**Kayla Wright**

**Jupyter, Matplotlib, and Pandas Tutorial Documentation**

**CTEC 298**

*Jupyter Notebook for Beginners:* jupyter notebook is used for developing and presenting data

science projects. A notebook integrates code and its output into a single document that combines visualizations, narrative text, mathematical equations, and other rich media

*Matplotlib Python Tutorial:* This tutorial explains matplotlib’s way of making plots in simplified

parts so you gain the knowledge and a clear understanding of how to build and modify full featured matplotlib plots.

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

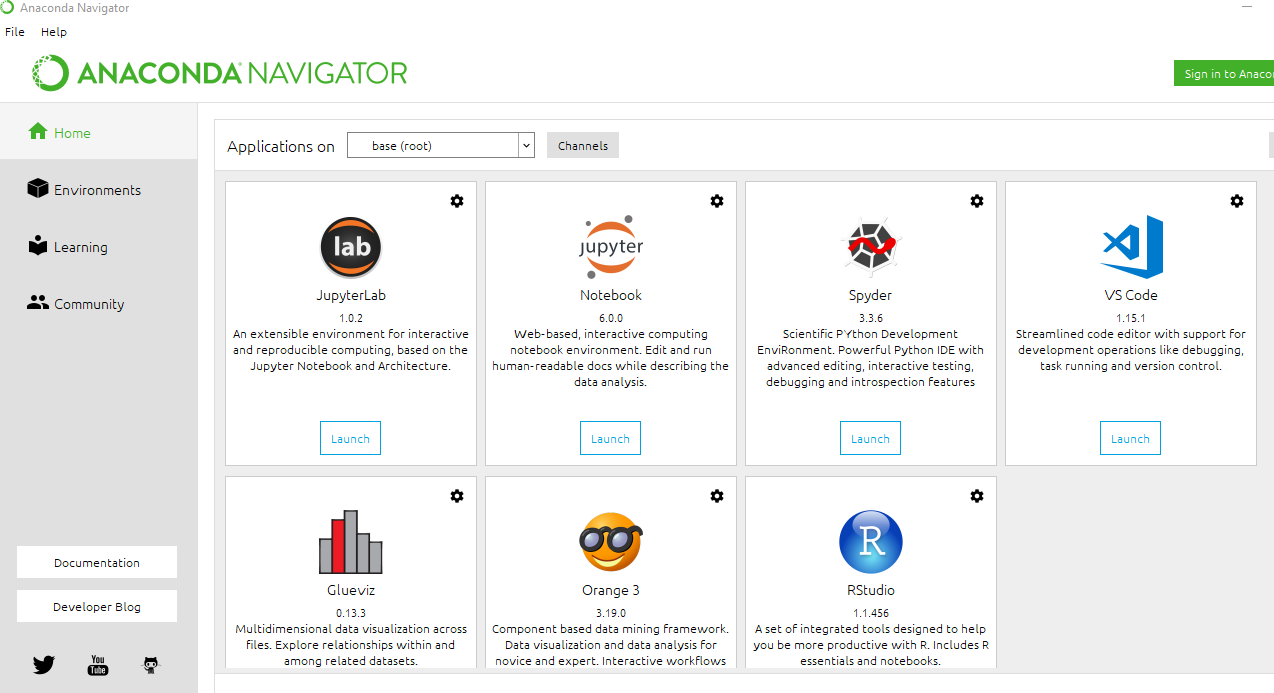
is usually imported as plt . It is the core object that contains the methods to create all sorts of charts and features in a plot.

*Pandas Tutorials*: for the python pandas tutorial, each group was given the instruction of creating

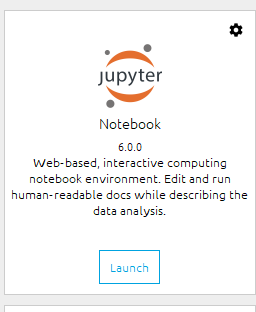
a powerpoint that would teach the importance of python pandas, as well as how to create a python pandas structure. Python Pandas has most features that you need for data wrangling and analysis, it allows you to clean, transform, and manipulate your raw data.

