CS 520 Data Integration Warehousing
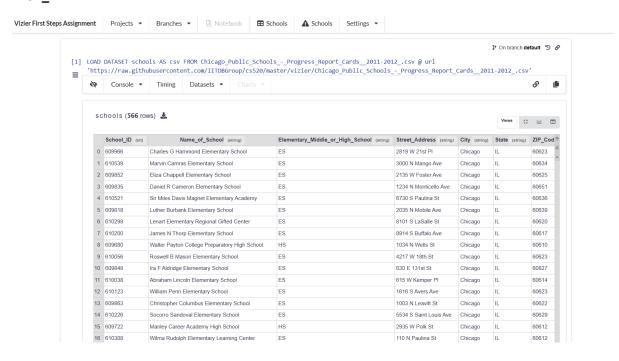
Group Members

Girish Rajani (A20503736)

Shriya Prasanna (A20521733)
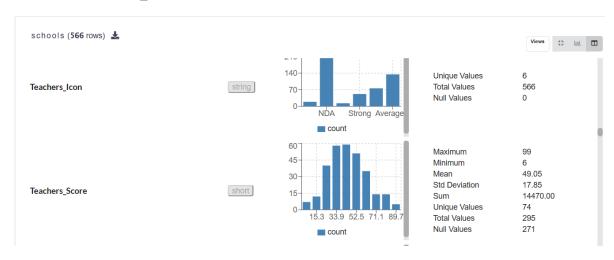
Subramanya Ganesh (A20516250)

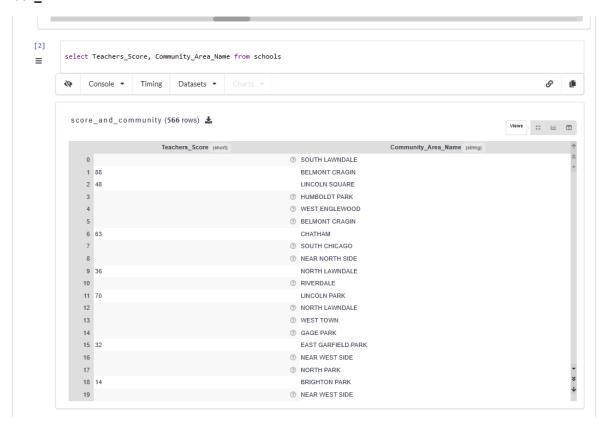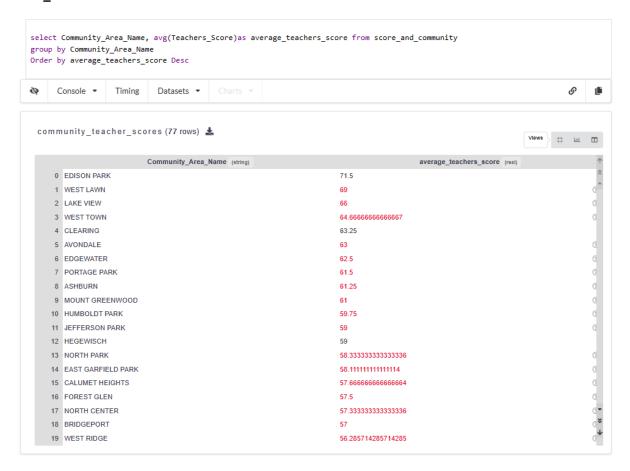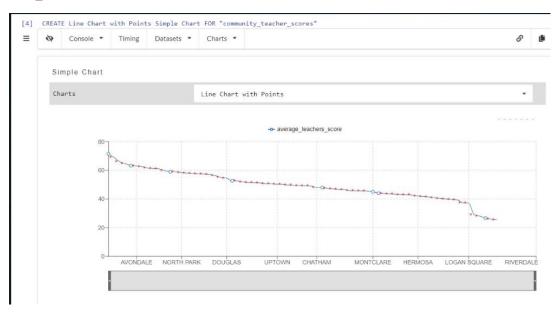## TASK_1

Task_2

Column View Teachers_Score

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		⊘ 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

```
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) ⬇

	Community_Area_Name (string)	average_teachers_score (real)
0	EDISON PARK	71.5
1	WEST LAWN	69
2	LAKE VIEW	66
3	WEST TOWN	64.666666666666667
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

```
[4]  CREATE Line Chart with Points Simple Chart FOR "community_teacher_scores"
```

| ≡ | 👁 | Console ▾ | Timing | Datasets ▾ | Charts ▾ | | 🔗 | 📋 |

Simple Chart

| Charts | Line Chart with Points ▾ |

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 ▾

▬▬ average_teachers_score

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
```