Challenge 0: Setting Up Your Workspace For A Successful OpenHack

Background

Before the team can begin working on Machine Learning and Data Science tasks, everyone needs to have a development environment that will work well with common Python libraries.

Challenge

Set up an environment that is conducive for Machine Learning tasks. It should include:

* Python 3.5
* Jupyter or JupyterHub access
* pip (Python package manager)
* See additional (optional) tools [here](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai" \l "optional" \t "_blank)

This environment can take advantage of cloud Data Science specific [Azure](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai#linux-dsvm) tooling or a [Local](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai#local-computer-alternative-to-dsvm-setup) Data Science setup on your machine.

In addition, the team is encouraged to set up:

* Code sharing capabilities
* A mode of team communication

Data Science Virtual Machine for Linux Ubuntu CSP

Ubuntu Data Science Virtual Machine (DSVM)

This setup has been found to help the team work together in a consistent environment.

We’ve commonly found the following setup to work very well:

* Ubuntu Data Science Virtual Machine (DSVM)
  + OS: Ubuntu
  + Size: *DS12 v2* Standard (4 cores / 28.00 GiB RAM / 56 GiB Temporary Storage) - may show up as *CSP*
  + Region: (Ask your coach for the appropriate region for your OpenHack)
  + Authentication type: *Password* (critical to logging in)
  + This will also include:
    - Python 3.5
    - Jupyterhub
  + Setting up one DSVM for the whole group and logging in with Jupyterhub is best to foster collaboration and consistency - ask your coach about your options
  + See [References](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai" \l "references" \t "_blank) for more guidance and help
* See data download instructions [here](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai" \l "data-downloads" \t "_blank)
* Determine whether any [optional](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai" \l "optional" \t "_blank) installs should be added to team members’ environments

Local Computer **Alternative to DSVM setup**

* Install Anaconda if you don’t have it for your system:
  + Installation information [Here](https://docs.anaconda.com/anaconda/install/" \t "_blank)
  + Create an environment with Python 3.5: conda create -n py35 python=3.5
  + Activate the environment. Windows: activate py35 Unix/Linux: source activate py35
  + You will be able to pip or conda install packages into this environment as needed going forward
* If *not* using Anaconda, but rather a system Python install, use the Python package manager (pip) to at least install Jupyter:
  + pip install jupyter
* Install other Data Science packages as the need arises
* See data download instructions [here](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai" \l "data-downloads" \t "_blank)
* Determine whether any [optional](https://openhacks.azurewebsites.net/labs/player/microsoft-open-hack-ai" \l "optional" \t "_blank) installs should be added to team members’ environments

Success Criteria

* Run 2 code cells, one with each of the following command blocks to ensure successful setup.

**Code cell 1**

On a Data Science Virtual Machine:

import sys

! {sys.executable} -m pip freeze

! {sys.executable} -m pip --version

Or if on a Local Setup:

! pip freeze

! pip --version

**Code cell 2**

Run this Python code:

import sys

sys.version

References

Ubuntu DSVM

* Create a Linux Data Science Virtual Machine (DSVM) and use JupyterHub to code with your team - [Video](https://www.youtube.com/watch?v=4b1G9pQC3KM" \t "_blank) or [Doc](https://docs.microsoft.com/azure/machine-learning/data-science-virtual-machine/linux-dsvm-walkthrough/?wt.mc_id=OH-ML-ComputerVision#jupyterhub)

**Important - Please Read**

* When provisioning the Ubuntu DSVM, ensure the team chooses *Password* as the Authentication type
* It’s recommended to *not* use Edge, but use a different browser
* The Jupyterhub is at an address that begins with **https** protocol
* To get to the Jupyterhub, one must click through the non-private connection warnings in browser - this is expected behavior
* The Jupyterhub is at port **8000** as the Video and Docs say - links above
* Use the **“Python 3”** kernel
* To install Python packages, example commands for the Ubuntu DSVM are shown in the [Docs](https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/linux-dsvm-walkthrough#jupyterhub)

Data Downloads

* For the cloud setup, with the DSVM, a convenient way to download the data is through OS commands within a Jupyter notebook, e.g.:

! curl -O https://challenge.blob.core.windows.net/challengefiles/gear\_images.zip

* For the local setup, download the gear dataset [by clicking here](https://challenge.blob.core.windows.net/challengefiles/gear_images.zip)

Optional

* Git [Download](https://git-scm.com/downloads)
* Azure ML CLI [Install](https://docs.microsoft.com/azure/machine-learning/preview/deployment-setup-configuration?wt.mc_id=OH-ML-ComputerVision)

Using installed tools

* Getting started with conda [Doc](https://conda.io/docs/user-guide/getting-started.html)
* Creating and activating a conda environment [Ref](https://conda.io/docs/user-guide/tasks/manage-environments.html" \t "_blank)
* Connecting a Jupyter Notebook to a specific conda environment [Ref](http://ipython.readthedocs.io/en/stable/install/kernel_install.html" \l "kernels-for-different-environments" \t "_blank)