Dhruv Dasadia
A20536411

CS 520

# Data Integration, Warehouse, and Provenance
# Vizier Assignment - 1

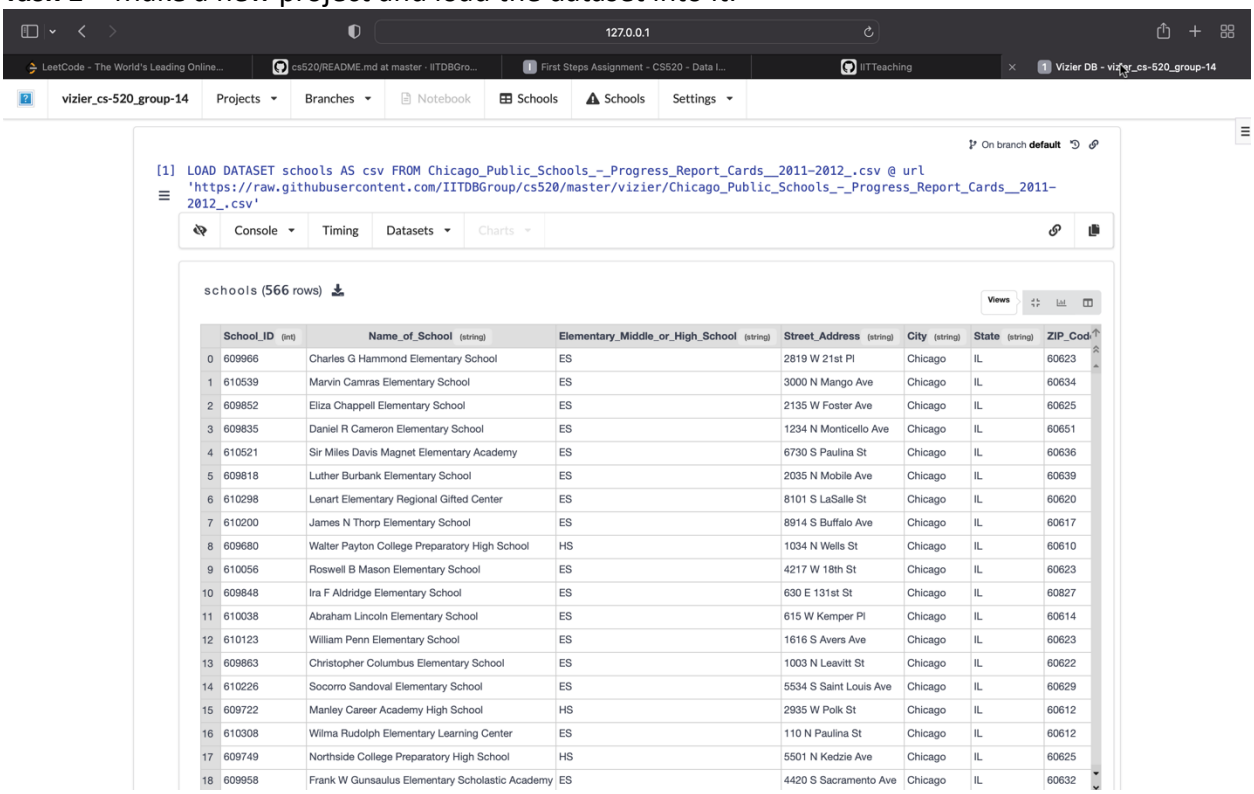
Intro to Vizier

For this assignment we have downloaded docker and downloaded the image. Configuration of Coursier and other dependency to use vizier.

Run the following Commands.

- docker pull iitdbgroup/vizier_iit_cs520_fall23:arm
- docker run  --name vizier --rm -v `pwd`:/vizier.db -p 5001:5001 -p 8089:8089 iitdbgroup/vizier_iit_cs520_fall23:arm -p 5001

