Assignment 01

Name : Kiran K. Patil ID:211070904

Sem : 06 Course : Machine Learning

**Kaggle :**

Kaggle is a platform where data scientists and machine learning developers can compete against each other to solve complex data science problems. Companies and organizations also use Kaggle to host competitions in order to crowdsource solutions to problems they are facing.

The platform provides a variety of tools for participants, including a cloud-based workbench for developing and running code, as well as a large dataset repository. Kaggle also hosts a community forum where participants can ask questions, share ideas and collaborate on projects. Additionally, Kaggle offers a number of educational resources, including tutorials, courses, and a blog to help users learn data science and machine learning. It is also one of the most popular platform for data scientist to showcase their skills and find job opportunities



**Pandas :**

Pandas is a open-source library for the Python programming language for data manipulation and analysis. It provides data structures such as dataframes and series that allow for easy manipulation and analysis of large datasets. Pandas is built on top of the NumPy library and is widely used in the data science community for its powerful data manipulation capabilities. Some of the main features of Pandas include.

Pandas is a powerful library with a wide range of functions for data manipulation and analysis. Here is a list of some of the most commonly used functions in Pandas, along with a brief description of what they do:

Pandas Functions :

* read\_csv(): reads a CSV file and returns a dataframe
* to\_csv(): writes a dataframe to a CSV file
* head(): returns the first n rows of a dataframe (default is 5)
* tail(): returns the last n rows of a dataframe (default is 5)
* shape: returns the dimensions of a dataframe (rows, columns)
* info(): returns information about a dataframe, including the data types of each column and the number of non-null values
* describe(): returns basic statistics for each numeric column in a dataframe
* columns: returns the column labels of a dataframe
* index: returns the index (row labels) of a dataframe
* value\_counts(): returns the frequency counts for each unique value in a column
* sort\_values(): sorts a dataframe by one or more columns
* groupby(): groups a dataframe by one or more columns and applies a function to each group
* merge(): merges two dataframes on one or more columns
* pivot\_table(): creates a pivot table from a dataframe
* melt(): "melts" a dataframe and returns a new, reshaped dataframe
* stack(): "stacks" the columns of a dataframe and returns a new, reshaped dataframe
* unstack(): "unstacks" the rows of a dataframe and returns a new, reshaped dataframe
* crosstab(): creates a cross-tabulation (frequency table) of two or more factors

**Seaborn :**

Seaborn is a Python data visualization library based on Matplotlib. It is built on top of Matplotlib and allows for easy creation of beautiful, informative, and highly-customizable statistical graphics. Some of the main features of Seaborn include:

Here is a list of some commonly used functions in Seaborn:

* sns.lineplot(): creates a line plot
* sns.barplot(): creates a bar plot
* sns.scatterplot(): creates a scatter plot
* sns.histplot(): creates a histogram
* sns.boxplot(): creates a box plot
* sns.violinplot(): creates a violin plot
* sns.catplot(): creates a categorical plot
* sns.pairplot(): creates a pair plot
* sns.jointplot(): creates a joint plot
* sns.heatmap(): creates a heat map
* sns.regplot(): creates a regression plot
* sns.kdeplot(): creates a kernel density estimate plot
* sns.lmplot(): creates a scatter plot with linear regression line
* sns.countplot(): creates a bar plot of counts
* sns.despine(): removes the top and right spines from the plot
* sns.set\_style(): sets the background theme of the plot
* sns.set\_context(): sets the context of the plot (paper, notebook, talk, poster)

**Matplotlib :**

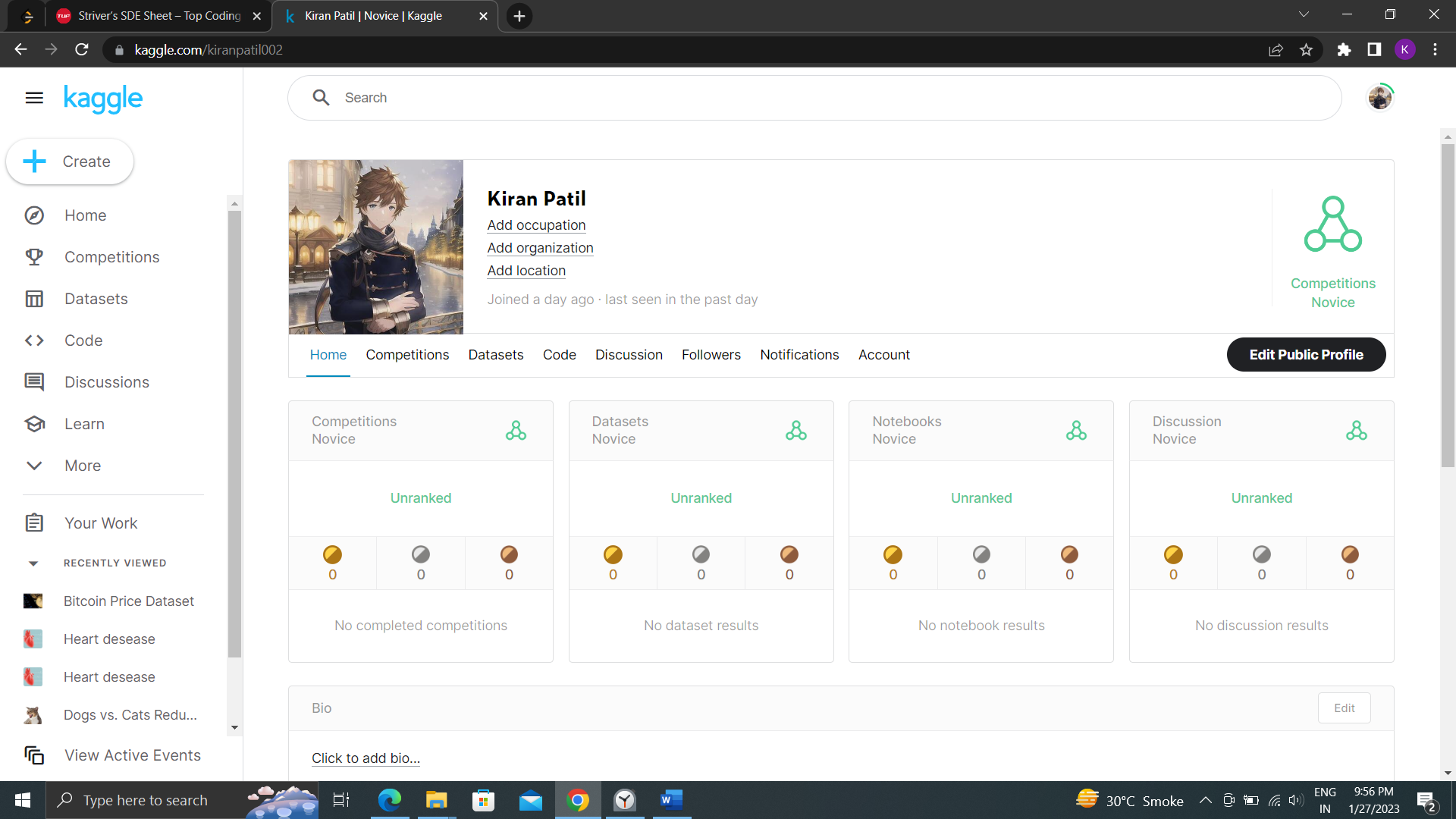
**Matplotlib** is a plotting library for the Python programming language and its numerical mathematics extension NumPy

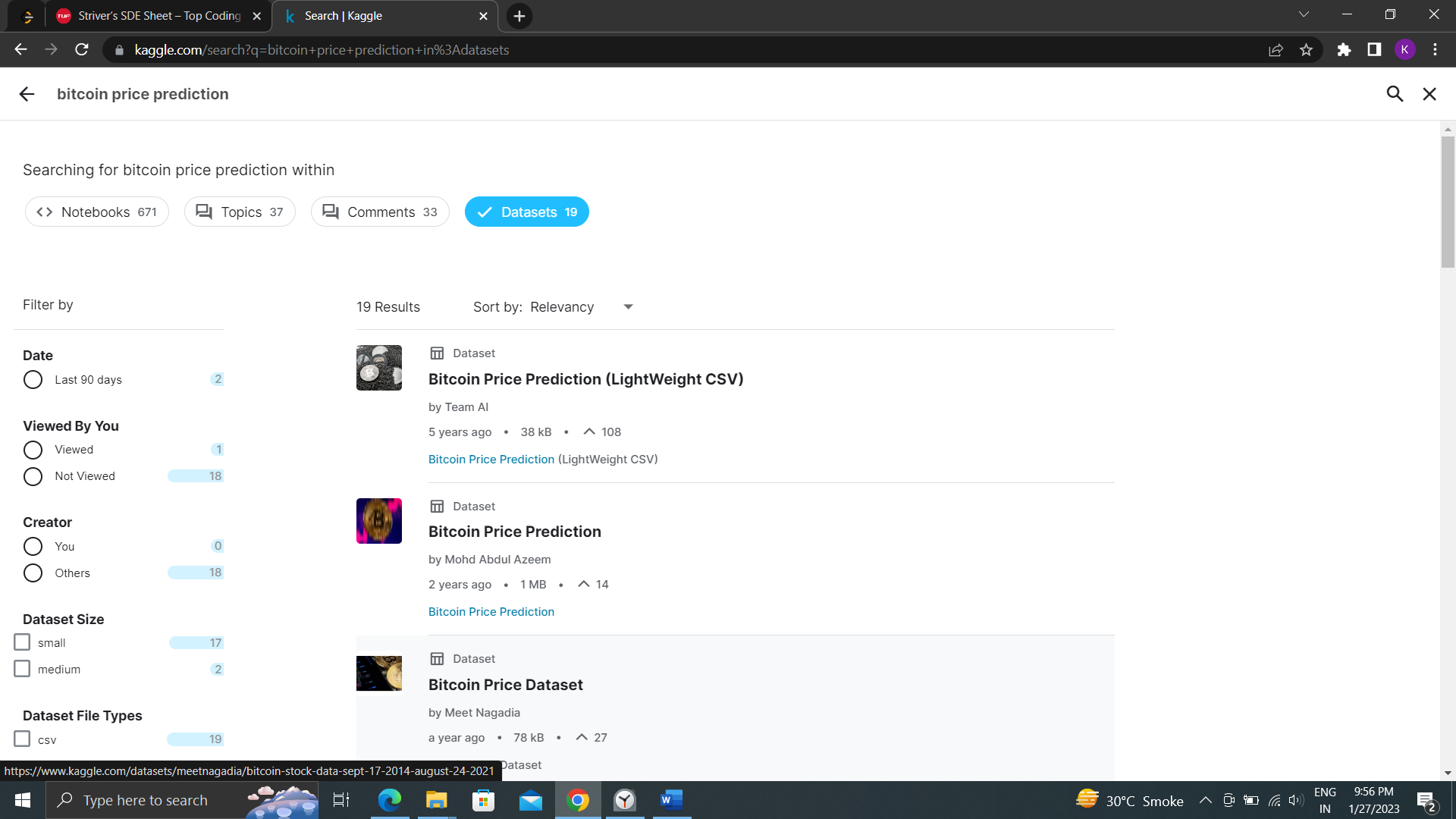
Matplotlib is a powerful data visualization library that provides a wide variety of customizable plots, but it can be complex to use for creating more advanced visualizations. Seaborn is built on top of Matplotlib and provides a simpler, more convenient interface for creating many common types of plots, especially those used in statistical data visualization.

