AMIA 2019 Informatics Summit
Workshop W07: Hands-On Full Life Cycle Data Science Workshop

Thank you for your interest in the workshop. This is a "hands-on" workshop which means that you will need to bring a laptop so that you can follow along, execute code and do short exercises as we work through aspects of a data science project. You can either 1) run it in the cloud or 2) install the environment on your laptop and execute it locally.

1. The advantage to running the environment in the cloud is that there is nothing to install! We will be using the Google Colab environment. One disadvantage is that the environment will shut down after 12 hours of use. That is fine for the purpose of this workshop, but probably is not what you'd want to use for a production data science environment.

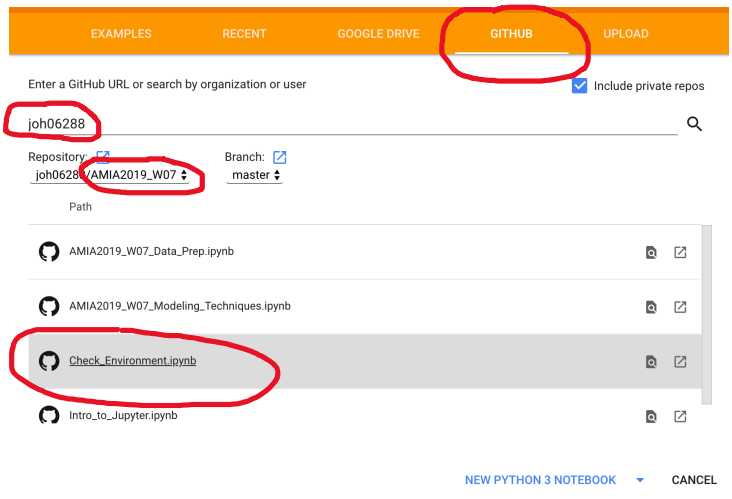
2. The advantage to installing it locally is that everything you need for a data science project is available on your computer. The disadvantage is that it is sometimes difficult and time consuming to install all of the elements of the environment. To install it locally, we will use the Anaconda Python distribution which has all of the packages that we will be using. It should install equally well on with Windows or Mac machines.

1. Access Jupyter notebooks in the cloud using Google Colab

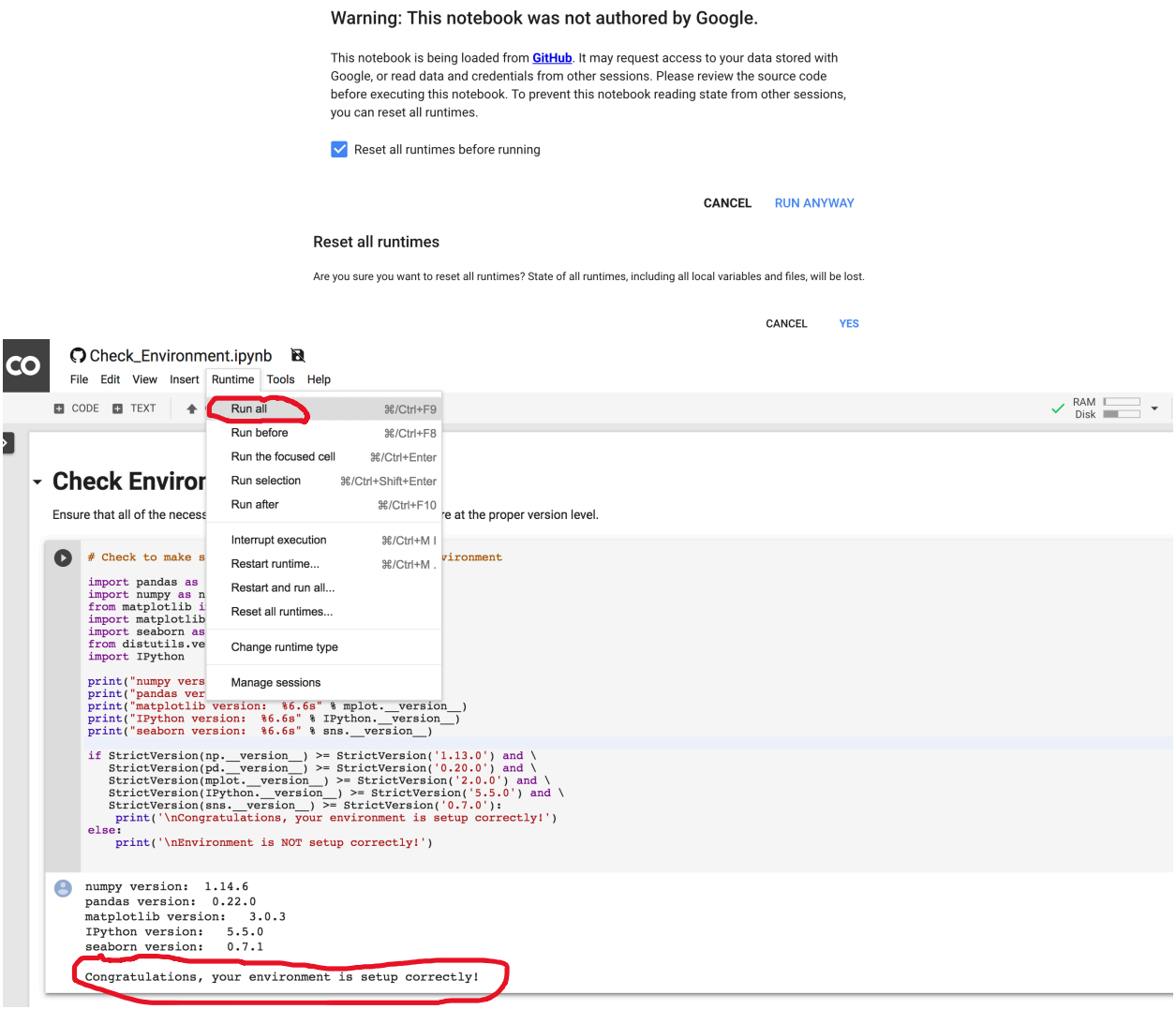
This is probably the easiest way to get started. You can access Jupyter notebooks in a shared cloud environment using a Google research project called Colab. This has the advantage of letting you try out a notebook without having to install anything on your local computer. But the disadvantage is that the session will timeout after 12 hours. It is not suitable for doing real data science work, but it is a good option for this workshop.

