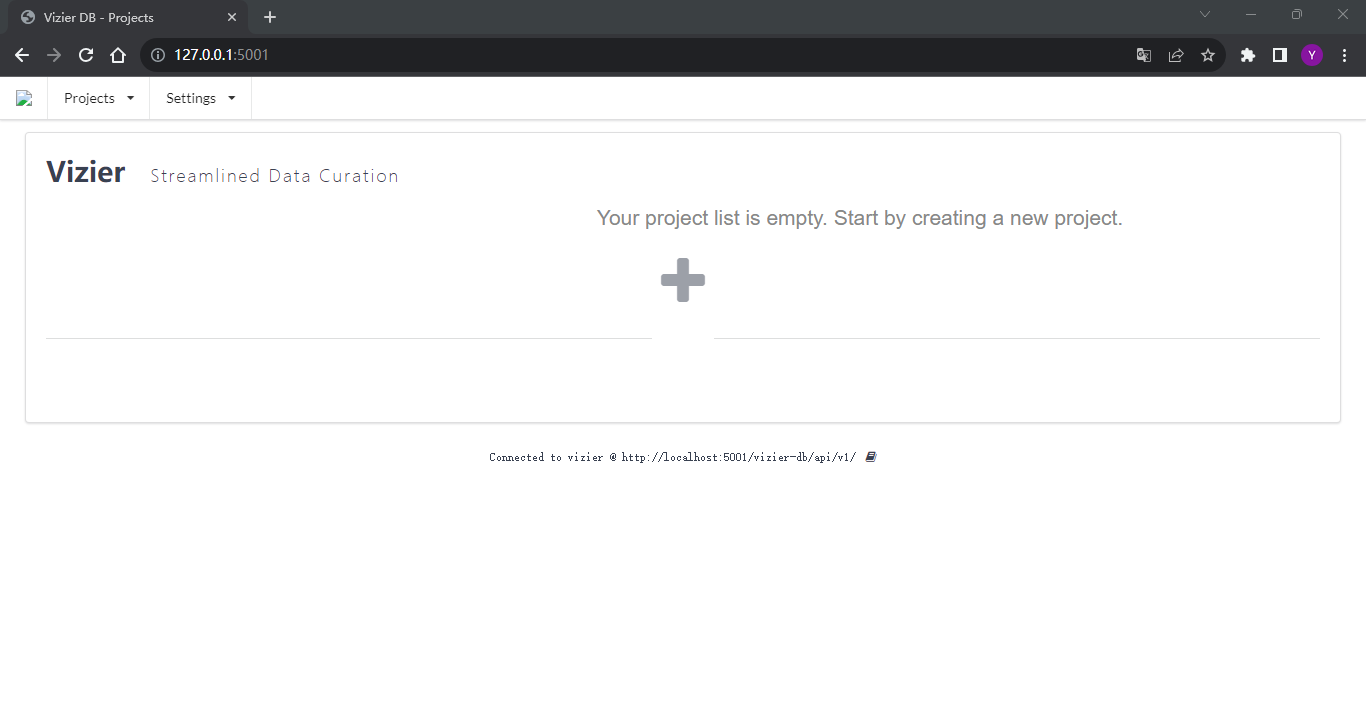
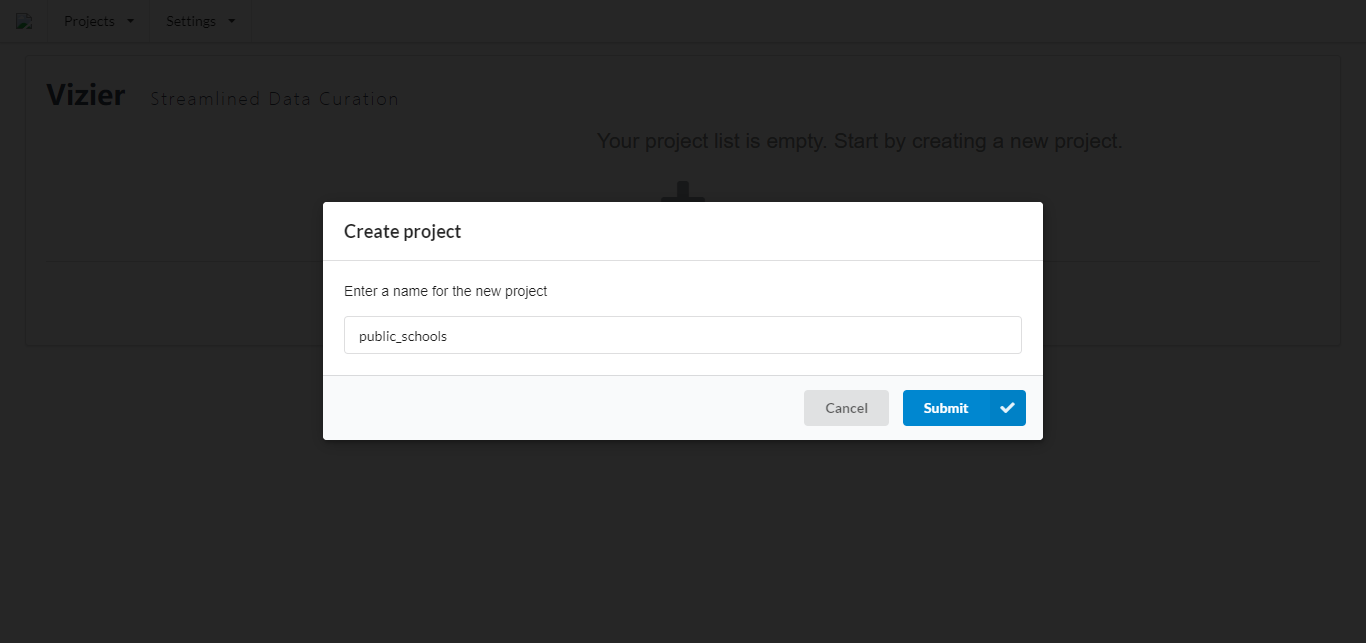
