***Use this word document to describe all your steps after each problem. Show all commands & screen pages and describe your steps. Points are deducted for missing steps.***

Describe all steps for each problem set and capture screen shots for each step with a JPG image within this document. Do not forget to put your name on the top of this file for your submitted homework. Also, please replace \_ALL\_ text in red with your information!

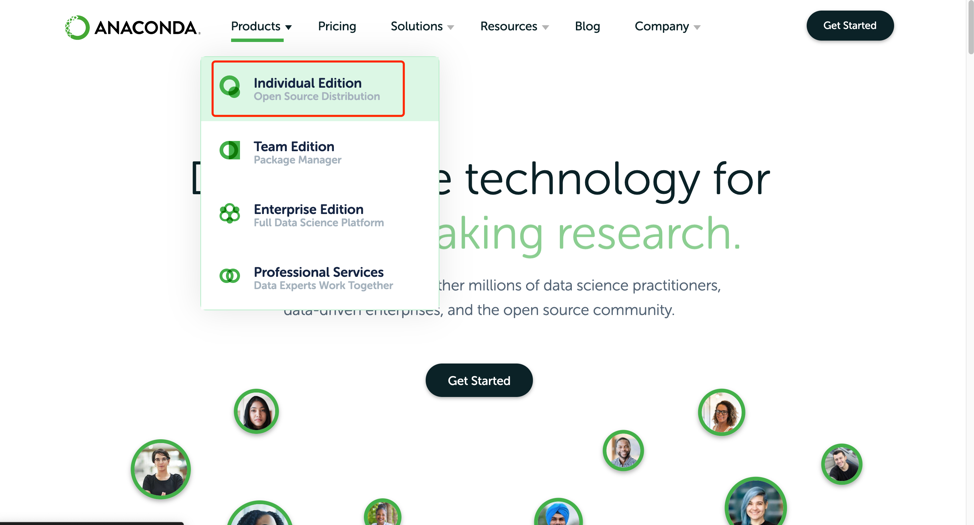
**Problem 1:**

1. Download/Install ~~Jupiter~~ Jupyter Notebook on your computer;
2. Provide **all steps** of installing it, thinks about it as you are making a manual for someone who has no idea on how to do so;

**Total points: 35**

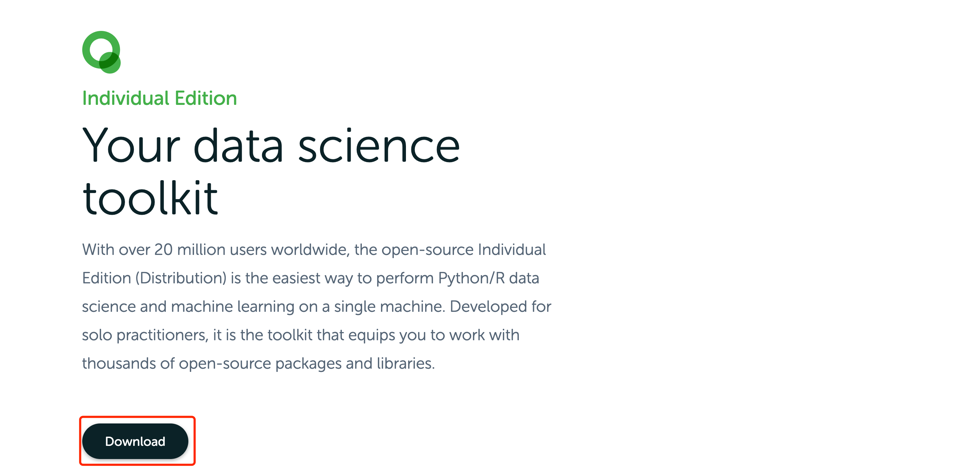
**Download/Install ~~Jupiter~~ Jupyter Notebook on your computer;**

To use the Jupyter Notebook, we have to download the Anaconda (a very handy data science toolkit platform) through the following link: <https://www.anaconda.com/>