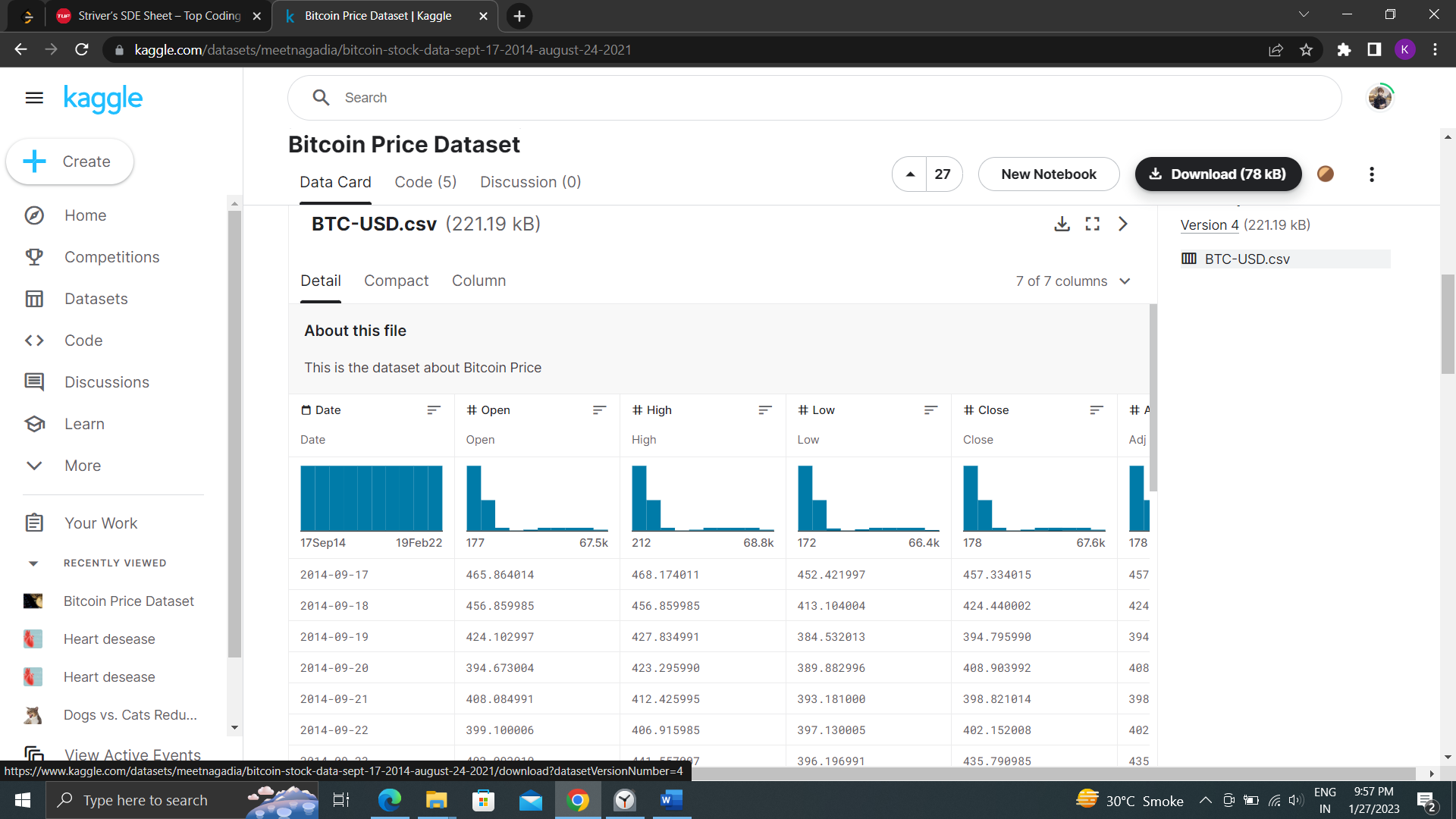
Here is a list of some commonly used functions in Matplotlib:

* plt.plot(): creates a line plot
* plt.scatter(): creates a scatter plot
* plt.bar(): creates a bar plot
* plt.hist(): creates a histogram
* plt.pie(): creates a pie chart
* plt.boxplot(): creates a box plot
* plt.stem(): creates a stem plot
* plt.quiver(): creates a vector field plot
* plt.semilogx(): creates a semilogarithmic plot of x-axis
* plt.semilogy(): creates a semilogarithmic plot of y-axis
* plt.loglog(): creates a log-log plot
* plt.polar(): creates a polar plot
* plt.contour(): creates a contour plot
* plt.imshow(): displays an image
* plt.colorbar(): adds a colorbar to a plot
* plt.xlabel(): adds a label to the x-axis
* plt.ylabel(): adds a label to the y-axis
* plt.title(): adds a title to the plot
* plt.legend(): adds a legend to the plot
* plt.show(): displays the plot

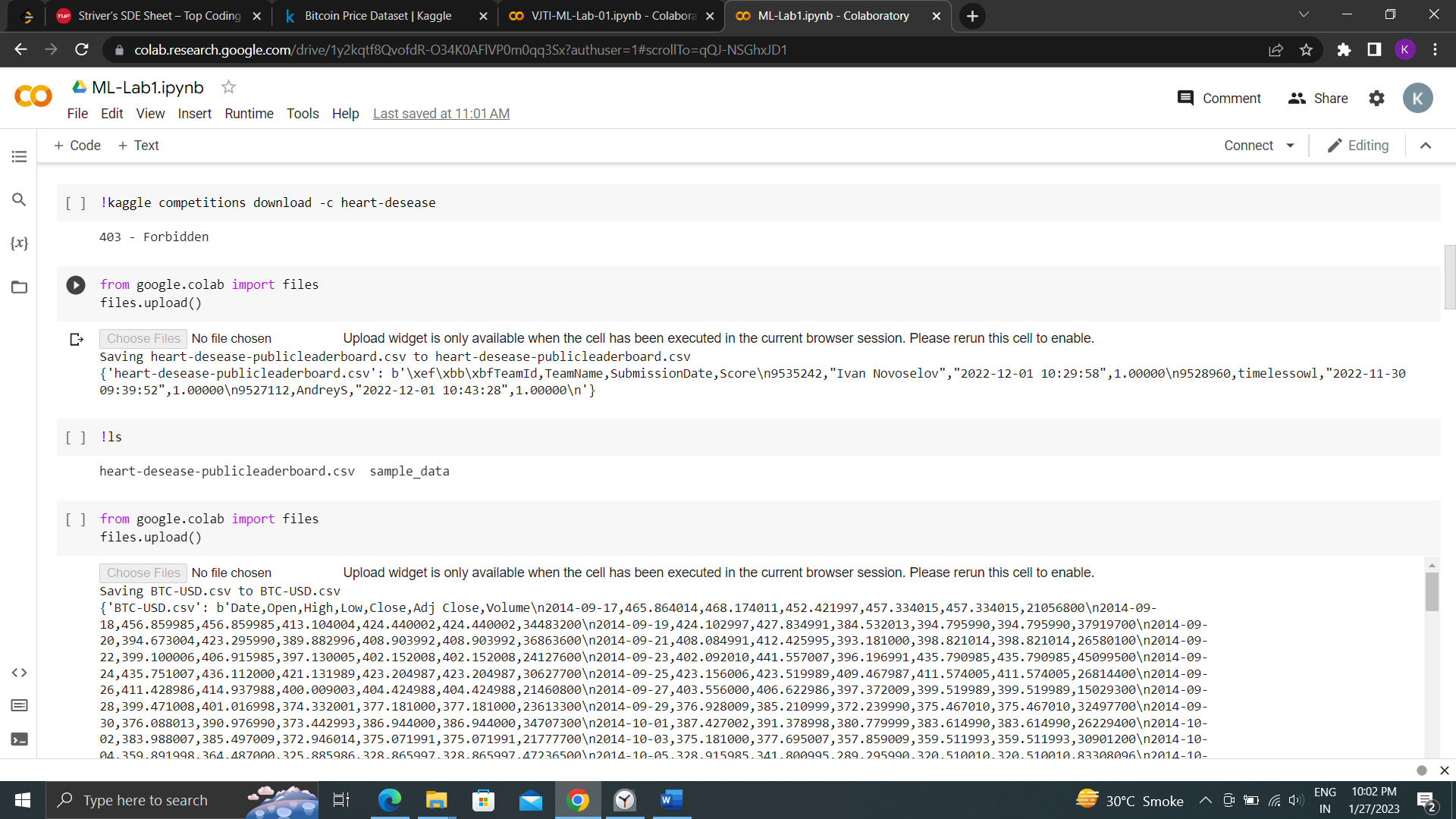
Setting up an account and download the dataset from Kaggle :

