Assignment Run Book

Pre-requisites and Installations

The document is prepared with respect to the RedHat Data Engineer Assignment.

Git Hub Link: ANKITBISANI/Comics-Assignment: Comics Assignment (github.com)

(**Download the branch on to desktop)

Tools/Technologies:

- Anaconda Jupyter Notebook (Python 3.7)
- MySQL (installation with root user on the desktop for localhost)
- Python IDLE (3.7) (as an alternate)

Modules used and installation:

- Requests
- Random
- Json
- PyMySQL
- Sys

If any of the module gives the error "No Module found error: $module name "after importing "import $module name", please install on windows:

Py -m pip install $module name

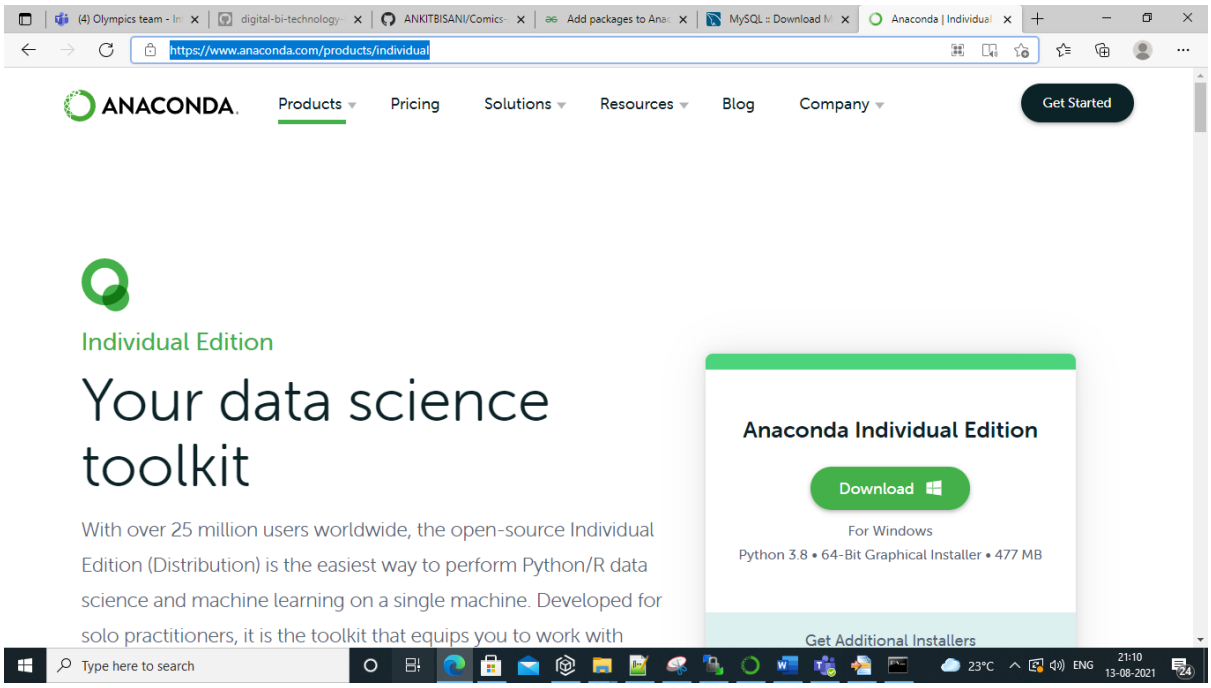
If using Anaconda(recommended):

(Please note MySQL module is not defaulted in Anaconda Jupyter and can be downloaded by below method. Other Modules mentioned are by default)**

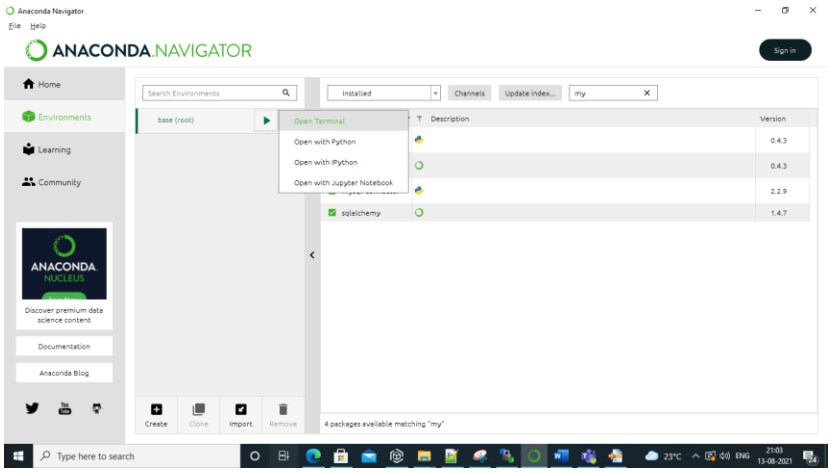
If you are using Anaconda Jupyter Notebook, follow below steps:

