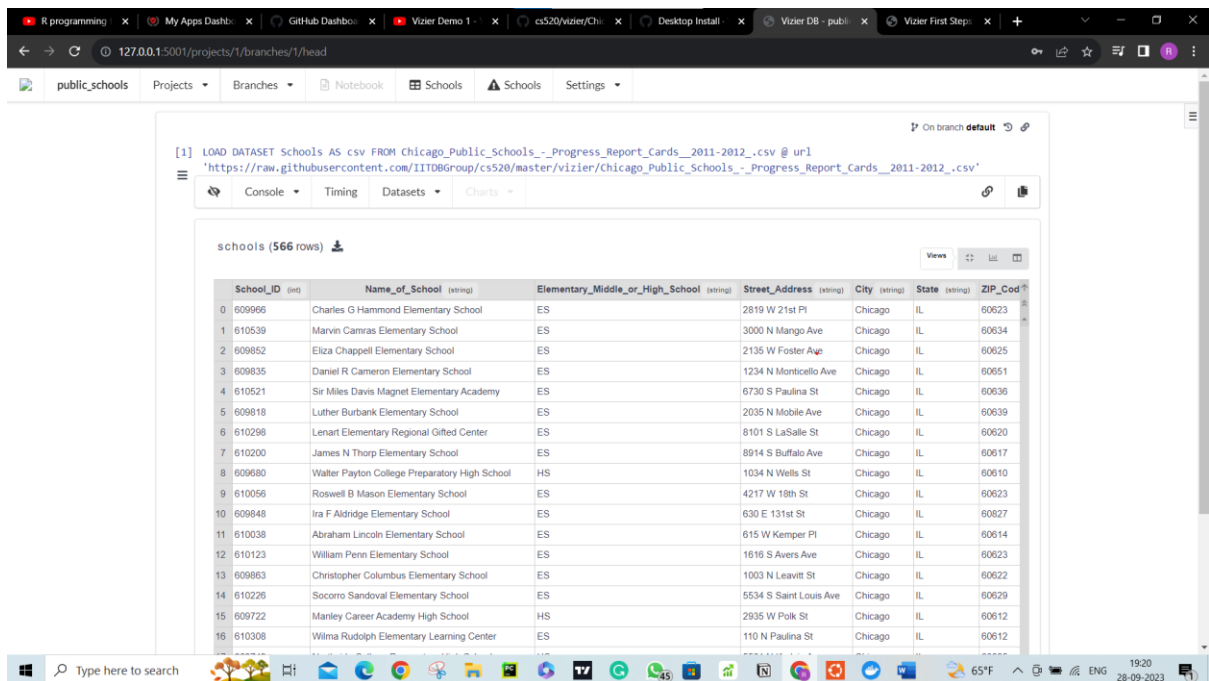


VIZIER FIRST STEPS ASSIGNMENT

DATA INTEGRATION, WAREHOUSING, AND PROVENANCE

CS-520-01 (FALL 2023) TEAM-16

