**A MINI PROJECT REPORT**

**ON**

**“COUNTRY DATA ANALYSIS”**



**Sinhgad College of Engineering**

**Department of Computer Engineering**

## SKILL DEVELOPMENT LABOTATORY

**(T.E Computer Engineering)**

**SUBMITTED BY**

***Group No:- 47***

**NAME** **ROLLNO**

**1.** **Dheeraj Jadhav** **305178**

**2.** **Bikash Kumar Yadav** **305181**

**3.**  **Hasan Bohra** **305174**

**YEAR 2019-2020**

**CERTIFICATE**



**“COUNTRY DATA ANALYSIS”**

**Sinhgad College of Engineering**

**Department of Computer Engineering**

**NAME** **ROLLNO**

1. **Dheeraj Jadhav**  **305178**
2. **Bikash Kumar Yadav** **305181**
3. **Hasan Bohra** **305174**

**Prof. Sakeeb Sheikh**   **Prof. M. P. Wankhade**

Internal Guide Head of Department

Department of Computer Engineering Department of Computer Engineering

**Dr.S.D.Lokhande**,

Principal

SCOE, Pune

**ACKNOWLEDGEMENT**

First and foremost we take the opportunity to extend our deep heartfelt gratitude to our guide **Prof. S.H.Sheikh** for guiding us throughout the entire project and for his kind and valuable suggestions, without which this idea won’t have executed. We also humbly thank **Prof. M.P.Wankhade, Head of Department of Computer Engineering, SCOE, Pune** for his indispensable support, his priceless suggestions and for his valuable time. We would also like to take privilege in expressing our gratitude towards other teaching & non-teaching staff for the advice and kind cooperation.

We also thank our parents and all my colleagues for encouraging me with their valuable suggestions and motivating us from time to time. Finally, last but not the least, we would thank the almighty without whose care this work would have not complete.

**INDEX**

List of figures i

Abstract ii

**1. Introduction**

1.1. Introduction to Country Data Analysis 1

1.2. Purpose 2

1.3. Scope 2

1.4. Problem Statement 3

**2. Project Planning and Management**

2.1. Software Requirements 4

2.2. Hardware Requirements 4

2.3. Proposed System 4

**3. Analysis and design**

3.1. Process model in visual form 5

**4. Implementation and coding**

4.1. Methodology 6

4.2. Algorithms/Flow Charts 7

4.3.GUI design/screen shots 8

**5. Graphs Shown As Example** 9

**6. Conclusion** 17

**Reference**

**List of Figures**

1. Visualization of demographics 8
2. Graphs Showing:
   1. Top 5 GDP countries 9
   2. Top 5 most populated countries 10
   3. Top 5 countries with higher Infant Mortality Rate 11
   4. Top 5 most literate countries 12
   5. Comparing Agriculture, Industry& service with GDP 13
   6. Comparing Net migration with population 14
   7. Comparing Infant mortality with Birth rate 15
   8. Showing Birth and Death rate of African countries 16

**ABSTRACT**

Data analysis is the process of evaluating data using analytical and statistical tools to discover useful information and aid in business decision making. There are a several data analysis methods including data mining, text analytics, business intelligence and data visualization. Our topic is completely based on the analysis of Country Data. Data analysis is a part of a larger process of deriving business intelligence.

Text analytics is the process of deriving useful information from text. It is accomplished by processing unstructured textual information, extract meaningful numerical indices from the information and make the information available to statistical and machine learning algorithms for further processing. This project includes analysis of all 193 countries. The Analysis comprises of 10-12 major points of multiple countries and regions. Entries include information on many of the following: Birth & Death Rate; Population; Population Density Area; Net Migration; Infant Mortality Rate; Literacy Rate; GDP; Agriculture; Service; Industry; Arable %; Area Population and more.