**Task 1** – Make a new project and load the dataset into it.

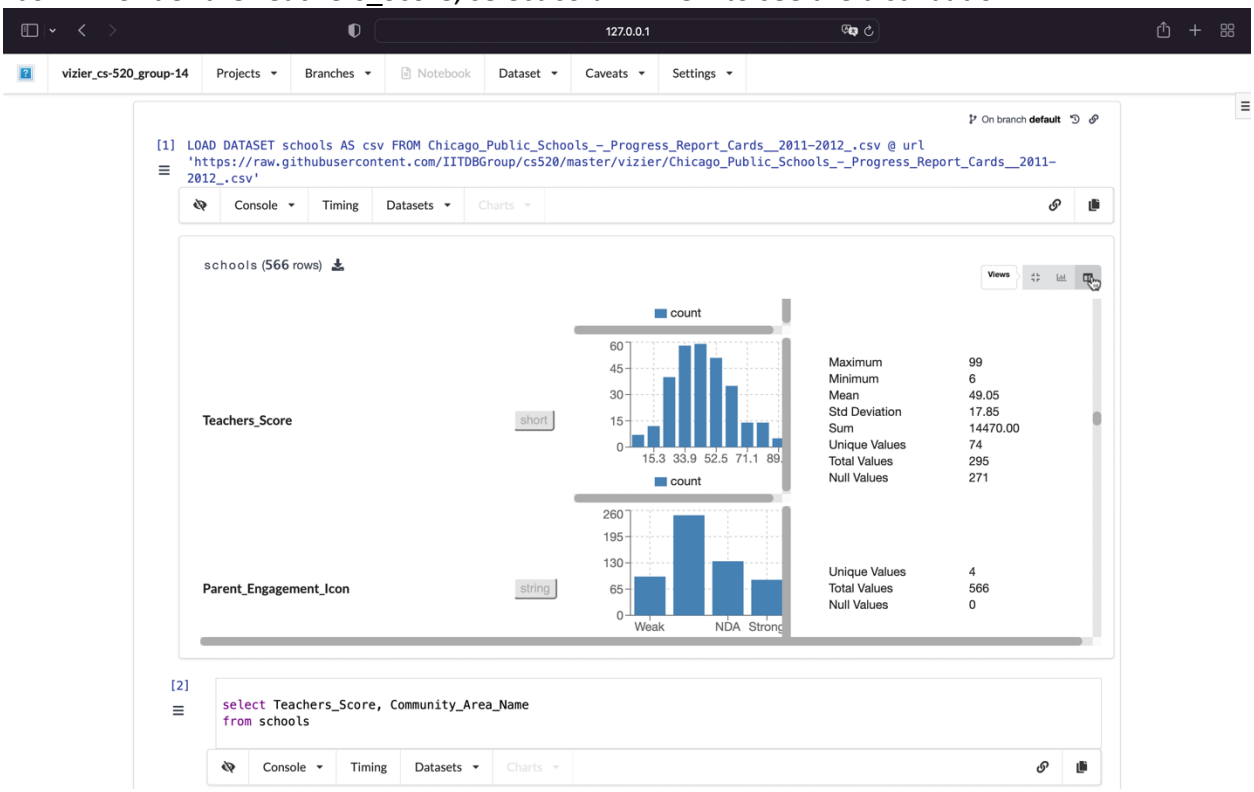



| | School_ID (int) | Name_of_School (string) | Elementary_Middle_or_High_School (string) | Street_Address (string) | City (string) | State (string) | ZIP_Cod |
|---|---|---|---|---|---|---|---|
| 0 | 609966 | Charles G Hammond Elementary School | ES | 2819 W 21st Pl | Chicago | IL | 60623 |
| 1 | 610539 | Marvin Camras Elementary School | ES | 3000 N Mango Ave | Chicago | IL | 60634 |
| 2 | 609852 | Eliza Chappell Elementary School | ES | 2135 W Foster Ave | Chicago | IL | 60625 |
| 3 | 609835 | Daniel R Cameron Elementary School | ES | 1234 N Monticello Ave | Chicago | IL | 60651 |
| 4 | 610521 | Sir Miles Davis Magnet Elementary Academy | ES | 6730 S Paulina St | Chicago | IL | 60636 |
| 5 | 609818 | Luther Burbank Elementary School | ES | 2035 N Mobile Ave | Chicago | IL | 60639 |
| 6 | 610298 | Lenart Elementary Regional Gifted Center | ES | 8101 S LaSalle St | Chicago | IL | 60620 |
| 7 | 610200 | James N Thorp Elementary School | ES | 8914 S Buffalo Ave | Chicago | IL | 60617 |
| 8 | 609680 | Walter Payton College Preparatory High School | HS | 1034 N Wells St | Chicago | IL | 60610 |
| 9 | 610056 | Roswell B Mason Elementary School | ES | 4217 W 18th St | Chicago | IL | 60623 |
| 10 | 609848 | Ira F Aldridge Elementary School | ES | 630 E 131st St | Chicago | IL | 60827 |
| 11 | 610038 | Abraham Lincoln Elementary School | ES | 615 W Kemper Pl | Chicago | IL | 60614 |
| 12 | 610123 | William Penn Elementary School | ES | 1616 S Avers Ave | Chicago | IL | 60623 |
| 13 | 609863 | Christopher Columbus Elementary School | ES | 1003 N Leavitt St | Chicago | IL | 60622 |
| 14 | 610226 | Socorro Sandoval Elementary School | ES | 5534 S Saint Louis Ave | Chicago | IL | 60629 |
| 15 | 609722 | Manley Career Academy High School | HS | 2935 W Polk St | Chicago | IL | 60612 |
| 16 | 610308 | Wilma Rudolph Elementary Learning Center | ES | 110 N Paulina St | Chicago | IL | 60612 |
| 17 | 609749 | Northside College Preparatory High School | HS | 5501 N Kedzie Ave | Chicago | IL | 60625 |
| 18 | 609958 | Frank W Gunsaulus Elementary Scholastic Academy | ES | 4420 S Sacramento Ave | Chicago | IL | 60632 |

