

CS 520 Data Integration Warehousing

Group Members

Girish Rajani (A20503736)

Shriya Prasanna (A20521733)

Subramanya Ganesh (A20516250)

TASK\_1

Vizier First Steps Assignment

Projects

Branches

Notebook

Schools

Schools

Settings

On branch default

[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'

ConsoleTimingDatasetsCharts

schools (566 rows)

Views

	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

Task\_2

Column View Teachers\_Score

schools (566 rows)

Views

Teachers\_Icon

string

Category	Count
NDA	150
Strong	50
Average	140

Unique Values6  
Total Values566  
Null Values0

Teachers\_Score


short


Score Range	Count
15.3	10
33.9	40
52.5	60
71.1	50
89.7	15

Maximum99  
Minimum6  
Mean49.05  
Std Deviation17.85  
Sum14470.00  
Unique Values74  
Total Values295  
Null Values271

### Task\_3

### Cell Annotations




**Comment**  
Could not cast 'NDA' to ShortType (in schools)




### Task\_4

[2]  
≡

```
select Teachers_Score, Community_Area_Name from schools
```

Console Timing Datasets Charts

score\_and\_community (566 rows) 

Views   

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

Task\_5

```
select Community_Area_Name, avg(Teachers_Score)as average_teachers_score from score_and_community
group by Community_Area_Name
Order by average_teachers_score Desc
```

Console ▾

Timing

Datasets ▾

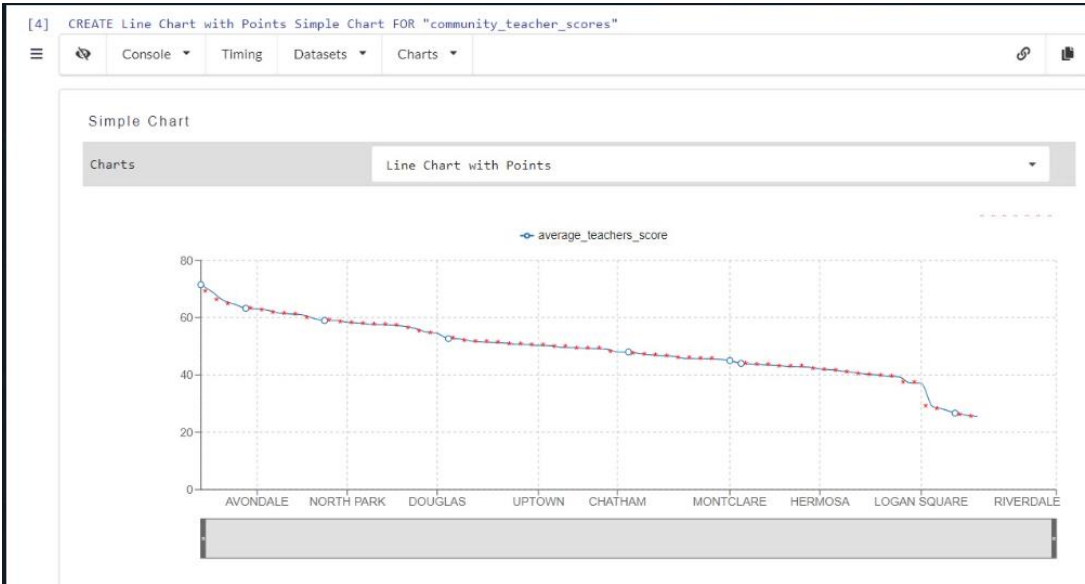
Charts ▾

community\_teacher\_scores (77 rows)

Views

	Community_Area_Name (string)	average_teachers_score (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

Task\_6



## Task\_7

[\*]

### Impute Missing Values

Dataset	score_and_community	
Columns	Column	Model
	Teachers_Score	Mean
		<Pick One For Me>

Change Command

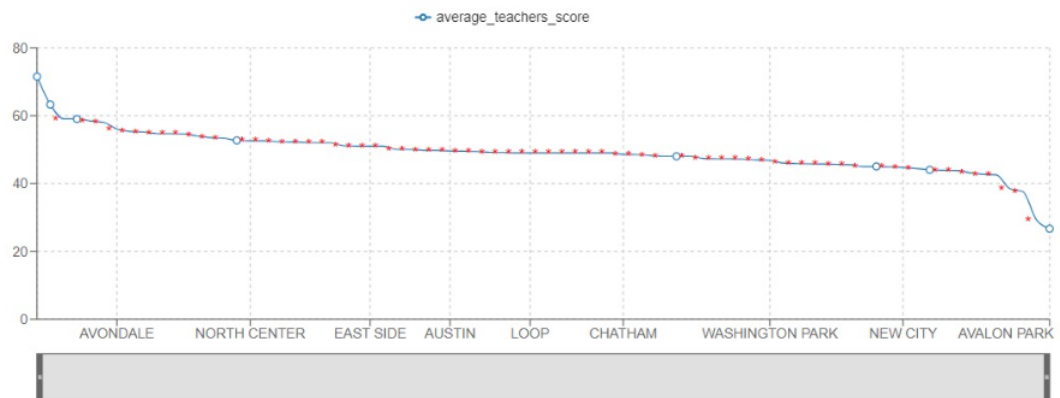
Dismiss

Submit

### Simple Chart

Charts

Line Chart with Points



## Task\_8

[6]



```
def print_avg_teachers():  
  
    ds = vizierdb.get_dataset('community_teacher_scores')  
    for row in ds.rows:  
        print(row.get_value('average_teachers_score'))  
vizierdb.export_module(print_avg_teachers)
```



Console ▾

Timing

Datasets ▾

Charts ▾

[7]



```
vizierdb.get_module("print_avg_teachers")  
print_avg_teachers()
```



Console ▾

Timing

Datasets ▾

Charts ▾

```
71.5  
63.25  
59  
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  
52.09090909090909  
52  
52  
51.1  
51
```

## Task\_9

```
df = vizierdb.get_data_frame('community_teacher_scores')
df=df[(df['average_teachers_score']<30)]
print(df)
```



Console ▾

Timing

Datasets ▾

Charts ▾

	Community_Area_Name	average_teachers_score
75	ROGERS PARK	29.333333
76	AVALON PARK	26.666667