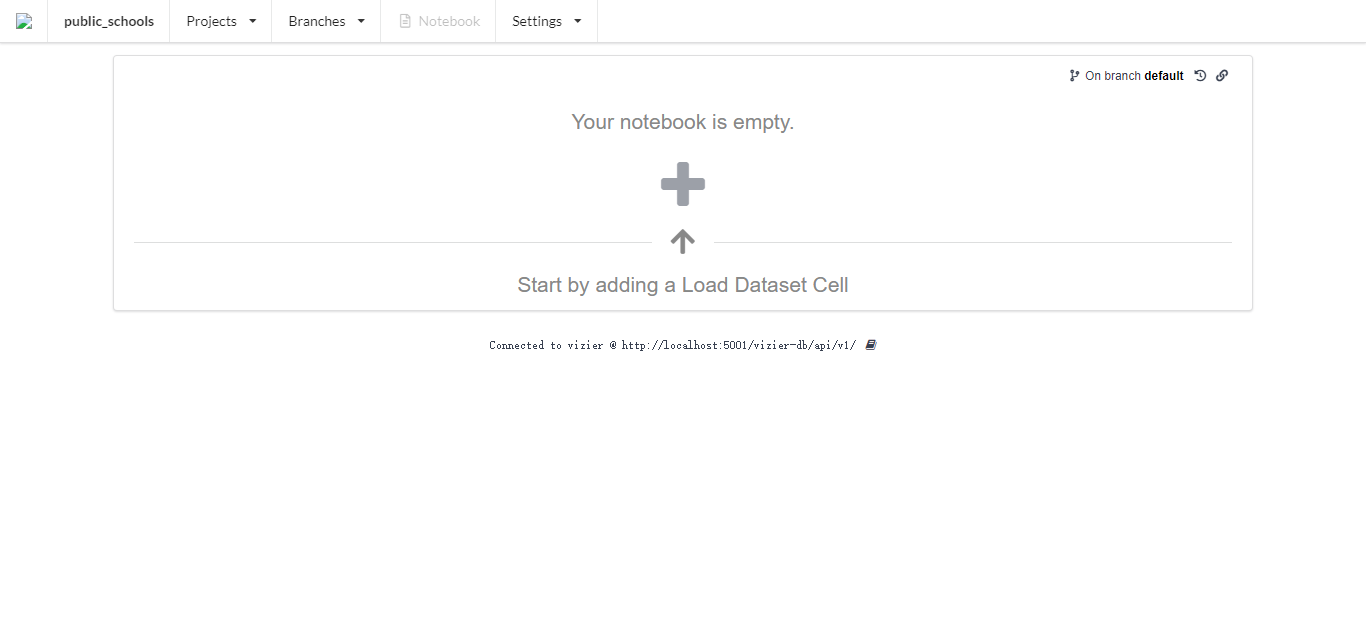
CS520 Project, Group 4