This analysis tells us about the development status of countries. Country Analysis involves the examination and interpretation of a nation’s economic, social and political environment. With Country Analysis we can choose the most suitable country for living. The countries providing highest quality of living can be determined from Country Data Analysis only. To implement Country Data Analysis we first learn about libraries of python such as Numpy where array interface is the best and the most important feature and pandas used for data manipulation and analysis.

1. **INTRODUCTION**

Country Data Analysis describes about different categories or aspects in countries which shows its status. Categories includes population, gdp, agriculture, arable land, industry, net migration rate, birth and death rate.

1. **Population** :- It is defined as all the inhabitants of a particular town, area, or country**.**
2. **Gross Domestic Product** :- Gross domestic product (GDP) is the market value of all final goods and services from a nation in a given year.
3. **Net Migration Rate** :- The net migration rate is the difference between the number of immigrants (people coming into an area) and the number of emigrants (people leaving an area) throughout the year.
4. **Literacy** :- Literacy is traditionally defined as the ability to read and write.
5. **Birth Rate** :- The birth rate (technically, births/population rate) is the total number of live births per 1,000 in a population in a year or period.
6. **Death Rate** :- Death rate, is a measure of the number of deaths (in general, or due to a specific cause) in a particular population, scaled to the size of that population, per unit of time.
7. **Agriculture** :- Agriculture is the science and art of cultivating plants and livestock.
8. **Arable** :- Arable land**,** according to one definition, land capable of being ploughed and used to grow crops.

**1.2. Purpose:-**

The purpose of Country report is to provide economic and business analysis of 226 different countries. It also helps us to analyse different international strategies among countries and compare each country’s status. Another purpose for developing this report is to lookout the outside view of each country.

**1.3. Scope:-**

The scope of this project is the system report, i.e. the project is analysed on different aspects and it will work for each country to view its current status and later expand to benefit in each and every category.

**1.4. Problem Statement :**

Initially we store data of 226 countries in .csv file and according to different categories as( population, literacy, birth and death rate, industry, agriculture and arable land, etc)

we created dataframe and performed data cleaning.

We sort them with respective categories and plot the graph according to requirements.

**2. PROJECT PLANNING & MANAGEMENT**

**2.1. Software Requirements :**

* Pycharm python editor(For Development)
* Anaconda-navigator
* HTML(For GUI)

**2.2. Hardware Requirements :**

* Intel® Core™  i5 7th Gen
* 8 GB of RAM
* Disk space:2 to 3 GB
* Operating system: Windows 10 or Linux

**2.3. Proposed System**

Our aim to analyse and show different countries data and plot graphs. It covers most major emerging markets and a number of financially challenged developed countries.

**3.ANALYSIS AND DESIGN**

**Process Model in Visual Form:**

Retrieve data

Displaying graphs according to categories.

Sort according to categories(top 5)

Convert into dataset

**4. IMPLEMENTATION & CODING**

**4.1. Methodology :**

Network contributors and full time analysts gather information from around the world.

Globally based analysts teams verify different datas whether it be economic or local health objective development with correspondent to ensure accuracy and timeliness. Experts use complex methodologies and provide detailed analysis explaining the assumptions to our baseline projections.

.

**4.3. Algorithms/Flowcharts :**

Country Data Analysis

1. The countries data consisting all the categories that we included was retrieved from a data file i.e. csv file and was loaded for the same.

2. Then dataframe was created using the libraries and the data categories were included for all the remaining groups.

3. Cleaning of data was done in which all the commas were replaced with full stop and all the NAN values were replaced with zero value.

