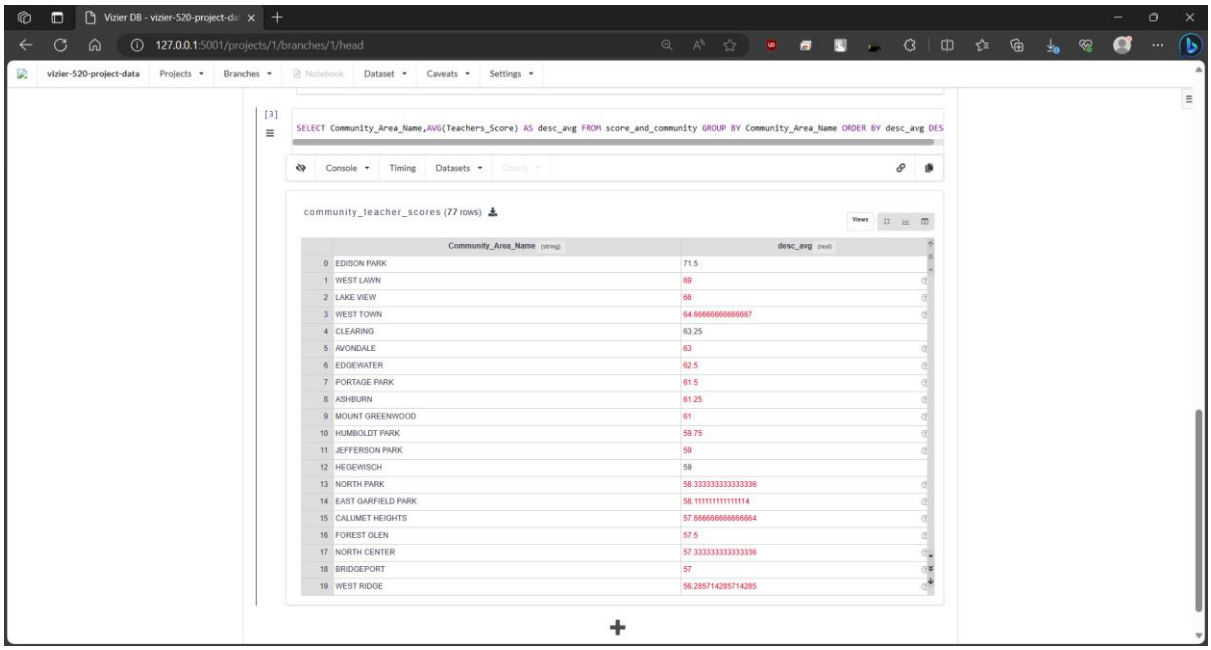


Dharmik Patel
A20526771

Task 4: 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.

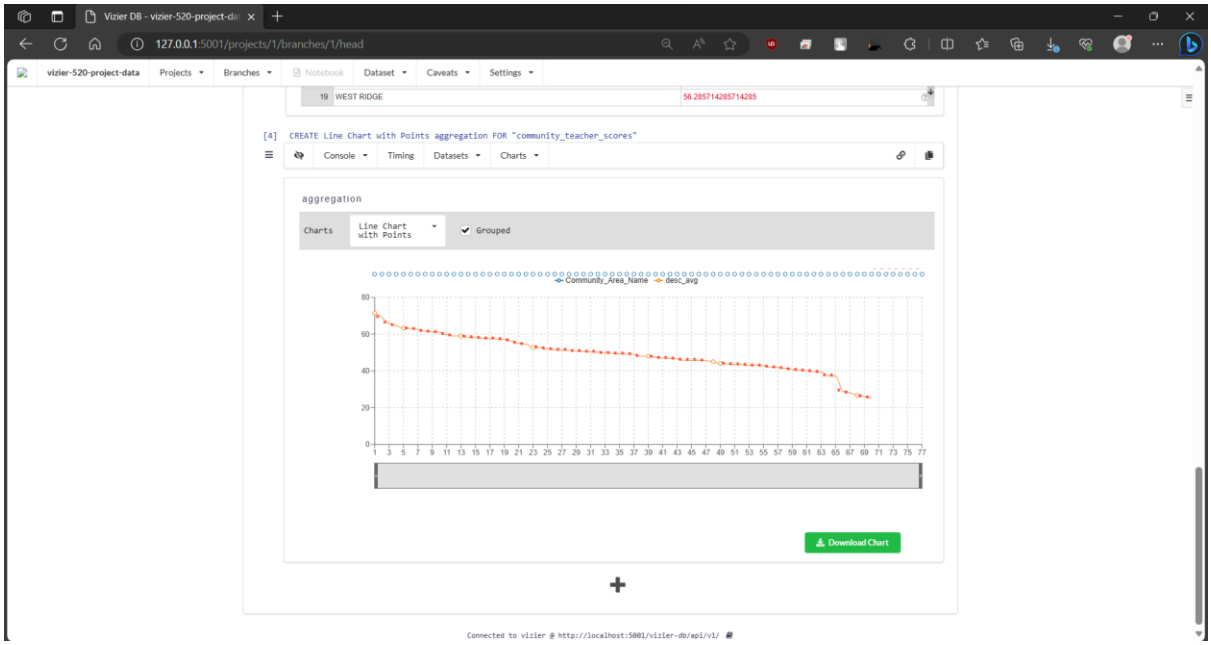


Task 5: Create a SQL cell and write a query over the over the score_and_community dataset that computes the result as described above. Call the result dataset community_teacher_scores. And take a screenshot of the result.