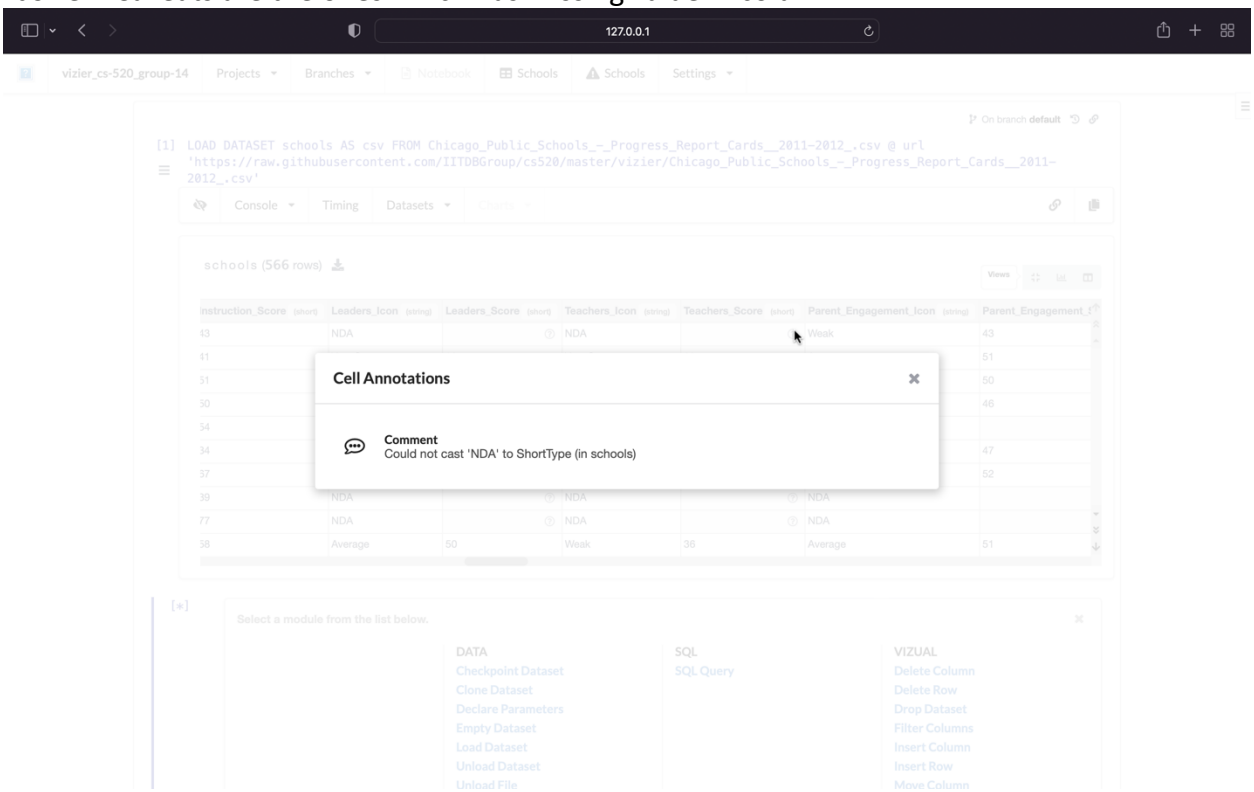
Dhruv Dasadia
A20536411



**Task 2** – Under the Teachers_Score, select column view to see the distribution.



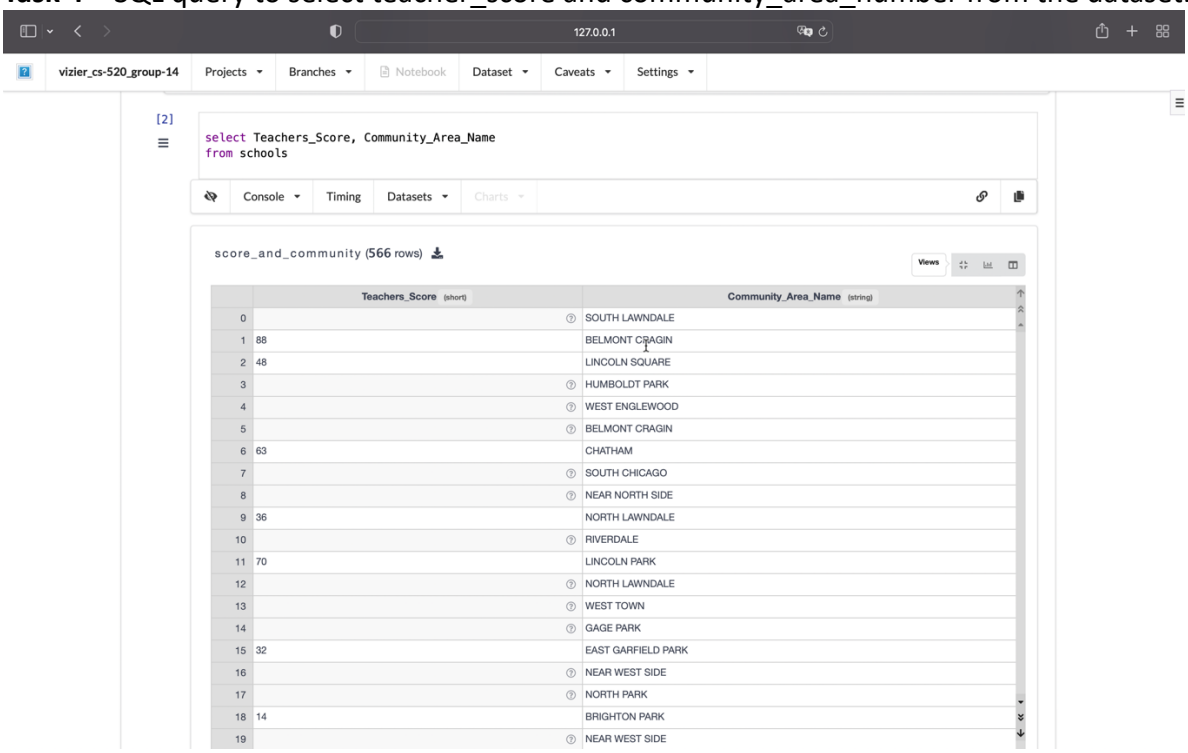

**Task 3** – Caveats are the ones which has missing value in column.



Dhruv Dasadia
A20536411

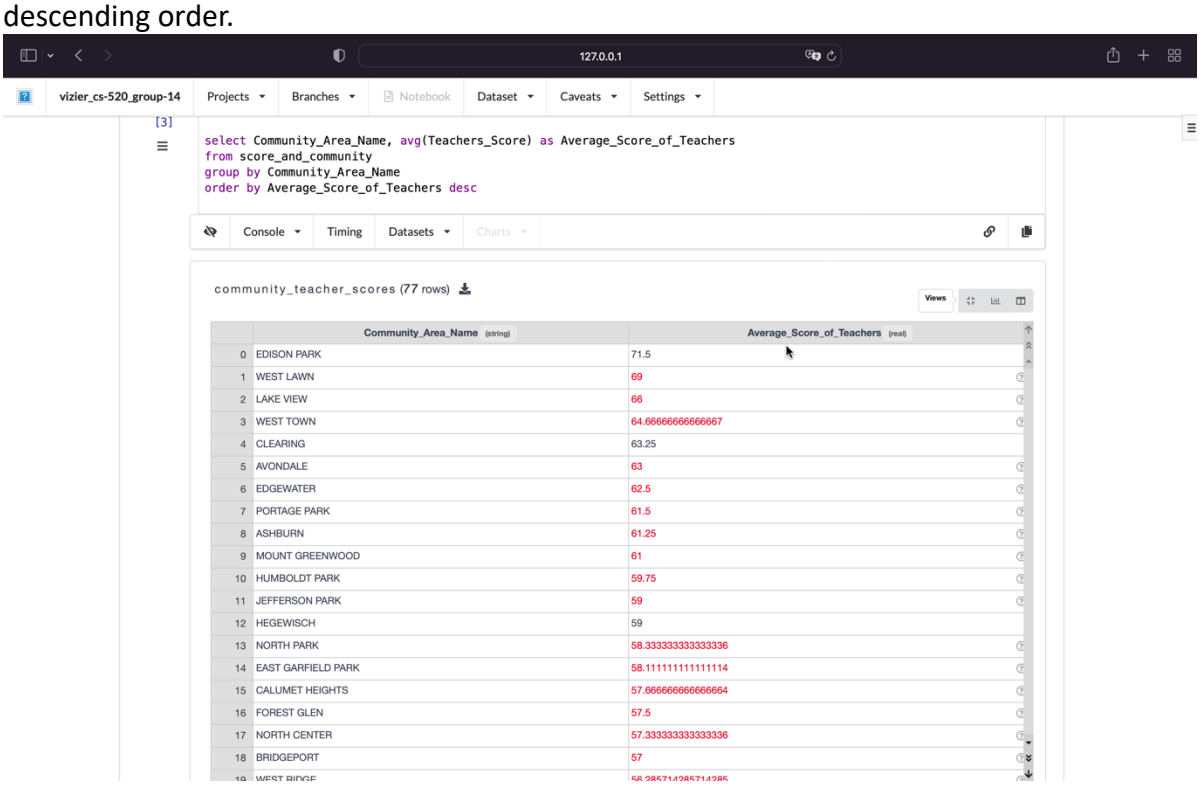


**Task 4** – SQL query to select teacher_score and community_area_number from the dataset.




| | Teachers_Score (short) | Community_Area_Name (string) |
|---|---|---|
| 0 | | SOUTH LAWNDALE |
| 1 | 88 | BELMONT CRAGIN |
| 2 | 48 | LINCOLN SQUARE |
| 3 | | HUMBOLDT PARK |
| 4 | | WEST ENGLEWOOD |
| 5 | | BELMONT CRAGIN |
| 6 | 63 | CHATHAM |
| 7 | | SOUTH CHICAGO |
| 8 | | NEAR NORTH SIDE |
| 9 | 36 | NORTH LAWNDALE |
| 10 | | RIVERDALE |
| 11 | 70 | LINCOLN PARK |
| 12 | | NORTH LAWNDALE |
| 13 | | WEST TOWN |
| 14 | | GAGE PARK |
| 15 | 32 | EAST GARFIELD PARK |
| 16 | | NEAR WEST SIDE |
| 17 | | NORTH PARK |
| 18 | 14 | BRIGHTON PARK |
| 19 | | NEAR WEST SIDE |

**Task 5** – SQL query for dataset score and community which we created in task 4, where we have to select the Community Area Name and Average scores of Teachers where the score is in descending order.

