**Web Scrapping Tool**

**Objective:**

For given full name and email address of a person, build a tool in Python that identifies the following information:  
- gender  
- degree(s) earned (MS/BS/ BA, etc)  
- universities/ schools attended

**Softwares required:**

* Postgresql
* Python
* Matplotlib and beautifulsoup libraries

**Tool Description:**

To develop this tool, python script (web\_scrapy. pi) is written to get personal information of a person. A URL of google sheet is opened using urllib library. It reads data in HTML format. Then, full name and email address are taken as an input from the user. After that, python’s another library called beautifulsoup is used to find gender, degree and university data.

Beautifulsoup uses an HTML parser to parse HTML tags and content to extract information. Generally web scraping tool uses the same technique to get useful or required information from a webpage. So by using “find” function of beautifulsoup, important information is fetched.

If the sheet is having provided name and email address then this tool gather all the data related to that person, but it doesn’t match with provided name and email address then it will show “match not found error” and ask the user whether he/she wants to fetch more information or not. It handles one more exception, i.e. if given name and email address mismatch(handled cross validation), then also it will show the same error as above.

Once data is fetched, the tool will run an SQL query to store data in Postgres DB. Here, connection was opened in the beginning of code to eliminate the connection establishment in every iteration. The Database is having “id” as the primary key and it is automatically getting incremented with the insertion of new rows.

The tool generates a pie chart from stored data once user doesn’t want to fetch more data from sheet.

**Flow chart:**

Parse google sheet link

Get name and email address as an input from user

Try to find gender,highest degree earned and university information of the person in sheet using beautifulsoup library

Is information found?

