## *CA682 Data management and visualization*

|  |  |
| --- | --- |
| Name | Priyanka Agnes David |
| Student Number | 17210753 |
| Programme | MCM(Data Analytics) |
| Module Code | CA682 |
| Assignment Title | Data Visualisation |
| Submission date | 10/12/2017 |
| Module coordinator | Suzanne Little |

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found recommended in the assignment guidelines.

Name: Priyanka Agnes David Date: 10-12-2017

***INTRODUCTION:***

Population growth is the biggest problem the world is facing at the moment. Resource reduction and environmental degradation are apparently making the crisis more vulnerable. There are approximately 80 million humans added to the Earth every year. The Resource (especially oil) scarcity can lead to the decline in economic growth. [[3]](http://www.businessinsider.com/population-growth-must-stop-2010-7?IR=T) In 1970 Norman Borlaug won the Nobel Peace Prize for his work on developing new plant strains that formed the basis for the Green Revolution that began in the 1960s. However, in his Nobel acceptance speech Borlaug perceptively commented that "There can be no permanent progress in the battle against hunger until the agencies that fight for increased food production and those that fight for population control unite in a common effort. Fighting alone, they may win temporary skirmishes, but united they can win a decisive and lasting victory to provide food and other amenities of a progressive civilization for the benefit of all mankind."

At this rate of population growth Scientists predict that it can cause a catastrophe which can cause the death of billions of people whereas scholars contradict the same and say that this may reach a stable and reasonable population limit. These facts interested me in digging up for more information on over population. The United Nations have predicted the growth of population in numbers from the present till the year 2011. I have been analyzing the dataset and visualized the growth of population continent wise as the dataset and the countries are vast in numbers.

***DATA COLLECTION:***

* The dataset that I have used is and open dataset from the Population division of the United Nations. The Data was credible as it was from a reliable source.
* The dataset is an excel file and is geographic, structured and tabulated but could not be judged as accurate or not because the content or the values were prediction of the future population values for every year.
* The datasets consists of various tabs of data of which I have used the medium fertility variant of data.
* The Medium variant tab consists of total population estimates of people (both sex-men and women) by Region, sub-region and countries from 2015 to 2100.
* Dataset link-<https://esa.un.org/unpd/wpp/Download/Standard/Population/>

***DATA PROCESSING AND VISUALIZATION:***

I chose Python libraries to visualize my data. In order to get a lean and stable data frame for my visualization I carried out the following as a part of processing the dataset.

* I imported the datasets using the pandas package and created the graph plot using matplotlib.

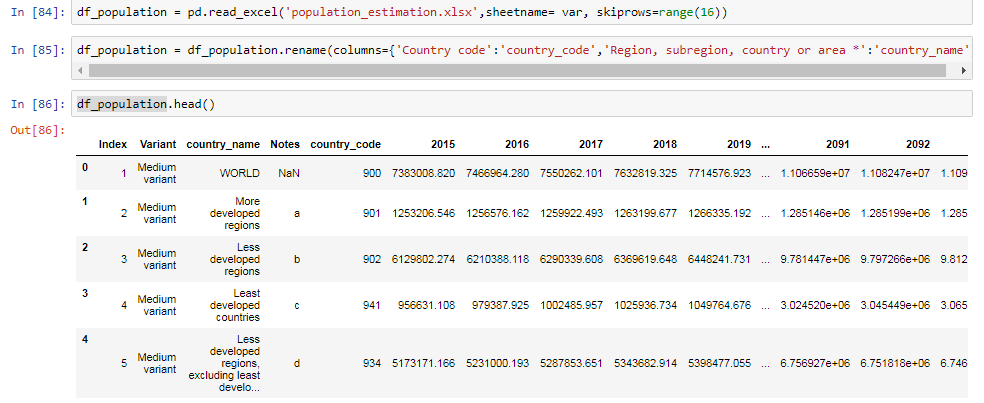


Fig 1:Original Dataset imported and informal column names renamed

* I multiplied all the numeric value columns in the original dataset by 1000 so that the population units are in 1 instead of 1000.



Fig 2: Use of lambda function to convert the unit of numeric values

* For the visualization I created a sub dataset from the original tdset containing only the continents and its subsequent values.

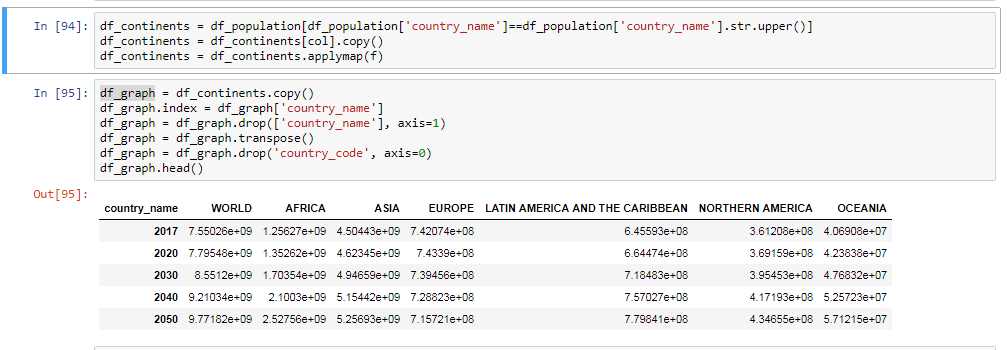


Fig 3: Continents of the world created as a separate data frame

* I visualized the estimation of population of continents of the world from 2017 to 2100 with 2020 to 2100 having a 10 year interval.