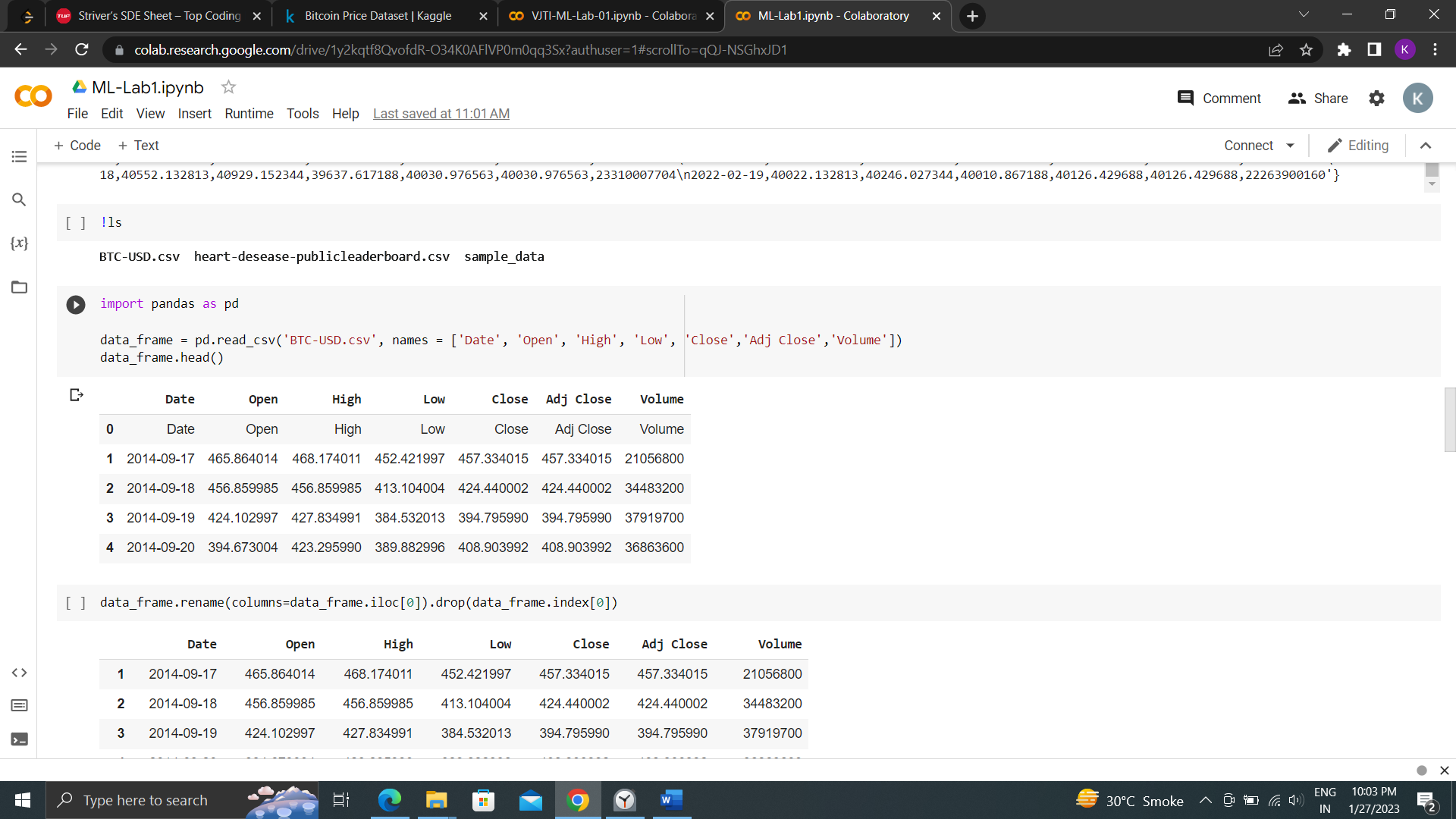


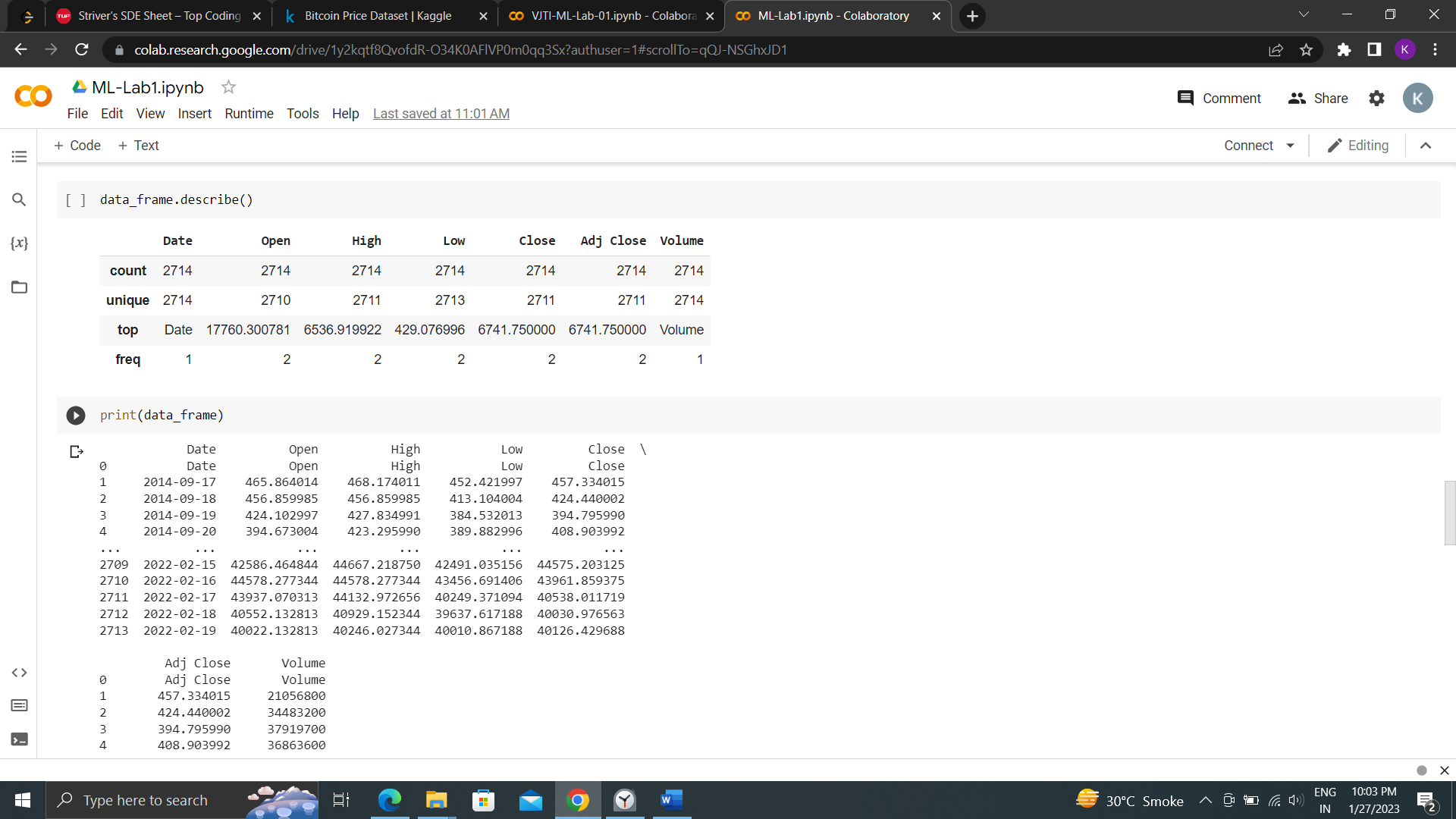


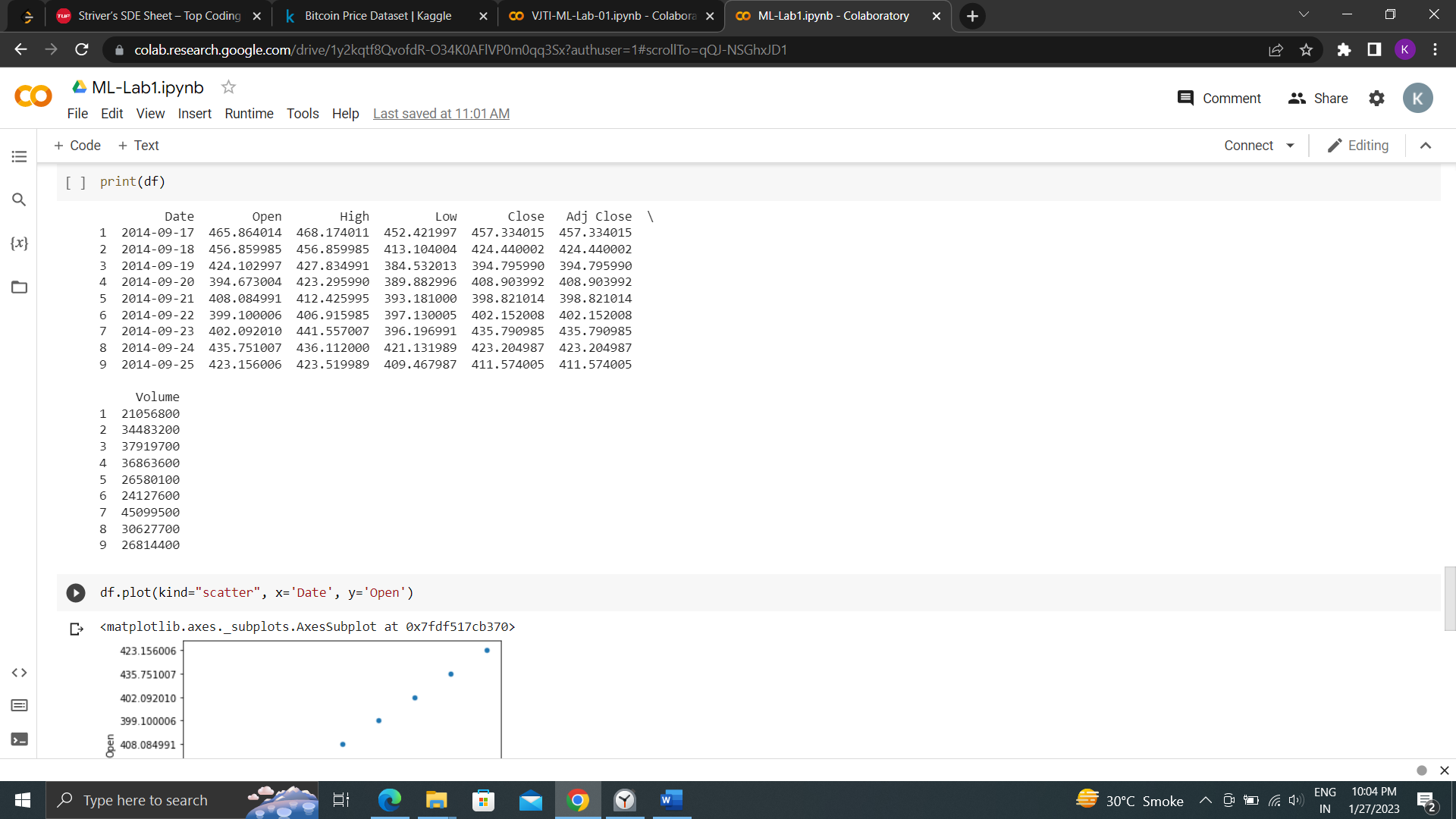


