

CS520- Data Integration, Warehousing, and Provenance- 2023 Fall

Vizier Assignment – Group 11

11#1 - Satya Nandikeswara rao, Chinta, schinta6@hawk.iit.edu

11#2 – Bavith Kumar Reddy, Komtiredy, bkomtiredy@hawk.iit.edu

11#3 – Ahaan, Kareem, akareem1@hawk.iit.edu

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

public_schools Projects Branches Notebook Dataset Caveats Settings

[1] Load Dataset

From Local Machine From the Internet

URL

Download file from URL

https://raw.githubusercontent.com/IITDBGGroup/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

Console Timing Datasets Charts

public_schools Projects Branches Notebook Dataset Caveats Settings

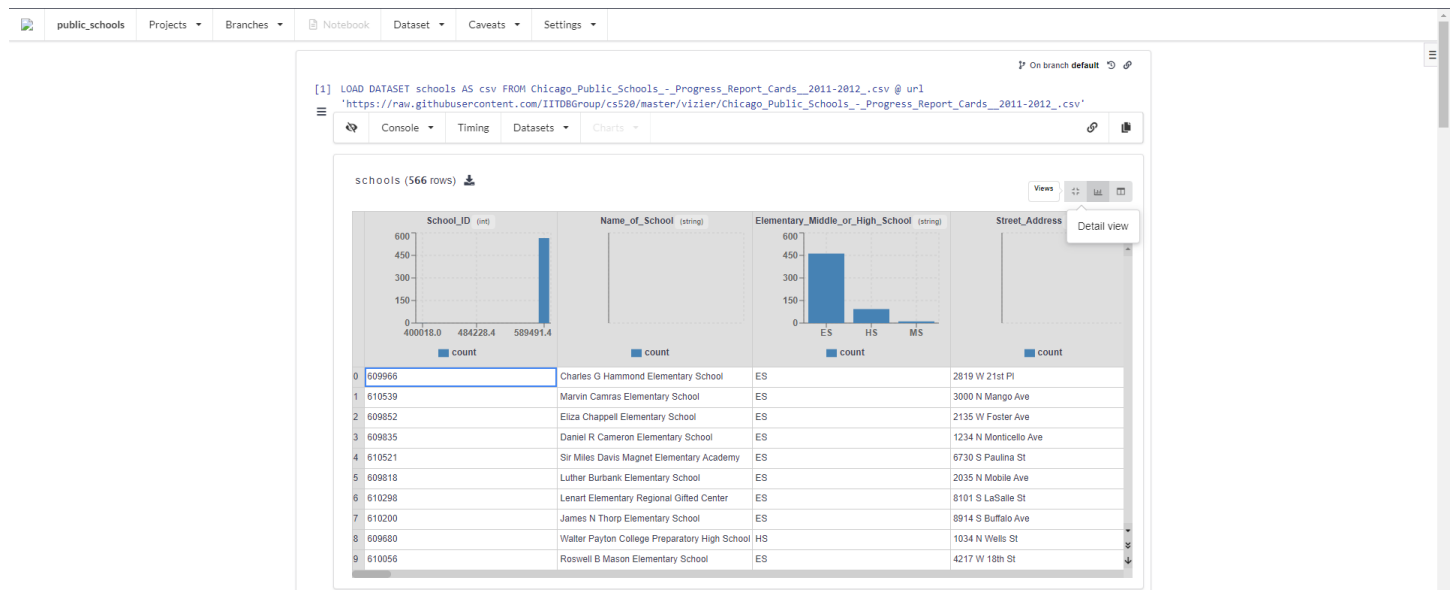
[1] LOAD DATASET schools AS csv FROM Chicago_Public_Schools_-_Progress_Report_Cards_2011-2012_.csv @ url
'https://raw.githubusercontent.com/IITDBGGroup/cs520/master/vizier/Chicago_Public_Schools_-_Progress_Report_Cards_2011-2012_.csv'

schools (566 rows)