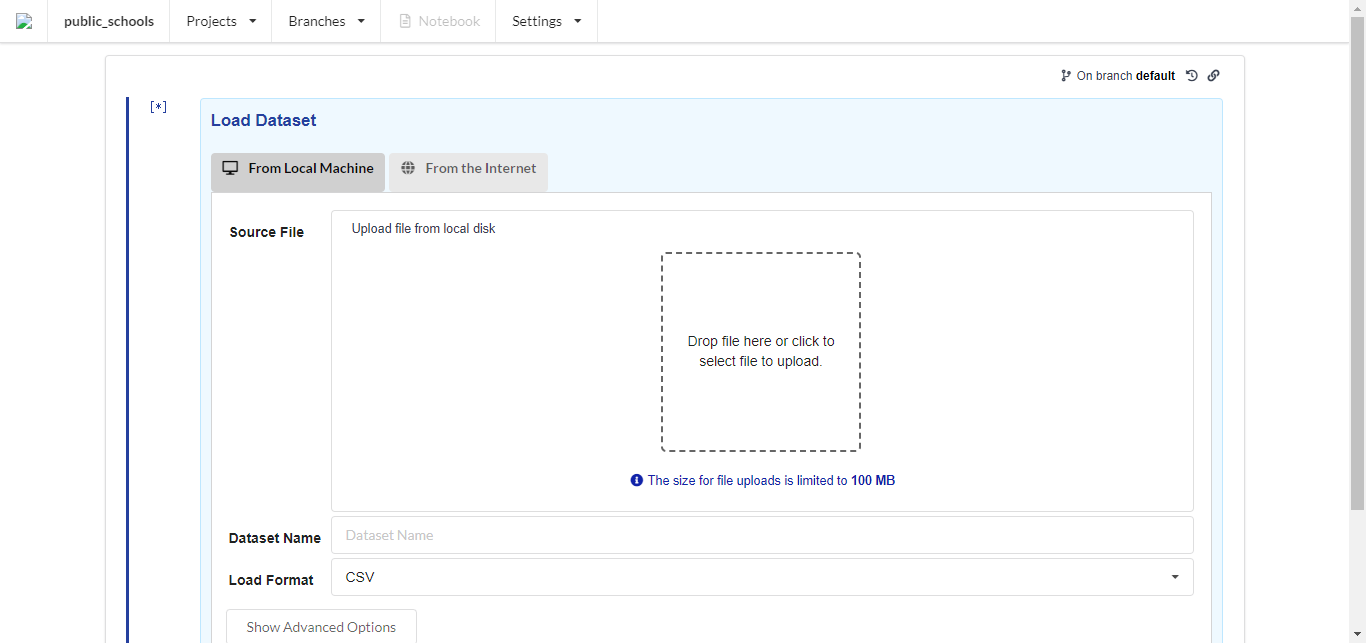
Fan Tianyin CWID: , Li Ziyin CWID: , Liu Yuhai CWID：A20498516

