**Michelle Zokou**

**Tutorial Documentation**

**CTEC 298**

**Professor Bemely**

***Visualization:*** is the presentation of data in a pictorial or graphic format.

**Anaconda:** is a free and open-source distribution of the Python and R programming languages for scientific computing data science, machine learning applications, large-scale data processing, predictive analytics, that aims to simplify package management and deployment.

***Matplotlib Tutorial:*** *Python Plotting*: Python plotting (pyplot), in this tutorial, matplotlib. Pyplot

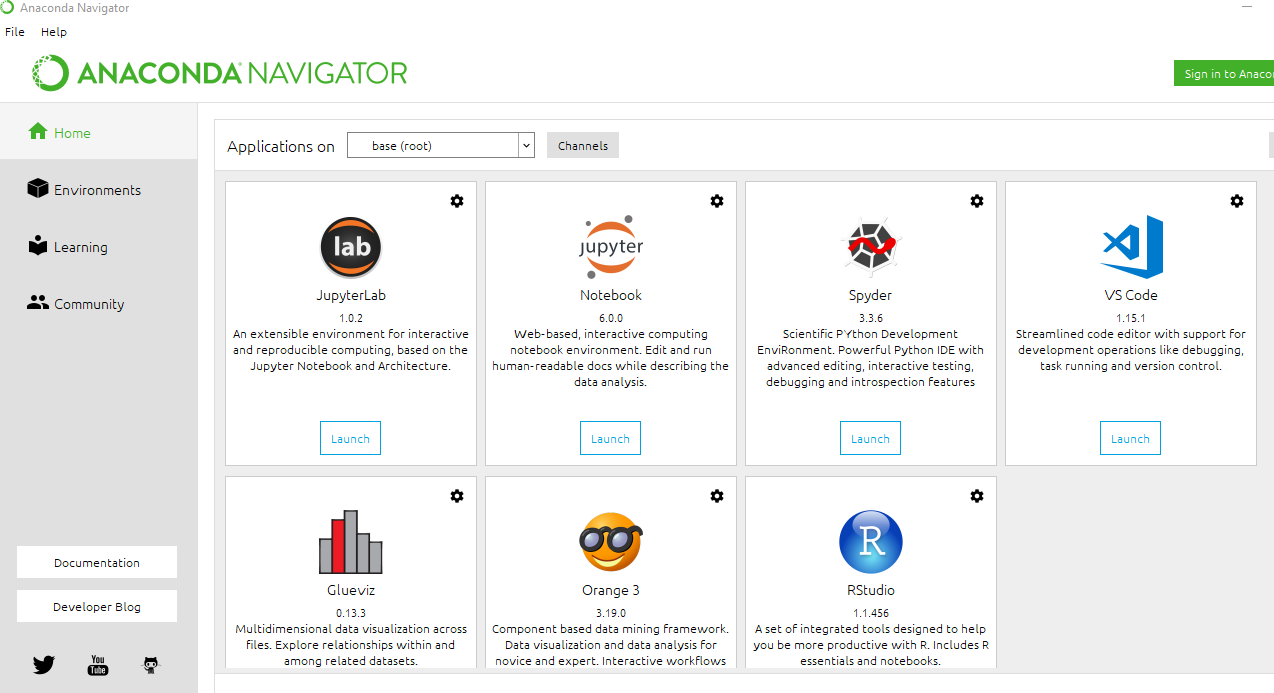
is usually imported as plt . It is the core object that contains the methods to create all sorts of charts and features in a plot.

***Pandas Tutorials*:** for the python pandas’ tutorial, each group was given the instruction of creating a PowerPoint that would teach the importance of python pandas, as well as how to create a python pandas’ structure. Python Pandas has most features that you need for data wrangling and analysis, it allows you to clean, transform, and manipulate your raw data.

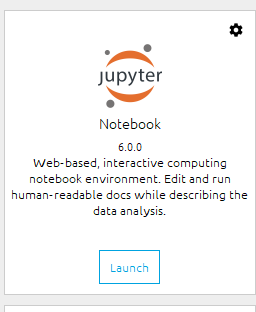
**Pandas data frame:** is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns.