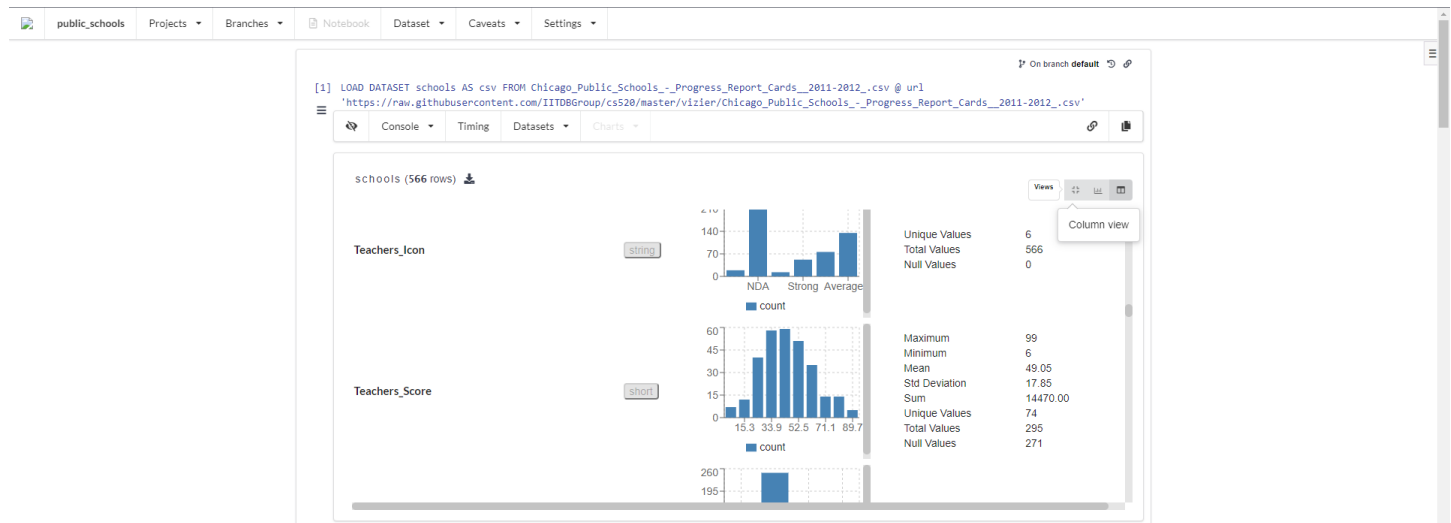
School_ID (int)	Name_of_School (string)	Elementary_Middle_or_High_School (string)	Street_Address (string)	City (string)	State (string)	ZIP_Code (int)
0 609966	Charles G Hammond Elementary School	ES	2819 W 21st Pl	Chicago	IL	60623
1 610539	Marvin Camras Elementary School	ES	3000 N Mango Ave	Chicago	IL	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	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	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
19 610121	Washington Irving Elementary School	ES	749 S Oakley Blvd	Chicago	IL	60612

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.