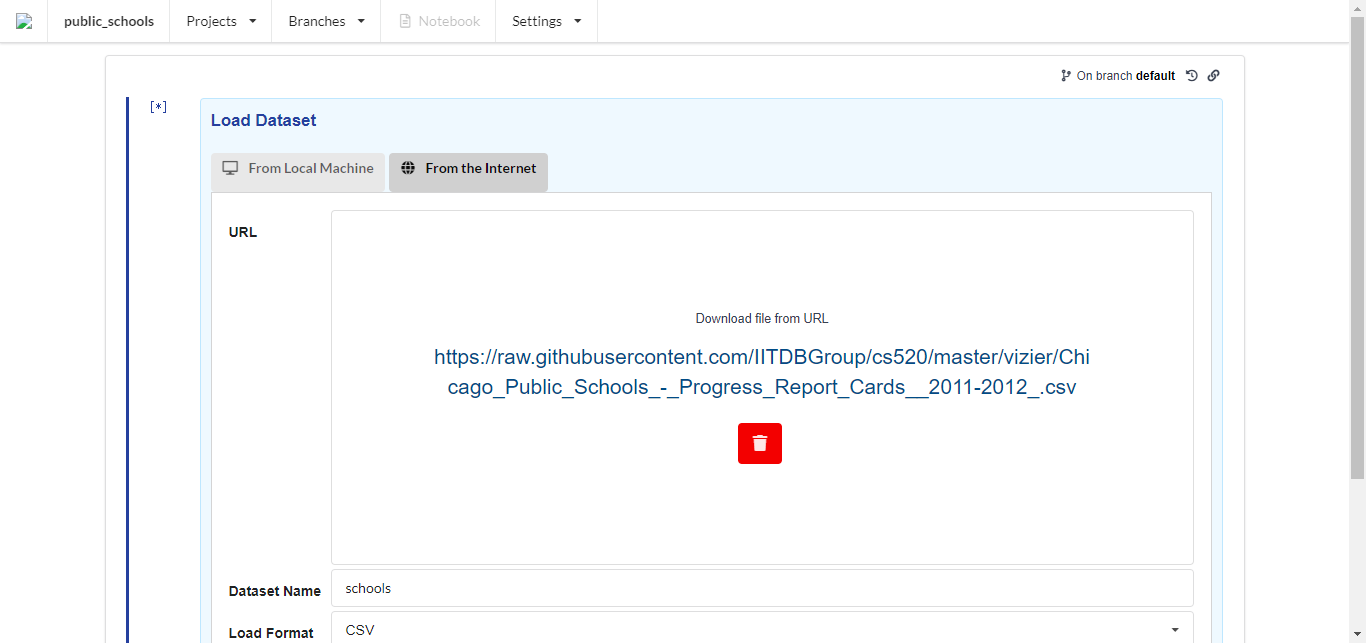
## **Load the public school dataset**



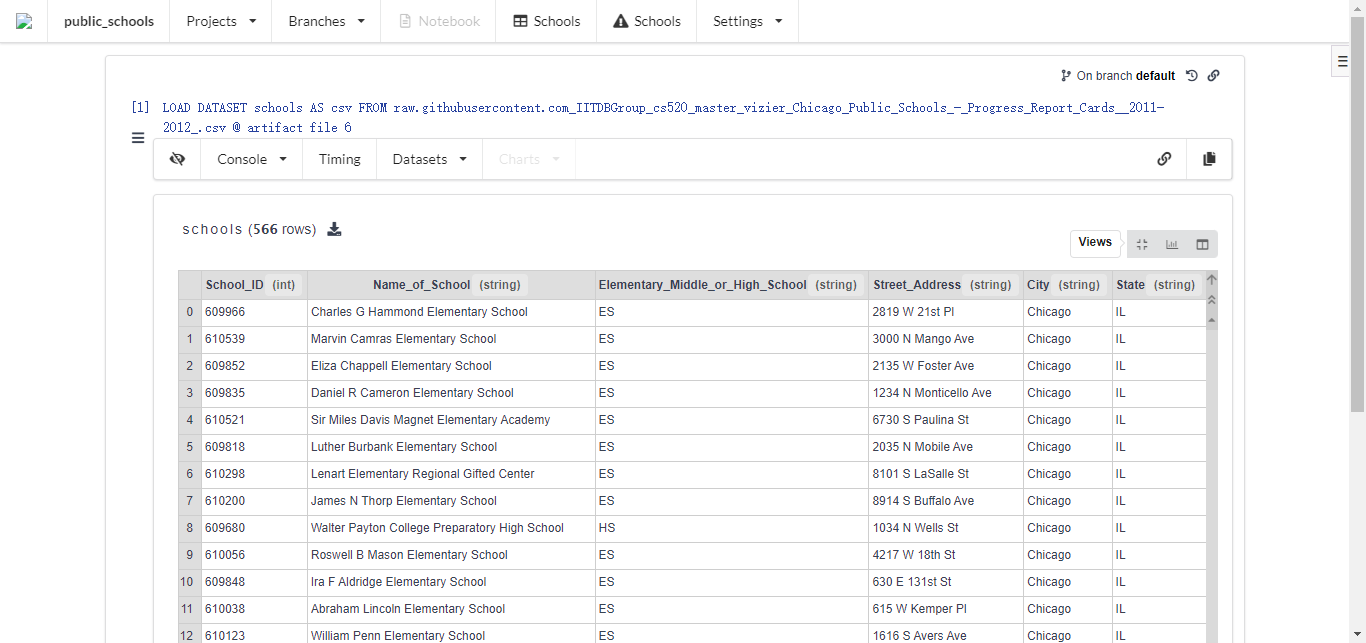








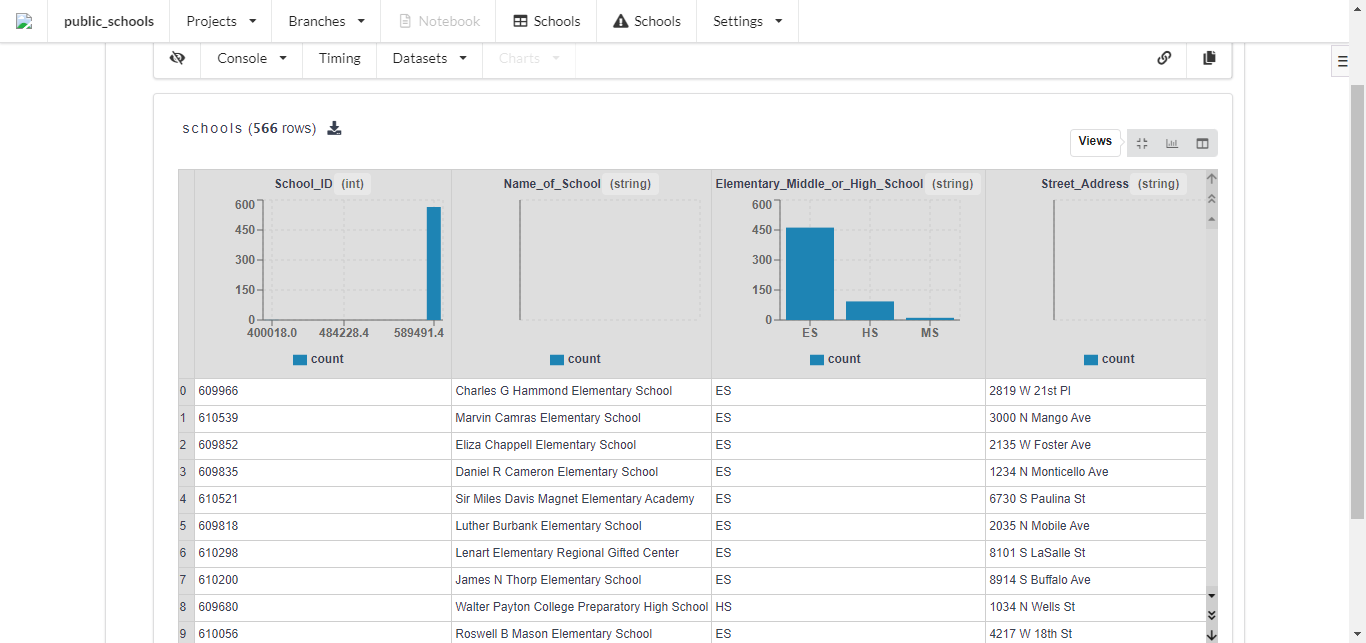
* ****Task 1****: load a dataset and take a screenshot of the result