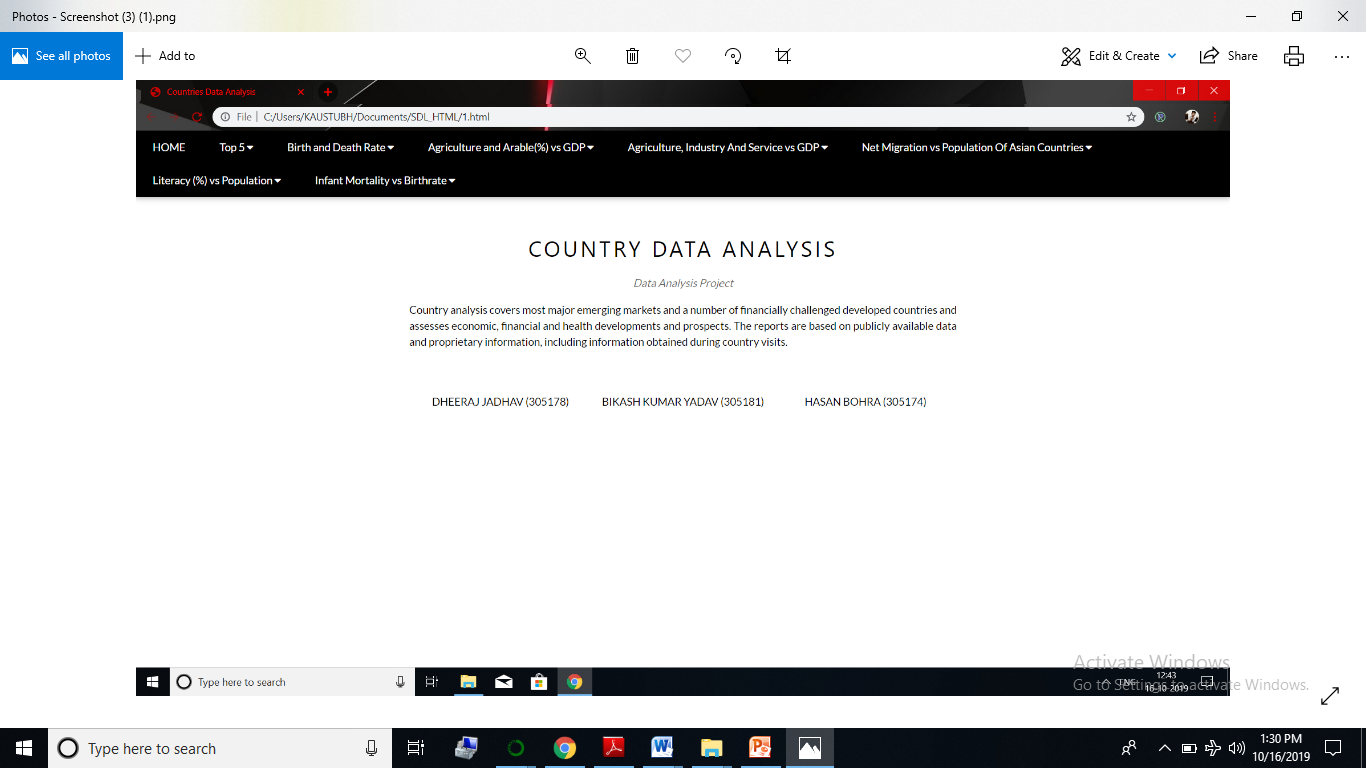
4. The data acquired consisted 226 countries with all its categories as(population, area, gross domestic product, industry, agriculture, service, arable land, literacy) so the countries were divided according to its group into regions.

5. Finally, the graphs were plotted according to regions of specific domains.

6. Exit.

**4.4. GUI design/screen shots :**

**4.4.1. Visualization of demographics**

****

Above image shows the front end of the project (Country Data Analysis).