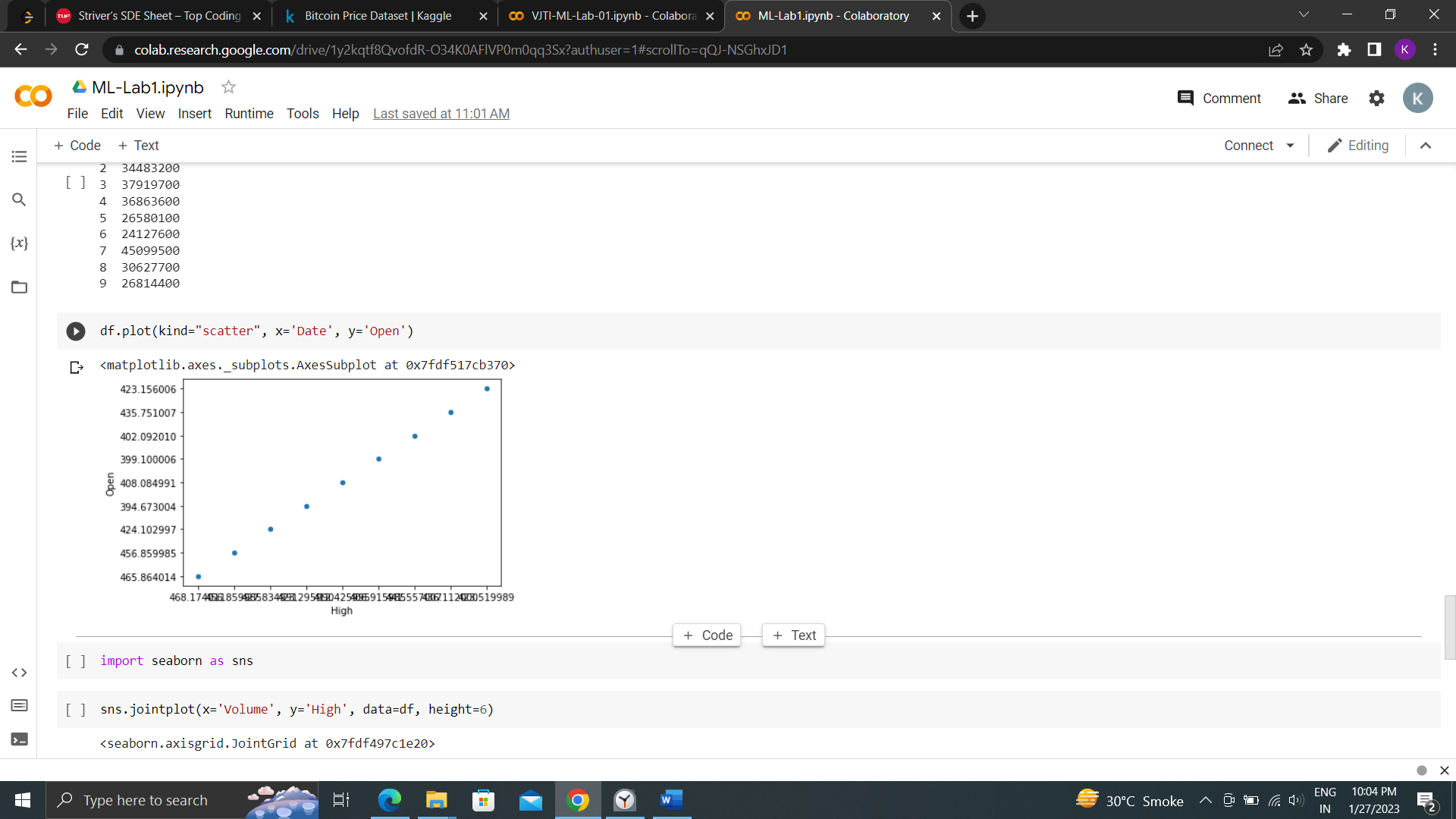


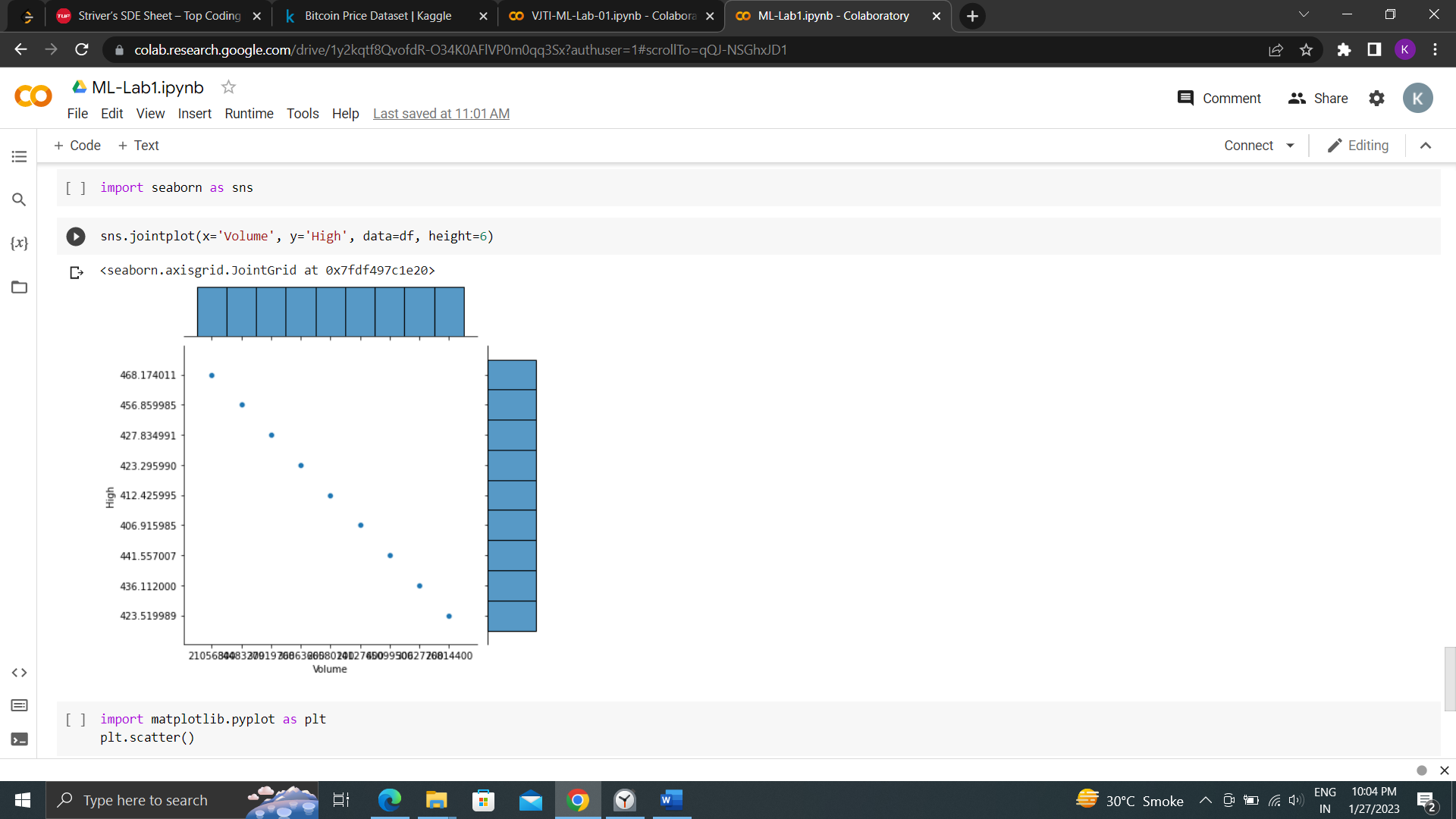




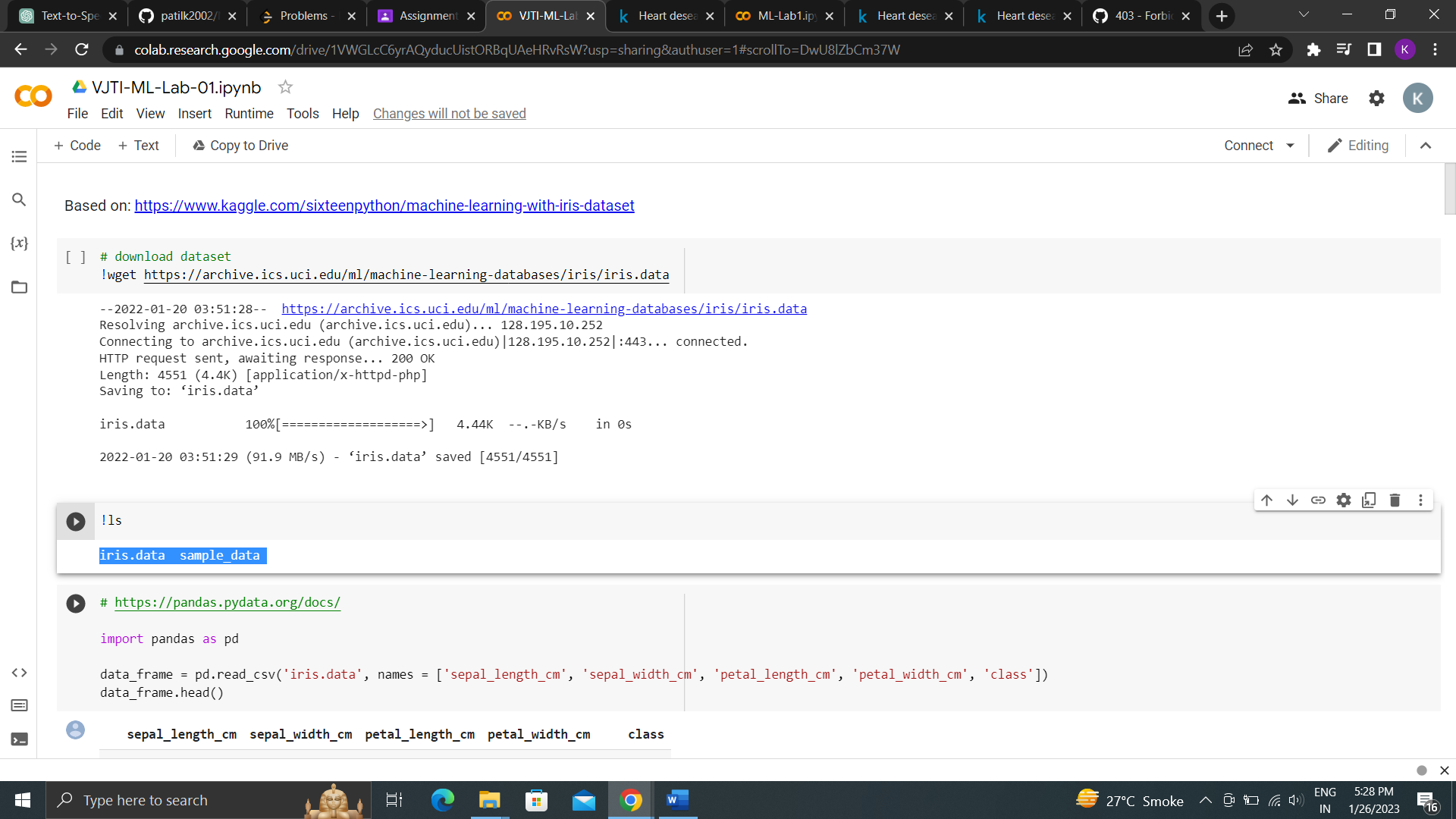