To access the system, go to the following URL in a browser (Chrome works best): https://colab.research.google.com/ This will launch a new virtual environment where you can play with the workshop notebooks. Check to make sure the environment is working by doing the following:

1. When you go https://colab.research.google.com, a pop-up screen will ask you which notebook to open. Click the "GITHUB" tab and then put in joh06288 for the Github URL and find the AMIA2019_W07 repository and select the "Check_Environment" notebook:



2. Navigate to the main menu and select "Runtime -> Run all". You will be warned that the notebook is not from Google and that your runtimes will be reset. Pick "Run anyway" and "Yes" to run the notebook. If



3. The final message should say “Congratulations, your environment is setup correctly!”

4. If you have problems, there will be time at the Workshop to help you troubleshoot. Please come to the workshop meeting room 30 minutes early on the day of the workshop and we can help resolve the problem.

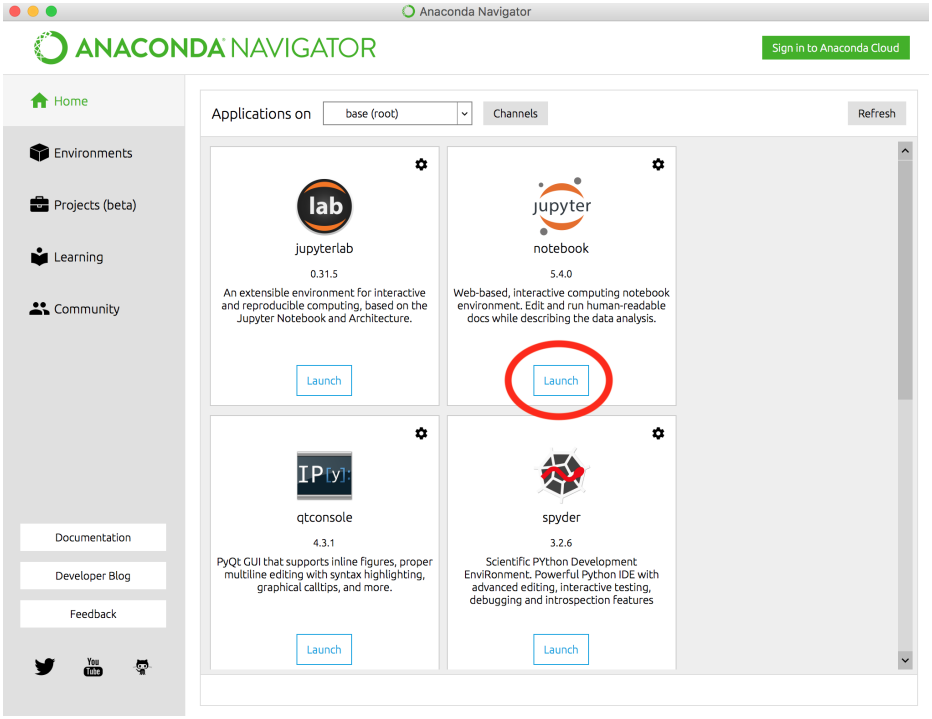
5. If you are interested, you can take a look at the “Intro_to_Jupyter” notebook which is a quick overview of some of the libraries that we will be using during the workshop.

