Practice 2

Nguyễn Trần Minh Quang - 20127298

Contents

1 Progress

	Task	Progress
1	Read data from the given Covid-19 cases CSV file	100%
2	Visualize data with atleast 3 graphs using matplotlib	100%
3	Give some comments, analytic… from the corresponding graphs	100%

2 The program

2.1 Data input

- Covid-19 cases CSV files
- Major samples files

2.2 Data output

- Vietnam's covid cases and death graph
- Covid case by country pie chart
- Men and women in each major category chart

2.3 Usage

1. Navigate to the source folder
2. Create a virtual environment

```
$ virtualenv venv
```

3. Activate the virtual environment

```
$ source venv/bin/activate
```

4. Install the required dependencies

```
# On windows or osx
$ pip install -r requirements.txt
# On linux
$ pip install -r requirements_linux.txt
```

5. Run the program

```
$ python main.py
```

2.4 Output graph

2.4.1 Vietnam's covid cases and death graph

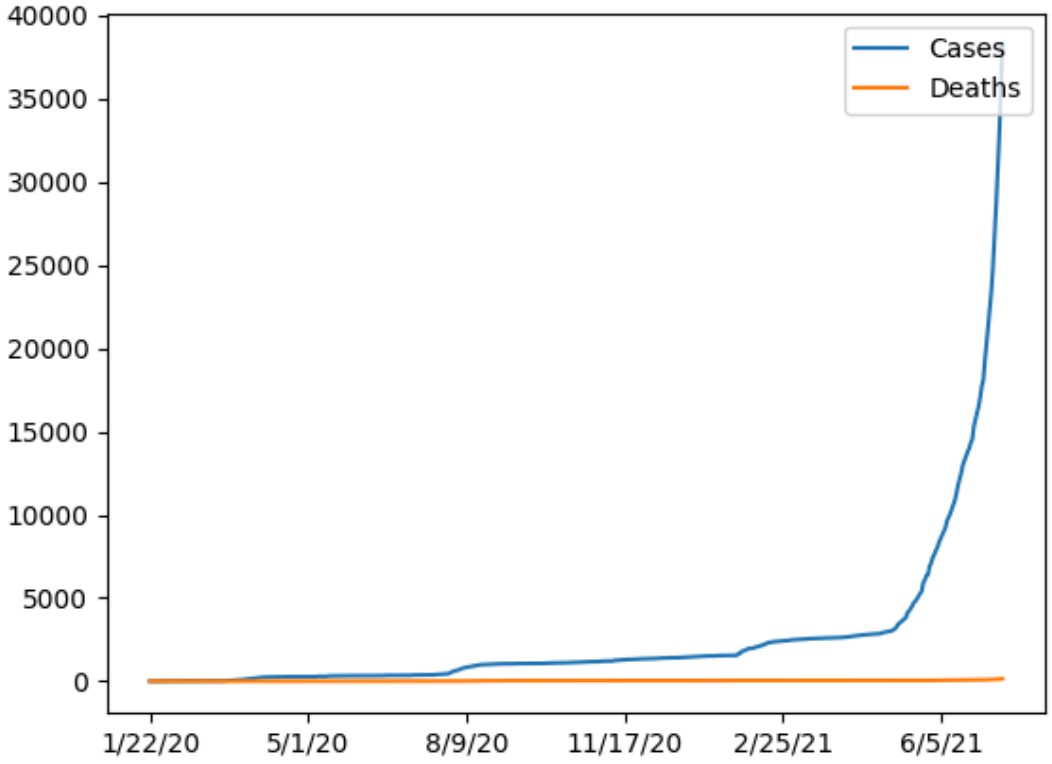

Figure 1: Vietnam's covid cases and deaths from Jan 22nd 2020 to May 6th 2021

2.4.1.1 Implementation

1. Read the content from covid cases csv and covid deaths csv using `pandas`
2. Filter row with the `Country/Region` column is `Vietnam`
3. Concat 2 data frames
4. Plot it

2.4.1.2 Comment

The graph surge around August 2020, January 2021 and April 2021 also indicate the 3 covid "waves" in Vietnam.

