

Dharmik Patel  
A20526771

## CS 520

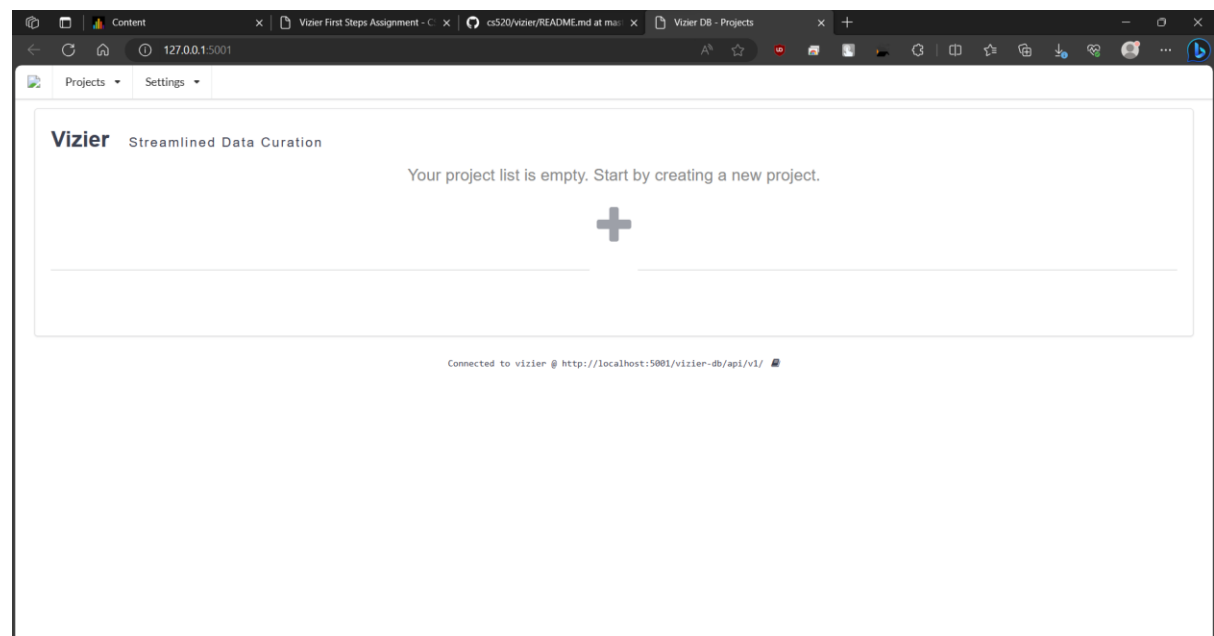
### Data Integration, Warehousing, and Provenance

#### 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

Load Dataset

From Local Machine From the Internet

URL

Download file from URL

[https://raw.githubusercontent.com/IITDBGroup/cs520/master/vizier/Chicago\\_Public\\_Schools\\_-\\_Progress\\_Report\\_Cards\\_\\_2011-2012\\_.csv](https://raw.githubusercontent.com/IITDBGroup/cs520/master/vizier/Chicago_Public_Schools_-_Progress_Report_Cards__2011-2012_.csv)