Steps 1: Download Anaconda individual addition

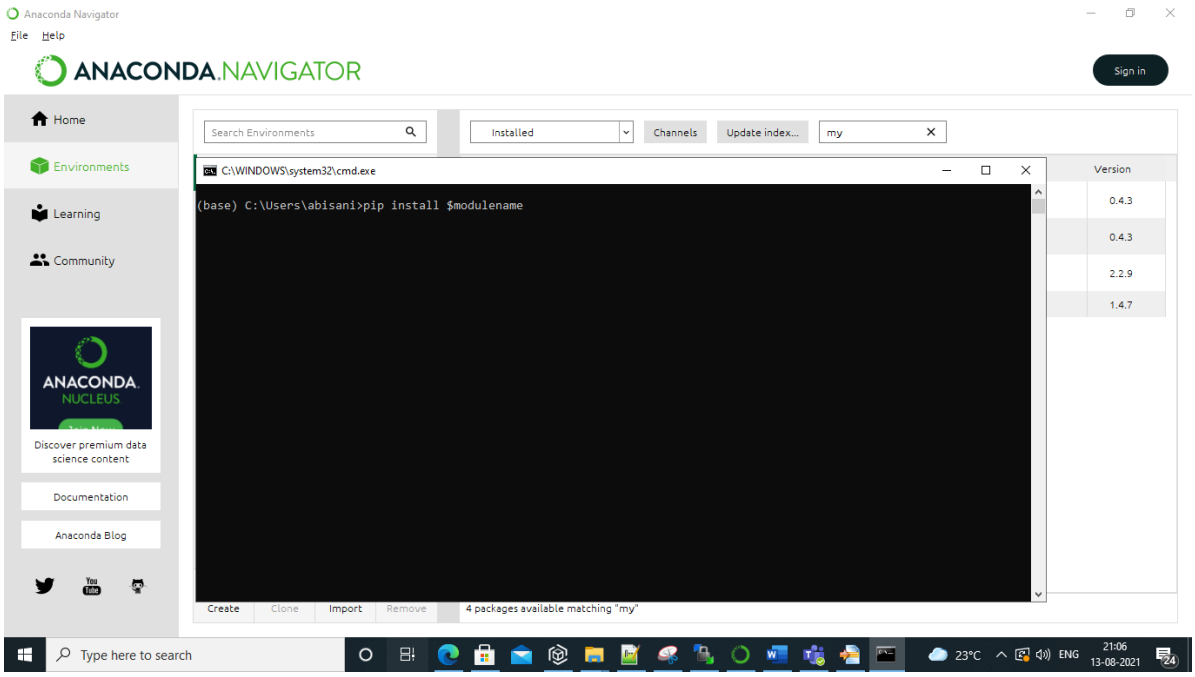
Anaconda | Individual Edition



Step 2: Go to Environments



Step 3: Install the modules as in screenshot



MySQL Installations:

Please Download: MySQL :: Download MySQL Installer from the link and follow the installation instructions

Open the MySQL workbench with the root password

**Please remember the root password as that would be used in scripts as per assignments

Solutions (1 &2)

1. GET 15 random comics and following details in using Python. a. Get the names of the comic b. Get the alt-text of the comic c. Get the number of the comic d. Get the link of the comic e. Get the image of the comics f. Get the image Link of the comics

2. Insert into MySQL - Please include SQL database schema(s) for any table(s) created in your GitHub repo

Step 1: create schema and user from root password on MySQL workbench

---

Create schema comics;