1. **Graphs Shown As Example**

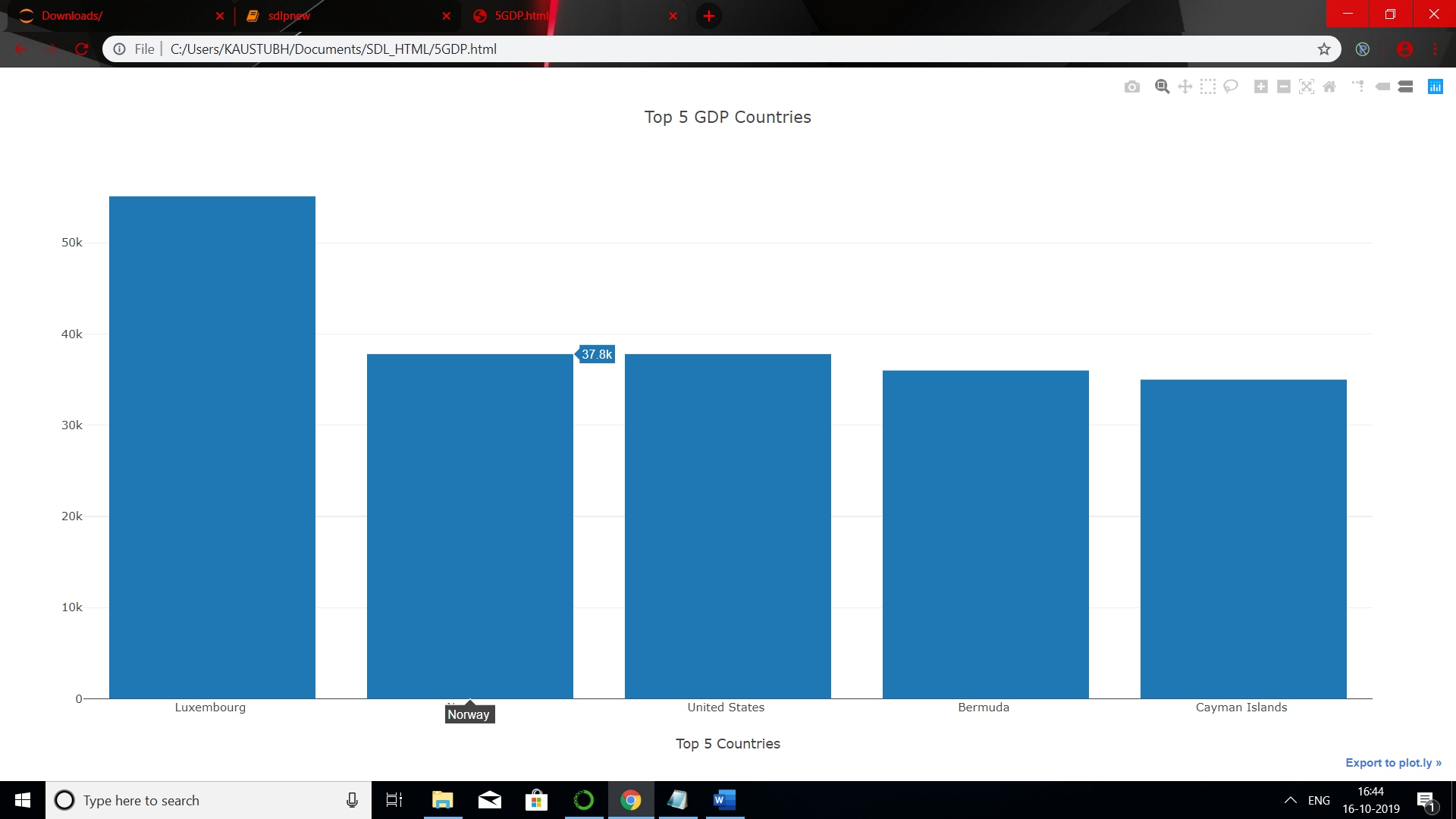
****

Fig.5.1.Top 5 GDP Countries

The bar diagram illustrates the data of GDP growth rate per year of top 5 countries. The graph summarized that, Luxembourg is the country with highest gross domestic product.