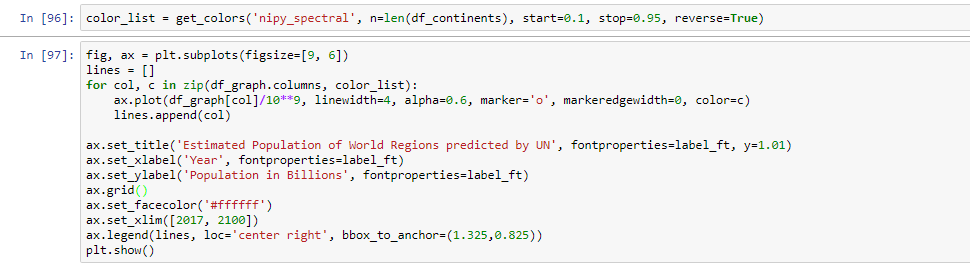


Fig 4: Code for plotting the graph with colored line aspects

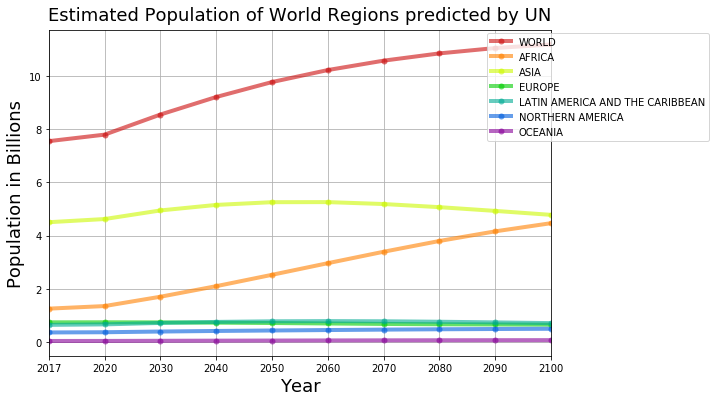


Fig 5: Estimated population of world regions as estimated by the UN

* In order to be specific about countries and its codes I used another dataset which I created from the information available in the Wikipedia.

Link - <https://en.wikipedia.org/wiki/ISO_3166-1_numeric>

* The second dataset had country codes and their names published by the [International Organization for Standardization](https://en.wikipedia.org/wiki/International_Organization_for_Standardization) (ISO) for representing countries.
* I filtered the original dataset with country codes present in the second dataset allowing me to create a new dataset which contains only countries, codes and the estimates of 10 years of population.



Fig 6: Country codes datasets imported and new data set with country names created

***INFERENCE BASED ON VISUALIZATION:***

* African population will grow by 3,211,320 people by 2100
* World population will grow by 3,634,106 people by 2100
* That is 422,785 new people outside of Asia
* Africa accounts for 88.4% of the projected growth

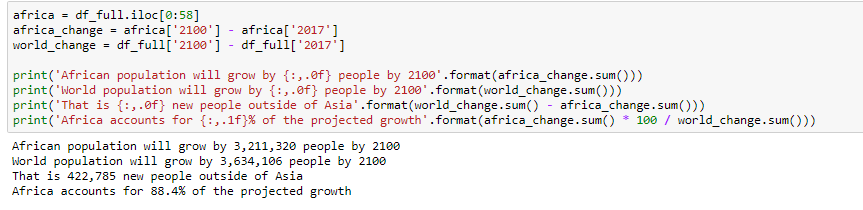


Fig 7: Code for inference based on the Dataset

* The most populated country in 2017 is China.



Fig 8: 10 most populated countries in 2017

* The most populated country in 2100 is predicted to be India

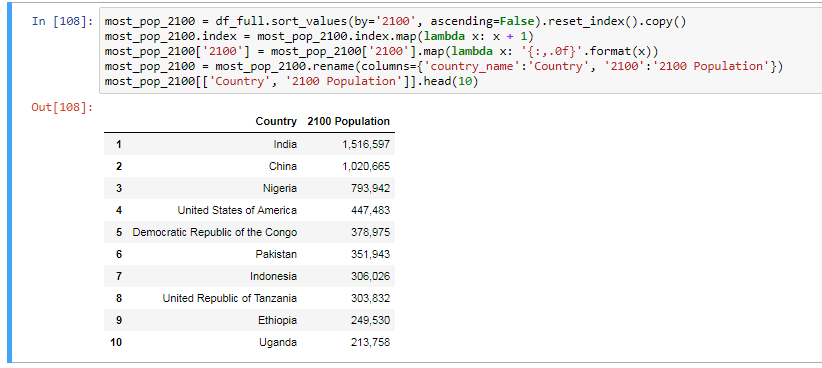


Fig 9: Prediction of 10 most populated countries in 2100

***CONCLUSION***

* I used Python libraries to process and visualize the data. I faced challenges in improvising and processing the data for an ideal clean visualization.
* I selected a line chart as my dataset was a collection of quantities over years (Temporal graph).
* Usage of color palettes(nippy spectral color list of python) have enhanced the visual aspect a little more than just a plain line graph.
* The Data I have visualized a high level of visualization that shows only a certain level of information on the continent level whereas the dataset is more informative on the country and sub region level too which is a aspect I could have improved on.

***REFERENCES***

1. <https://esa.un.org/unpd/wpp/Download/Standard/Population/>
2. <https://en.wikipedia.org/wiki/ISO_3166-1_numeric>
3. <http://www.businessinsider.com/population-growth-must-stop-2010-7?IR=T>
4. <http://www.yourarticlelibrary.com/essay/essay-on-the-problem-of-population-growth-657-words/27352>
5. <https://en.wikipedia.org/wiki/Human_overpopulation#Effects_of_human_overpopulation>