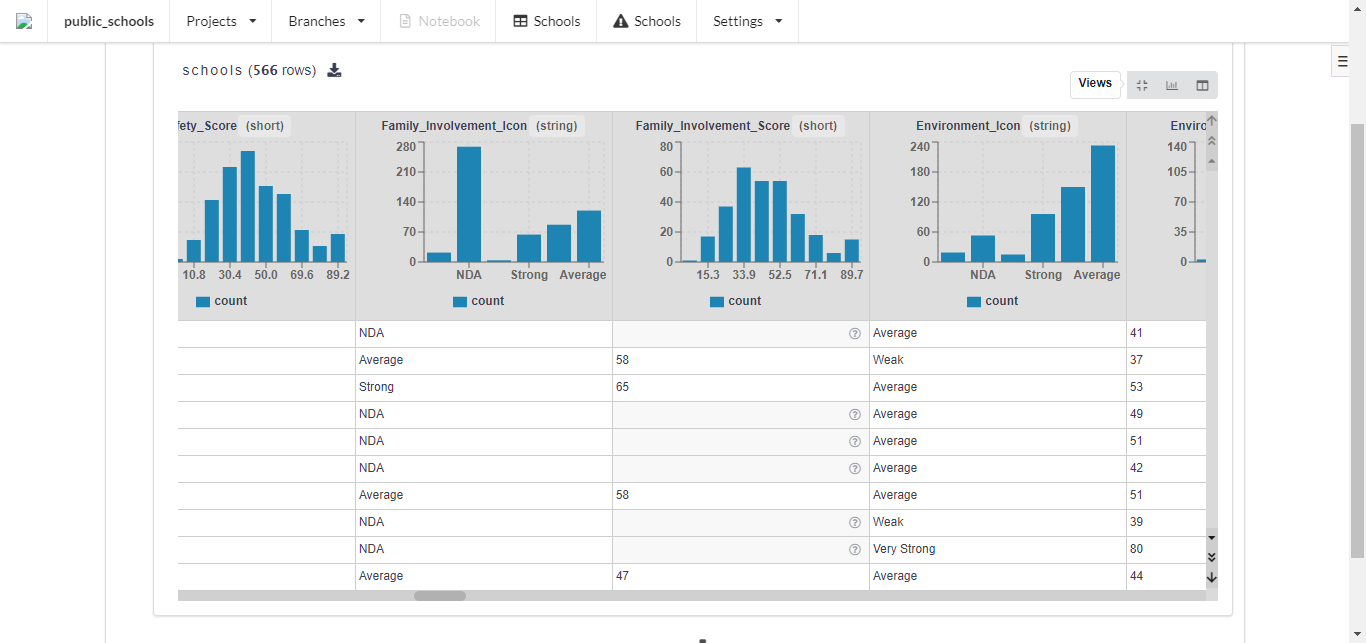


## **Explore the datasets and caveats**

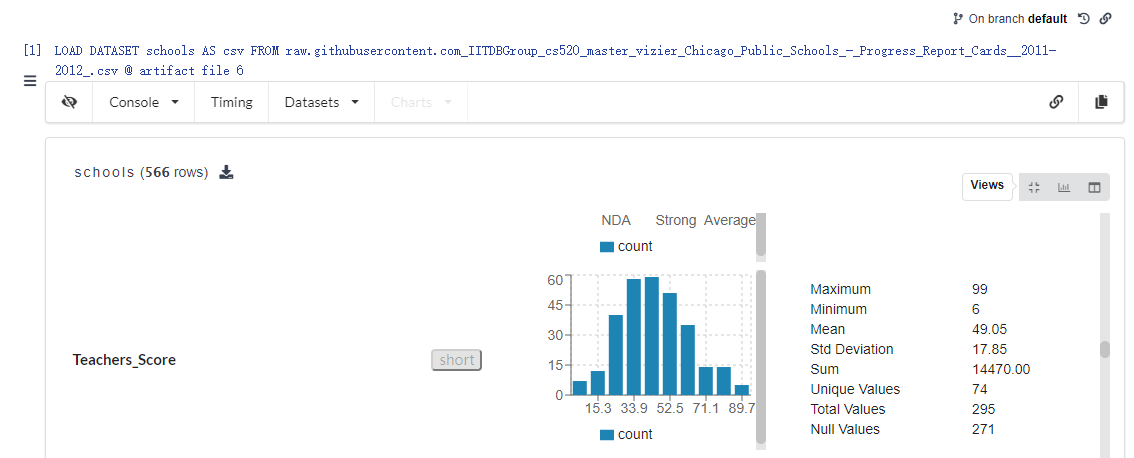
* ****Task 2****: Select the detail view and look at the distributions of some columns. Then look at the column view and take a screenshot of the distribution for column Teachers\_Score.

Detail view and the distributions of some columns.



