# #66DaysOfData – Day 4: Setting up my Jupyter Notebook

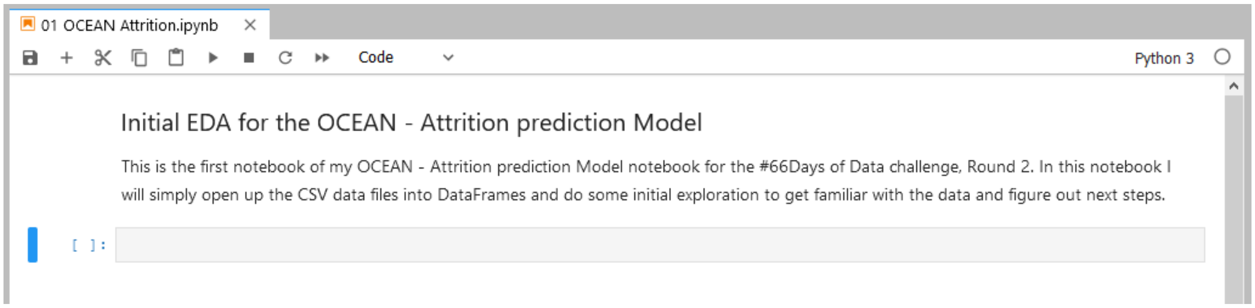
Since I already have some data to start playing around with, it is time to set up the environment I want to use for initial EDA (Exploratory Data Analysis).

I personally love working with Jupyter Notebooks for a simple reason, I get to write about the project, write the code, get instant feedback and all the visualization and graphics I need in just one simple place. Also, they are extremely easy to share for others to play with or just to check them out through Github.

I run Jupyter Lab through Anaconda, also a personal preference, since I can set up a VM environment for each project, which allows me to have the specific versions of any packages used, and if I always use that environment, the notebooks I use for the project should always work.

Since I will probably use a few different notebooks for this project, it is important to name them correctly. I would also suggest numbering them, so this one will be named 01 OCEAN Attrition.

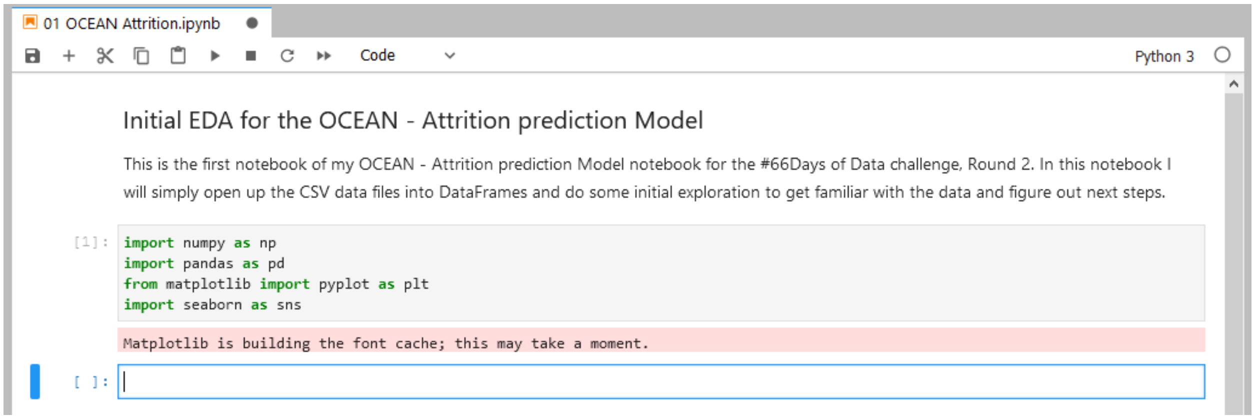
When setting up any notebook, be sure to do an initial introduction with markdown text, to make sure people know what the notebook is all about and what they can expect to find. For me this is vital, since it helps me remember quickly what I was supposed to be doing on a specific notebook.



Initial EDA for the OCEAN - Attrition prediction Model

This is the first notebook of my OCEAN - Attrition prediction Model notebook for the #66Days of Data challenge, Round 2. In this notebook I will simply open up the CSV data files into DataFrames and do some initial exploration to get familiar with the data and figure out next steps.

Then on the first coding cell, I like to import the libraries I am sure to use for this initial EDA. In this case there are 4, numpy, for all the numbers and statistics functions it has, pandas for dataframes, since the data is structured and it is not actually all that big, this will be ideal, then the pyplot functions from matplotlib for plotting which will help exploration and finally seaborn for some more interesting visualization options. These are my tools to go, but there are plenty of others you can choose from.