**Jupyter Notebook for Beginners Documentation**

Install Anaconda for Python 3 which could be found on the tutorial page for the Jupyter Notebook



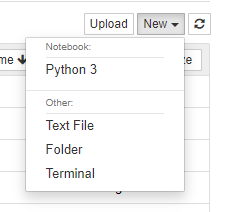
Once Anaconda is installed, you will have to create your first notebook. Locate the notebook and click launch



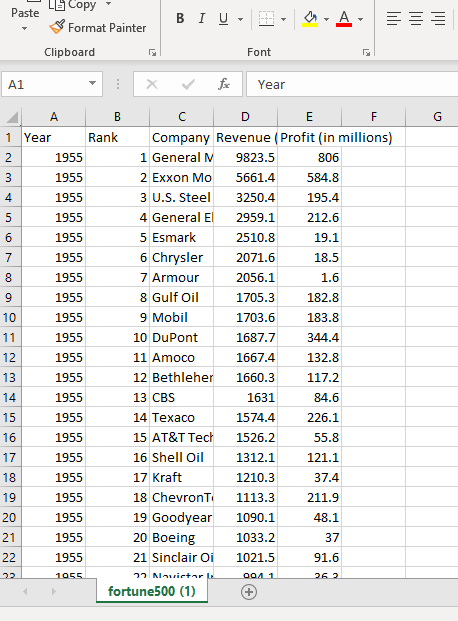
When clicking the launch button, the application will open up a web browser to open the notebook.



Creating a new notebook. On the right-hand side, click the new option and select Python 3



They have given a data set called fortune500, you have to download that excel spreadsheet from the tutorial page.



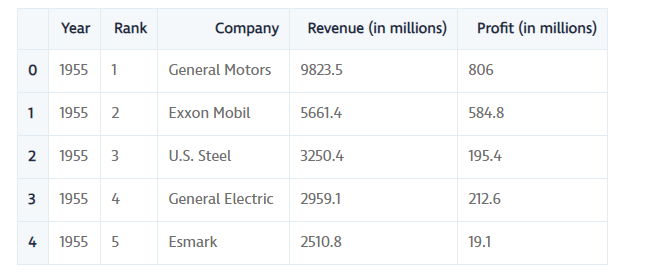
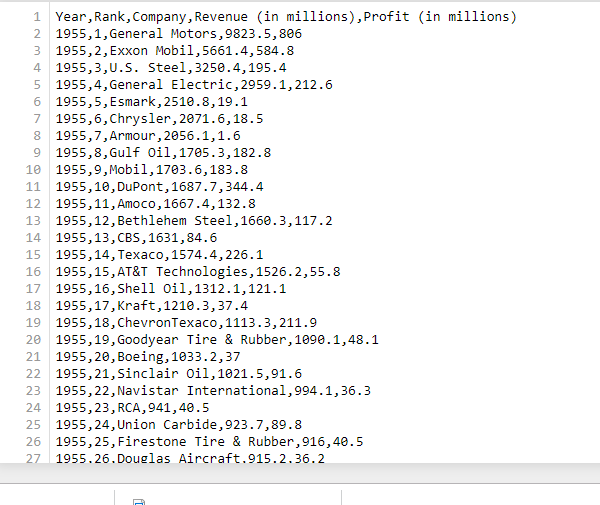
rename your notebook, and Click file > rename.



Starting your code

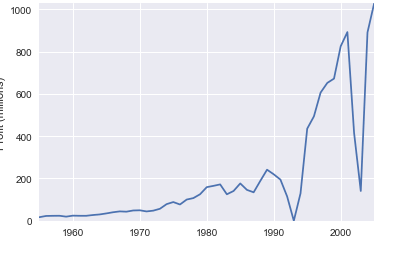
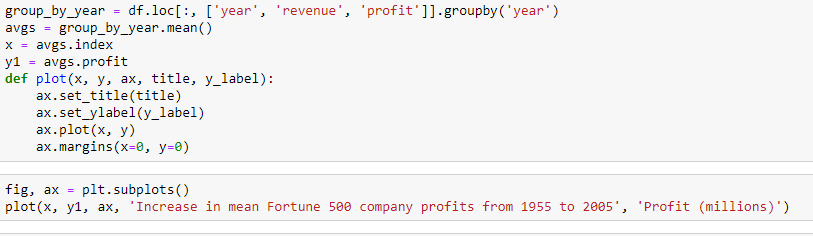


To upload your dataset from the fortune500 file the tutorial page gave you, you have to type “df = pd.read\_csv(‘fortune500.csv’). And save.



