## Introduction:

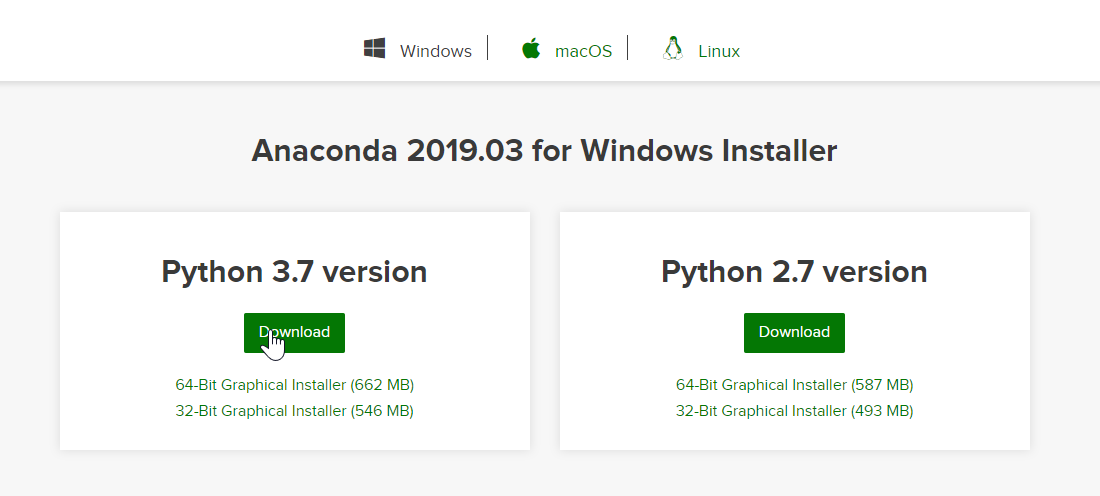
The purpose of this document is to walk through the steps

* To install and set up a python environment (with Anaconda distribution)
* To install the relevant python packages (python-snowflake connector & python snowflake-sqlalchemy connector)
* To execute the code from jupyter notebook
  + Note: It is possible to download the jupyter notebook as a native python file and execute in other ways as well

## Pre-Requisites:

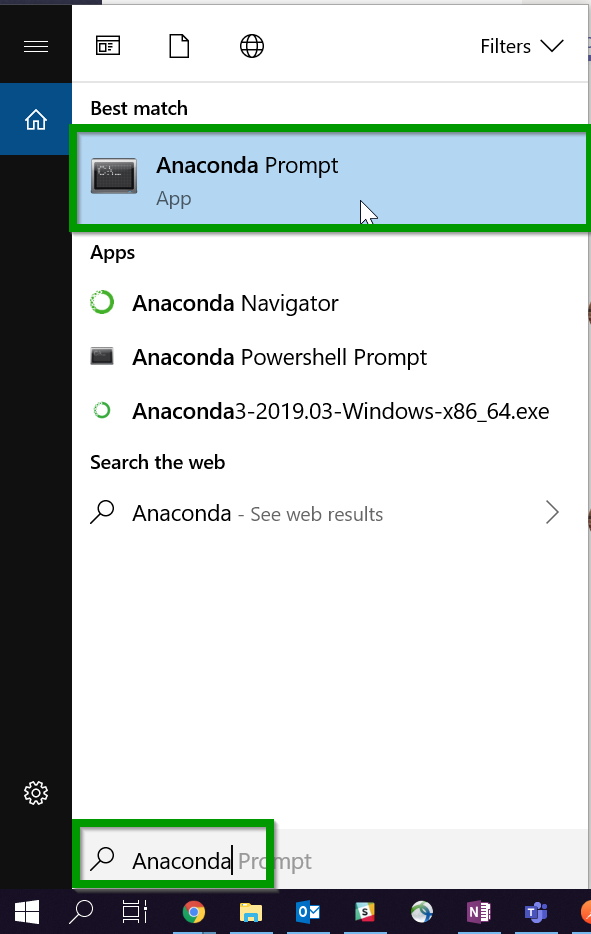
### Anaconda Installation

1. Install Anaconda distribution of python from <https://www.anaconda.com/distribution/>
2. Select python version 3 as python version 2 is slowly getting phased out. Select the appropriate installer (32 bit v/s 64 bit) based on the system configurations.



