USER GUIDE

LEI YANG[1] (LY888) & YIAN ZHANG[1](YZ2777)

CONTENTS

LIST OF FIGURES

PROJECT DESCRIPTION

Allow users to input their SAT scores (math, writing and critical reading) or ACT scores (English, math, writing and cumulative) to find their available US national universities. After entering either scores, our program will first output the top 10 available universities (if applicable), and then by choosing one school, the program will plot the school location on the US map in the browser, plot the proportion of the undergraduate students with age above 25 verus 25 or below, plot the world rankings and the US rankings from 2014 to 2016, plot the linear regression trend of criminal number in the campus based on the data from 2001 to 2014, plot the linear trend of earnings (unit: US$) after graduating from the school 6, 8 and 10 years by quantile, and print the summarized information orderly. Our database only contains the US universities which ranked top 1000 in the world based on the CWUR data.

Data used in this program including:

- *College Scorecard data* from `https://collegescorecard.ed.gov/data/`
- *Campus Safety and Security data* from `http://ope.ed.gov/campussafety/#/`
- *Center for World University Rankings: CWUR data* from `http://cwur.org/`

[1] *Department of Environmental Health - Biostatistics Division, New York University*

1 PREREQUISTIES FOR USING THE PROGRAM

- Python 3.5 or above installed
  - Download from `https://www.python.org/downloads/`
- Packages installation
  - Packages you need: *pandas, numpy, matplotlib, statsmodels, folium*
  - To install the packages you may need *Shell* such that open *Windows PowerShell* on Windows or open *Terminal* on a Mac
  - Enter the following command:
    » pip install pandas numpy matplotlib statsmodels folium
- Files structure
  - This program contains six python files (suffix is .py), a user guide and a folder named "data" with three Excel files and one html file in it. **MAKE SURE put them in one folder if you want to run the program**.
  - Python files:
    - ***clean_data.py:*** clean and merge raw datasets.
    - ***function.py:*** contain functions to justify the user input errors and search data based on user input
    - ***information_and_plot.py:*** contain a class which includes several functions to output plots such as linear regression trends, pie charts and US map and so on
    - ***test.py:*** program to do unit tests
    - ***test_data:*** contain the functions to generate data for unit tests
    - ***main.py:*** main program
  - Data folder:
    - ***cwurdata.xlsx:*** contain the university rankings data and campus crime data
    - ***Most-Recent-Cohorts-All-Data-Elements.xlsx:*** contain the college scorecard data
    - ***complete_data.xlsx:*** complete dataset after clean and merge by above two datasets using ***clean_data.py***
    - ***college_map.html:*** the program will store the map information after running
- Internet is required in order to show the location of a university on the map

2 USE THE PROGRAM

This program is easy to use. You can use *Shell* (e.g., PowerShell on Windows and Terminal on Mac) to run it. Following shows the steps (use Windows OS as an example) :

1. Assume the files of the program are stored in a folder named "py_project" and the path is *C:\py_project*. Then, the first thing is to change your working directory to this path. So, open the PowerShell and enter the following command:
   » cd "C:\py_project\"

2. Run the main program (main.py) by entering following command:
   » python main.py

3. Then, the following sentence will show up:
   - "Please enter the socre you use. 0 represents SAT and 1 represents ACT. Finish the program by entering quit".

4. The valid anwser for above question is 0, 1 or quit. If your input is invalid, the program will ask you to input again until your input is correct.
   - Input 0, the program will ask you to input SAT scores by following order:
     a) "Please enter your SAT reading socre. The score range from 0 to 800. Finish the program by entering quit";
     b) "Please enter your SAT math socre. The score range from 0 to 800. Finish the program by entering quit";
     c) "Please enter your SAT writing socre. The score range from 0 to 800. Finish the program by entering quit".
   - Input 1, the program will ask you to input ACT scores by following order:
     a) "Please enter your ACT English socre. The score range from 0 to 36. Finish the program by entering quit";
     b) "Please enter your ACT math socre. The score range from 0 to 36. Finish the program by entering quitt";
     c) "Please enter your ACT writing socre. The score range from 0 to 36. Finish the program by entering quit";
     d) "Please enter your ACT cumulative socre. The score range from 0 to 36. Finish the program by entering quit".
   - Input quit, the program will be quited and shows following sentence:
     – "You have already quited the program. Thanks for using!".