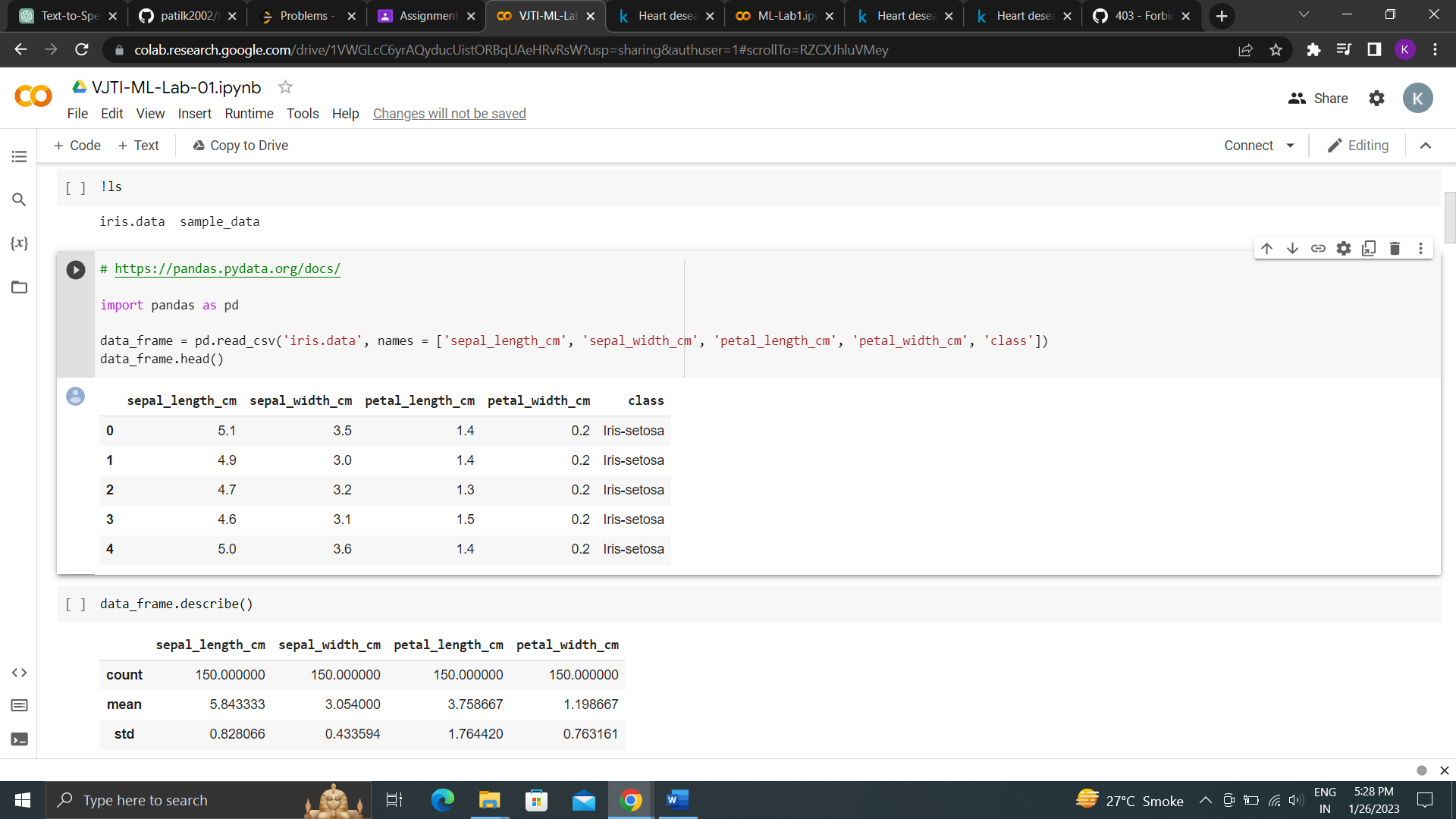


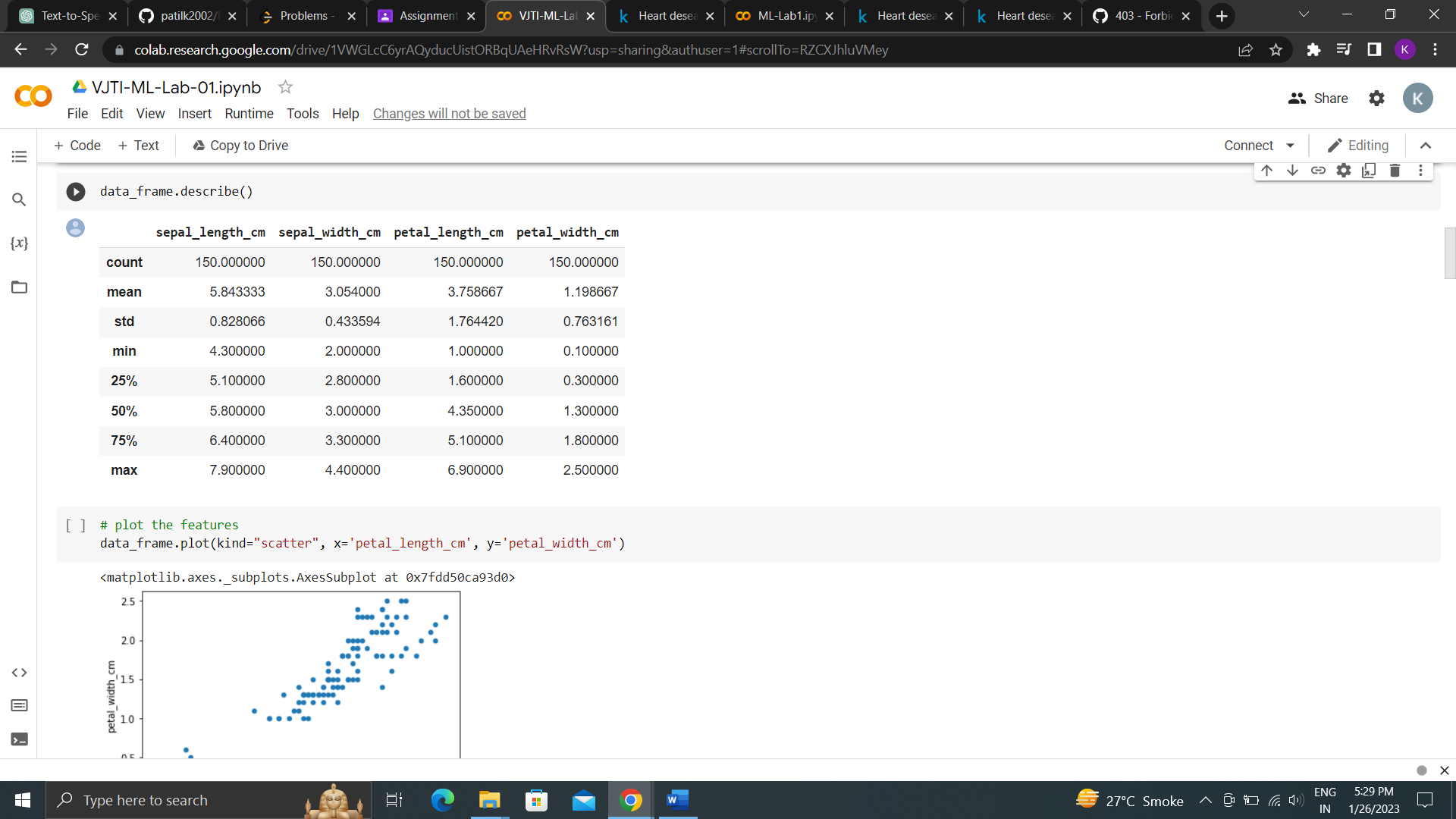




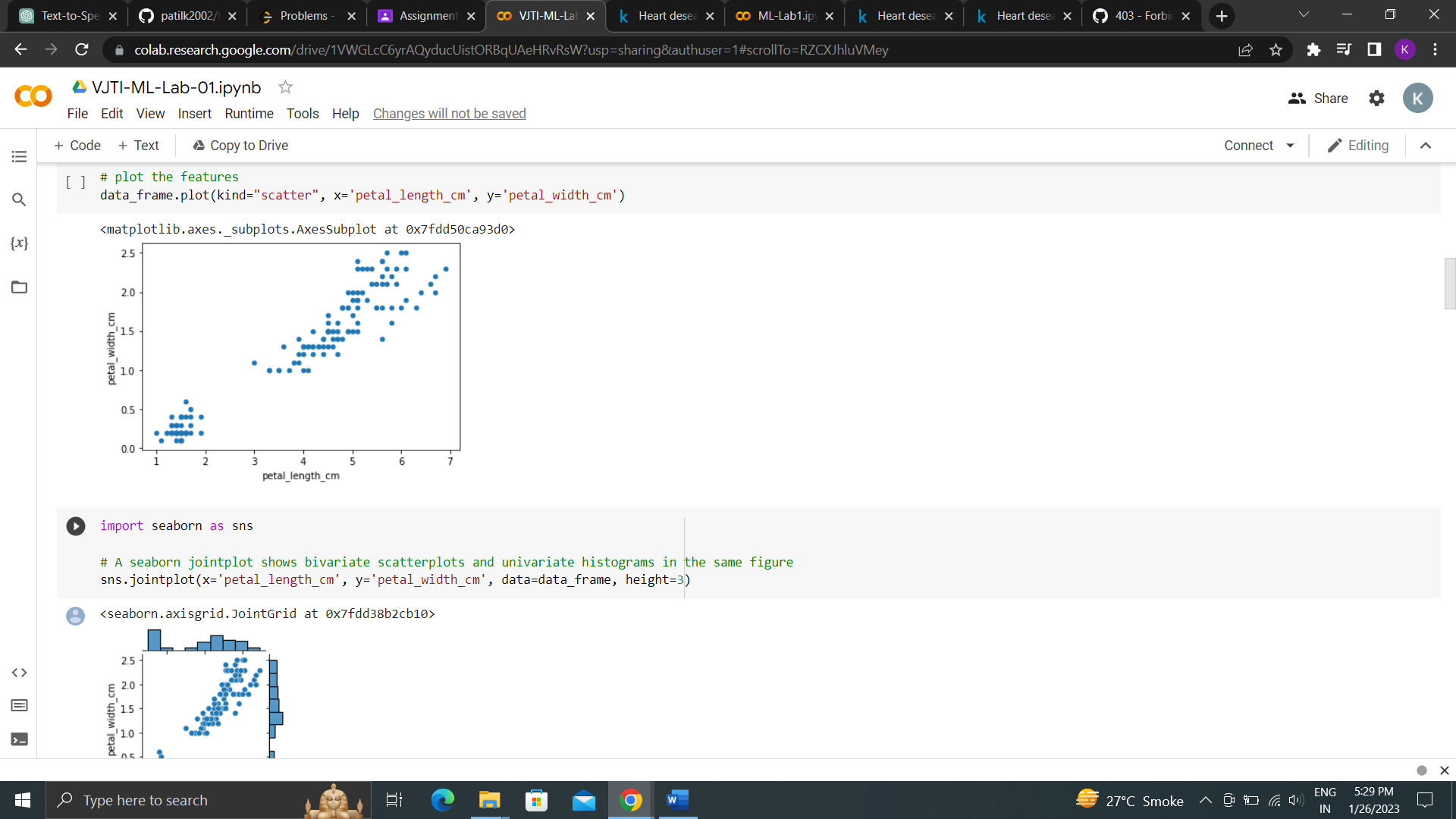