NO

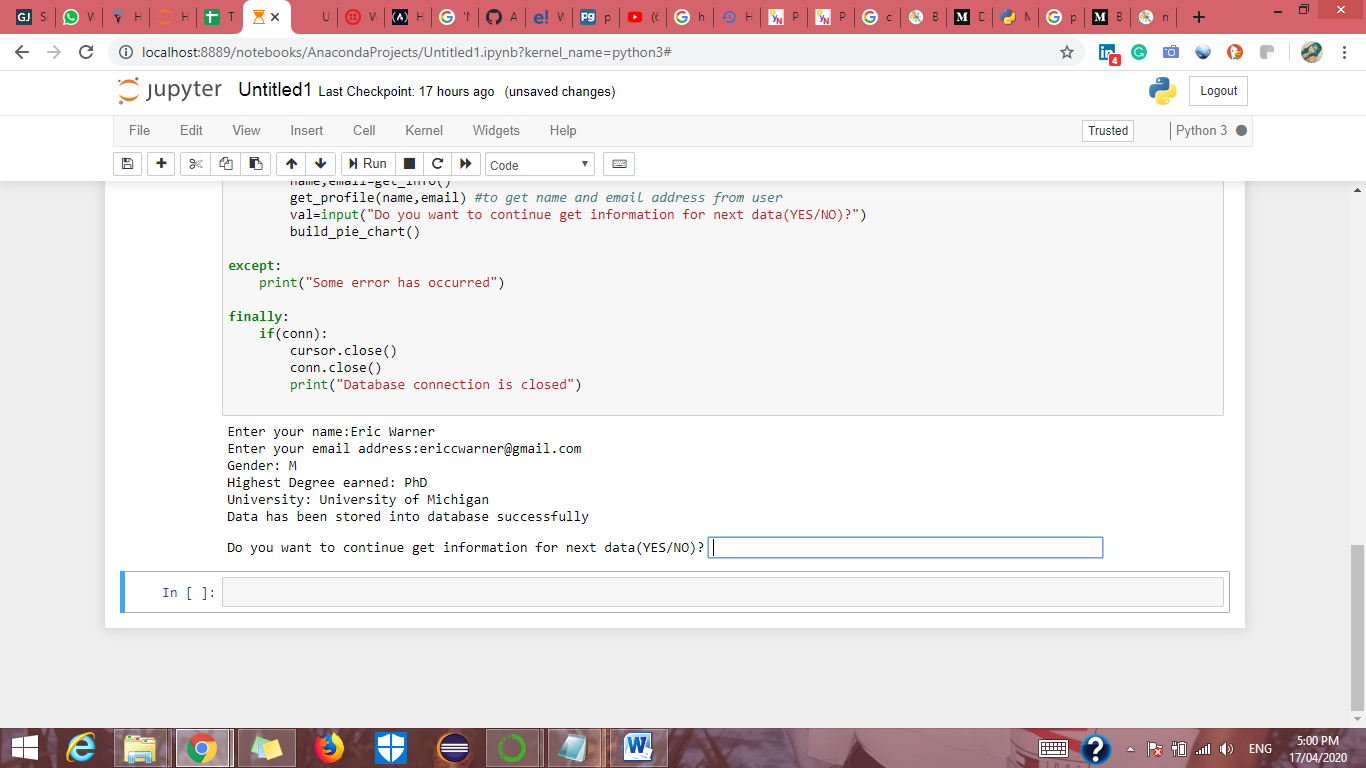
YES

Print and store data into Database

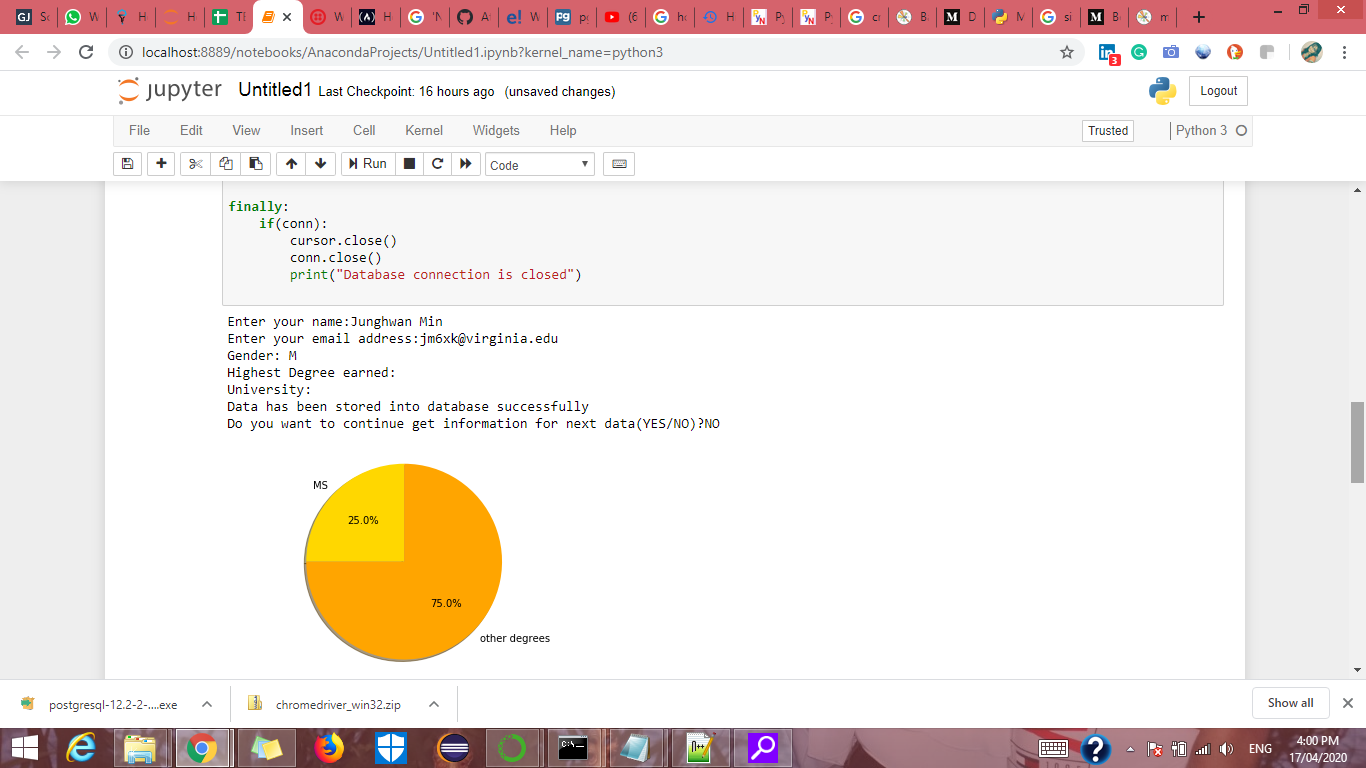
Build pie chart to know the people having “MS” as highest degree

**Output:**

**Code output on terminal:**



**Pie chart of people with a Master’s degree:**



**Data stored in Postgres DB:**

