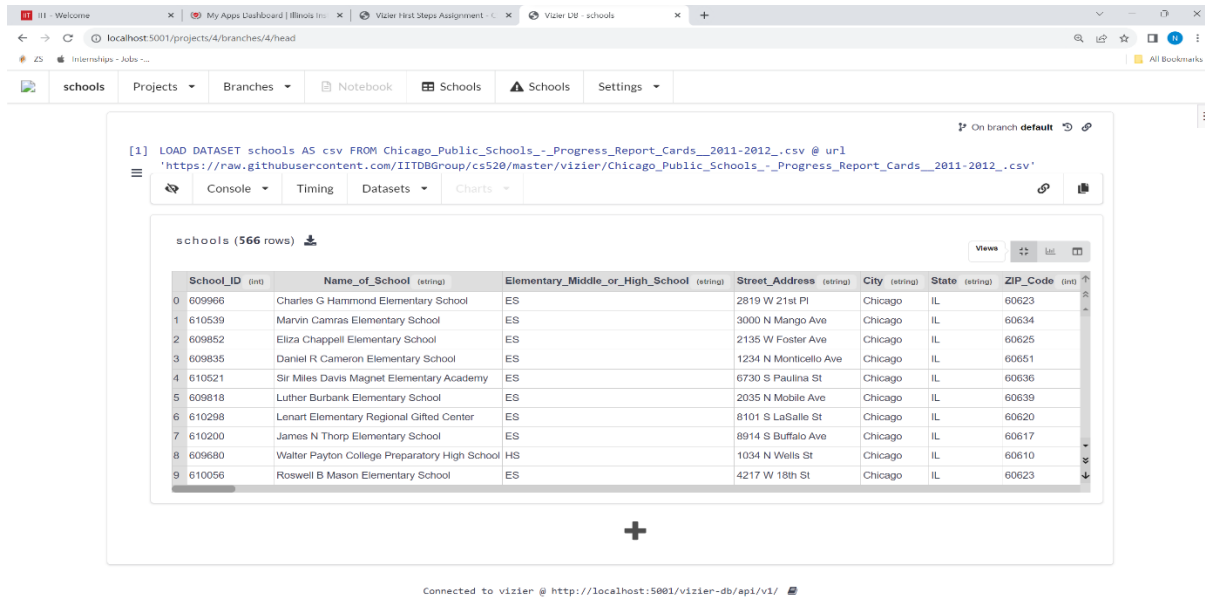


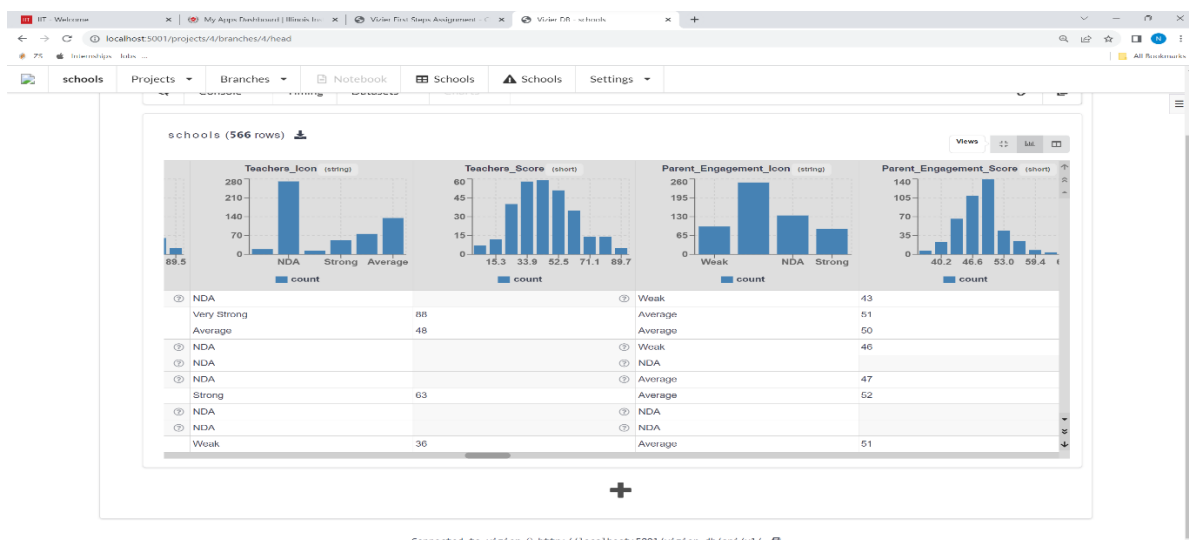
Vizier First Steps Assignment

Group 21
Nithin. R
Lalitha. V
Samantha. D

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



- 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.