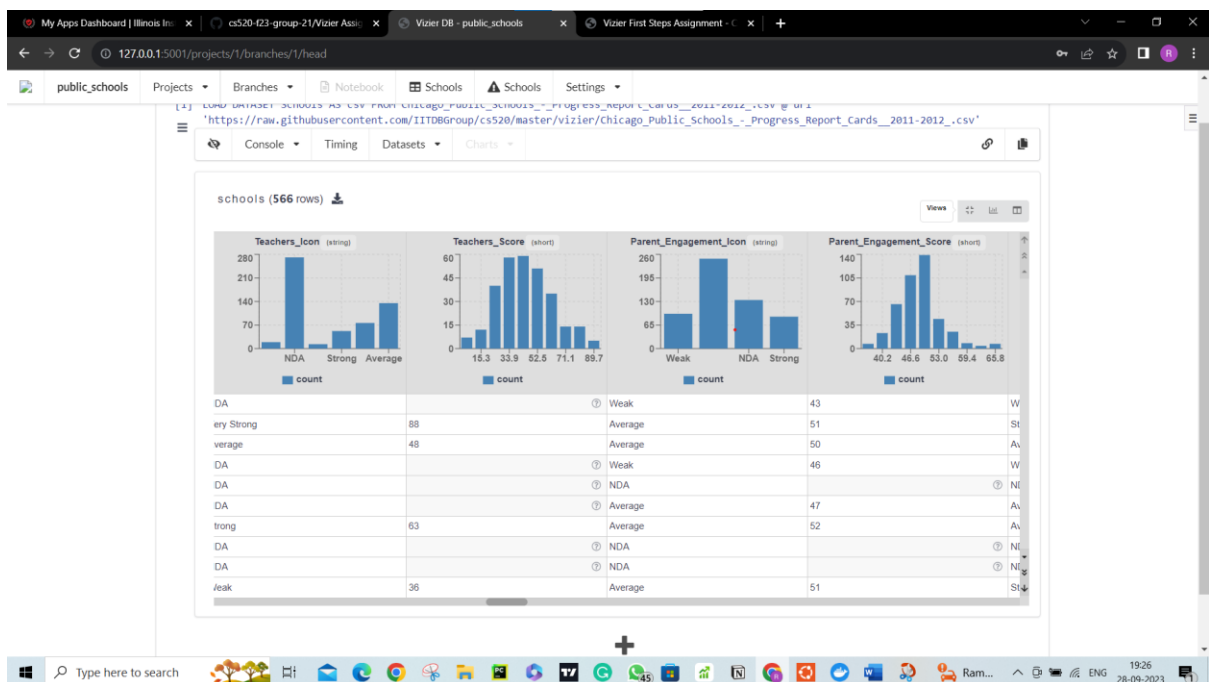
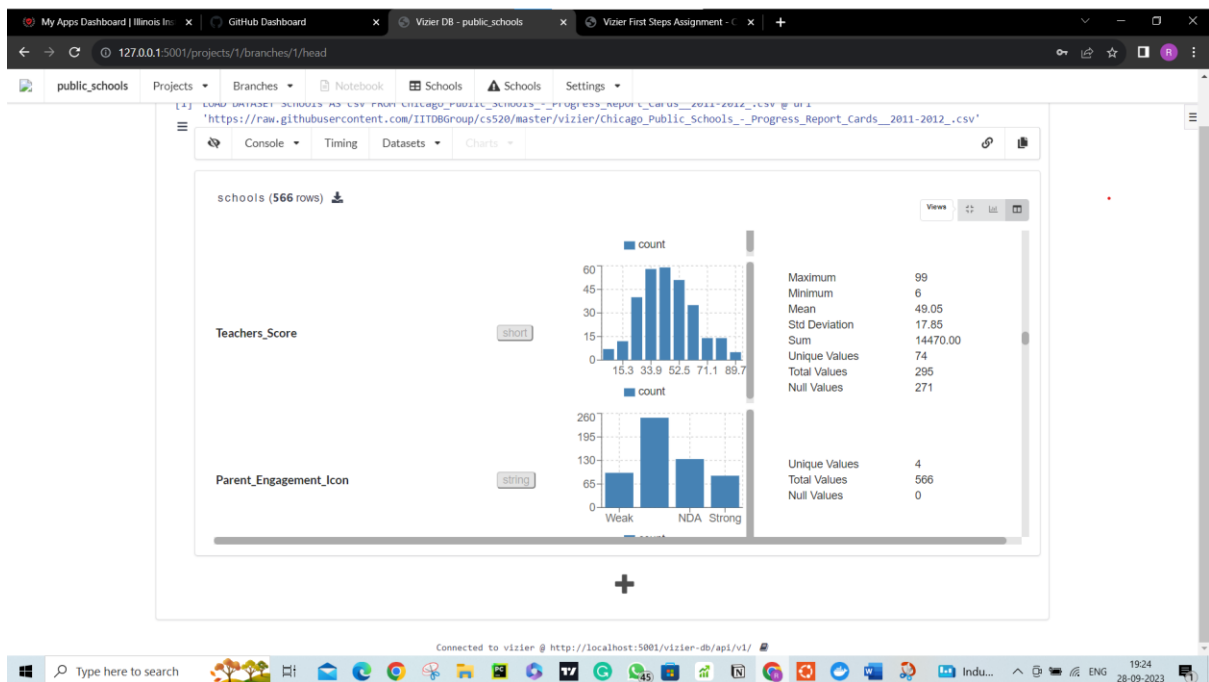
TASK1: Load a dataset and take a screenshot of the result