### Python Packages Installation (Snowflake & SQLAlchemy)

1. Once the installation for Anaconda is complete, start the Anaconda prompt
   1. Click on windows search and type Anaconda, select Anaconda prompt from the search results



1. Execute the following command to install the python snowflake connector – “**pip install --upgrade snowflake-connector-python**”
2. On successful installation we should see a message stating that the package was successfully installed



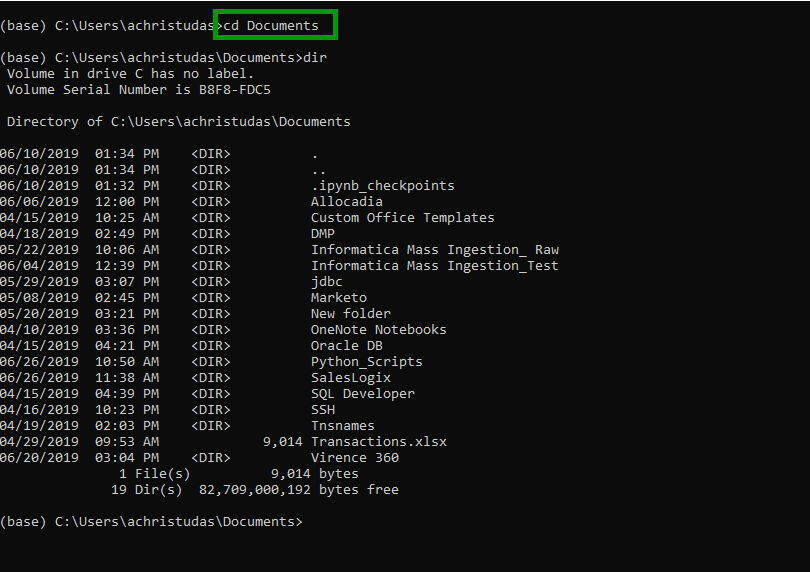
1. Now, execute the following command to install the python sqlalchemy connector =- “**pip install --upgrade snowflake-sqlalchemy**”
2. On successful installation we should see a message stating that the package was successfully installed



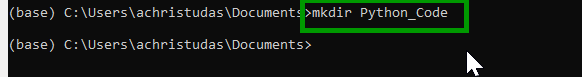
1. This completes the pre-requisites of the tools and the packages needed to execute the data ingestion code.

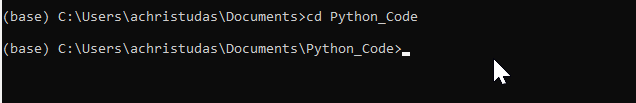
### Python Directory set up

1. Start Anaconda prompt and CD to a directory of choice (e.g. My Documents)

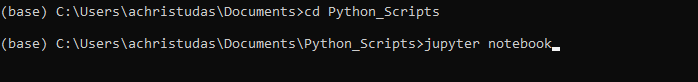


1. Create a new directory or chose one of the existing directories to store all the python code
   1. To create a new directory execute the following command – “mkdir <directory/folder name>. E.g. mkdir Python\_Code