The screenshot shows the Vizier web application interface. At the top, there's a navigation bar with tabs for 'schools', 'Projects', 'Branches', 'Notebook', 'Schools', and 'Settings'. Below this is a table with columns: 'Leaders_Score', 'Teachers_Icon', 'Teachers_Score', 'Parent_Engagement_Icon', 'Parent_Engagement_Score', and 'Parent_Environment_Icon'. A 'Cell Annotations' dialog box is open over the table, showing a comment: 'Could not cast 'NDA' to ShortType (in schools)'. The dialog has a close button (X) in the top right corner.

- **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 web application interface. At the top, there's a navigation bar with tabs for 'schools', 'Projects', 'Branches', 'Notebook', 'Dataset', 'Caveats', and 'Settings'. Below this is a SQL cell with the query: `SELECT Teachers_Score, Community_Area_Name FROM schools`. The result is displayed as a table with 16 rows, showing Teachers_Score and Community_Area_Name. The table is titled 'temporary_dataset (566 rows)'.

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

- **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 VIZIR interface with a SQL query in cell [3] and its result in a temporary dataset.

SQL Query:

```
SELECT Community_Area_Name, AVG(Teachers_Score) AS Average_Teacher_Score
FROM score_and_community
GROUP BY Community_Area_Name;
```

temporary_dataset (77 rows)

	Community_Area_Name (string)	Average_Teacher_Score (real)
0	BRIGHTON PARK	50.2
1	LINCOLN PARK	51.8
2	MONTCLARE	45
3	HERMOSA	42
4	LOOP	
5	BELMONT CRAGIN	45.5
6	ROSELAND	45.714285714285715
7	OAKLAND	
8	WASHINGTON PARK	43.5
9	GREATER GRAND CROSSING	40.333333333333336
10	AUBURN GRESHAM	49.57142857142857
11	SOUTH SHORE	52.6
12	SOUTH LAWNDALE	49.15384615384615
13	GARFIELD RIDGE	50.666666666666664
14	LINCOLN SQUARE	43.333333333333336

- **Task 6:** Create a line chart of the aggregation result by creating a plot cell and take a screenshot of the result.

The screenshot shows the VIZIR interface with a lens creation step and a line chart visualization.