CREATE TABLE `comics_details` (

`names` varchar(500) DEFAULT NULL,

`alt_text` varchar(500) DEFAULT NULL,

`number` int DEFAULT NULL,

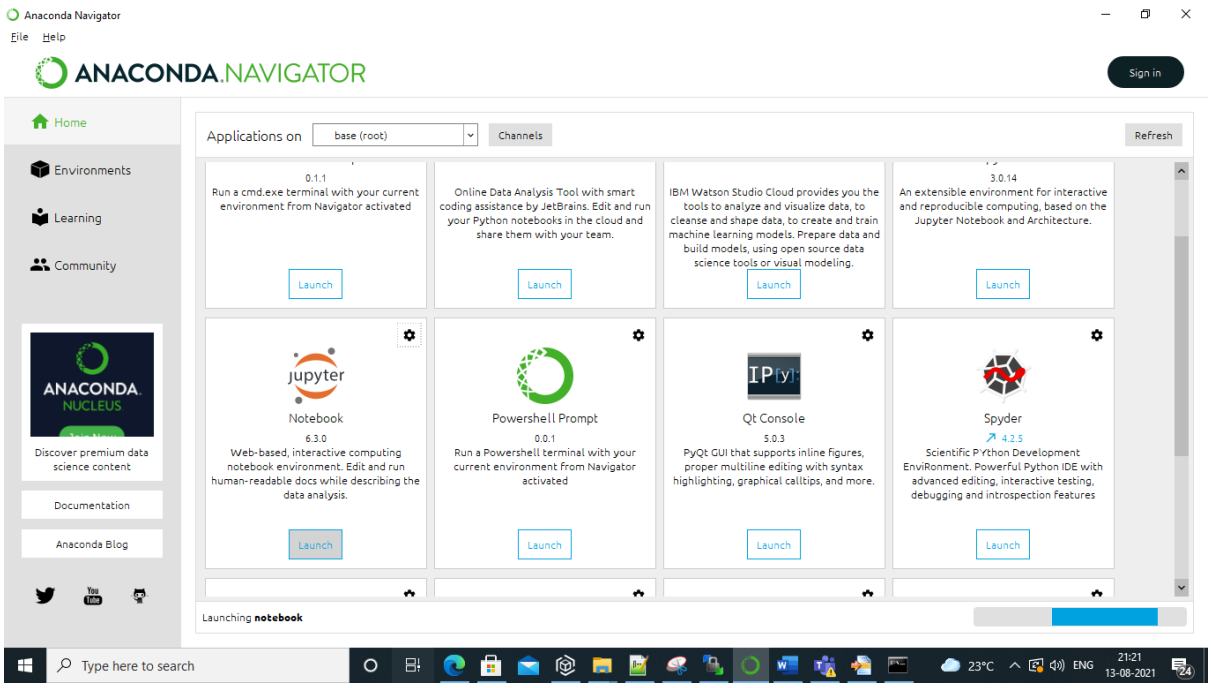
`link` varchar(500) DEFAULT NULL,

`image` varchar(500) DEFAULT NULL,

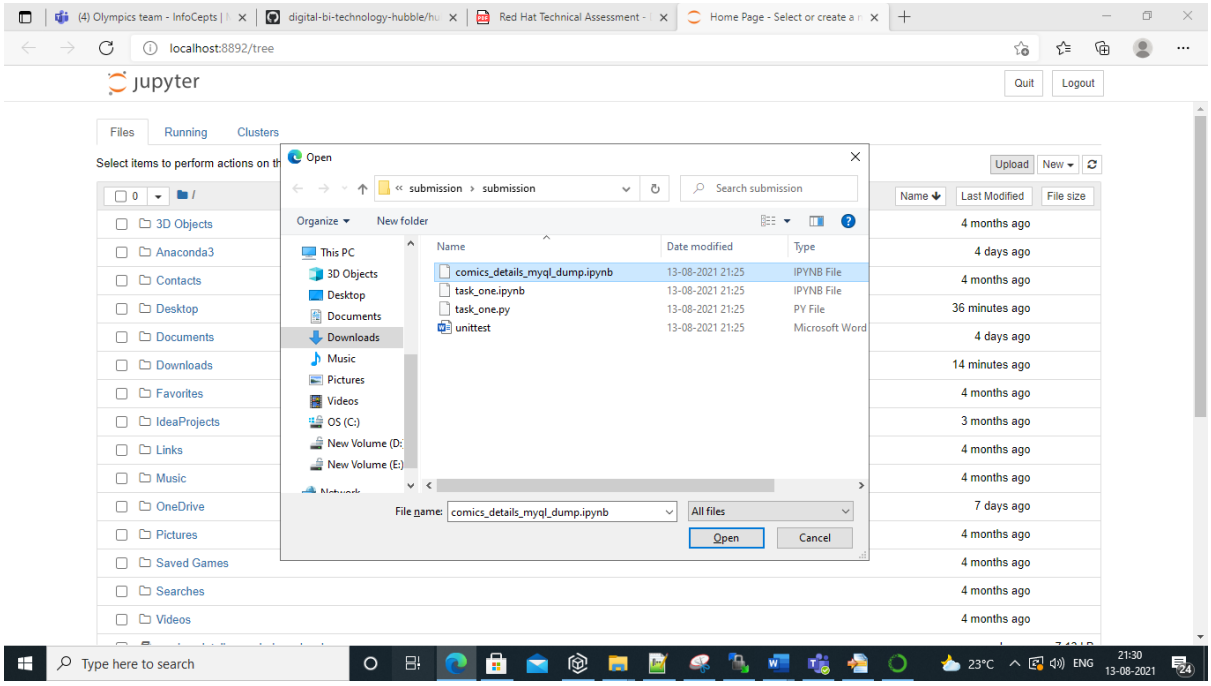
`image_link` varchar(500) DEFAULT NULL