**Jupyter Notebook for Beginners Documentation**

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



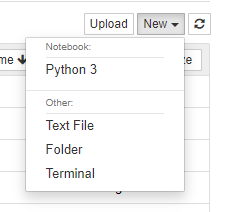
Once Anaconda is installed, you will have to create your first notebook. Locate the notebook and click launch



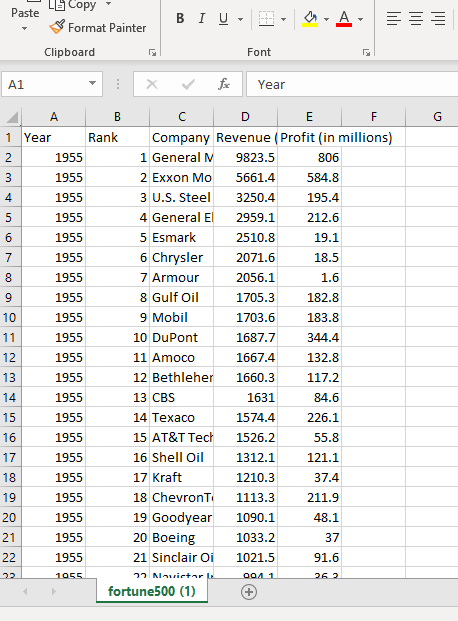
When clicking the launch button, the application will open up a web browser to open the notebook.



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



The goal is to our goal is to find out how the profits of the largest companies in the US changed historically. They have given a data set called fortune500, you have to download that excel spreadsheet from the tutorial page.



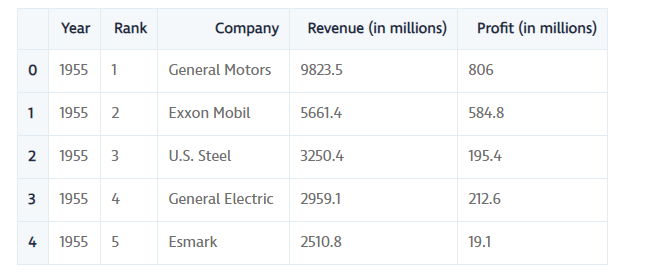
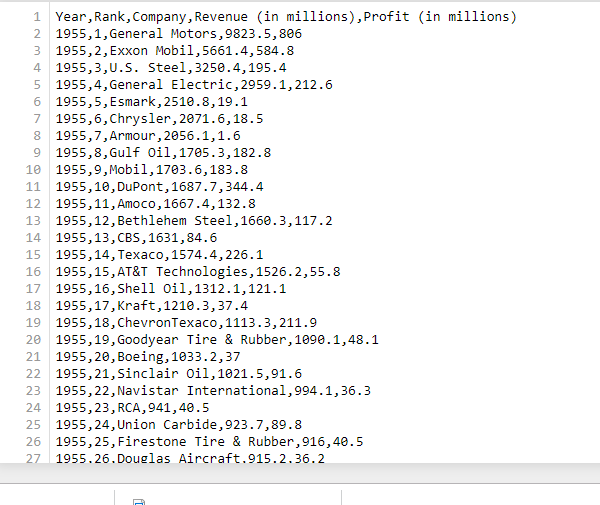
The next thing you will have to do is rename your notebook, and you cannot rename your notebook white it is running. Click file > rename.



Starting your code



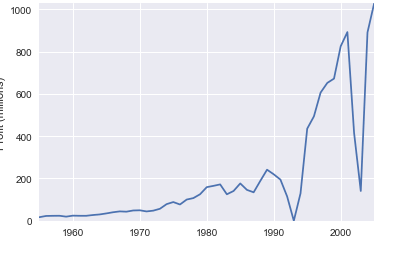
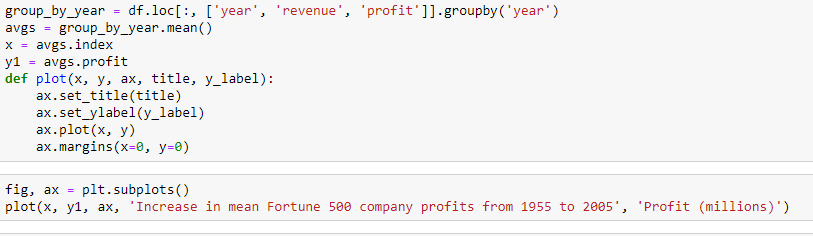
To upload your dataset from the fortune500 file the tutorial page gave you, you have to type “df = pd.read\_csv(‘fortune500.csv’). And save.



