

MedInfo 2019

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. For this workshop we will be running all of the notebooks in the cloud.

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.

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. Make sure you have a Google login and, if you are not already logged in, you will be prompted do so. 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 “medinfo2019” repository and select the “Check_Environment” notebook:
- 2.

The screenshot shows the Google Colab interface. At the top, there are tabs: EXAMPLES, RECENT, GOOGLE DRIVE, GITHUB, and UPLOAD. The GITHUB tab is selected and highlighted with a red box. Below the tabs, there is a search bar with the text "Enter a GitHub URL or search by organization or user". The search bar contains the text "joh06288" and is highlighted with a red box. To the right of the search bar is a checkbox labeled "Include private repos". Below the search bar, there is a dropdown menu for "Repository:" showing "joh06288/medinfo2019" and a dropdown menu for "Branch:" showing "master". Below these, there is a list of notebooks under the heading "Path". The first notebook, "Check_Environment.ipynb", is highlighted with a red box. Below it are "Intro_to_Jupyter.ipynb", "Medinfo2019_Data_Prep.ipynb", and "Medinfo2019_Modeling_Techniques.ipynb". At the bottom right, there are buttons for "NEW PYTHON 3 NOTEBOOK" and "CANCEL".

3. 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

Warning: This notebook was not authored by Google.

This notebook is being loaded from [GitHub](#). It may request access to your data stored with Google, or read data and credentials from other sessions. Please review the source code before executing this notebook. To prevent this notebook reading state from other sessions, you can reset all runtimes.

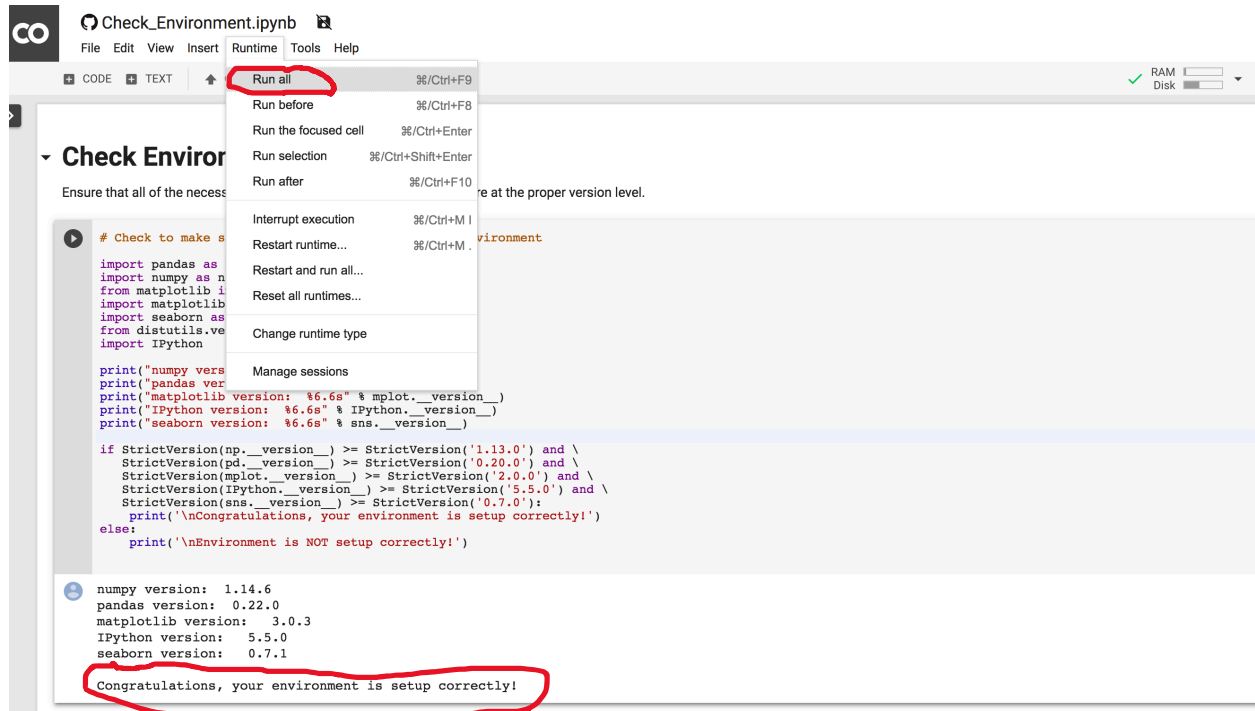
☒ Reset all runtimes before running

CANCEL RUN ANYWAY

Reset all runtimes

Are you sure you want to reset all runtimes? State of all runtimes, including all local variables and files, will be lost.

CANCEL YES



4. The final message should say “Congratulations, your environment is setup correctly!”
5. 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.
6. 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.