Functions performed using Pandas :



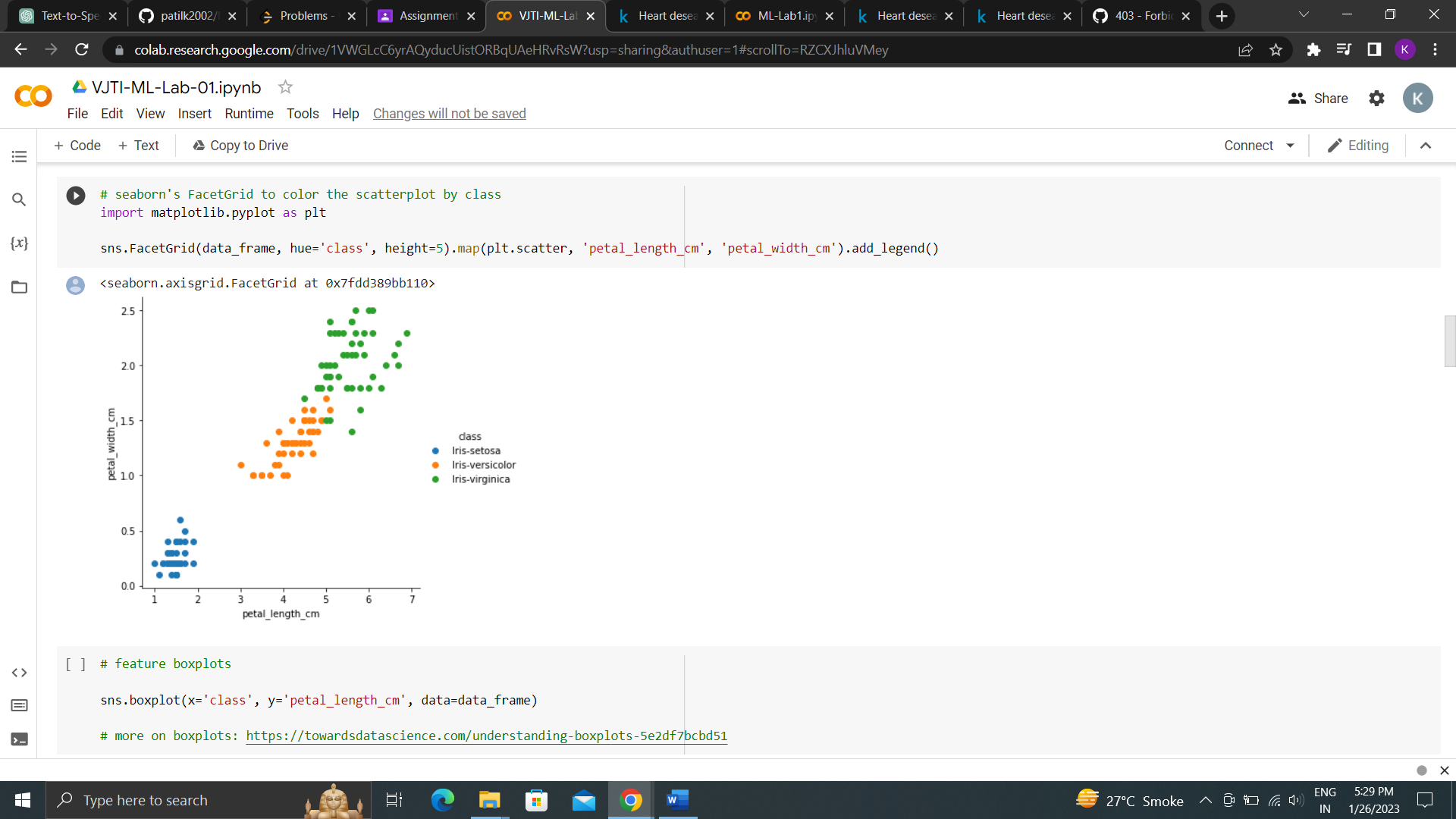




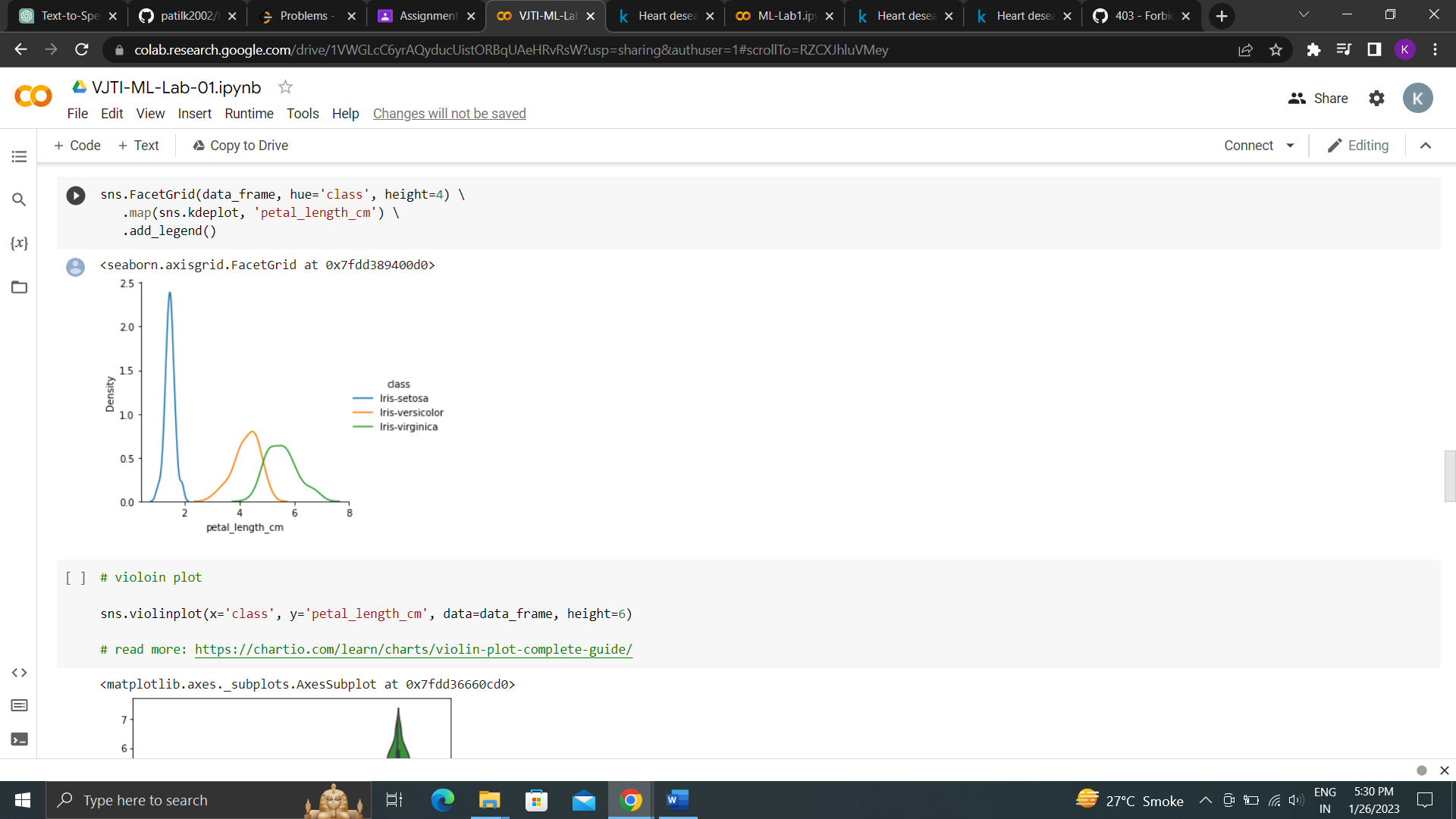
Functions Performed Seaborn :