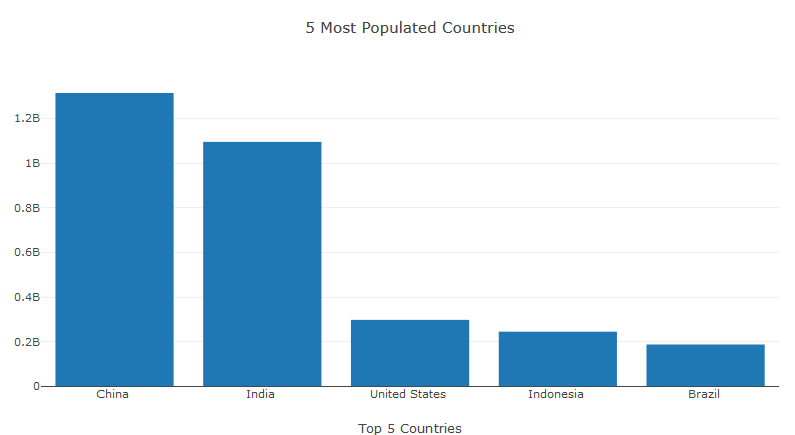
****

Fig.2.Top 5 Most Populated Countries

The bar diagram illustrates the top five most populated countries. It can be seen that China and India are the only countries in the world with population more than 1 billion.

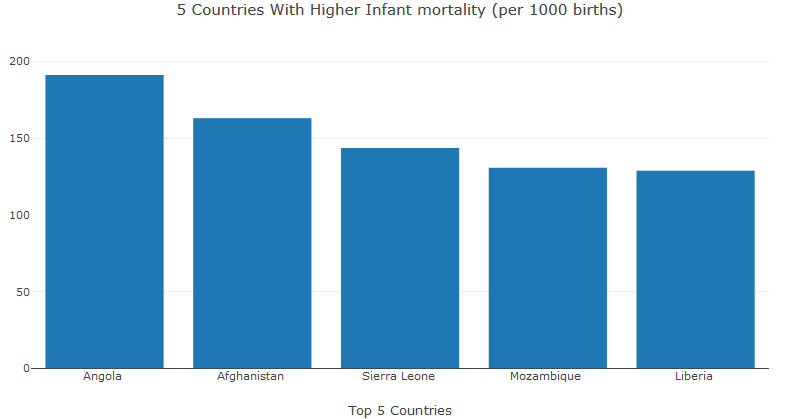
****

Fig.3.Top 5 Countries with higher Infant Mortality

The graph above shows the top 5 countries with higher infant mortality rate in which number of deathsper 1,000 live births of children under one year of age is shown.

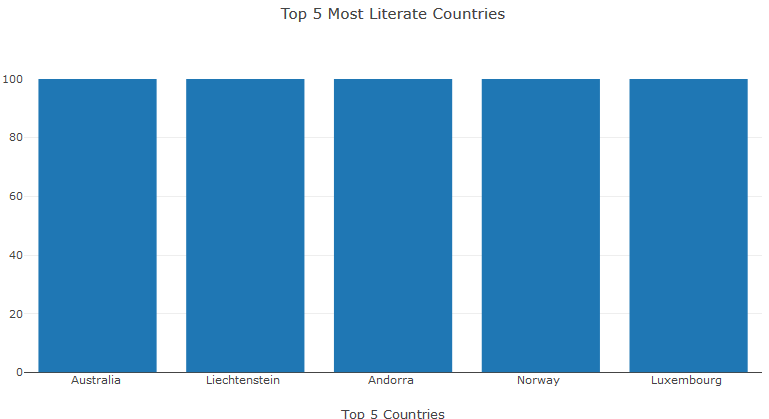
****

Fig.4.Top 5 Most Literate Countries

The graph shows the top 5 countries which has higher literacy rate i.e. the ability to read and write.