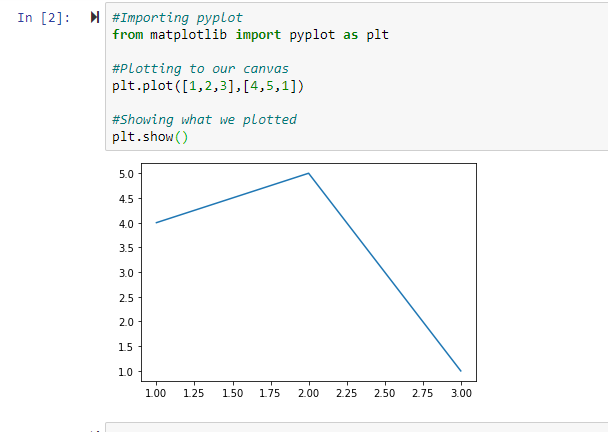
Your columns include the year, rank, company, revenue, profit. Which would be set up in your program as df.column= [‘year’, ‘rank’, ‘company’,’revenue’, ‘profit’]

**For plotting in the notebook**



**Matplotlib Python Tutorial**

Matplotlib is the most popular plotting library in python. Using matplotlib, you can create pretty much any type of plot.

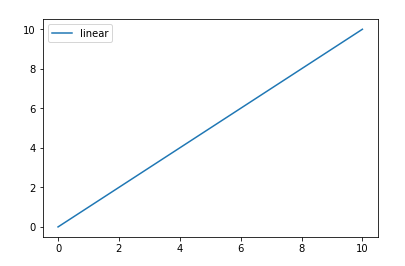
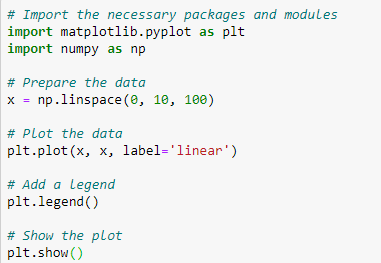


Matplotlib provides two convenient ways to create customized multi-subplots layout, which would include; plt.subplot2grid & plt.GridSpec

**Matplotlib Tutorial: Python Plotting**

You import the pyplot module of the matplotlib library under the alias plt.

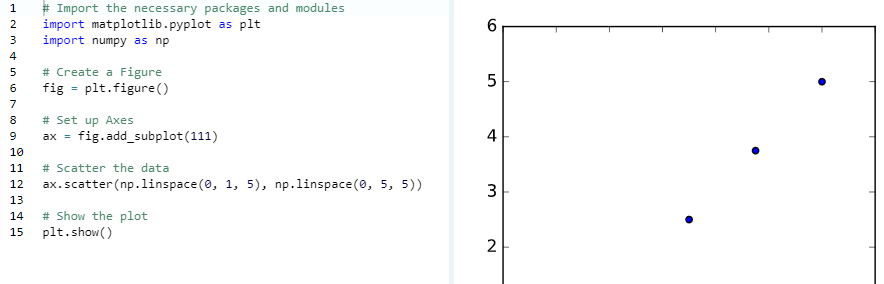
pyplot is a module in the matplotlib package. That’s why you often see matplotlib.pyplot in code.



The axis is the area on which the data is plotted with functions such as plot() and scatter() and that can have ticks, labels, etc. associated with it.

Each axis has an x-axis and a y-axis, which contain ticks, which have major and minor tick lines and tick labels. There’s also the axis labels, title, and legend to consider when you want to customize your axes,

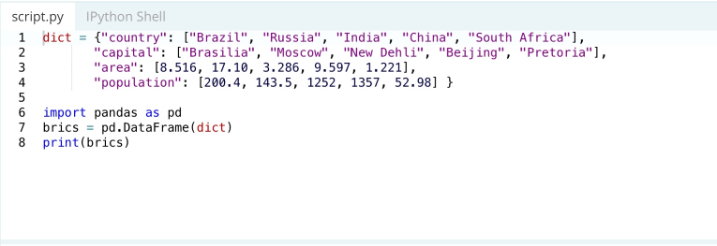
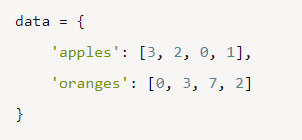
You use subplots to set up and place your axis on a regular grid.When you do call subplot to add axis to your figure, do so with the add\_subplots() function.



The line plt.show() says indeed that you want to see the plot. If you execute this line, you’ll see a window popping up. And you’ll see if it looks like what you had in mind.

**Python Pandas Tutorial**

pandas is a software library written for the Python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. Pandas will extract the data from that CSV into a DataFrame — a table, basically — then let you do things like Store the cleaned, transformed data back into a CSV, other file or database, ask you for the average, median, max, or min of each column, it will do a column A correlate with column B and lastly,



Each (key, value) item in data corresponds to a column in the resulting DataFrame.