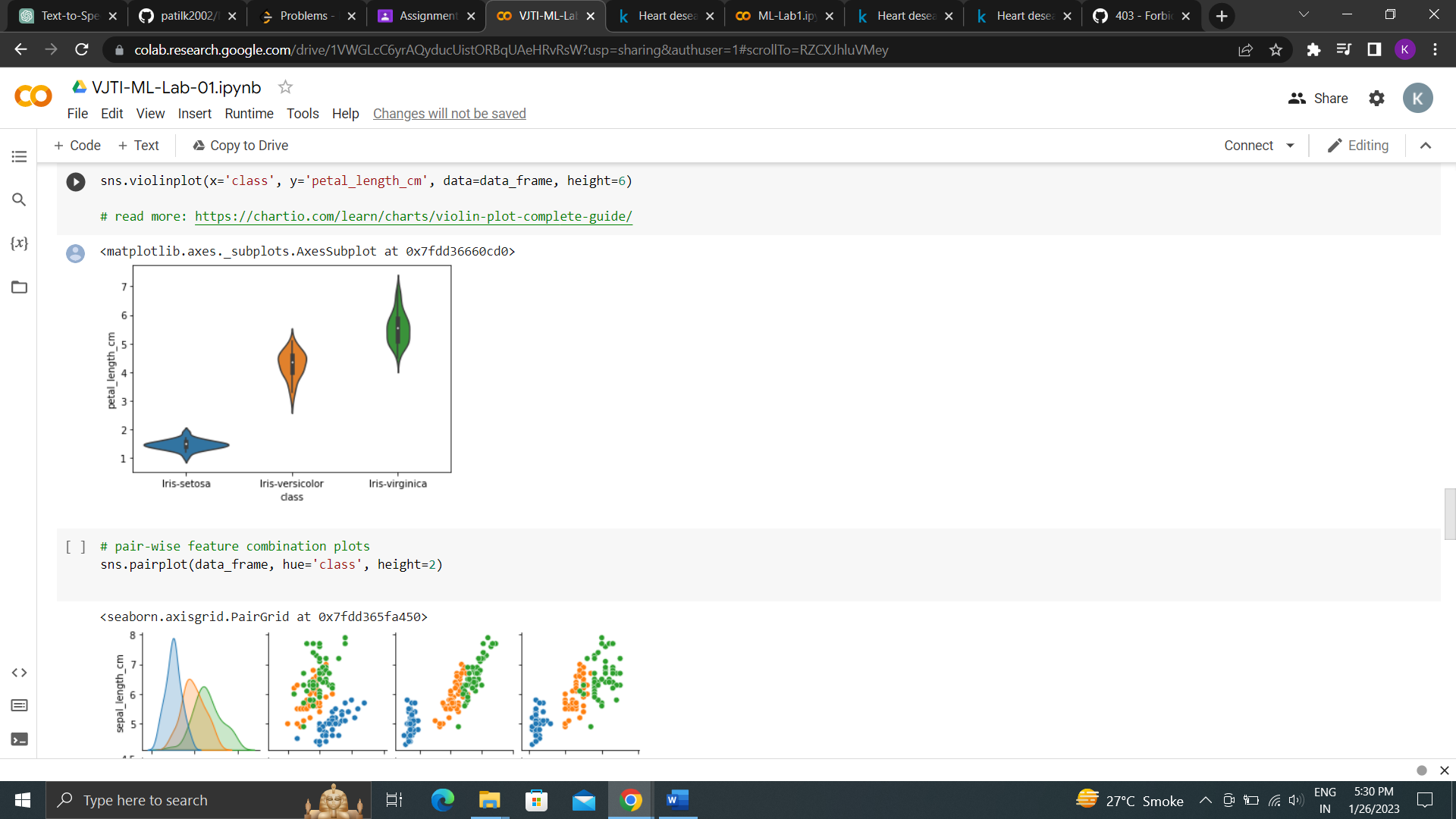


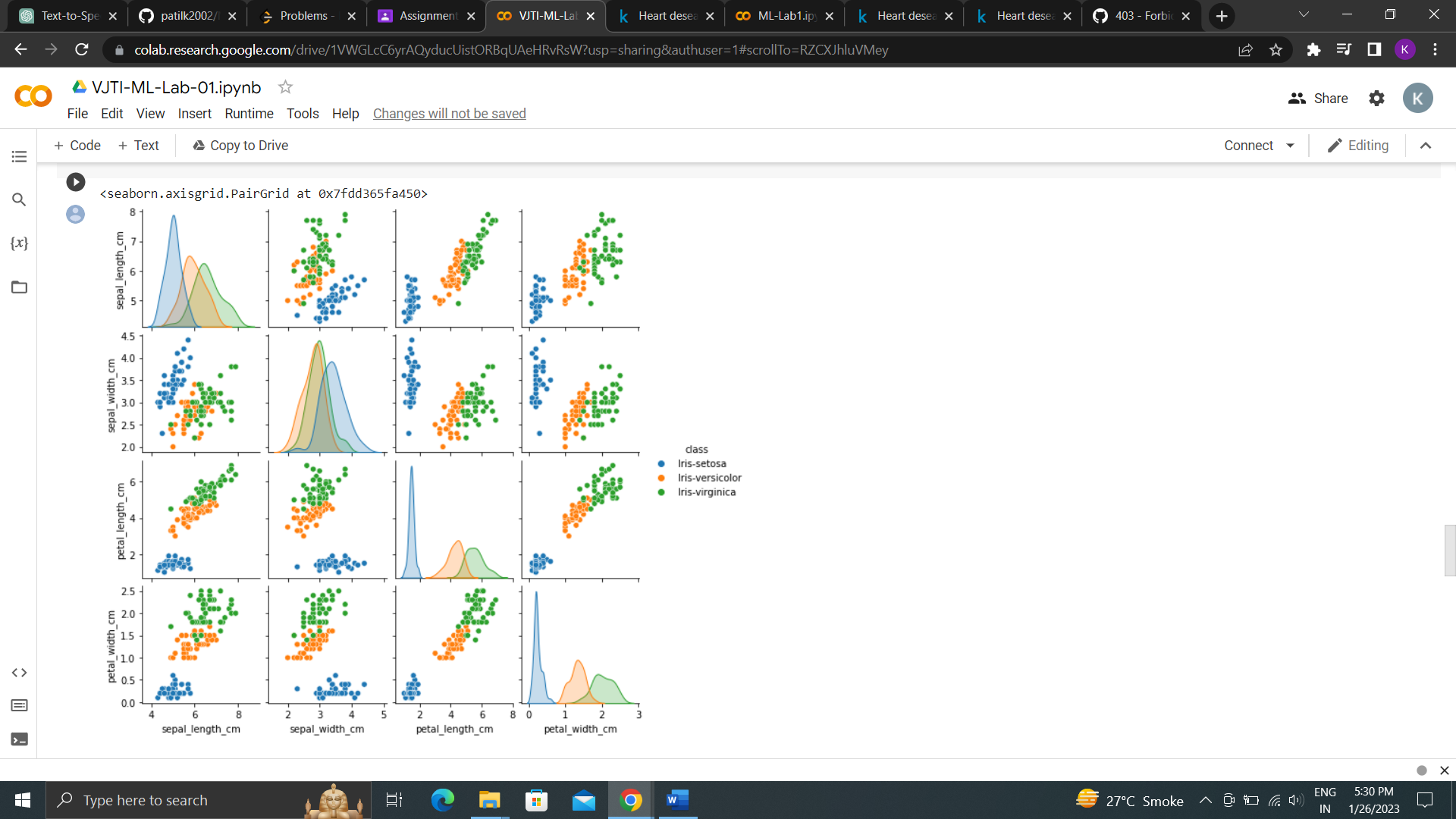
Functions performed using matplotlib

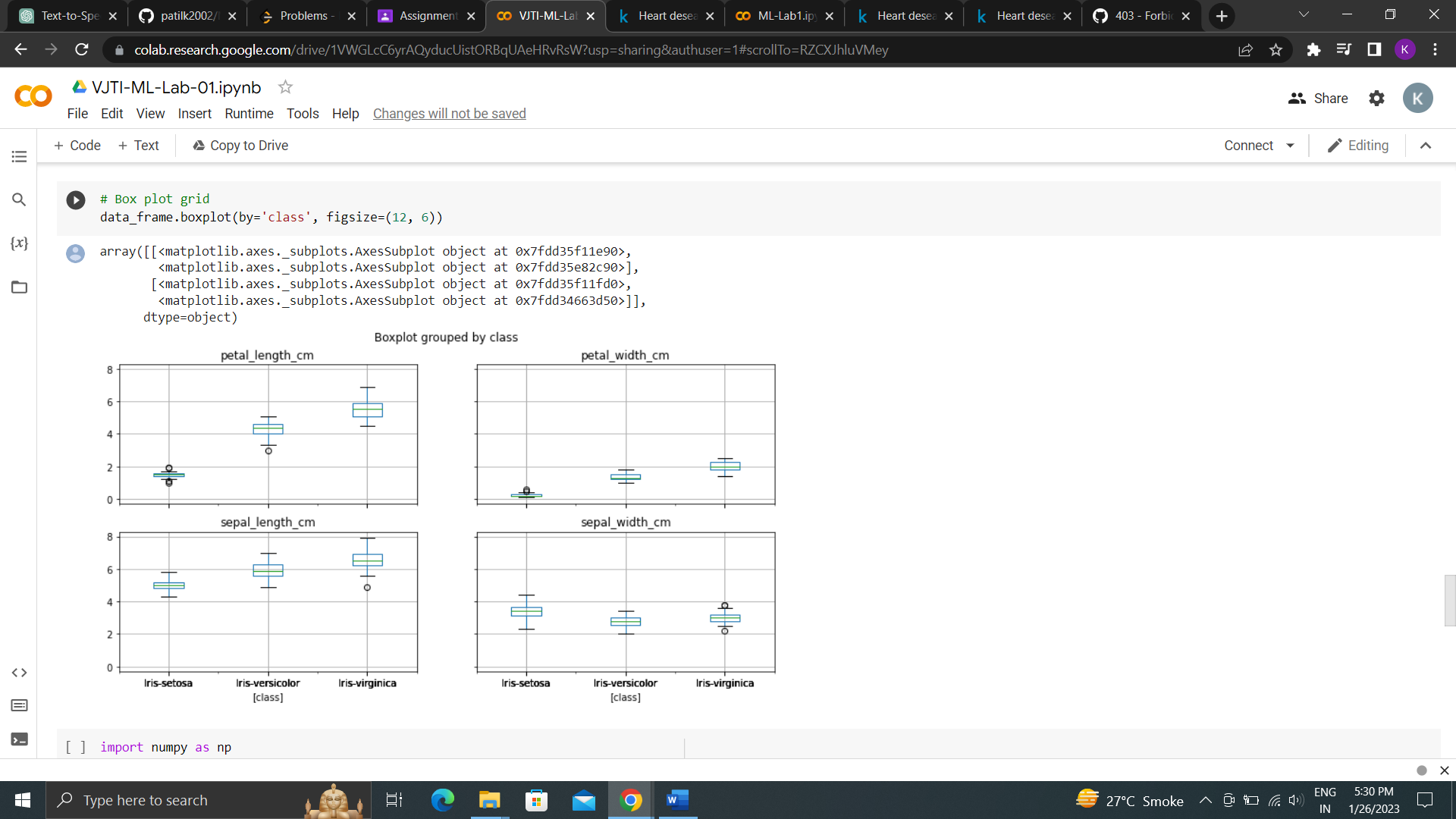






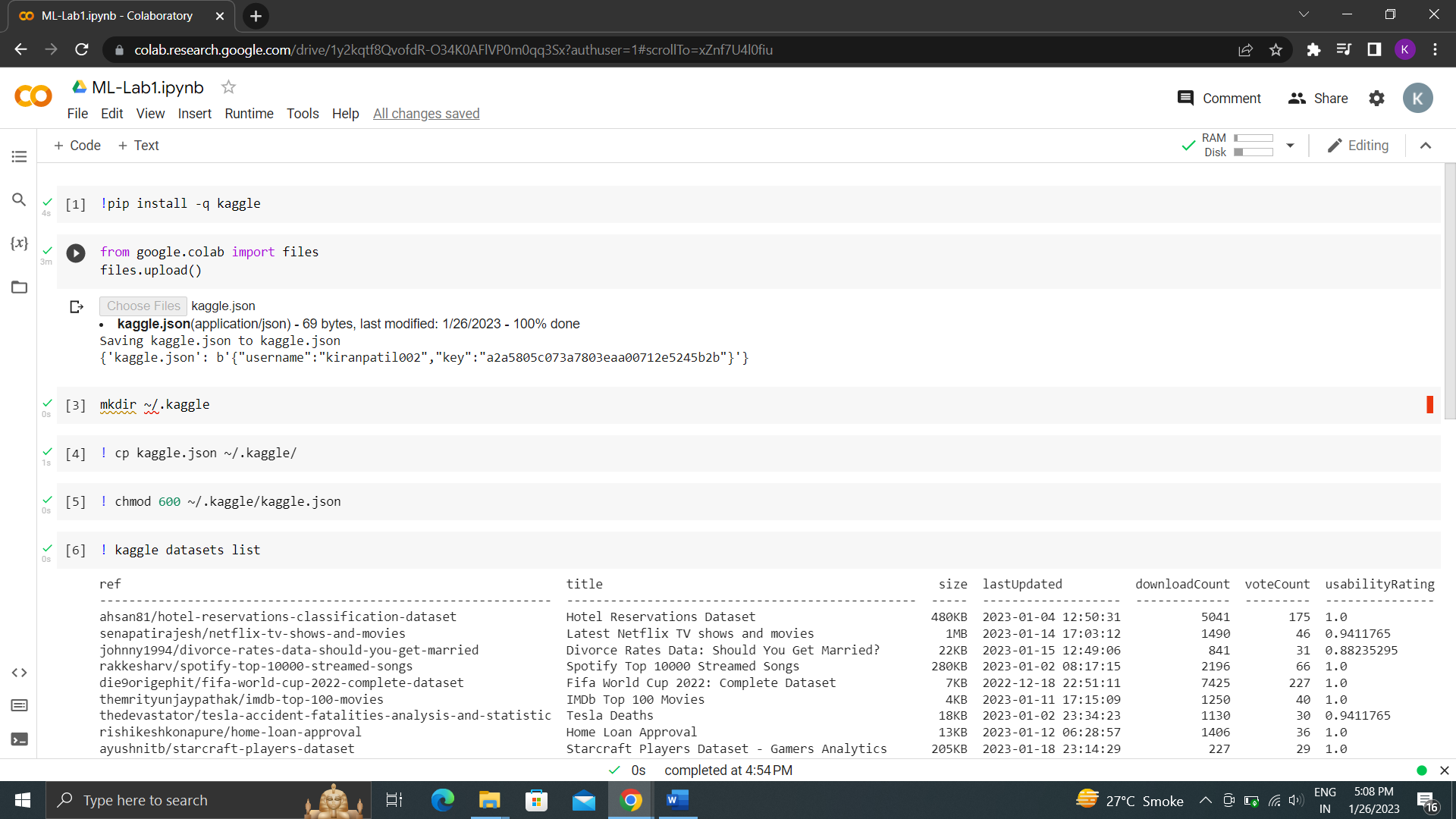