The screenshot shows a web browser window with a Vizio dashboard. The dashboard has a top navigation bar with tabs for 'public_schools', 'Projects', 'Branches', 'Notebook', 'Schools', and 'Settings'. The main content area displays a dataset titled 'schools (566 rows)'. The dataset is shown as a table with the following columns: School_ID, Name_of_School, Elementary_Middle_or_High_School, Street_Address, City, State, and ZIP_Cod. The table contains 17 rows of data, showing various schools in Chicago, including Charles G Hammond Elementary School, Marvin Camras Elementary School, Eliza Chappell Elementary School, Daniel R Cameron Elementary School, Sir Miles Davis Magnet Elementary Academy, Luther Burbank Elementary School, Lenart Elementary Regional Gifted Center, James N Thorp Elementary School, Walter Payton College Preparatory High School, Roswell B Mason Elementary School, Ira F Aldridge Elementary School, Abraham Lincoln Elementary School, William Penn Elementary School, Christopher Columbus Elementary School, Socorro Sandoval Elementary School, Manley Career Academy High School, and Wilma Rudolph Elementary Learning Center.

School_ID	Name_of_School	Elementary_Middle_or_High_School	Street_Address	City	State	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

TASK2: 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.



TASK3: Click on one of the question marks for values in the teachers column and take a screenshot.

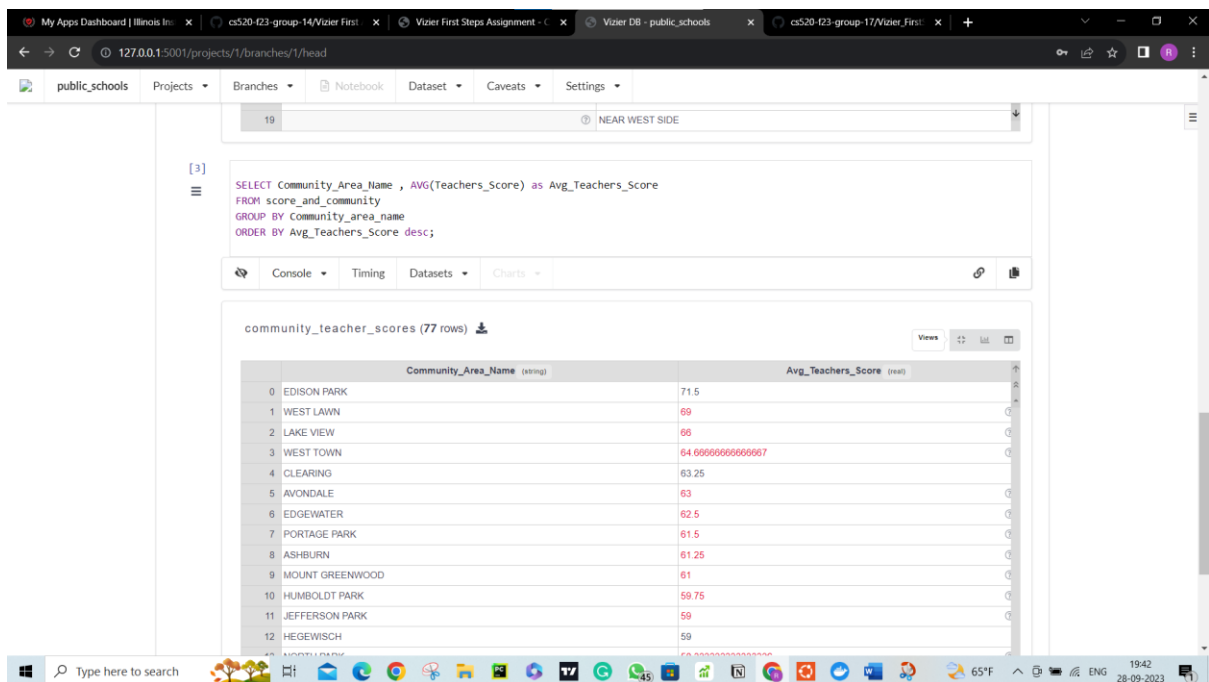
The screenshot shows the Vizier web application interface. At the top, there are browser tabs for 'My Apps Dashboard | Illinois In...', 'cs520-f23-group-21/Vizier Assign...', 'Vizier DB - public_schools', and 'Vizier First Steps Assignment'. The address bar shows '127.0.0.1:5001/projects/1/branches/1/head'. The main content area displays a dataset named 'schools' with 566 rows. The columns are: 'id', 'Leaders_Score', 'Teachers_Score', 'Parent_Engagement_Score', and 'Community_Area_Name'. A 'Cell Annotations' dialog box is open over the 'Teachers_Score' column, showing a comment: 'Could not cast 'NDA' to ShortType (in schools)'. The Windows taskbar at the bottom shows the time as 19:29 on 28-09-2023.

TASK4: 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.