Dataset Name schools

Load Format CSV

Show Advanced Options

Change Command Dismiss Submit

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

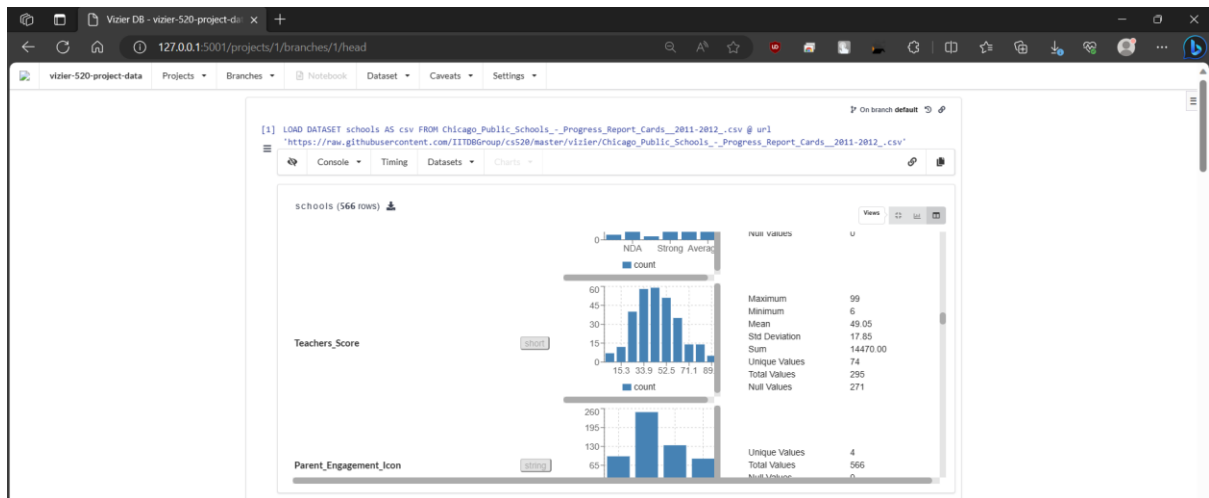
[1] LOAD DATASET schools AS csv FROM Chicago\_Public\_Schools\_-\_Progress\_Report\_Cards\_\_2011-2012\_.csv @ url 'https://raw.githubusercontent.com/IITDBGroup/cs520/master/vizier/Chicago\_Public\_Schools\_-\_Progress\_Report\_Cards\_\_2011-2012\_.csv'

schools (566 rows)

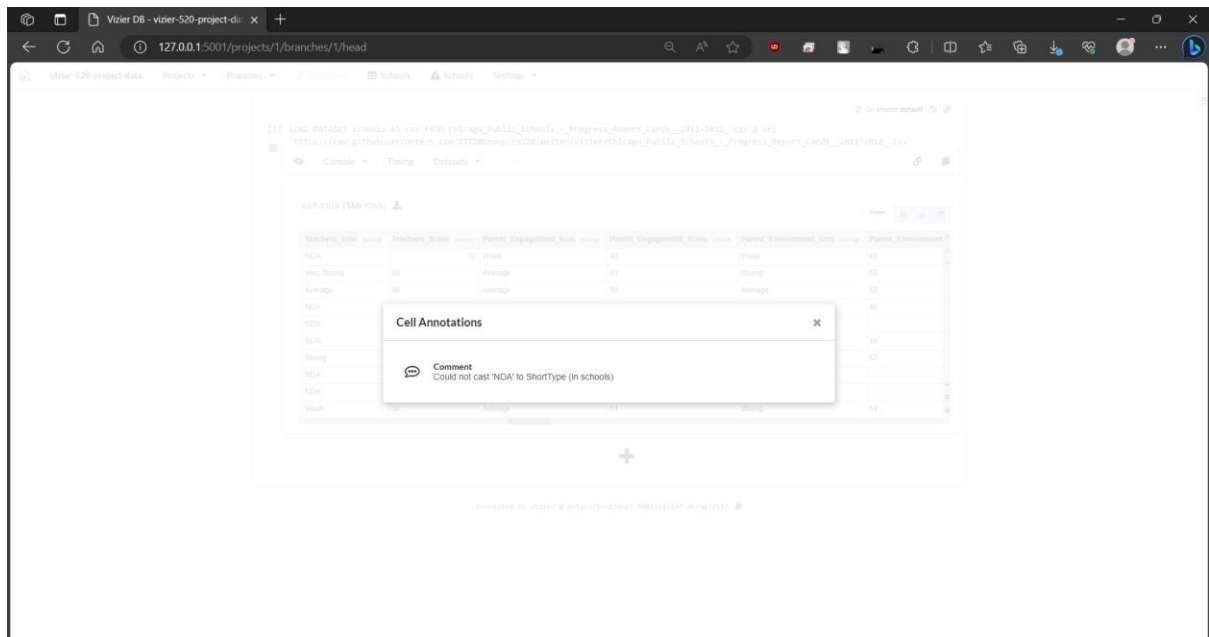
School_ID	Name_of_School	Elementary_Middle_or_High_School	Street_Address	City	State	ZIP_Code
0 609966	Charles O Hammond Elementary School	ES	2819 W 21st Pl	Chicago	IL	60623
1 610639	Marvin Camras Elementary School	ES	3080 N Mango Ave	Chicago	IL	60634
2 609852	Eliza Chappell Elementary School	ES	2135 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 609810	Luther Burbank Elementary School	ES	2035 N Mobile Ave	Chicago	IL	60639
6 610290	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	Roosevelt 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	60627
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	60632
14 610326	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 Gonzauals Elementary Scholastic Academy	ES	4420 S Sacramento Ave	Chicago	IL	60632
19 610121	Washington Irving Elementary School	ES	749 S Oakley Blvd	Chicago	IL	60612