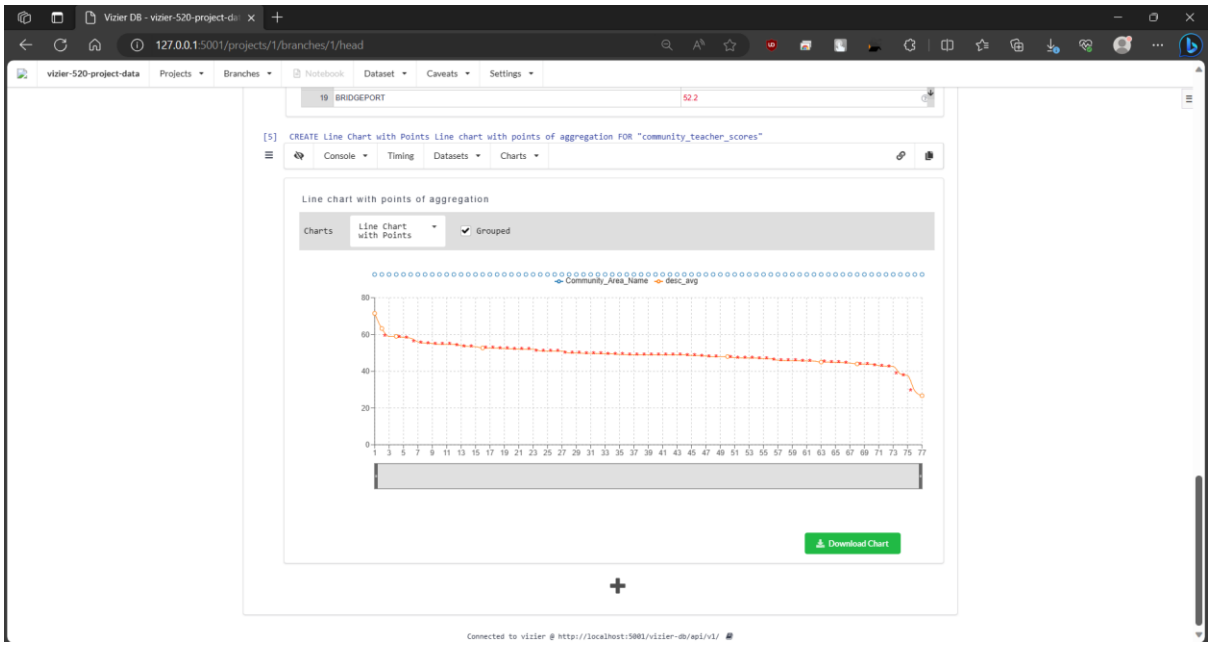


Dharmik Patel
A20526771

Task 6: Create a line chart of the aggregation result by creating a plot cell and take a screenshot of the result.

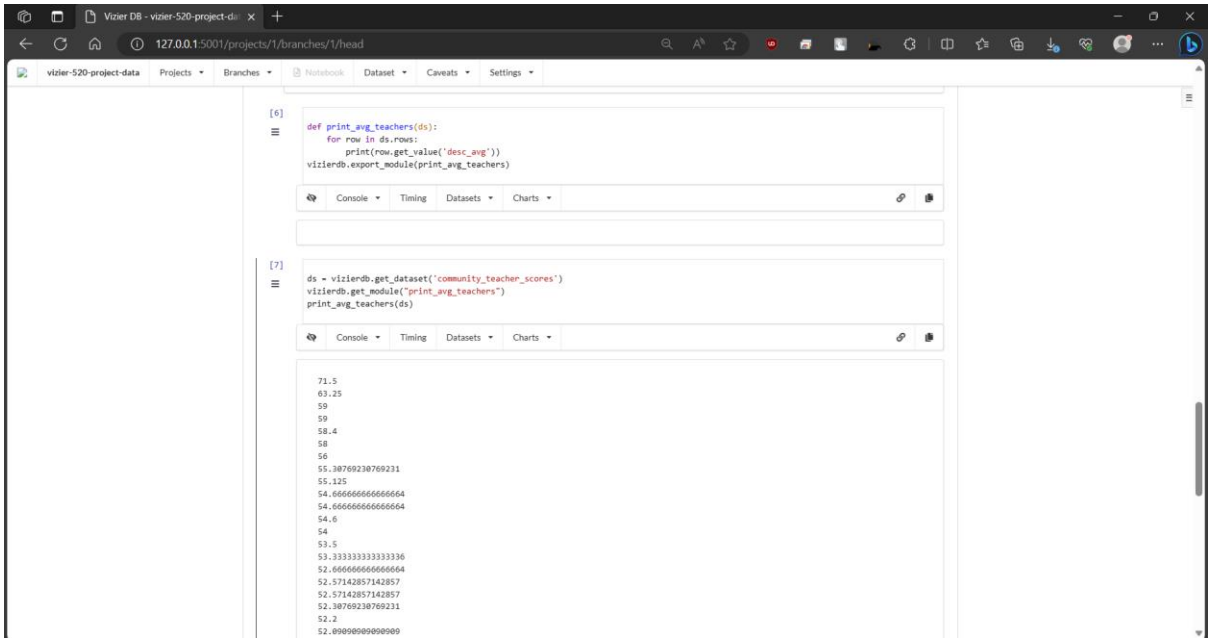


Task 7: Insert a new cell above the SQL cell that computes the average teacher scores (notebooks in Vizier are executed top down) by pressing the three bars below the cell number. Select "Impute Missing Values", select the score_and_community dataset and Teachers_Score as the column to be imputed, and select mean as the imputation method and take a screenshot of the updated line chart.



Dharmik Patel
A20526771

Task 8: Create a Python cell at the end of the notebook and create a function called print_avg_teachers that uses Vizier's API to get a handle for this dataset and print all values of the avg_teacher_score column. Hint: use the "Show Code Examples" buttom to see example Vizier API usage and see here for the API documention. 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.