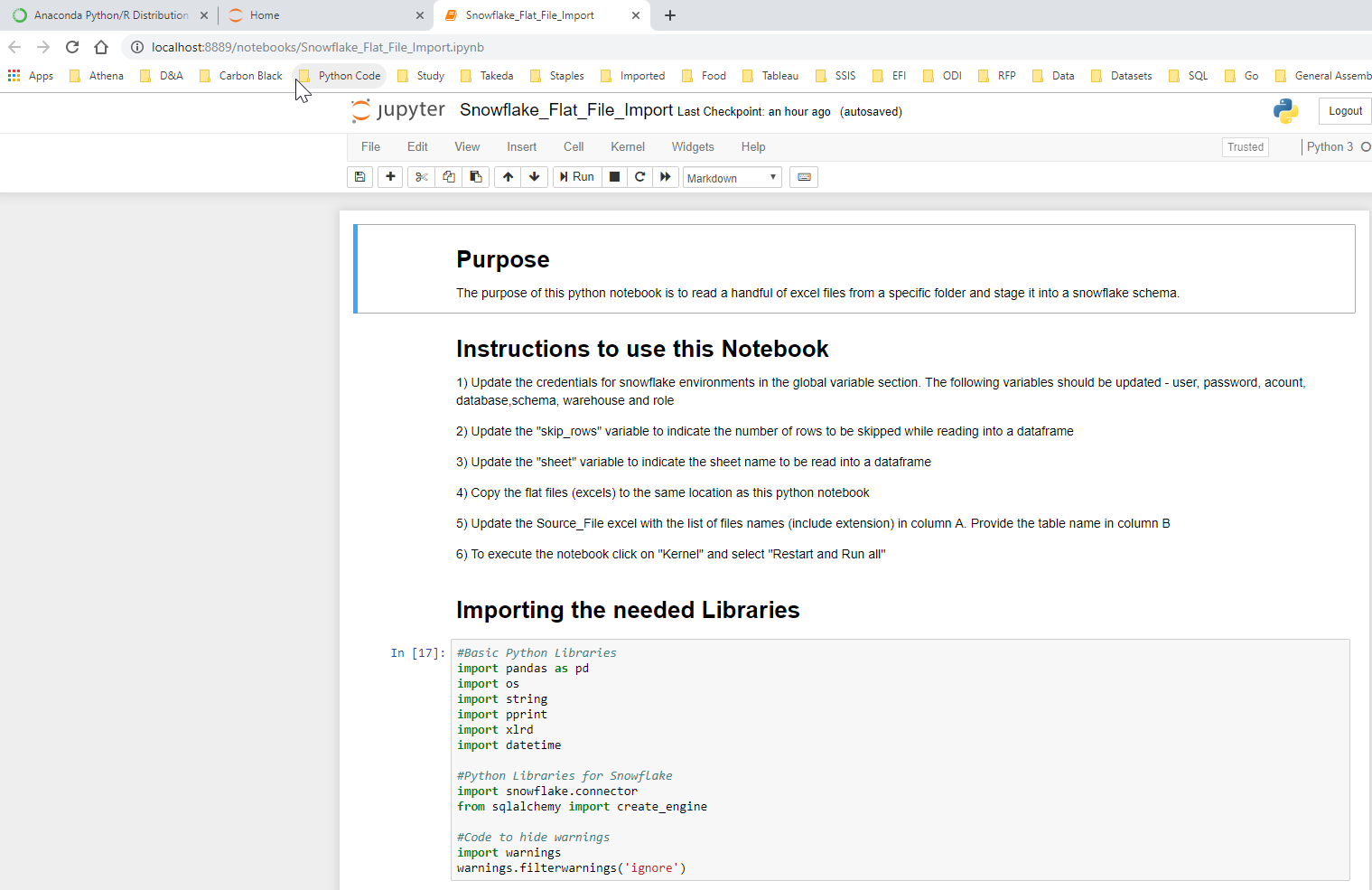


* 1. To navigate to an existing directory, execute the following command – “cd <directory/folder name>. E.g. cd Python\_Code
  2. 

1. Copy the Ipython notebook and/or the python file into this folder (Snowflake\_Flat\_File\_Import.ipynb)
   1. The copy operation need not be done via command line, just use the regular windows explorer ctrl+c, ctrl+v method 😊
2. Copy the Source\_File.xlsx into the same folder location
3. From the anaconda prompt execute the following command – “jupyter notebook”



1. This should open up the browser window with the contents of the folder. Select the relevant ipython notebook by clicking on it.