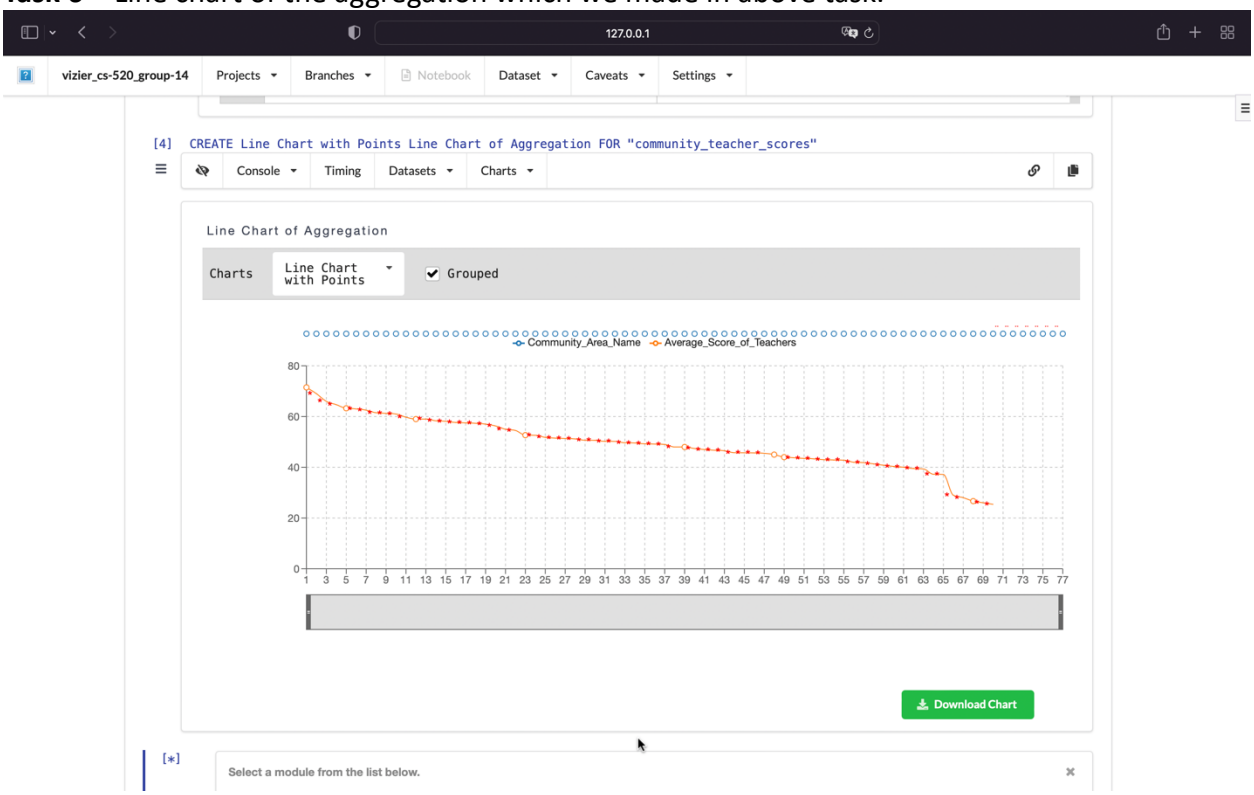



| | Community_Area_Name (string) | Average_Score_of_Teachers (real) |
|---|---|---|
| 0 | EDISON PARK | 71.5 |
| 1 | WEST LAWN | 69 |
| 2 | LAKE VIEW | 66 |
| 3 | WEST TOWN | 64.66666666666667 |
| 4 | CLEARING | 63.25 |
| 5 | AVONDALE | 63 |
| 6 | EDGEWATER | 62.5 |
| 7 | PORTAGE PARK | 61.5 |
| 8 | ASHBURN | 61.25 |
| 9 | MOUNT GREENWOOD | 61 |
| 10 | HUMBOLDT PARK | 59.75 |
| 11 | JEFFERSON PARK | 59 |
| 12 | HEGEWISCH | 59 |
| 13 | NORTH PARK | 58.333333333333336 |
| 14 | EAST GARFIELD PARK | 58.111111111111114 |
| 15 | CALUMET HEIGHTS | 57.666666666666664 |
| 16 | FOREST GLEN | 57.5 |
| 17 | NORTH CENTER | 57.333333333333336 |
| 18 | BRIDGEPORT | 57 |
| 19 | WEST RIDGE | 56.285714285714285 |