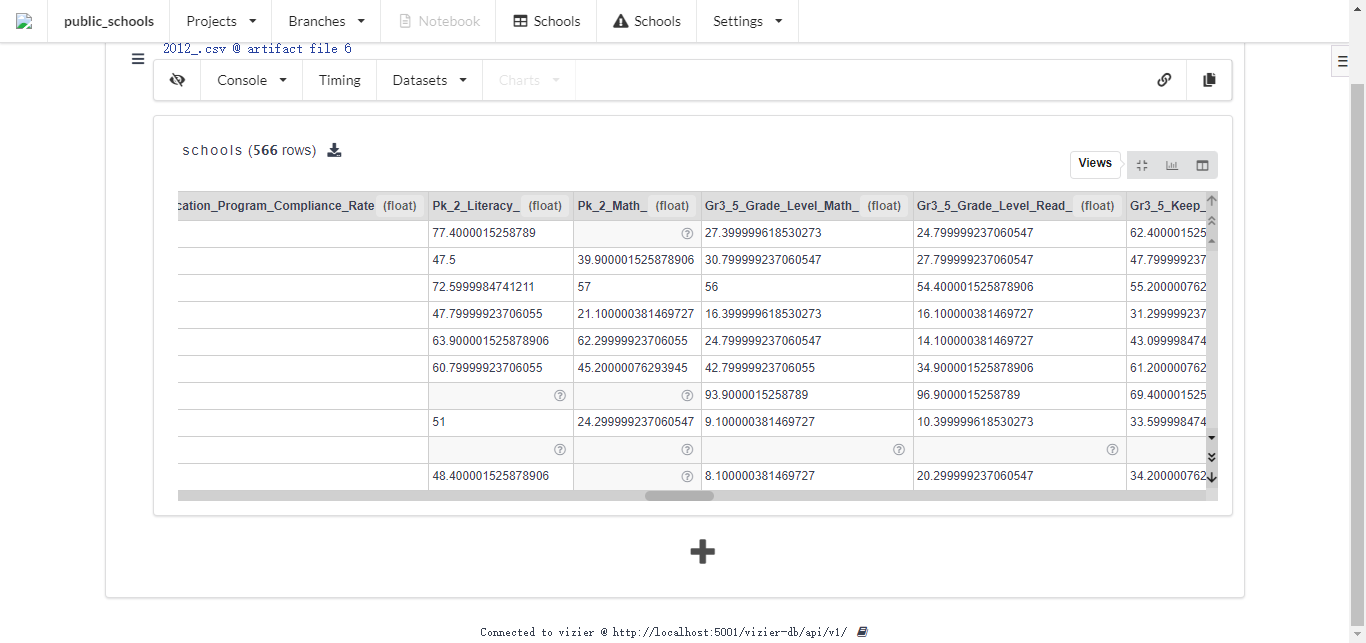


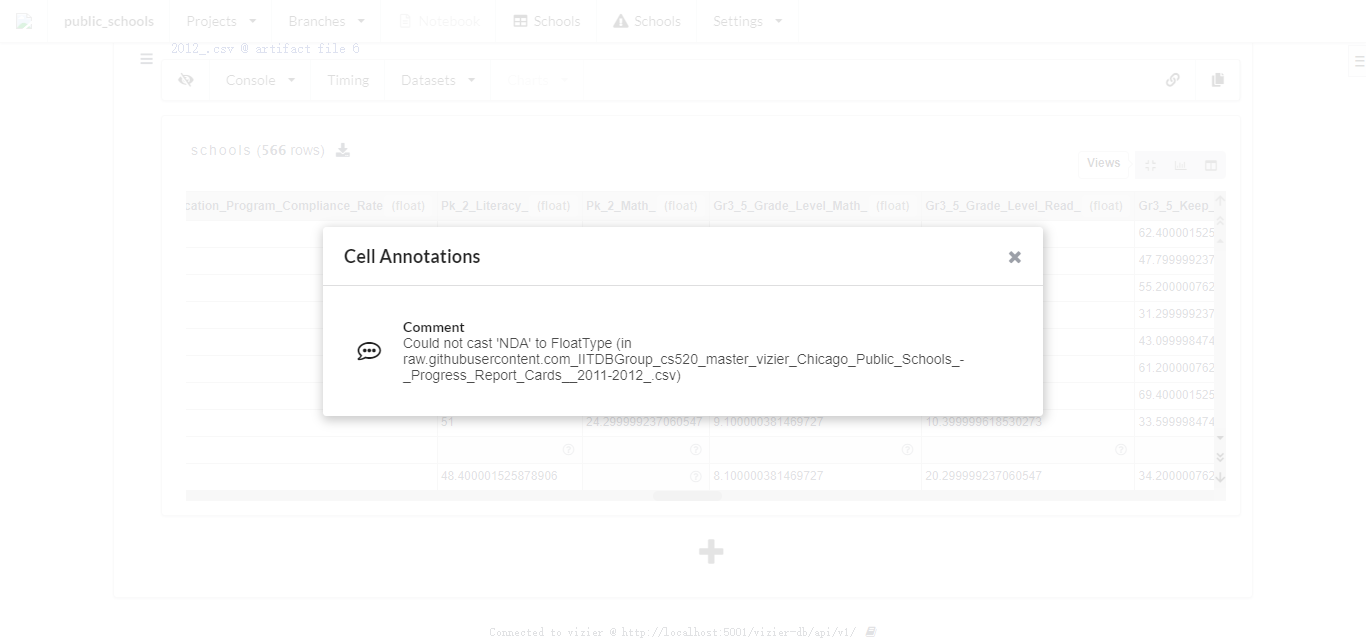
Columns view and distribution for ‘Teachers\_Score’ column.