1. Copy the excel files to be migrated into snowflake into the same folder location
2. Open the Source\_File.xlsx, and update column A with the list of files to be migrated (use the extension as well). Update column B with the corresponding tables names to be used

## Code Execution

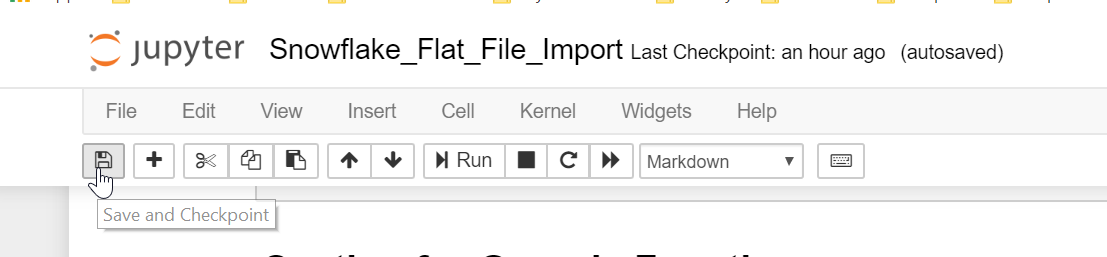
### Update Global Variables

1. In the Ipython notebook update the following global variables for the Athena Snowflake account
   1. User - <Provide Snowflake user name>
   2. Password - <Priovide Snowflake password>
   3. Account - <Provide Snowflake account name. e.g. athenahealth>
   4. Database - <Provide Snowflake database name>
   5. Schema - <Provide Snowflake schema name>
   6. Warehouse - <Provide Warehouse name>
   7. Role - <Provide Role Name>

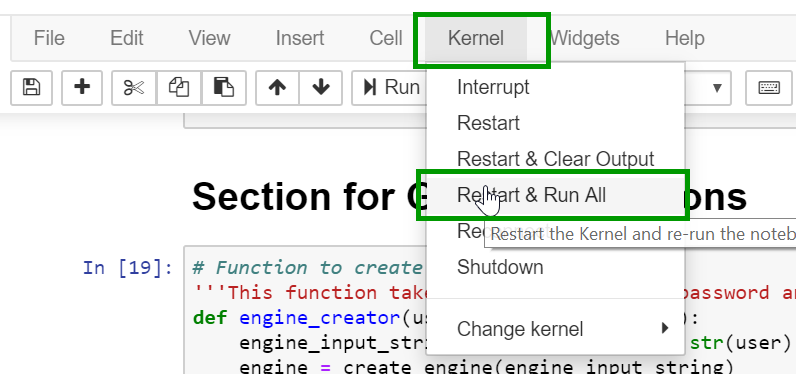
1. Update the following variables related to the excel file
   1. Skip\_rows – <Provide the number of rows to skip before reading the data>
      1. E.g. In some cases, if the data starts only in row 10, provide 9 as the value for this variable. If the data starts in row 1, provide 0 as the value for this variable
   2. Sheet - <Provide the name of the sheet to be read from the excel file>
   3. **Note:** *It is important to standardize the sheet name for all the files that are getting loaded, and also standardize the rows at which the data is starting.*

### Execution

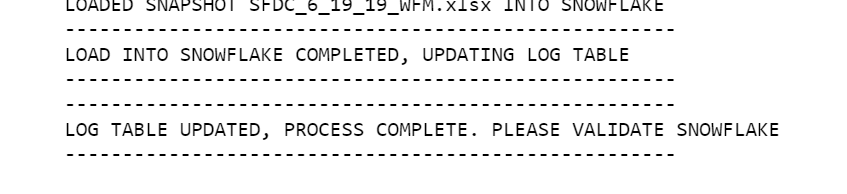
1. Save the changes to the notebook



1. Click on Kernel and select “Restart and Run All” to run the notebook



1. The process loops through the list of files provided in the Source\_File.xlsx and loads them into snowflake. The print statements in the code, lets us know the table that is getting processed.
2. At the end of the execution, the following message is displayed



1. This message signals the end of the process and that all the files provided were migrated to snowflake.

### Validation

1. Log into snowflake with the appropriate credentials. Navigate to the appropriate database. Query the tables under the appropriate schema
2. The process also creates an audit log table – “IMPORT LOG” to store basic log metrics like
   1. Source\_File Name
   2. Source File Count
   3. Target Table Name
   4. Target Table Count
   5. Load Datetime
3. The log table is in the same schema were the files are getting migrated.