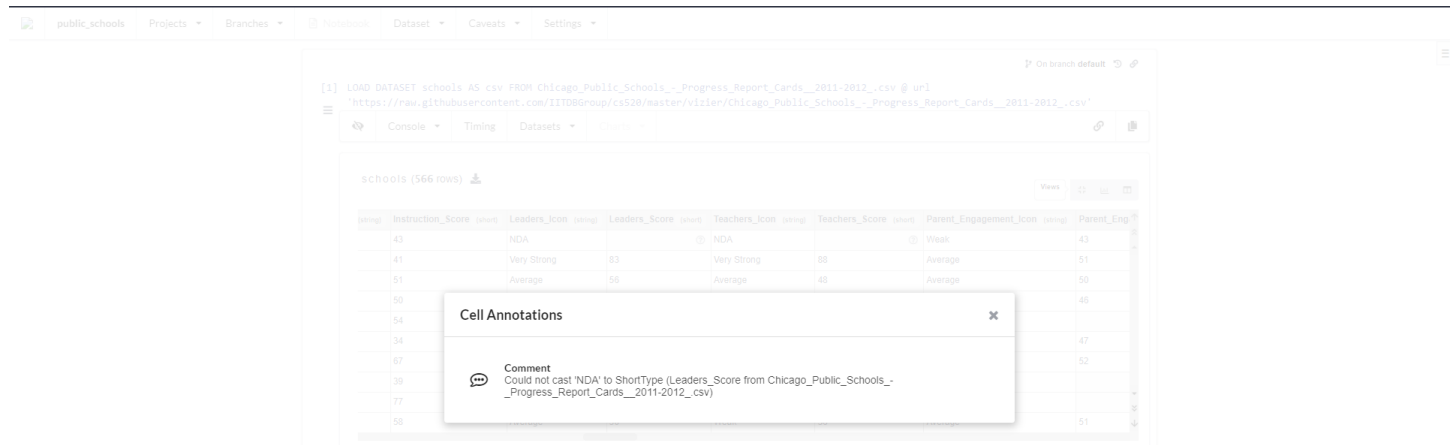
Detail View:



Column View of Teachers_Score:



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



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.

public_schoolsProjectsBranchesNotebookDatasetCaveatsSettings

[2]
≡

```
select Teachers_score, Community_Area_Name
from schools;
```

ConsoleTimingDatasetsCharts

score_and_community (566 rows)

Views

	Teachers_score (short)	Community_Area_Name (string)
0		SOUTH LAWINDALE
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 LAWINDALE
10		RIVERDALE
11	70	LINCOLN PARK
12		NORTH LAWINDALE
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 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.

public_schoolsProjectsBranchesNotebookDatasetCaveatsSettings

[3]
≡

```
select avg(Teachers_Score) as avg_score, Community_Area_Name
from score_and_community
group by Community_Area_Name
order by avg_score desc;
```