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.

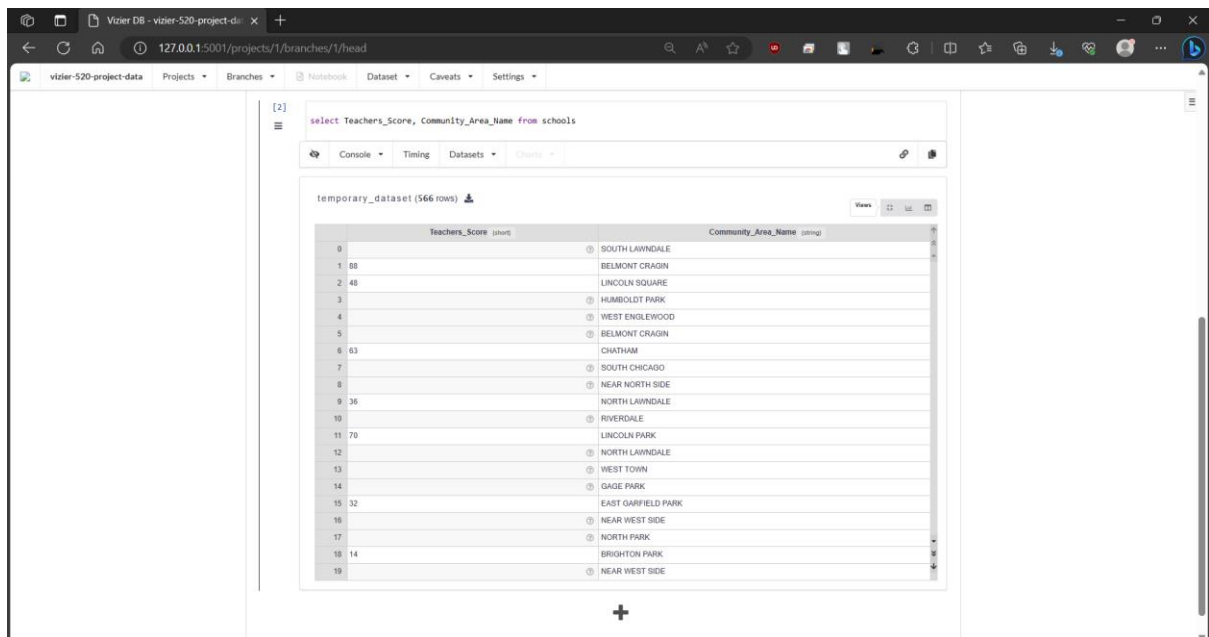


Task 3: Click on one of the question marks for values in the teachers column and take a screenshot.