### **Caveats**

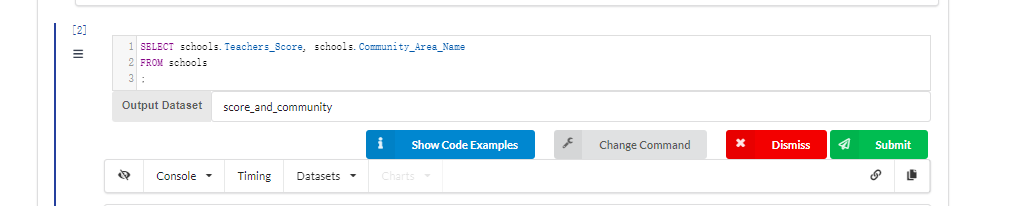
* ****Task 3****: Click on one of the question marks for values in the teachers column and take a screenshot.

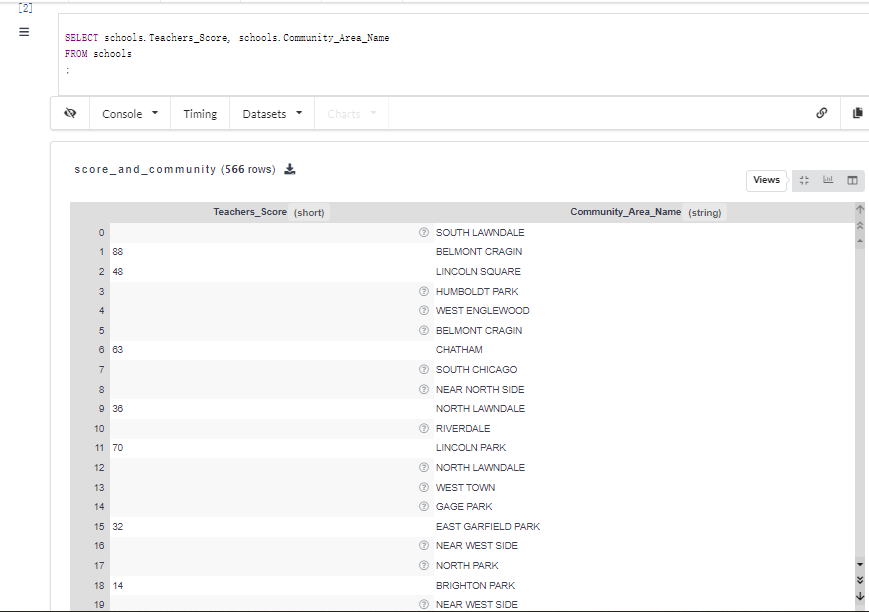




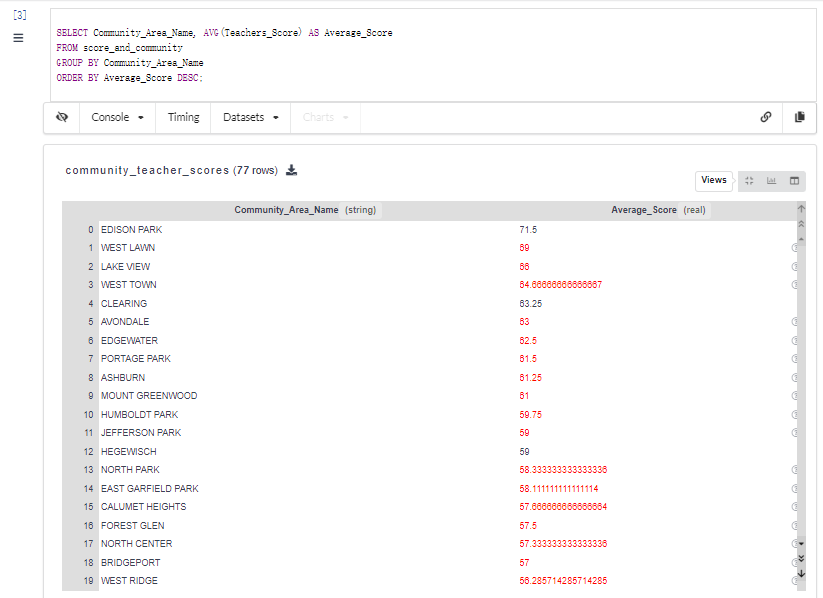
### **SQL cells and caveat propagation**

* ****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.





* ****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.