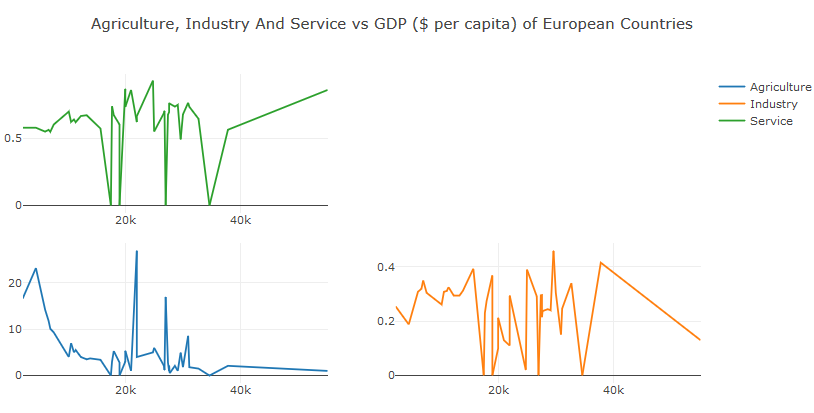
****

Fig.5.Agriculture, Industry and Service vs GDP

The above graph shows the comparison in sectors as agriculture industry and service of European countries with gross domestic product.

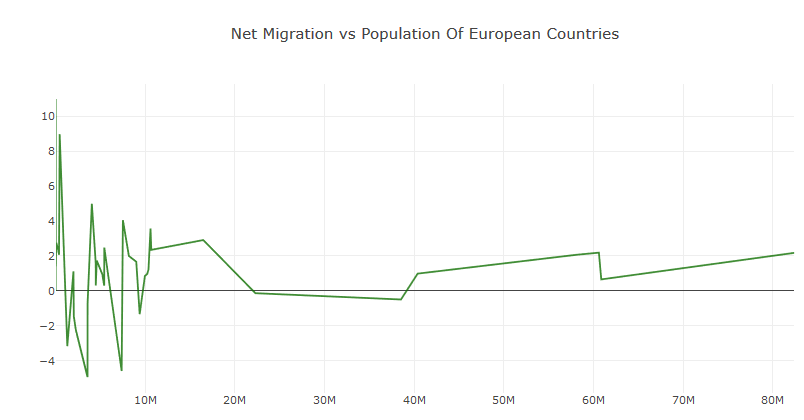
****

Fig.6. Net Migration vs Population

The above figure shows the net migration i.e. the number of immigrants and number of emigrants comparing with population of European countries.