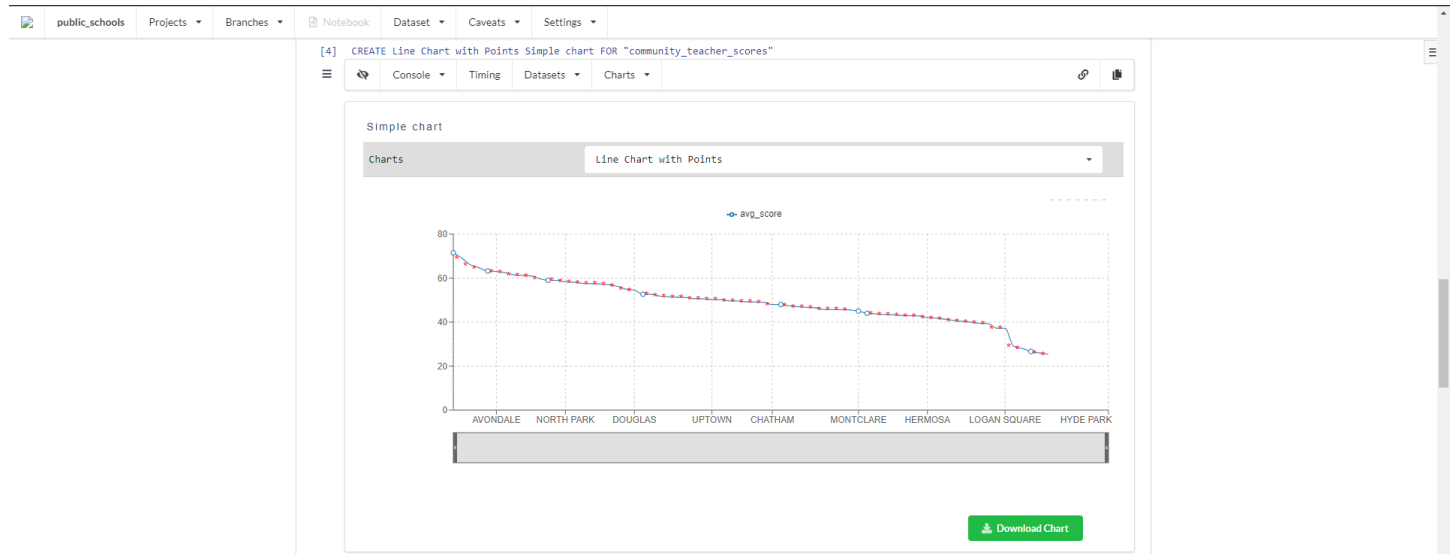
ConsoleTimingDatasetsCharts

community_teacher_scores (77 rows)

Views

	avg_score (real)	Community_Area_Name (string)
0	71.5	EDISON PARK
1	69	WEST LAWN
2	66	LAKE VIEW
3	64.66666666666667	WEST TOWN
4	63.25	CLEARING
5	63	AVONDALE
6	62.5	EDGEWATER
7	61.5	PORTAGE PARK
8	61.25	ASHBURN
9	61	MOUNT GREENWOOD
10	59.75	HUMBOLDT PARK
11	59	JEFFERSON PARK
12	59	HEGEWISCH
13	58.33333333333333	NORTH PARK
14	58.11111111111111	EAST GARFIELD PARK
15	57.66666666666666	CALUMET HEIGHTS
16	57.5	FOREST GLEN
17	57.33333333333333	NORTH CENTER
18	57	BRIDGEPORT
19	56.285714285714285	WEST RIDGE

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.

[3] CREATE LENS ON score_and_community IMPUTE MISSING VALUES ON COLUMN 0

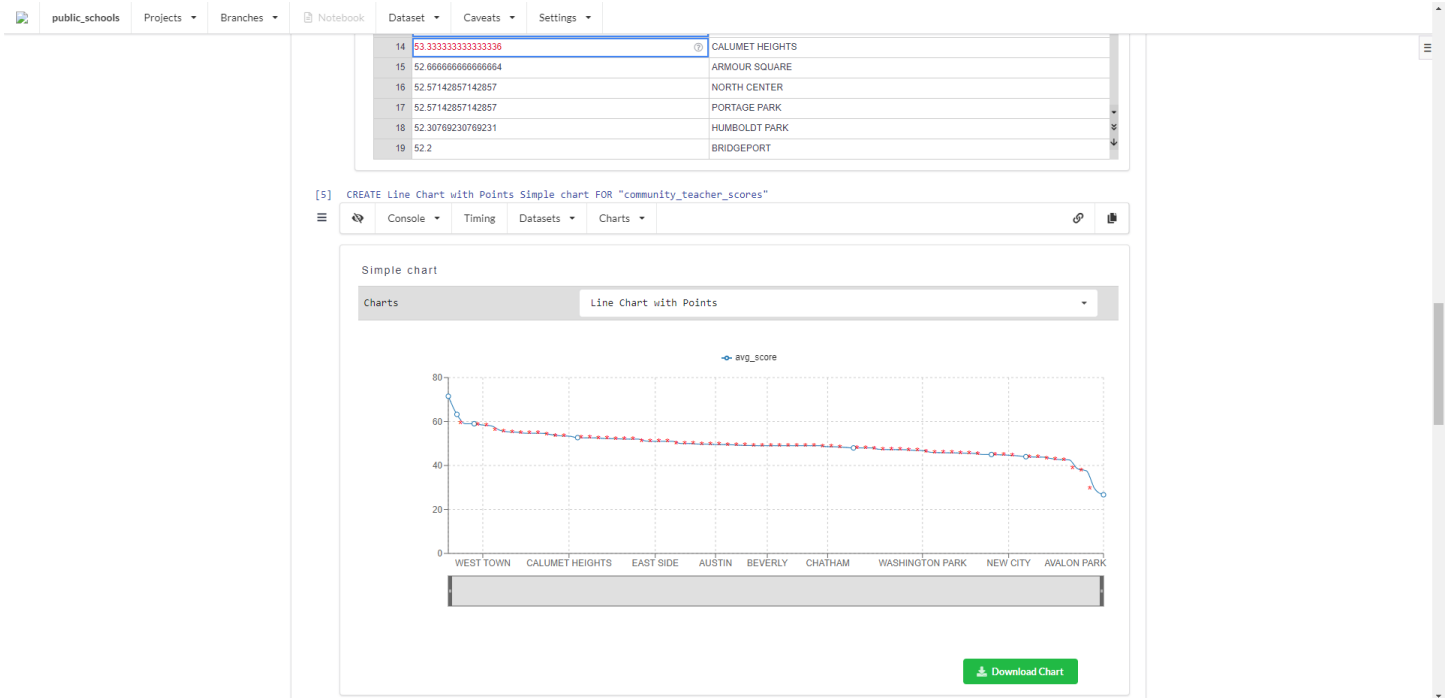
Created Impute Missing Values Lens on score_and_community