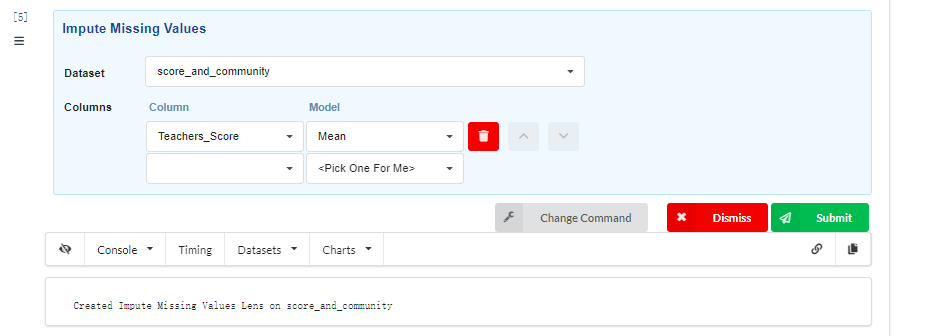
### **Plotting data**

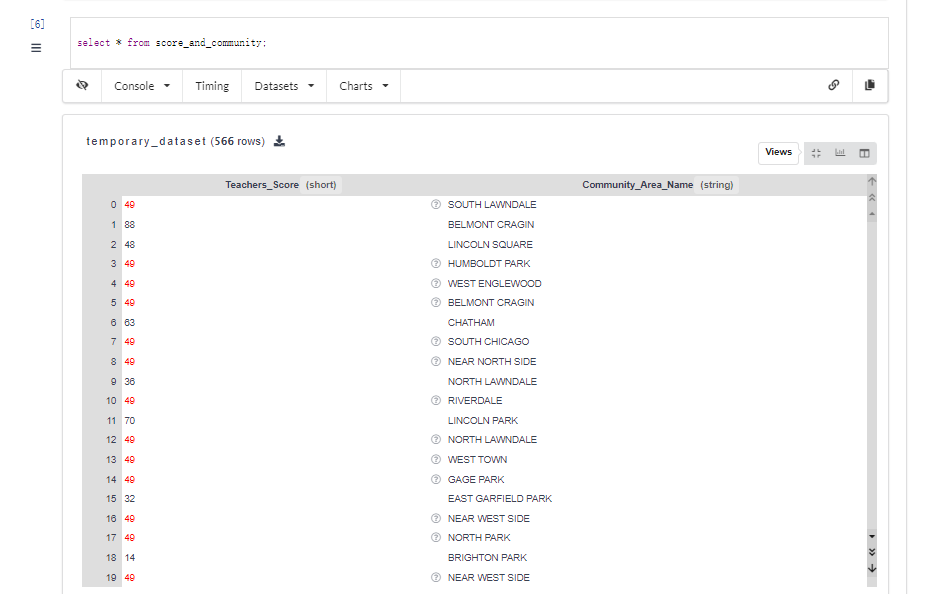
* ****Task 6****: Create a line chart of the aggregation result by creating a plot cell and take a screenshot of the result.



### **Lenses**

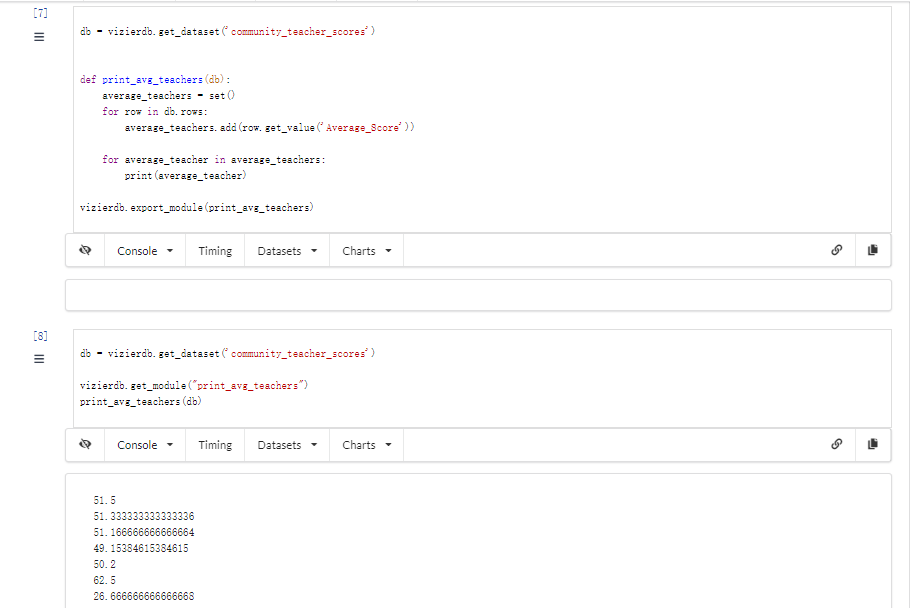
* ****Task 7****: Insert a new cell above the SQL cell that computes the average teacher scores (notebooks in Vizier are executed top down) by pressing the three bars below the cell number. Select "Impute Missing Values", select the score\_and\_community dataset and Teachers\_Score as the column to be imputed, and select mean as the imputation method and take a screenshot of the updated line chart.





### **Using Python**

* ****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" buttom to see example Vizier API usage and see [here](https://github.com/VizierDB/vizier-scala/wiki/Cell-Python) 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.



#### **Get Vizier dataset as a Pandas DataFrame**

* ****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.