5. After inputing the correct SAT scores or ACT scores, you may have a list of top 10 available universities that you can choose (**Figure 1 on page 5** is an example). If your scores are not good enough, the number of schools which you can choose may less than 10. Or even worse, the program will return a sentence such that "There is no available world ranking top 1000 collges available for you, please try next year. Fighting!"

6. If the program return a list like **Figure 1 on page 5**, it will continue to ask you to input the following information:
   - "Please enter the order of your favorite college. Order from 1 to 10. Finish the program by entering quit."
   - By entering an order number, for example, 1 in **Figure 1 on page 5** (i.e., Harvard University), the program will do the following analyses one by one:

a) The location of the university will show up on the map in your browser and the map is available to zoom in or zoom out (**Figure 2 on the next page**);

b) A pie chart to show the proportion of the undergraduate students with age above 25 verus 25 or below in the university (**Figure 3 on page 6**);

c) After closing the pie chart, a trend plot will show the university's world rankings and the US rankings from 2014 to 2016 (**Figure 4 on page 6**). The two lines may be overlap (e.g., Harvard is the top 1 in the US and also the top 1 in the world);

d) By closing the ranking plot, a linear regression trend plot will display the crime trend in the campus based on the data from 2001 to 2014 (**Figure 5 on page 7**);

e) By closing the crime plot, a linear regression trend plot will show the trend of earnings (unit: US$) after graduating from the school 6, 8 and 10 years by quantile (**Figure 6 on page 7**)

f) Final, a summary will be printed after you close the last figure (i.e., earnings plot). The content looks like, *" The college Harvard University has 2016 US rank 1 and 2016 world rank 1. The college is located in state MA, city Cambridge with zip code 2138. The average admission rate of this college is 0.059600 and annal cost for attending is around $59950.000000. From 2001 to 2014, the average crime number is 235.857143 with standard deviation 155.245362. For some further information, please visit the college website* `www.harvard.edu`*"*.

7. Then, the program will continue to ask you if you want to check another available schools. If so, enter the order number and the same analyses will be performed as in step 6. If not, you can quit the program by entering quit.

3 UNIT TEST

This program also includes the unit test file (test.py) to test whether the functions work properly. To do the test, enter the following command in the *Shell*:

» python test.py

Or,

» python -m unittest

4 FIGURES

	USER ORDER	INSTNM	UNITID
1	1	Harvard University	166027
2	2	Stanford University	243744
3	3	Massachusetts Institute of Technology	166683
4	4	Columbia University in the City of New York	190150
5	5	University of California-Berkeley	110635
6	6	University of Chicago	144050
7	7	Princeton University	186131
8	8	Yale University	130794
9	9	California Institute of Technology	110404
10	10	Cornell University	190415

Figure 1: An example of a list of top 10 available universities shows up on the program after the user entering the correct scores.

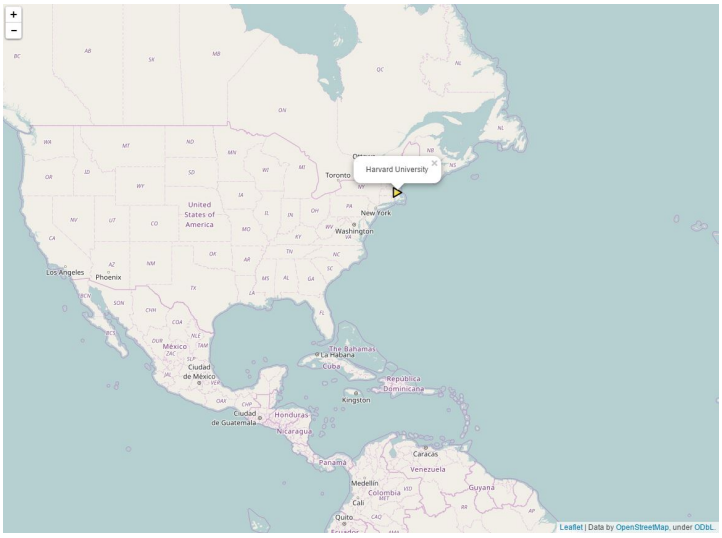

Figure 2: An example of the school location on the map

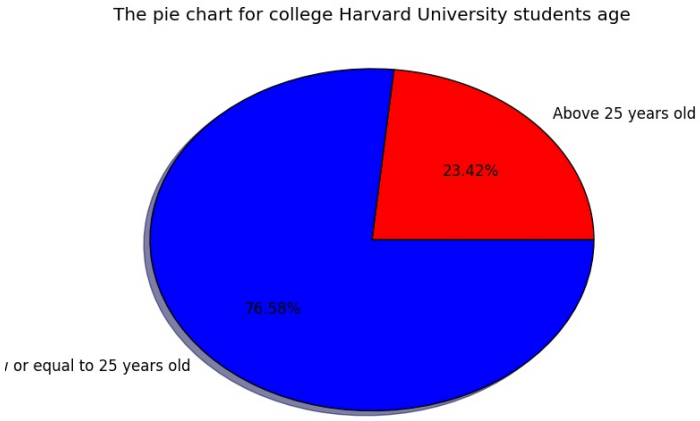

Figure 3: An example of the pie chart for the proportion of the undergraduate students with age above 25 verus 25 or below in the university