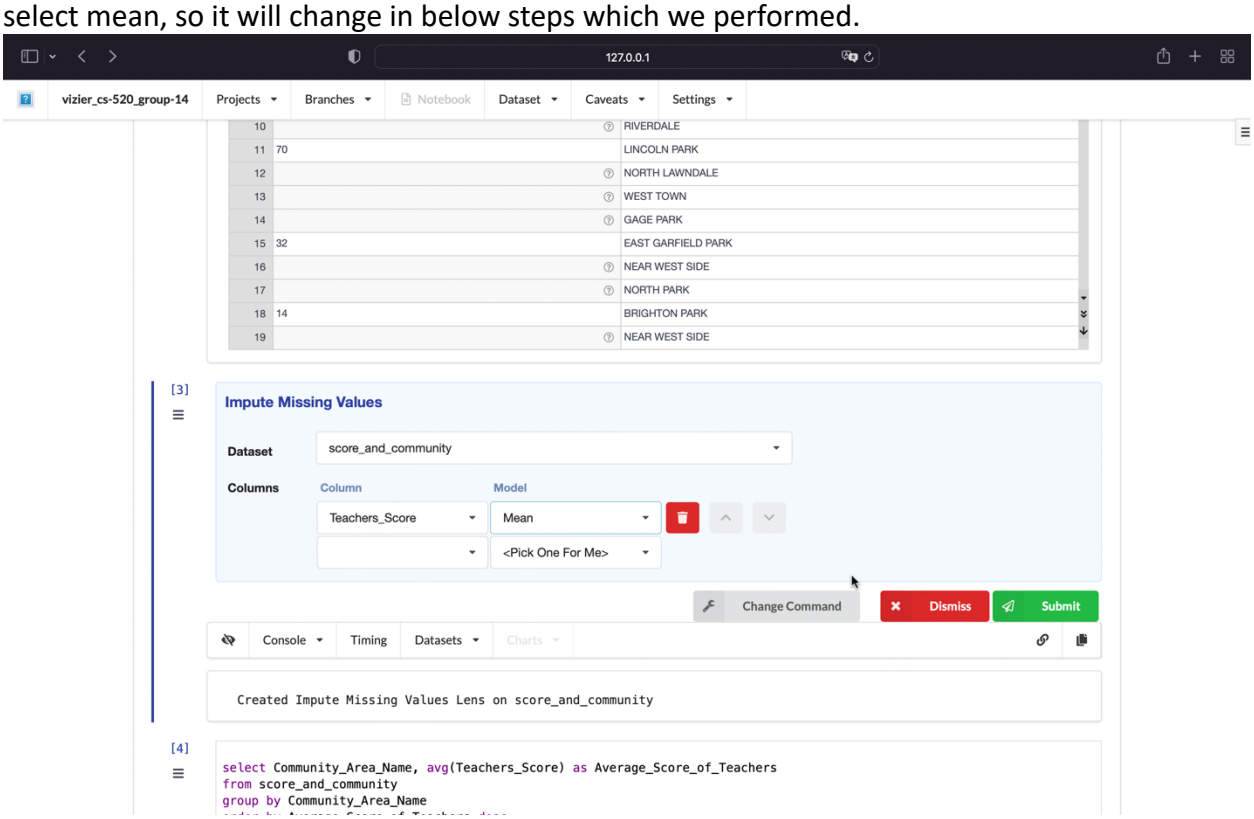
Dhruv Dasadia
A20536411

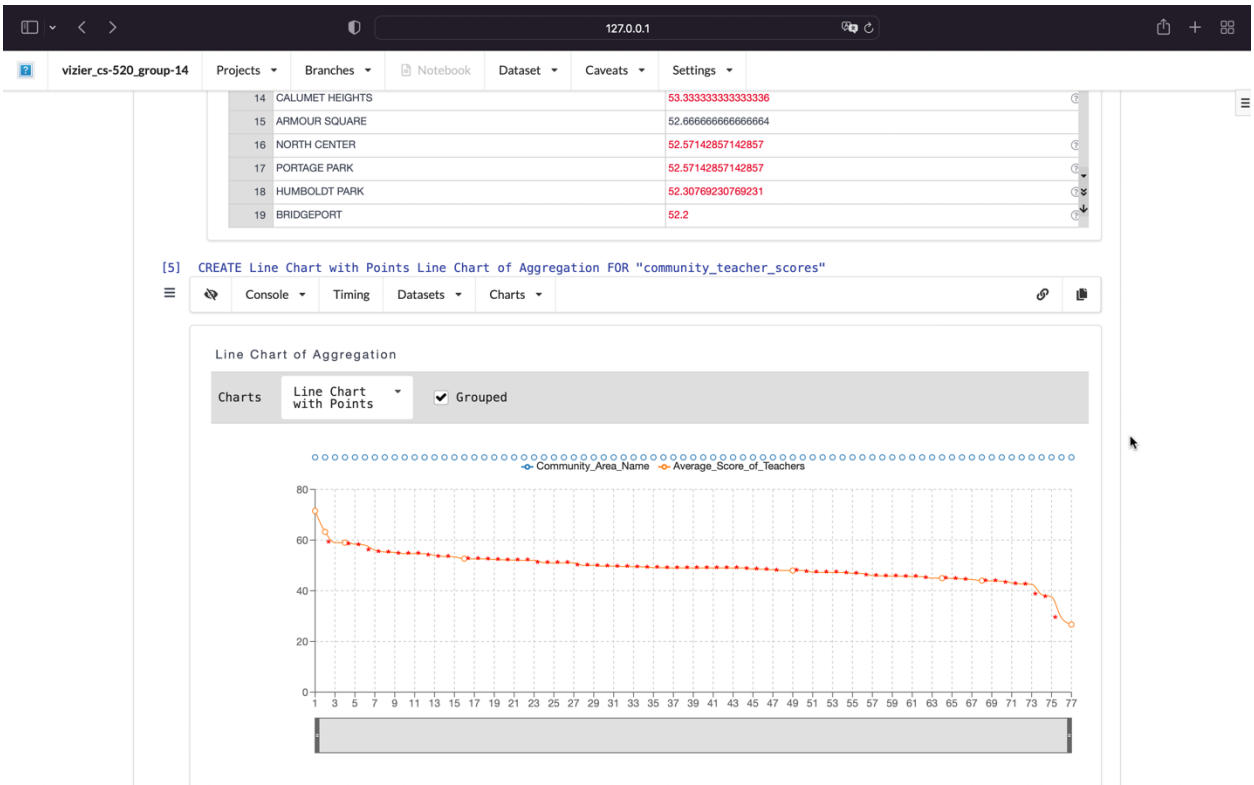


**Task 6** – Line chart of the aggregation which we made in above task.

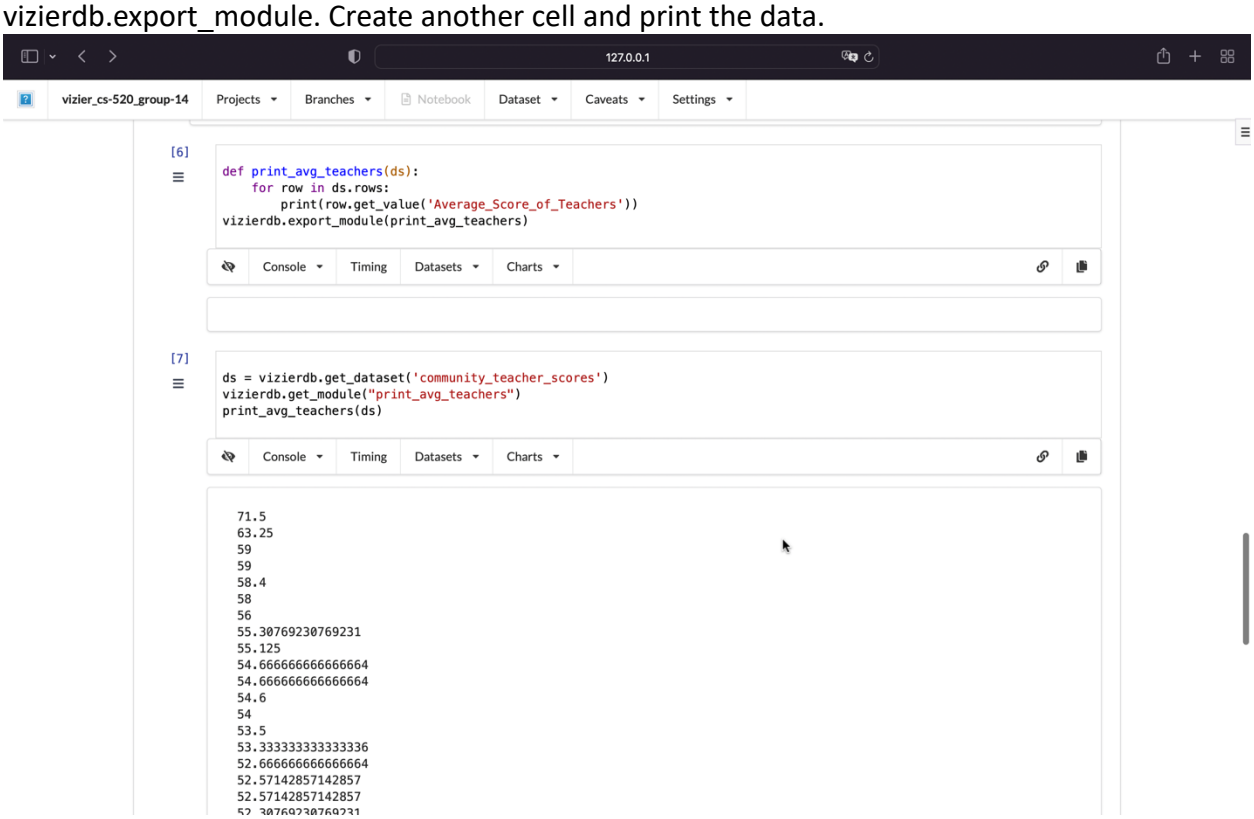



**Task 7** – Add a cell above the Average Scores of Teachers and add Impute Missing Values and select mean, so it will change in below steps which we performed.






**Task 8** – Create a Python cell to print all the rows of average scores of teacher using vizierdb.export_module. Create another cell and print the data.



**Task 9 –** Python cell to filter the rows where the average score of teachers should be more than 30.0.

127.0.0.1

vizier_cs-520_group-14 | Projects | Branches | Notebook | Dataset | Caveats | Settings

```
46
45.84615384615385
45.76470588235294
45.714285714285715
45.6
45.5
45
45
44.857142857142854
44.69230769230769
44.4
44
43.857142857142854
43.8
43.125
42.666666666666664
42.54545454545455
38.5
37.57142857142857
29.333333333333332
26.666666666666668
```

[8]

```
df = vizierdb.get_data_frame('community_teacher_scores')
print(df[df.Average_Score_of_Teachers < 30.0])
```

Console | Timing | Datasets | Charts

```
Community_Area_Name  Average_Score_of_Teachers
75         ROGERS PARK                 29.333333
76         AVALON PARK                 26.666667
```

Connected to vizier @ http://localhost:5001/vizier-db/api/v1/