

ENSF 592 Spring 2021 – Project Report

Melissa Liao

The terminal-based application starts with prompting the user to input a country name. The country name must be any of the countries in the file above. It will then print out information from two user-defined functions about the chosen country for years 2005, 2010 and 2015. It also includes a brief summarized information of the chosen country's sub-region from the latest year to compare against the country.

I have created a dictionary list and converted into a two-dimensional array to rename the columns of my main dataset for easier indexing matrix between the years and data information type. User will be prompt again to input a value from the menu provided in the message and a year. From the user's input, it will find the index and column name from the main dataset to print out the basic statistics of the specific chosen year and sub-dataset.

A third user-defined function will plot out and graph a comparison trend usage between cellphones and internet for each region. Lastly, a mini analysis question prints out at the end of the output.

For the project, I have chosen five different datasets (read in lines 100, 101, 104, 105, 115). It is located in two folders of the repository:

- Country Tech Use:
 - num_cell_phones_per_100_people.xlsx
 - total_cell_phones_by_country.xlsx
 - percentage_population_internet_users.xlsx
- UN Population Datasets:
 - UN Codes.xlsx
 - UN Population Dataset 1.xlsx

After merging and cleaning these datasets, it creates 194 rows and 12 columns (without counting indexes). Then, I added 3 more columns (lines 119 – 121) to the main merged data. The rest of the criteria for the project are found in the following lines:

- Merge/join operations (lines 102, 111, 114, 116)
- Duplicate rows deleted (taken care automatically by merging indexes)
- Sorted indexes (line 116)
- Comments and docstrings (across data and under each user-defined function)
- Describe method (line 206)
- Aggregation computation (lines 56, 58, 77, 78, 217)
- Masking operation (line 217)
- Groupby operation (line 56)
- Pivot table (line 58, 77, 78)
- User-defined functions:
 - country_stats(data, country):
 - From line 13 to 40
 - subreg_stats(data, country):
 - From line 43 to 64
 - region_cell_int_info(data):
 - From line 67 to 93
- Required user input (lines 126, 175, 186)
- Exception error message included (lines 124 – 131, lines 167 – 180, lines 184 – 191)

Screenshots of expected execution (screenshot_run1.png and screenshot_run2.png) and plot graph (cellphones_vs_internet_usage_trend.png) are uploaded in the repository. I have also included and commented out the export for the main merged file at the end of my code. The file is exported as well uploaded to the repository, called 'ProjectDataExport.xlsx'.

Online Dataset References:

Cell phones (per 100 people), Free data from World Bank via GAPMINDER.ORG, CC-BY LICENSE, June 2021. [Online]. Available: <https://www.gapminder.org/data/>

Cell phones (total), Free data from World Bank via GAPMINDER.ORG, CC-BY LICENSE, June 2021. [Online]. Available: <https://www.gapminder.org/data/>

Individuals using the Internet (% of population), Free data from International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database via GAPMINDER.ORG, CC-BY LICENSE, June 2021. [Online]. Available: <https://www.gapminder.org/data/>

SYB62_246_201907_Population Growth, Fertility and Mortality Indicators, Derived from total population. Population source: (1) United Nations Population Division. World Population Prospects: 2019 Revision, (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Reprot (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme., June 2021. [Online]. Available: <http://data.un.org/>

UNSD — Methodology (M49), The United Nations Statistical Division, June 2021. [Online]. Available: <https://unstats.un.org/unsd/methodology/m49/>