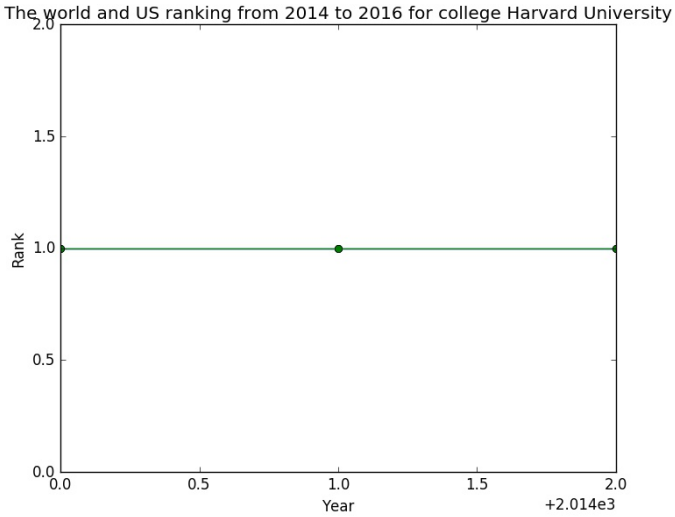

Figure 4: An example of the world rankings and the US rankings from 2014 to 2016 trend plot

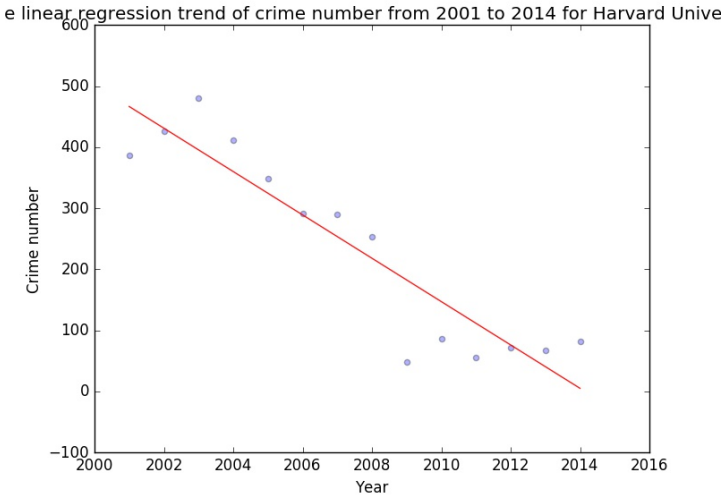

Figure 5: An example of the linear regression trend of criminal number in the campus based on the data from 2001 to 2014

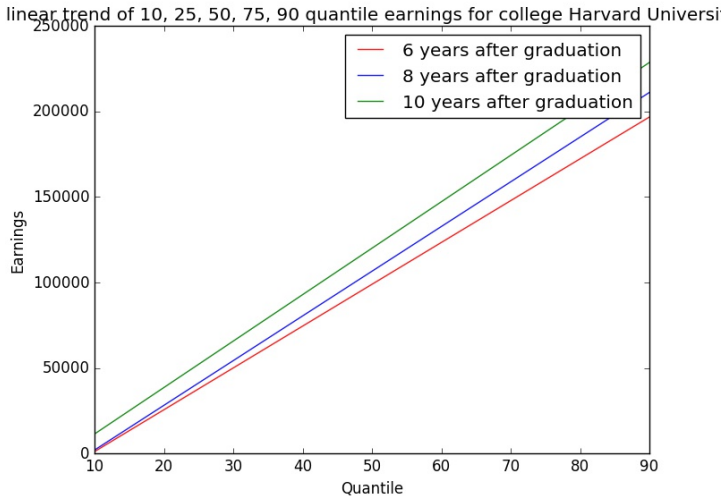

Figure 6: An example of the linear trend of earnings (unit: US$) after graduating from the school 6, 8 and 10 years by quantile