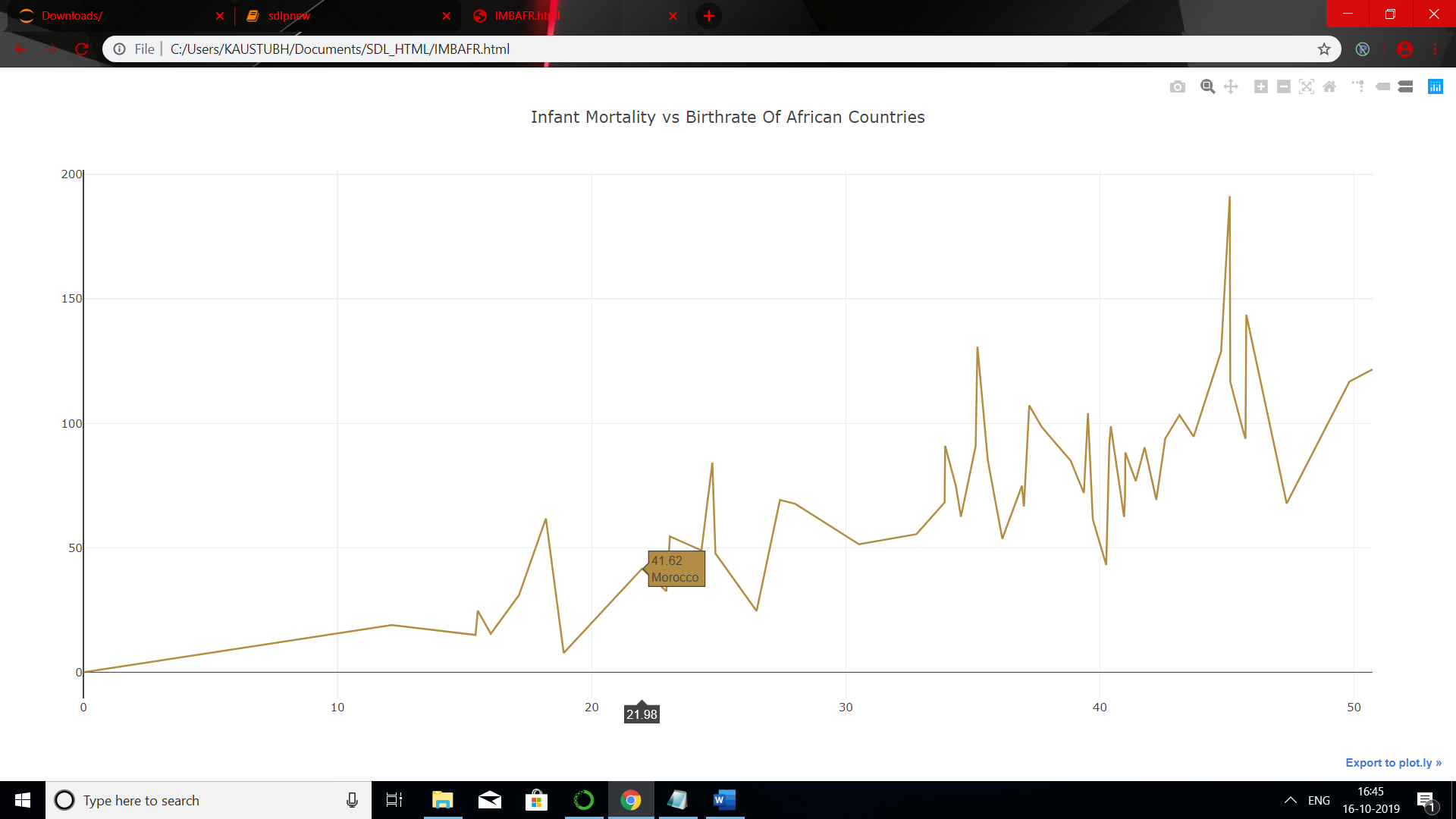
****

Fig.7.Infant Mortality vs Birth rate of African Countries

The above graph shows the infant mortality comparing with birth rate of African countries. In this, x-axis deals with infant mortality and y- axis with birth rate.

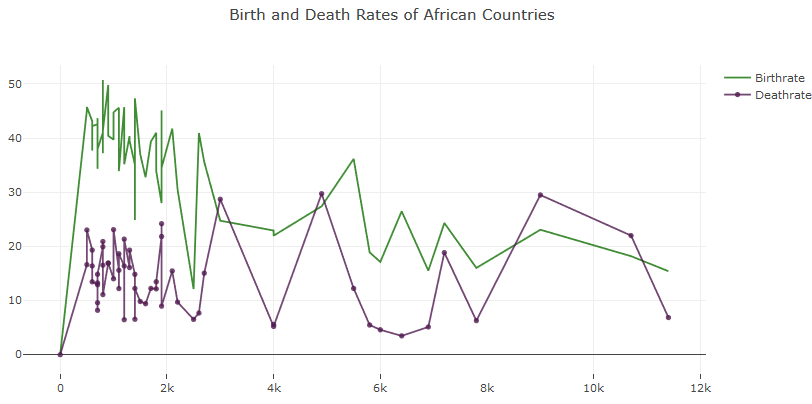


Fig.8. Birth and Death Rates of African Countries

The above graph shows the birth and death rates of African countries. In this, the green line indicates birth rate and purple indicates death rate.

**6. CONCLUSION**

First we retrieved data and created data-frame in different formats to plot graphs of different categories. We used matplotlib for creating different graphs. Sorting of data was done on all basis and for creating dashboard we used HTML as application.

**REFERENCES**

**BOOKS**:

1.Python for data analysis

2.OpenCv –python(tutorials Point)

1. [www.eiu.com](http://www.eiu.com/)
2. <https://libguides.mit.edu/country>
3. <https://data.worldbank.org/country>