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’]

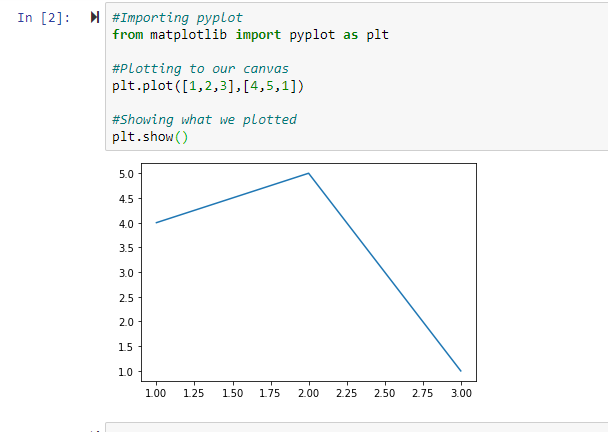


For plotting in the notebook



**Matplotlib Python Tutorial**

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

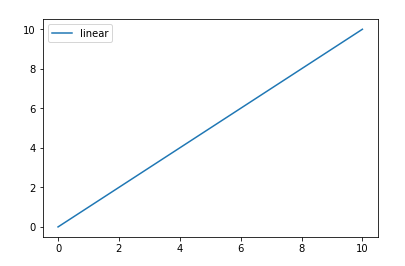
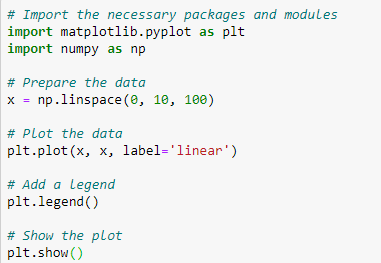


Matplotlib provides two convenient ways to create customized multi-subplots layout, which would include; plt.subplot2grid & plt.GridSpec

**Matplotlib Tutorial: Python Plotting**

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

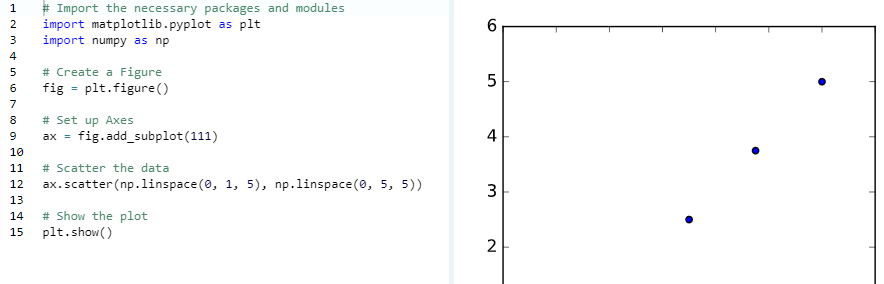
pyplot is a module in the matplotlib package. That’s why you often see matplotlib.pyplot in code.The module provides an interface that allows you to implicitly and automatically create figures and axes to achieve the desired plot.



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**

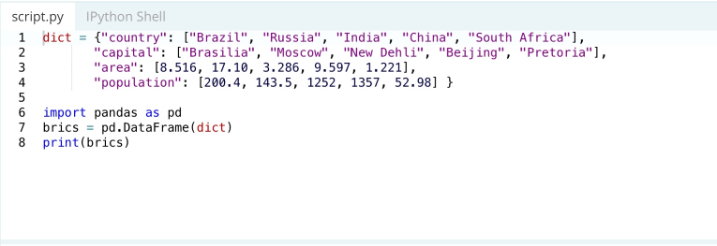
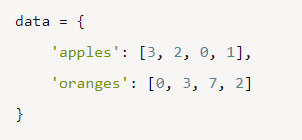
With the use of python pandas, this tool is essentially your data’s home. Through pandas, you get acquainted with your data by cleaning, transforming, and analyzing it. If you want to explore a dataset stored in a CSV on your computer. Pandas will extract the data from that CSV into a DataFrame — a table, basically — then let you do things like:

* Store the cleaned, transformed data back into a CSV, other file or database
* What's the average, median, max, or min of each column?
* Does column A correlate with column B?
* visualize the data with help from Matplotlib. Plot bars, lines, histograms, bubbles, and more.

Open a new notebook in Jupyter and type in !pip install pandas. The exclamation mark at the beginning runs the cell like it is in a terminal. To import pandas, you type in import pandas as pd.

The primary two components of pandas are the Series and DataFrame. A Series is essentially a column, and a DataFrame is a multi-dimensional table made up of a collection of Series.

There are many ways to create a DataFrame from scratch, but a great option is to just use a simple dict.



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