) ;

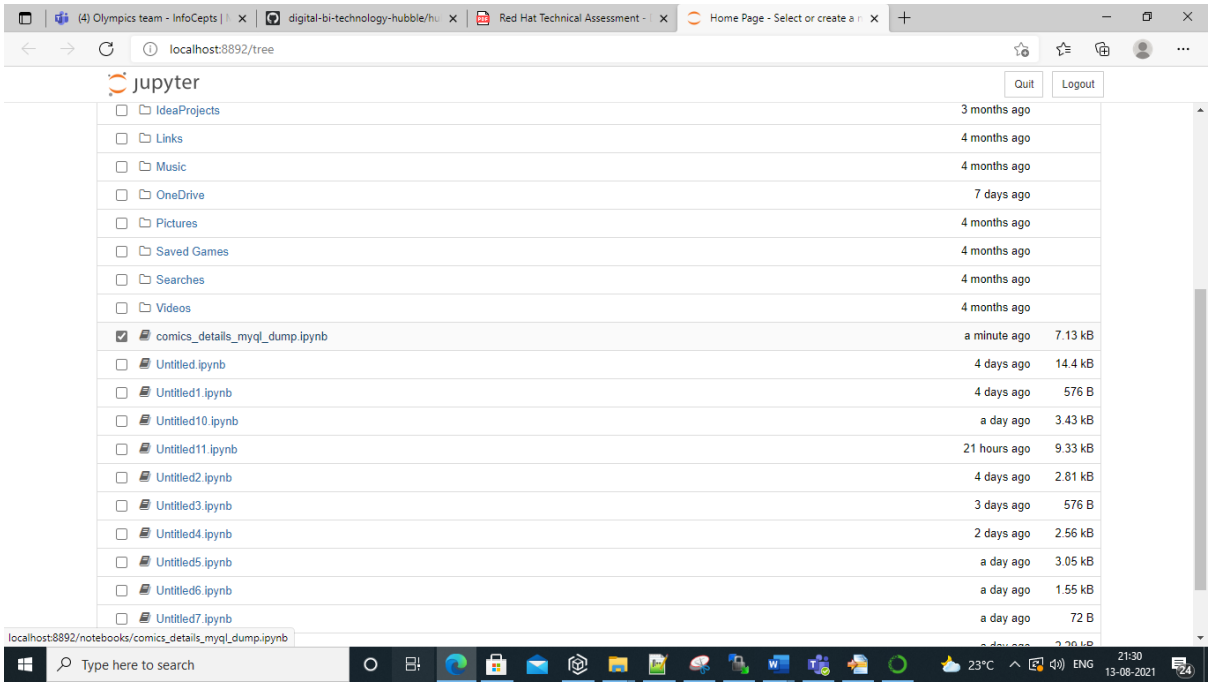
Step 2 : Launch Jupyter Notebook



Step 3 : Uplaod the downloaded file : comics_details_myql_dump.ipynb



Step 4: Open the file: comics_details_myql_dump.ipynb



Step 5 : Replace the password and replace with localhost root password in the file and run the file