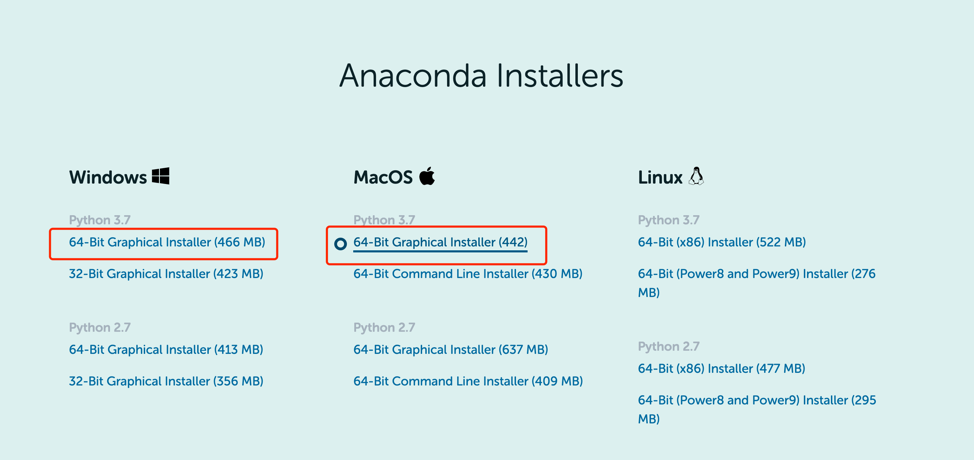


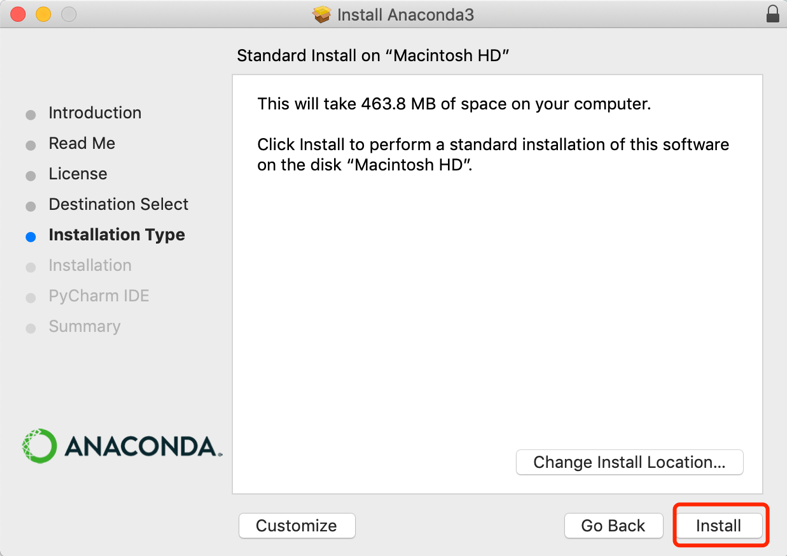
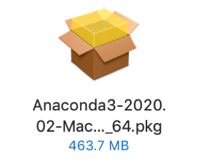
(choose the individual edition)

Then click the download button ↓



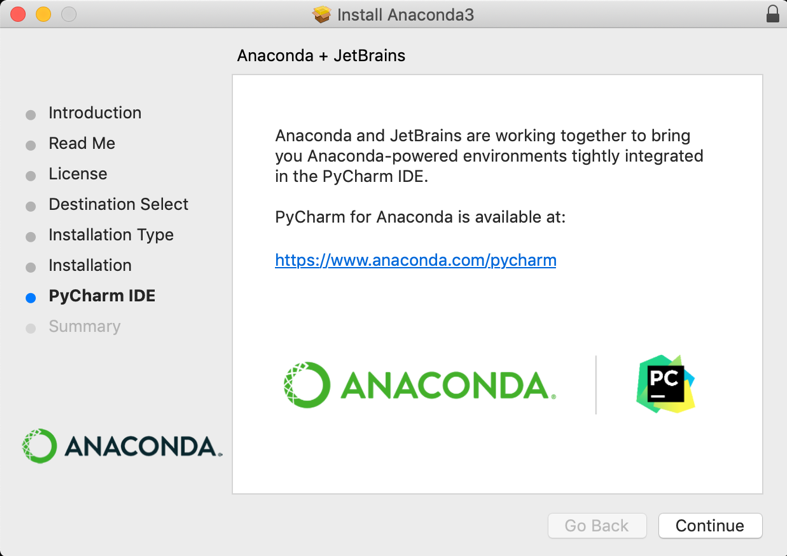
Choose the appropriate version for your PC, I personally use MacOS so I downloaded Python 3.7 64-Bit Graphical Installer ↓