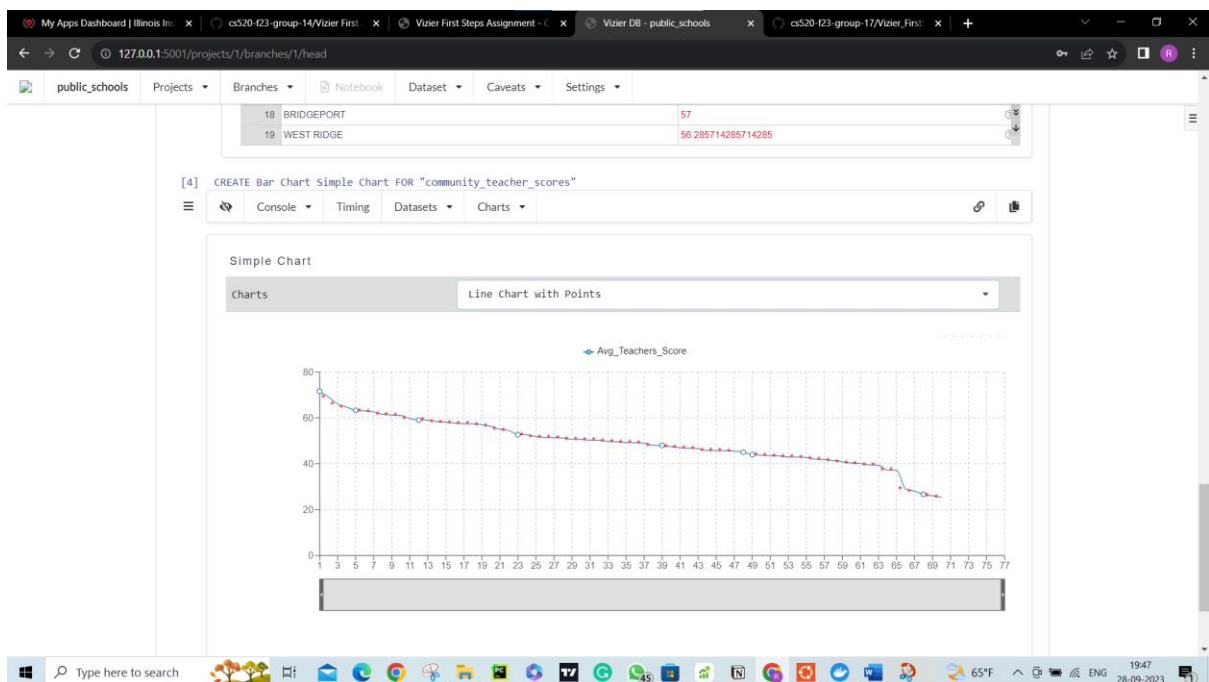
The screenshot shows the Vizier web application interface with a SQL query executed. The query is: `SELECT Teachers_Score, Community_Area_Name FROM schools`. The result is a new dataset named 'score_and_community' with 566 rows. The columns are 'Teachers_Score (short)' and 'Community_Area_Name (string)'. The data is displayed in a table with 17 rows visible. The Windows taskbar at the bottom shows the time as 19:37 on 28-09-2023.

	Teachers_Score (short)	Community_Area_Name (string)
0		SOUTH LAWDALE
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 LAWDALE
10		RIVERDALE
11	70	LINCOLN PARK
12		NORTH LAWDALE
13		WEST TOWN
14		GAGE PARK
15	32	EAST GARFIELD PARK
16		NEAR WEST SIDE
17		NORTH PARK

TASK5: 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.



TASK6: Create a line chart of the aggregation result by creating a plot cell and take a screenshot of the result.



TASK7: 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.