conn = pymysql.connect(

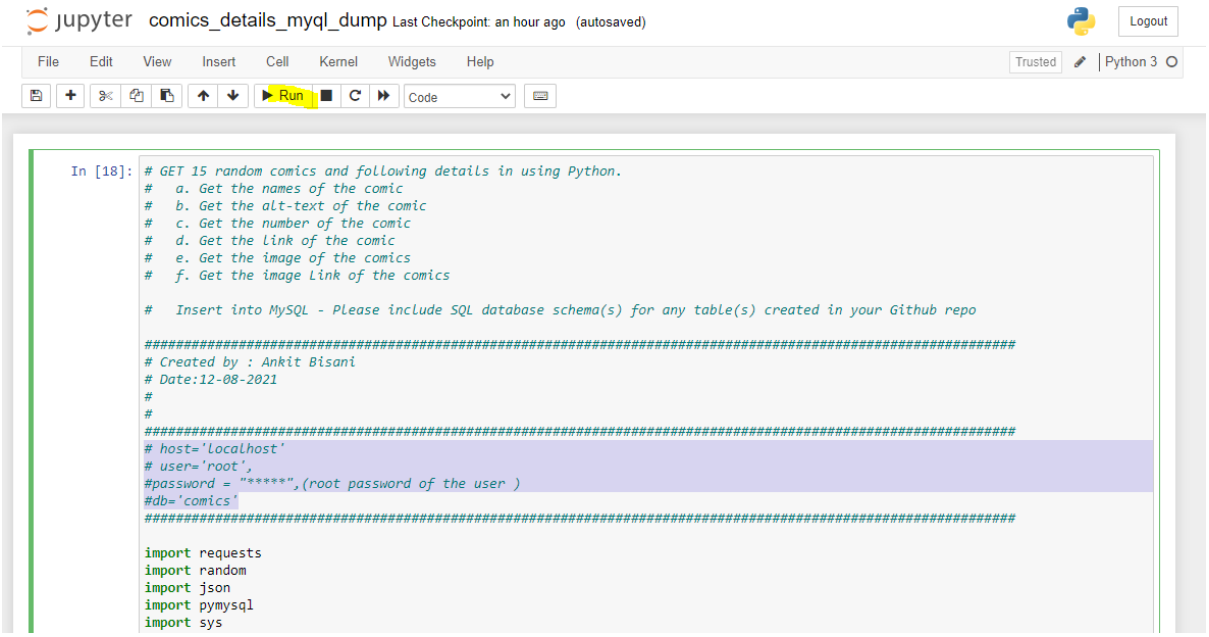
host='localhost',

user='root',

password = "March@1988", ##Need to replace

db='comics',

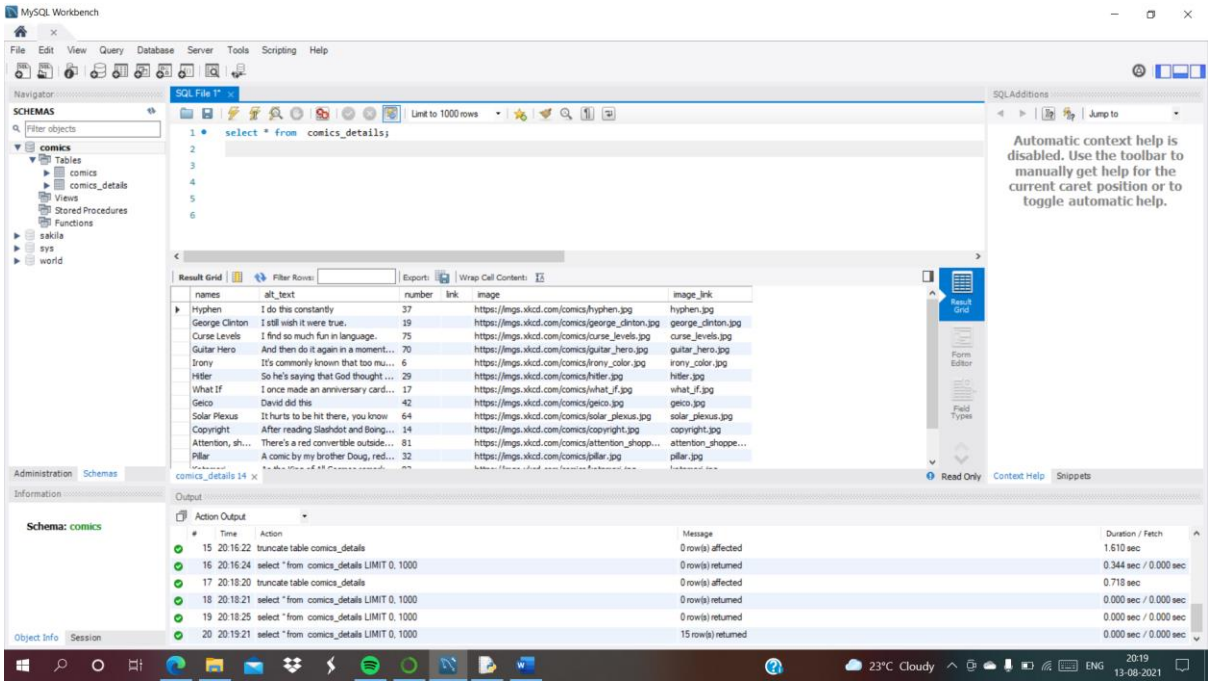
)