2. Install locally using the Anaconda distribution

If you don't want to use the cloud, follow the instructions to install the full Anaconda Distribution (https://www.anaconda.com/download), which includes Python, the Jupyter Notebook system, and other commonly used packages for scientific computing and data science. Make sure you install the latest Python 3 version of the distribution. After a successful install, you can run the following command at the Terminal (Mac/Linux) or Command Prompt (Windows) to start the Jupyter system:

```
jupyter notebook
```

You can also launch the Jupyter Notebook system from the Anaconda Navigator:



Check to make sure the environment is installed correctly by doing the following:

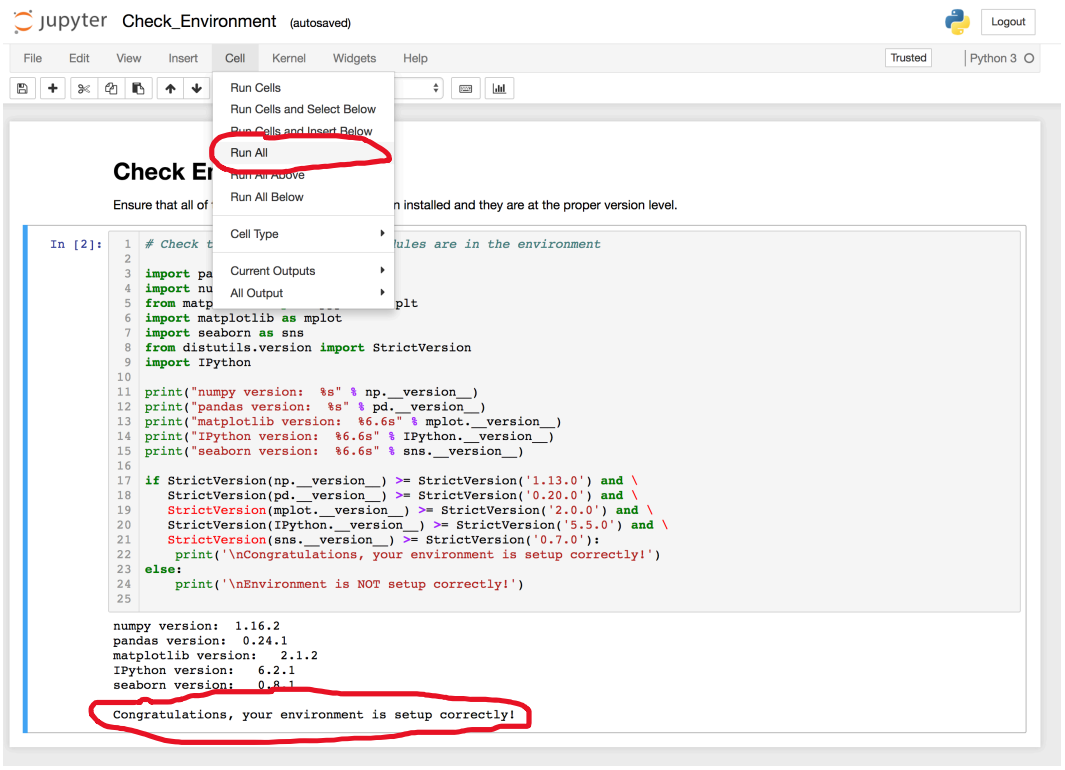
1. In a CMD prompt (Windows) or Terminal prompt (OSX) create a work directory that you will put the Workshop content in.
   ```
   mkdir workdir
   ```

2. Retrieve the workshop content from the github server using the following command:
   ```
   git clone https://github.com/joh06288/AMIA2019_W07.git
   ```

3. From the browser window that Jupyter notebook opened for you when it started, navigate to your workdir and go into the AMIA2019_W07 folder.

4. Select the "Check_Environment" notebook. A new browser window will open up.

5. Select "Cell -> Run All" from the Jupyter Notebook menu. The following output should be produced:



6. The final message should say “Congratulations, your environment is setup correctly!”

7. If you have problems, there will be time at the Workshop to help you troubleshoot. Please come to the workshop meeting room 30 minutes early on the day of the workshop and we can help resolve the problem.

8. If you are interested, you can take a look at the “Intro_to_Jupyter” notebook which is a quick overview of some of the libraries that we will be using during the workshop.