TASK8: 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.

```
[5] CREATE Bar Chart Simple Chart FOR "community_teacher_scores"

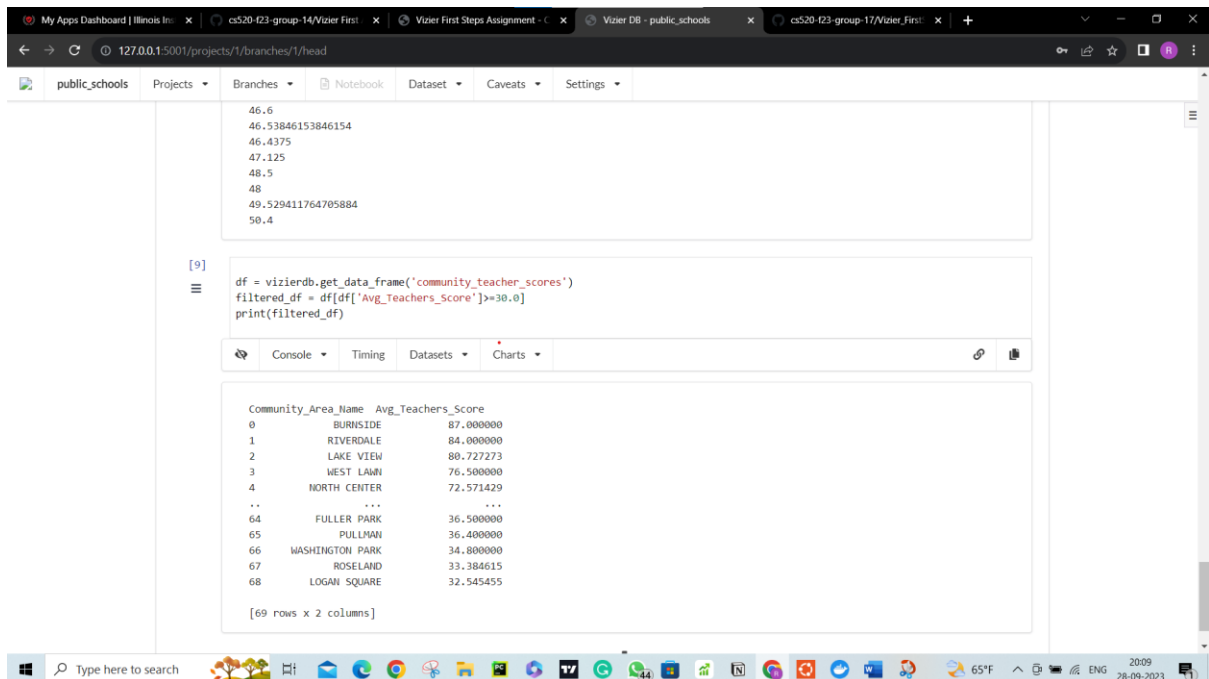
[6] def print_avg_teachers():
    ds = vizierdb.get_dataset("community_teacher_scores")
    AverageScores=set()
    for row in ds.rows:
        AverageScores.add(row.get_value('Avg_Teachers_Score'))
    for score in AverageScores:
        print(score)
    vizierdb.export_module(print_avg_teachers)

[8] print_avg_teachers()
```

The screenshot shows the Vizier interface with two code cells. The first cell, labeled [5], contains a SQL query to create a bar chart. The second cell, labeled [6], contains a Python function `print_avg_teachers()` that retrieves the 'community_teacher_scores' dataset, calculates the average of the 'Avg_Teachers_Score' column, and prints the result. The third cell, labeled [8], calls the `print_avg_teachers()` function. The output of the function is displayed in the console area below the code cells, showing a list of average scores.

```
51.5
51
52
6
16
19.142857142857142
26.666666666666668
27.75
28.5
29
26
32.54545454545455
33.38461538461539
```

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 a Jupyter Notebook interface with the following components:

- Browser Tabs:** My Apps Dashboard | Illinois In..., cs520-f23-group-14/Vizier First..., Vizier First Steps Assignment - ..., Vizier DB - public_schools, cs520-f23-group-17/Vizier First...
- Address Bar:** 127.0.0.1:5001/projects/1/branches/1/head
- Navigation Bar:** public_schools, Projects, Branches, Notebook, Dataset, Caveats, Settings
- Code Cell:**

```
[9]  
df = vizierdb.get_data_frame('community_teacher_scores')  
filtered_df = df[df['Avg_Teachers_Score'] >= 30.0]  
print(filtered_df)
```
- Output:**

```
46.6  
46.53846153846154  
46.4375  
47.125  
48.5  
48  
49.529411764705884  
50.4
```

Community_Area_Name	Avg_Teachers_Score
0 BURNSIDE	87.000000
1 RIVERDALE	84.000000
2 LAKE VIEW	80.727273
3 WEST LAMN	76.500000
4 NORTH CENTER	72.571429
...	...
64 FULLER PARK	36.500000
65 PULLMAN	36.400000
66 WASHINGTON PARK	34.800000
67 ROSELAND	33.384615
68 LOGAN SQUARE	32.545455

[69 rows x 2 columns]
- Taskbar:** Windows taskbar with search bar, application icons, and system tray showing 65°F, 20:09, and 28-09-2023.