Dharmik Patel  
A20526771

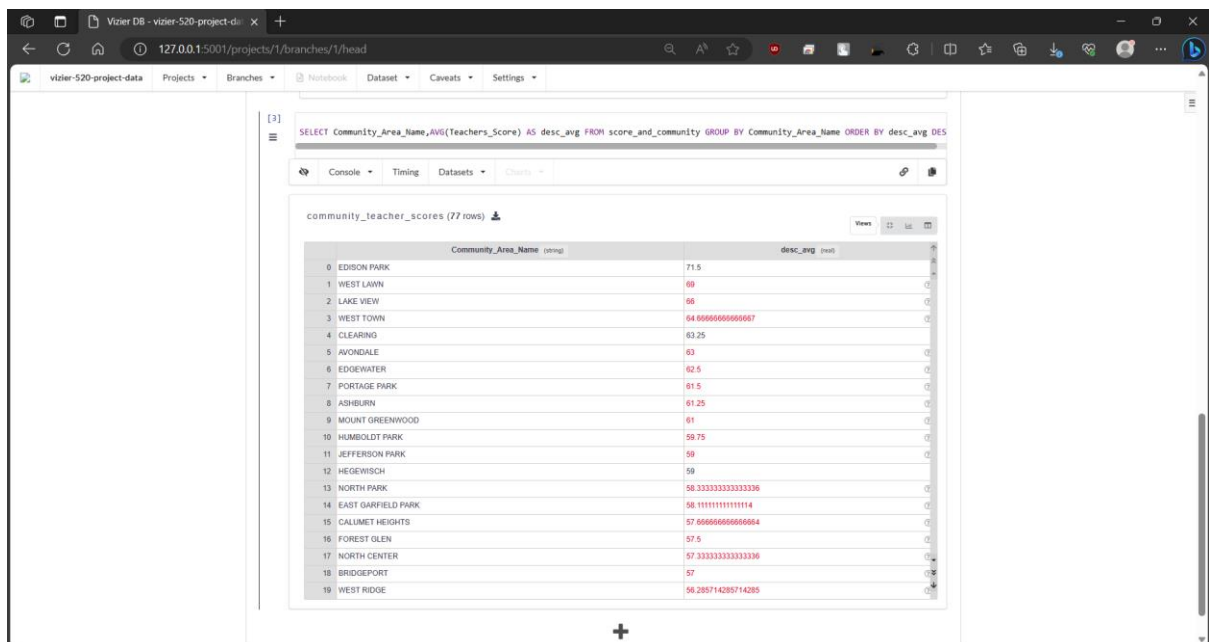
Task 4: Create a SQL cell and write a query that returns columns Teachers\_Score and Community\_Area\_Name. SQL results can be stored as new datasets in Vizier. Call the result dataset score\_and\_community. And take a screenshot of the result.



The screenshot shows the Vizier interface with a SQL query in a cell. The query is: `select Teachers_Score, Community_Area_Name from schools`. Below the query, the results are displayed in a table titled "temporary\_dataset (566 rows)". The table has two columns: "Teachers\_Score (float)" and "Community\_Area\_Name (string)".

Teachers_Score (float)	Community_Area_Name (string)
9	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: 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.