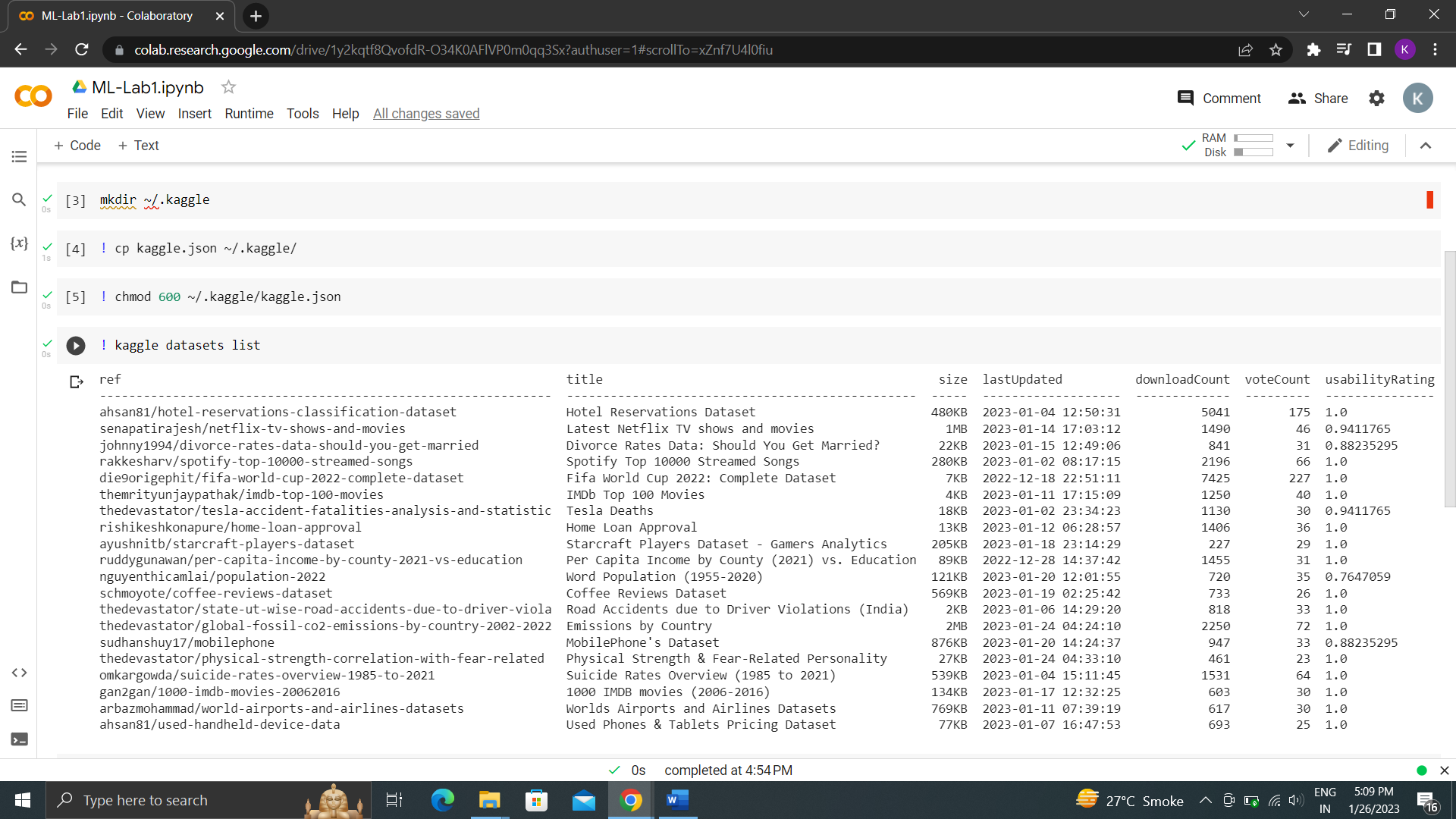












Conclusion : From this experiment we learned to Download a dataset from Kaggle and Performed data exploration using Pandas, Seaborn, Matplotlib python libraries