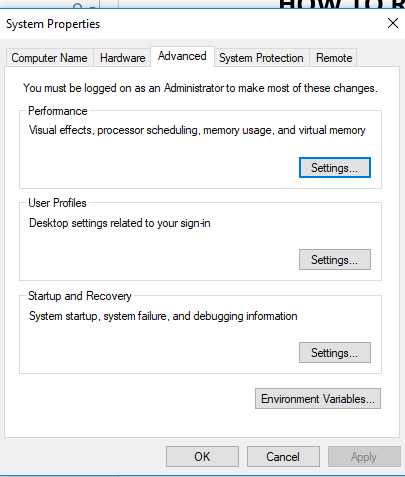
**HOW TO RUN**

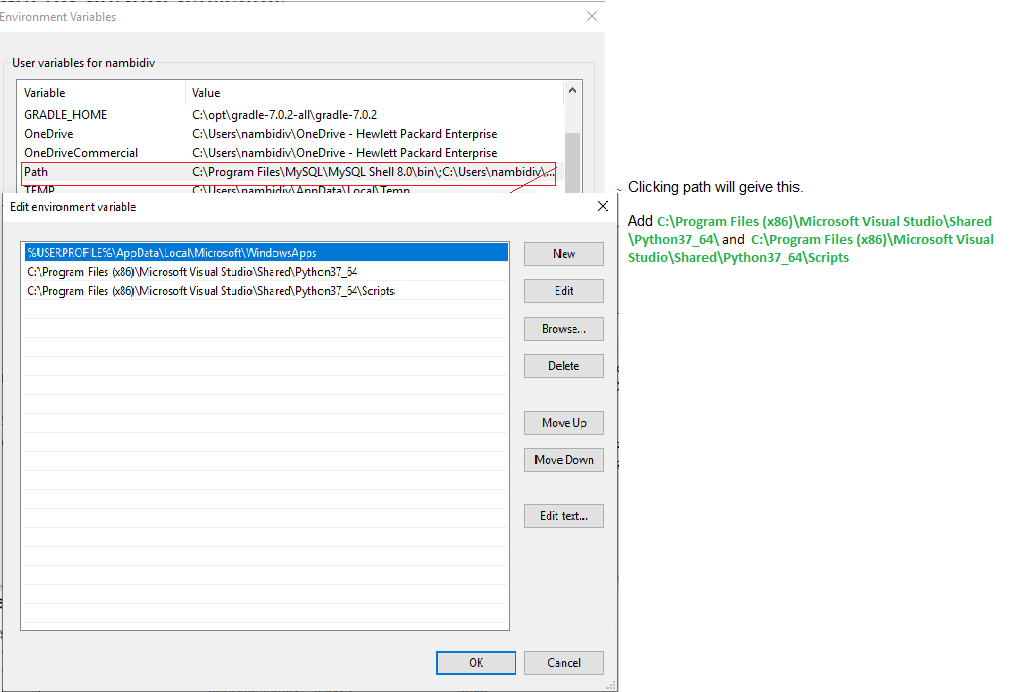
1. Install python 3.7.x or above and add python path in system environment variable.

Download python form <https://visualstudio.microsoft.com/vs/professional/> [vs\_professional\_\_301478306.1627117146\_\_rid.30004.exe] and install python from same visual studio, as C++ libraries required for some for python packages. Generally it will get installed here: C:\Program Files (x86)\Microsoft Visual Studio\Shared\Python37\_64\

Add this in System Environment variables as below. [Please check the installed directory and add the python path from your computer. Path can vary based on exes].



Add the python and script path a below.



1. Got to cmd ( or power shell, Gitbash etc)

Just type ‘where python’. It should show you below output if you set the environment variable correctly.



Also ‘where pip’ should you the below output.



[If you are using other OS, or git bash in any OS including windows the commands are which python and which pip].

1. Run the command ‘pip install -r requirements.txt’.
2. Unzip DataCleaningSolution
3. cd <path where you copied DataCleaningSolution>/DataCleaningSolution
4. Run ‘python DataCleaning.py’
5. This will generate ‘SIS\_Faculty-List-cleaned.xlsx’.

Also we can open DataCleaning.ipynb in any Jupiter notebook or google [Colaboratory and run each code by clicking ctrl + enter](https://colab.research.google.com/notebooks/)

<https://www.upgrad.com/blog/data-cleaning-techniques/>

<https://towardsdatascience.com/clean-your-data-with-unsupervised-machine-learning-8491af733595>