## Initial EDA for the OCEAN - Attrition prediction Model

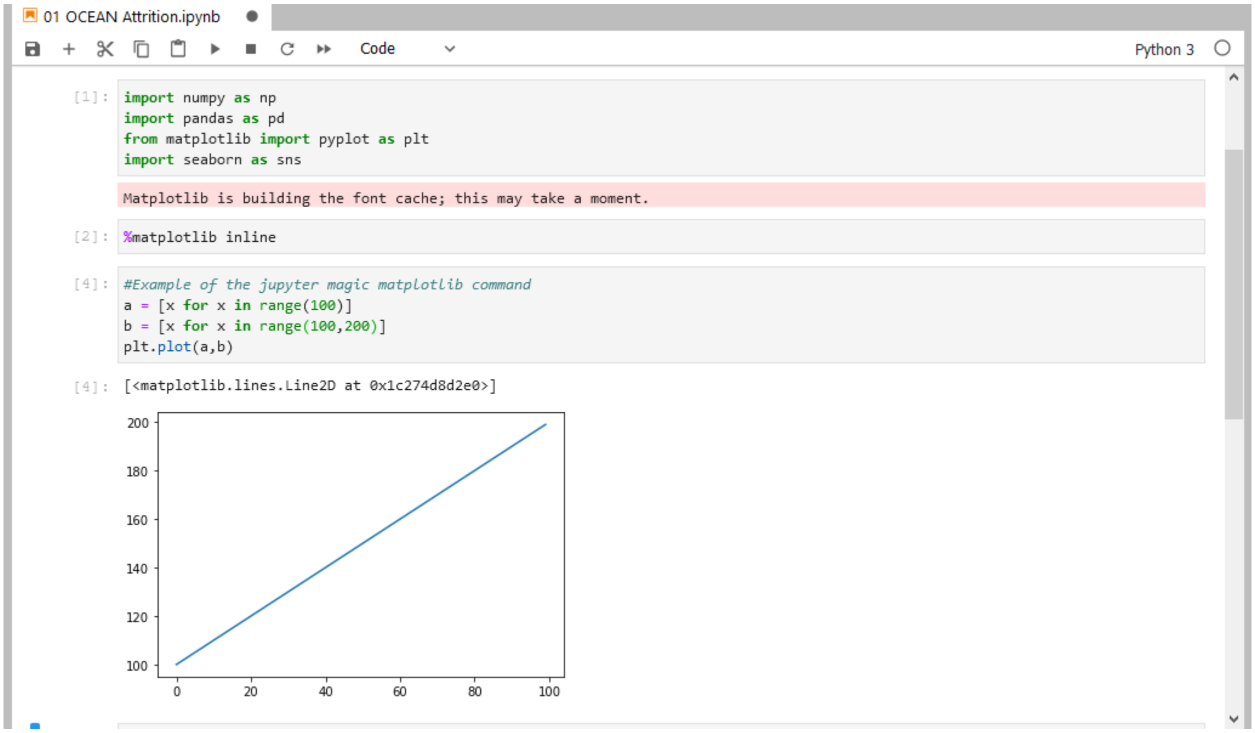
This is the first notebook of my OCEAN - Attrition prediction Model notebook for the #66Days of Data challenge, Round 2. In this notebook I will simply open up the CSV data files into DataFrames and do some initial exploration to get familiar with the data and figure out next steps.

```
[1]: import numpy as np
     import pandas as pd
     from matplotlib import pyplot as plt
     import seaborn as sns

     Matplotlib is building the font cache; this may take a moment.

[ ]:
```

And finally, the ever powerful magic command, %matplotlib inline, one of those incredible features I love about Jupyter Notebooks. This one allows you to plot visualizations right after the line of code, making all the info you need look good and well organized right there. A quick example of this for those not familiar with Jupyter:



```
[1]: import numpy as np
     import pandas as pd
     from matplotlib import pyplot as plt
     import seaborn as sns

     Matplotlib is building the font cache; this may take a moment.

[2]: %matplotlib inline

[4]: #Example of the jupyter magic matplotlib command
     a = [x for x in range(100)]
     b = [x for x in range(100,200)]
     plt.plot(a,b)

[4]: [<matplotlib.lines.Line2D at 0x1c274d8d2e0>]
```

Now we have our notebook set up and ready to look at some data.

## Next Time – Getting a sense for the data.

## Jack Raifer Baruch

**Follow me on Twitter: @JackRaifer**

**Follow me on LinkedIN: jackraifer**

## About the Road to Data Science - #66DaysOfData Series

Road to Data Science series began after I experienced the first round of Ken Jee´s #66DaysOfData challenge back in 2020. Since we are starting the second round of the challenge, I thought it would be a good idea to add small articles every day where I can comment my progress.

I will be sharing all the notebooks, articles and data I can on GitHub: https://github.com/jackraifer/66DaysOfData-Road-to-Data-Science

Please do understand I might have to withhold some information, including code, data, visualizations and/or models, because of confidentiality regards. But I will try to share as much as possible.

## Want to follow the #66DaysOfDataChallenge?

Just follow Ken Jee on twitter **@KenJee_DS** and join the #66DaysOfData challenge.

You can also reach out to me at any time through **LinkedIN** or **Twitter.**