2.4.2 Covid case by country pie chart

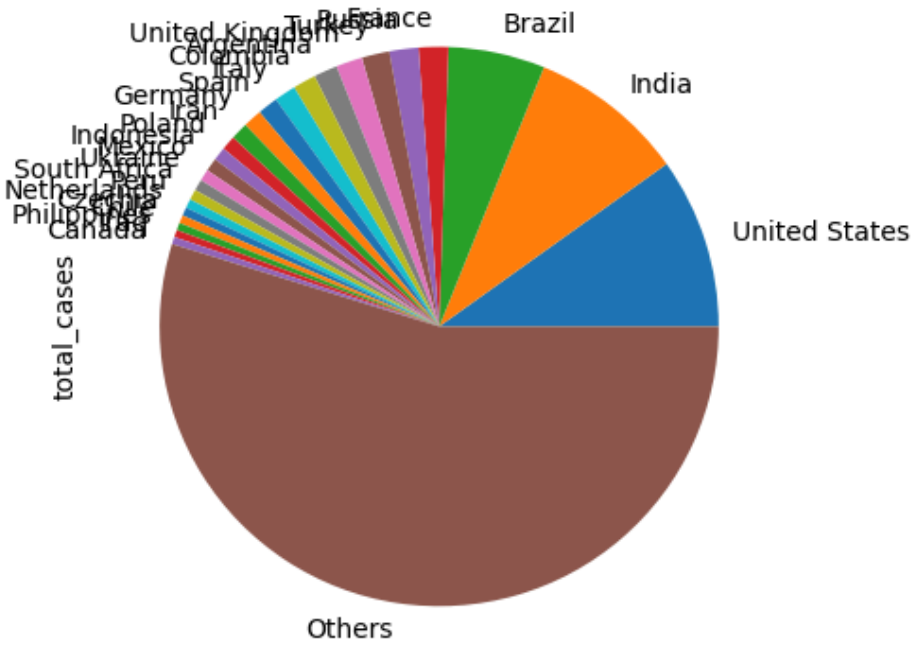

Figure 2: 25 countries with most covid cases

2.4.2.1 Implementation

1. Read the content from the latest covid data csv using `pandas`
2. Filter where the entry is not a country since this data also contains sum of continent and world data
3. Sort by `total_cases`
4. Get the top 20
5. Sum the other's country `total_cases` to the `Others` index
6. Concat it with the top 20 data frame
7. Plot it

2.4.2.2 Comment

Most of the cases on the later of the pandemic is mostly reported by the USA and India.

2.4.3 Men and women in each major category chart

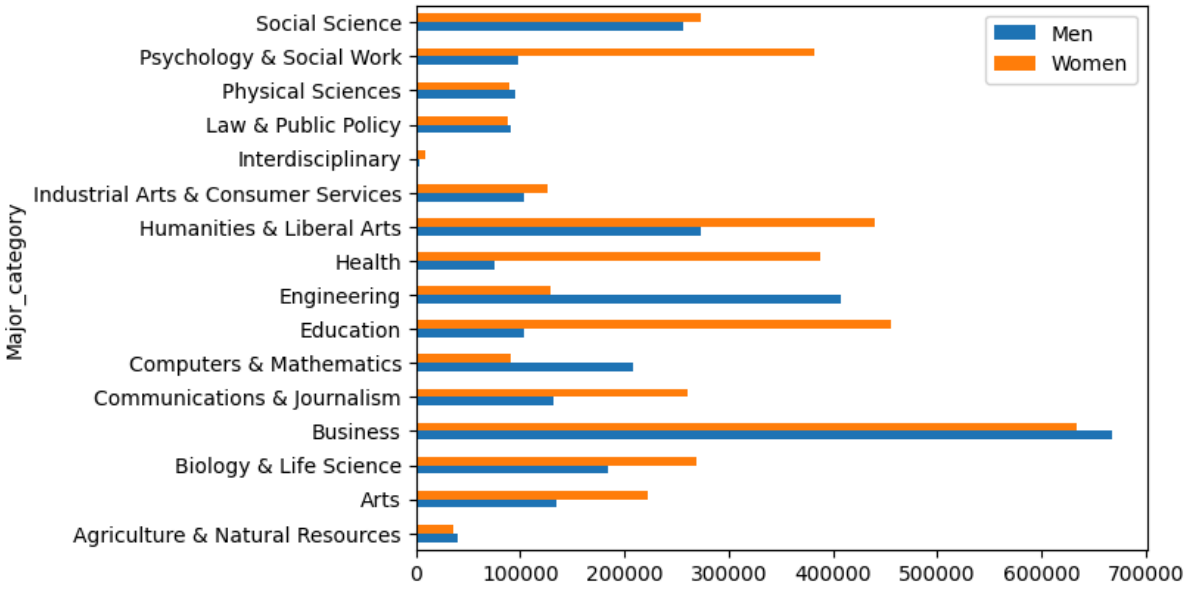

Figure 3: Men and womens distributions among top major categories

2.4.3.1 Implementation

1. Read the content from majors csv using `pandas`
2. Group by `Major_category` and get the sum of `Men` and `Women` columns
3. Plot it

2.4.3.2 Comment

There is an imbalance in the engineering-related fields which men had way higher ratio then women, also on social science-related field women had way higher ratio then men.