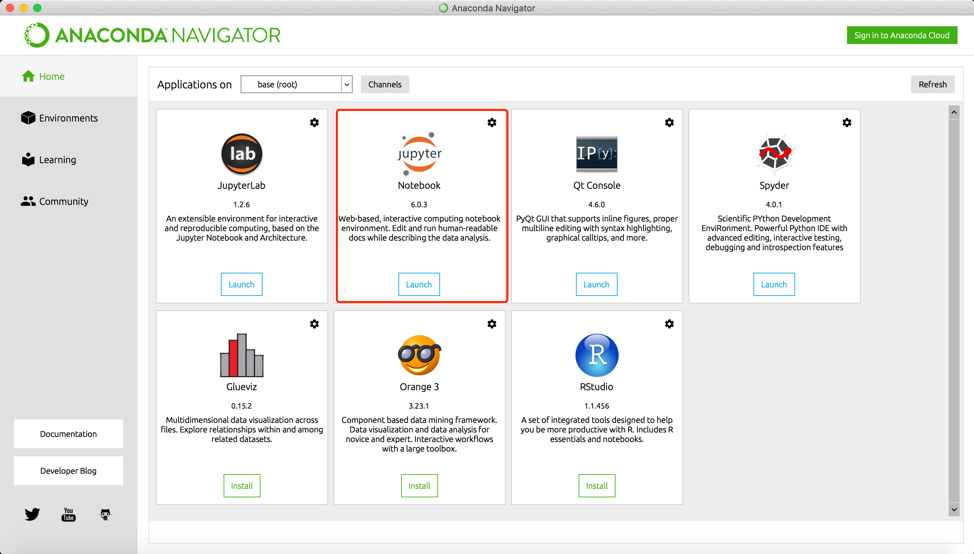


Then open the downloaded package to install Anaconda.

Simply click install and continue, DONE! ↓↓↓



this is the navigator that we’ll use to run the Jupyter notebook for writing python script.



On the home page, Jupyter is available now!

**Problem 2:**

1. Download and install Gutenberg corpus tool to your Jupyter Notebook; Provide **all steps** of installing it, thinks about it as you are making a manual for someone who has no idea on how to do so
2. Download and use source files (provided);
3. Create a table displaying relative frequencies with which “modals” (can, could, may, might, will, would and should) are used in all texts provided (source files)
4. For two modals with the largest span of relative frequencies (most used minus least used), select a text which uses it the most and the text that uses it the least. Compare usage in both texts by examining the concordances of those modals in two texts. Perhaps try to understand how are those words used in different texts;

**Total points: 35**

Put all your steps and screen captures here.

a). Launch Jupyter notebook and create a new python 3 file

A screenshot of a cell phone

Description automatically generated

**Gutenberg package**:

This is a package contains a wide variety of scripts to work with the “Project Gutenberg” body of public domain texts easier.

A screenshot of a social media post

Description automatically generated

The first step is to define the git function, then clone the Gutenberg repository by git clone.

To check if Gutenberg was installed ↓

Done.

**Problem 3:**

1. In the Inaugural corpus identify 10 most frequently used words longer than 7 characters;
2. Which one of those has the largest number of synonyms?
3. List all synonyms for those 10 words. Which one of those 10 words has the largest number of hyponyms?
4. List all hyponyms of those 10 most frequently used “long” words.

**Total points: 35**

**Please NOTE!!!**

**You literature, as well as a big help, for this assignment are chapters 1 and 2 of Natural Language Processing with Python book by Steven Bird et al.**