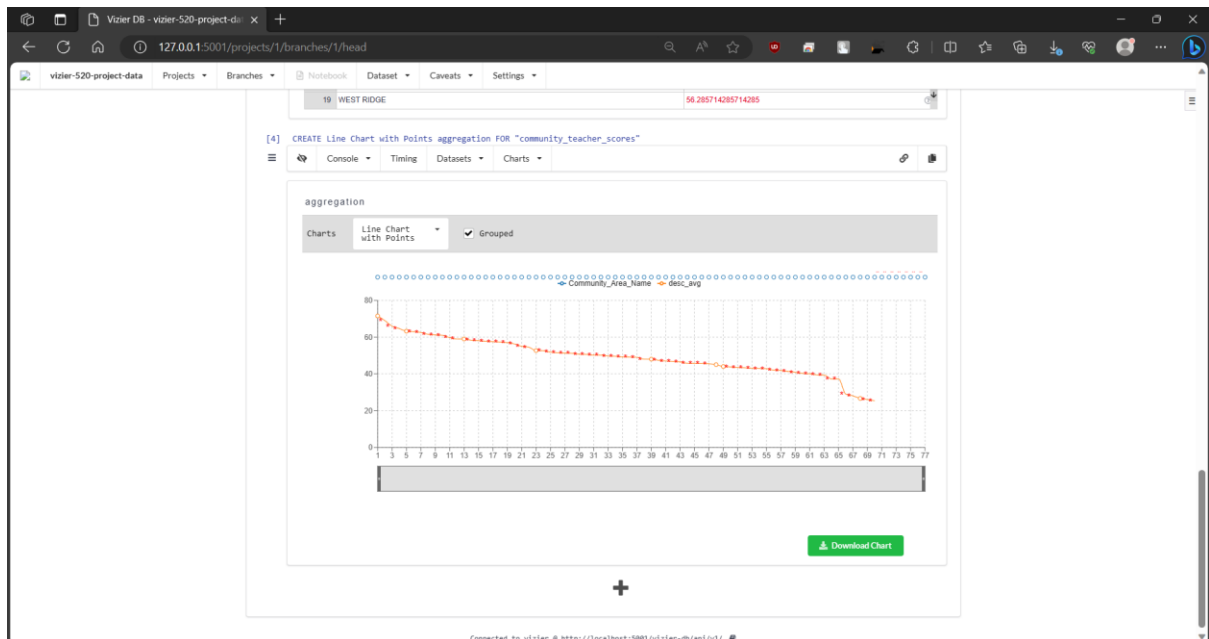


The screenshot shows the Vizier interface with a SQL query in a cell. The query is: `SELECT Community_Area_Name,AVG(Teachers_Score) AS desc_avg FROM score_and_community GROUP BY Community_Area_Name ORDER BY desc_avg DESC`. Below the query, the results are displayed in a table titled "community\_teacher\_scores (77 rows)". The table has two columns: "Community\_Area\_Name (string)" and "desc\_avg (avg)".

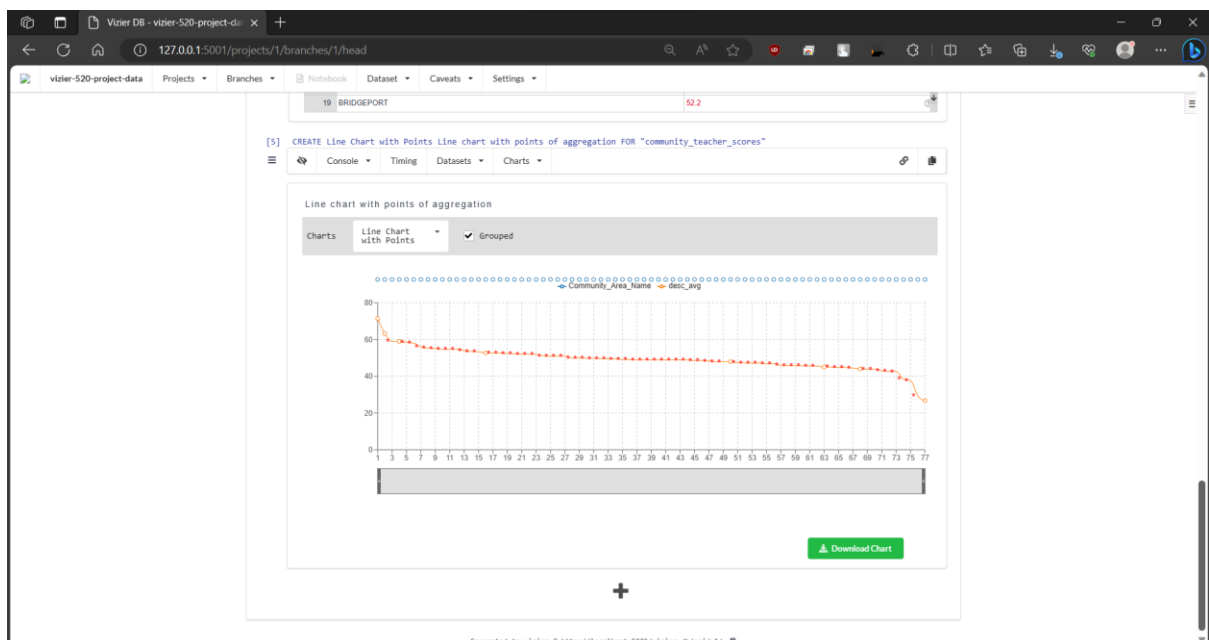
Community_Area_Name (string)	desc_avg (avg)
0 EDISON PARK	71.5
1 WEST LAWN	69
2 LAKE VIEW	58
3 WEST TOWN	54.66666666666667
4 CLEARING	63.25
5 AVONDALE	63
6 EDGEWATER	62.5
7 PORTAGE PARK	61.5
8 ASHBURN	61.25
9 MOUNT GREENWOOD	61
10 HUMBOLDT PARK	59.75
11 JEFFERSON PARK	59
12 HEGEWISCH	59
13 NORTH PARK	58.33333333333333
14 EAST GARFIELD PARK	58.11111111111111
15 CALUMET HEIGHTS	57.66666666666666
16 FOREST OLEN	57.5
17 NORTH CENTER	57.33333333333333
18 BRIDGEPORT	57
19 WEST RIDGE	56.285714285714285