[4]

```
select avg(Teachers_Score) as avg_score, Community_Area_Name
from score_and_community
group by Community_Area_Name
order by avg_score desc;
```

community_teacher_scores (77 rows)

	avg_score (real)	Community_Area_Name (string)
0	71.5	EDISON PARK
1	63.25	CLEARING
2	59	WEST LAWN
3	59	HEGEWISCH
4	58.4	WEST TOWN
5	58	MOUNT GREENWOOD
6	56	AVONDALE
7	55.30769230769231	EAST GARFIELD PARK
8	55.125	ASHBURN
9	54.666666666666664	FOREST GLEN
10	54.666666666666664	WEST RIDGE
11	54.6	NORTH PARK
12	54	JEFFERSON PARK
13	53.5	EDGEWATER
14	53.333333333333336	CALUMET HEIGHTS
15	52.666666666666664	ARMOUR SQUARE



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():
    df=vizierdb.get_dataset("community_teacher_scores")
    for row in df.rows:
        print(row.get_value("avg_score"))
    vizierdb.export_module(print_avg_teachers)

[7]
vizierdb.get_module('print_avg_teachers')
print_avg_teachers()
```

```
71.5
63.25
59
58.4
58
56
55.30769230769231
55.125
54.666666666666664
54.666666666666664
54.6
54
53.5
53.333333333333336
52.666666666666664
52.57142857142857
52.57142857142857
52.30769230769231
52.2
0.000000000000000
52
52
```

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.

public_schools

Projects

Branches

Notebook

Dataset

Caveats

Settings

38.5
37.57142857142857
29.333333333333332
26.666666666666668

[8]

```
df_new = vizierdb.get_data_frame('community_teacher_scores')
great30 = df_new[df_new['avg_score']>=30.0]
less30 = df_new[df_new['avg_score']<30.0]
print("Teachers with greater than or equal to 30 score: \n", great30)
print("Teachers with less than 30 score: \n",less30)
```

Console

Timing

Datasets

Charts

Teachers with greater than or equal to 30 score:

avg_score Community_Area_Name
0 71.500000 EDISON PARK
1 63.250000 CLEARING
2 59.000000 WEST LAWN
3 59.000000 HEGENISCH
4 58.400000 WEST TOWN
.. ..
70 43.125000 WOODLAWN
71 42.666667 WEST ELSDON
72 42.545455 LOGAN SQUARE
73 38.500000 FULLER PARK
74 37.571429 SOUTH CHICAGO

[75 rows x 2 columns]

Teachers with less than 30 score:

avg_score Community_Area_Name
75 29.333333 ROGERS PARK
76 26.666667 AVALON PARK

+

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