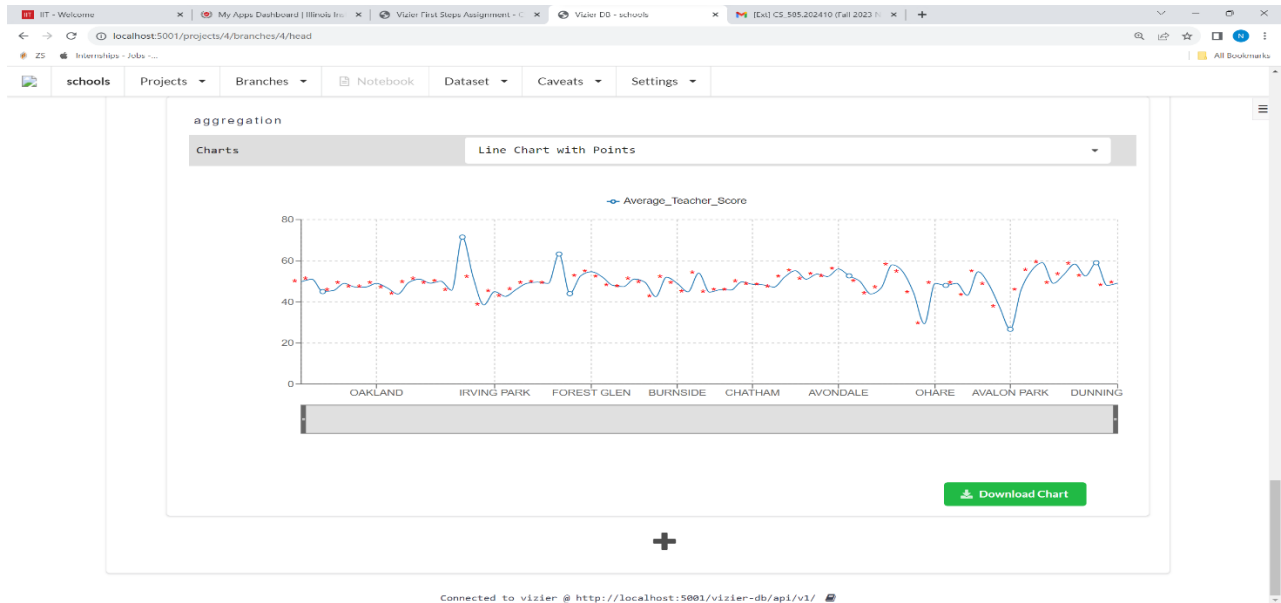
SQL Query:

```
SELECT Community_Area_Name, AVG(Teachers_Score) AS Average_Teacher_Score
FROM score_and_community
GROUP BY Community_Area_Name;
```

community_teacher_scores (77 rows)

	Community_Area_Name (string)	Average_Teacher_Score (real)
0	BRIGHTON PARK	49.75
1	LINCOLN PARK	51
2	MONTCLARE	45
3	HERMOSA	45.5
4	LOOP	
5	BELMONT CRAGIN	45.5
6	ROSELAND	45.714285714285715
7	OAKLAND	
8	WASHINGTON PARK	43.5
9	GREATER GRAND CROSSING	40.333333333333336
10	AUBURN GRESHAM	49.57142857142857
11	SOUTH SHORE	52.6
12	SOUTH LAWNDALE	49.15384615384615
13	GARFIELD RIDGE	50.666666666666664
14	LINCOLN SQUARE	43.333333333333336

- **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.



- **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 documention.* 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.

The screenshot shows the Vizier web interface with a Python code cell. The code defines a function `print_avg_teachers` that uses the Vizier API to fetch and print the 'Average_Teacher_Score' for each row in the 'community_teacher_scores' dataset. The output shows a table with 77 rows, displaying the 'Community_Area_Name' and the corresponding 'Average_Teacher_Score'.

	Community_Area_Name (string)	Average_Teacher_Score (real)
0	BRIGHTON PARK	49.75
1	LINCOLN PARK	51
2	MONTCLAIRE	45
3	HERMOSA	45.5
4	LOOP	49
5	BELMONT CRAGIN	47.25
6	ROSELAND	47.23076923076923
7	OAKLAND	49
8	WASHINGTON PARK	46.8
9	GREATER GRAND CROSSING	43.8

- **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.

The screenshot shows the Vizier web interface with a sidebar on the left containing navigation tabs: `schools`, `Projects`, `Branches`, `Notebook`, `Dataset`, `Caveats`, and `Settings`. The `schools` tab is active, displaying a table of school data:

Branches	Average Teacher Score
2 MONTCLARE	45
3 HERMOSA	45.5
4 LOOP	49
5 BELMONT CRAGIN	47.25
6 ROSELAND	47.23076923076923
7 OAKLAND	49
8 WASHINGTON PARK	46.8
9 GREATER GRAND CROSSING	43.8

Below the table is a Jupyter notebook cell with the following code:

```
[7]  
df = vizierdb.get_data_frame("community_teacher_scores")  
filtered_df = df[df["Average_Teacher_Score"] < 30.0]  
print(filtered_df)
```

Below the code cell is a console output showing the result of the filter query:

Community_Area_Name	Average_Teacher_Score
58 ROGERS PARK	29.333333
66 AVALON PARK	26.666667

At the bottom of the interface, a status bar indicates: `Connected to vizier @ http://localhost:5001/vizier-db/api/v1/`