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.

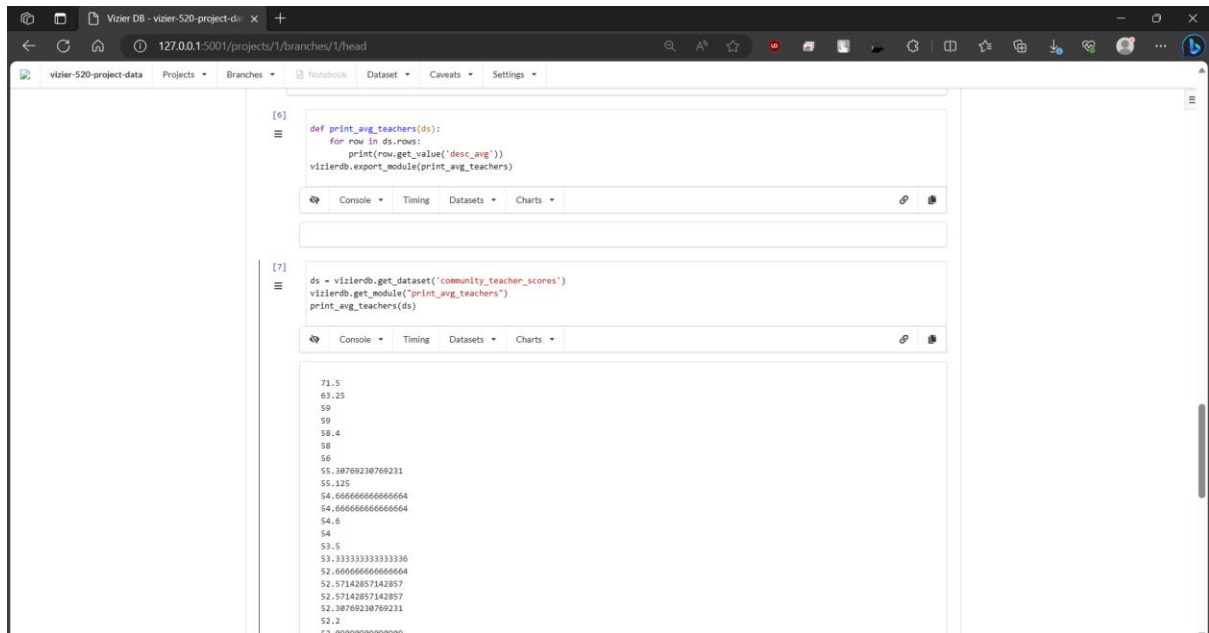


Task 7: Insert a new cell above the SQL cell that computes the average teacher scores (notebooks in Vizier are executed top down) by pressing the three bars below the cell number. Select "Impute Missing Values", select the score\_and\_community dataset and Teachers\_Score as the column to be imputed, 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" button to see example Vizier API usage and see here for the API documentation. 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.