Step 6:

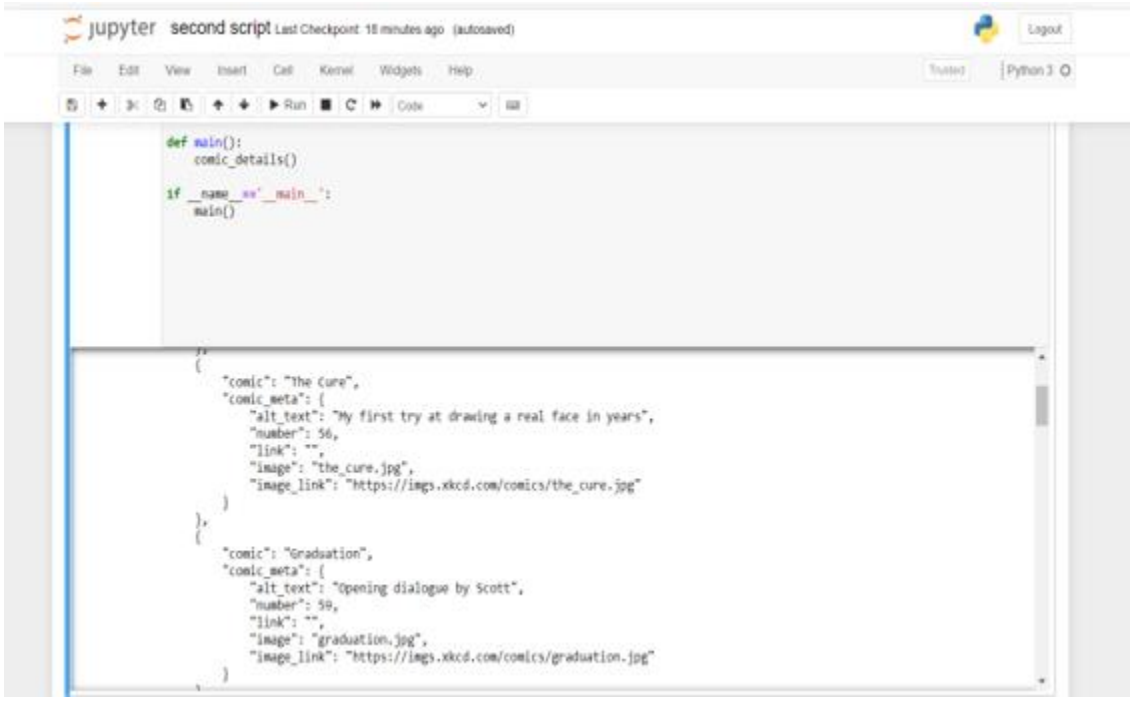
After successful run , go to MySQL workbench and login with root creds and check the table

select * from comics .comics_details :



3. Write a script called task_one.py that when called will output something like this to the console. (reference below)

Repeat the above steps 1-5 for task_one. ipynb and observe the results:



Alternatively, you can open task_one.py in Python IDLE (3.7) and run the scripts