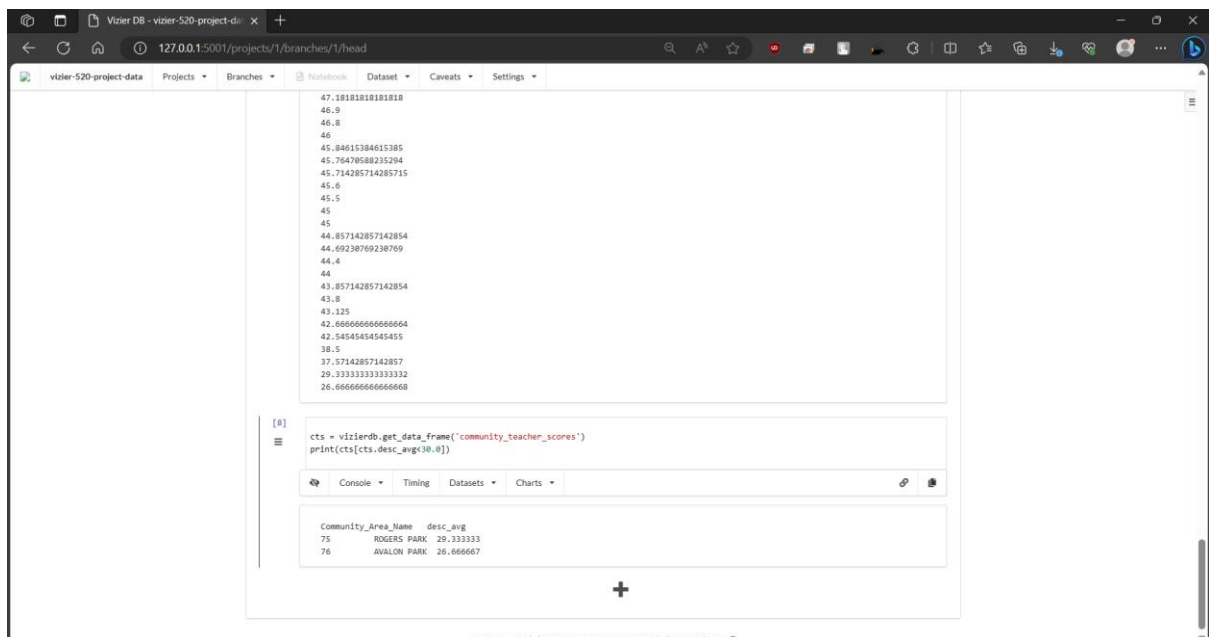


```
[6]
def print_avg_teachers(ds):
    for row in ds.rows:
        print(row.get_value('desc_avg'))
vizierdb.export_module(print_avg_teachers)

[7]
ds = vizierdb.get_dataset('community_teacher_scores')
vizierdb.get_module("print_avg_teachers")
print_avg_teachers(ds)
```

71.5  
63.25  
59  
59  
58.4  
58  
56  
55.38769238769231  
55.125  
54.66666666666666  
54.66666666666666  
54.6  
54  
53.5  
53.33333333333333  
52.66666666666666  
52.57142857142857  
52.57142857142857  
52.38769238769231  
52.2  
52.89898989898989

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.



```
[8]
cts = vizierdb.get_data_frame('community_teacher_scores')
print(cts[cts.desc_avg>=30.0])
```

Community_Area_Name	desc_avg
75 ROGERS PARK	29.333333
76 AVALON PARK	26.666667

Connected to vizier @ http://localhost:5881/vizier-db/api/v1/