VISUALISING AND PREDICTING

THE POPULATION DEMOGRAPHY OF INDIA

A PROJECT REPORT

Submitted by

MANOHAR [722819104076]

NIKHIL SRINIVASAN [722819104088]

PRAWIN R [722819104104]

RANJITH KUMAR A [722819104111]

In partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

SRI ESHWAR COLLEGE OF ENGINEERING

KINATHUKADAVU

COIMBATORE-641 202

ANNA UNIVERSITY::CHENNAI 600 025

SEPTEMBER 2021

VISUALISING AND PREDICTING THE POPULATION DEMOGRAPHY OF INDIA

A PROJECT REPORT

Submitted by

MANOHAR [722819104076]

NIKHIL SRINIVASAN [722819104088]

PRAWIN R [722819104104]

RANJITH KUMAR A [722819104111]

In partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

SRI ESHWAR COLLEGE OF ENGINEERING

KINATHUKADAVU

COIMBATORE-641 202

ANNA UNIVERSITY: CHENNAI 600 025

SEPTEMBER 2021

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"VISUALISING AND PREDICTING THE POPULATION DEMOGRAPHY OF INDIA"** is the bonafide work of

MANOHAR [722819104076]

NIKHIL SRINIVASAN [722819104088]

PRAWIN R [722819104104]

RANJITH KUMAR A [722819104111]

Who carried out the project work under my supervision.

---------- ----------

SIGNATURE **SIGNATURE**

Dr.R.Subha, ME (Ph.D.) Mr. Sampath Kumar.S, M.Tech(Ph.D.)

HEAD OF THE DEPARTMENT **SUPERVISOR**

Professor, Asst.Professor,

Department of CSE, Department of CSE,

Sri Eshwar College of Engineering, Sri Eshwar College of Engineering,

Coimbatore-641202 Coimbatore-641202

Submitted for the **ANNA UNIVERSITY** project work viva-voce held on ……

---------- ----------

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

The success of a work depends on a team and cooperation. We take this opportunity to express our gratitude and thanks to everyone who helped us in our project. We would like to thank the management for excellent facilities and the constant support provided by them to complete this project.

It is indeed our great honour bounded duty to thank our beloved Chairman **Mr.R. Mohanram,** for his academic interest shown towards the students. We are indebted to our Director **Mr. R. Rajaram,** for motivating and providing us with all facilities.

We wish to express our sincere regards and deep sense of gratitude to **Dr.SudhaMohanram, M.E, Ph.D.,** Principal, for the excellent facilities and encouragement provided during the course of the study and project.

We are indebted to **Prof. Dr.R.Subha, M.E., Ph.D.,** Head of Computer science and Engineering Department, for having permitted us to carry out this project and giving the complete freedom to utilize the resources of the department.

We express our sincere thanks to our project Co-coordinator **Mr. S. Sampath Kumar, MTech., (Ph.D.).,** Assistant Professor of Computer Science and Engineering Department for his valuable guidance and encouragement given to us for this project.

We solemnly express our thanks to all the teaching and non-teaching staff of the Computer Science and Engineering Department, family and friends for their valuable support which inspired us to work on this project.

DECLARATION

We,

MANOHAR

NIKHIL SRINIVASAN

PRAWIN R

RANJITH KUMAR A

Declare that the project entitled "**VISUALISING AND PREDICTING THE POPULATION DEMOGRAPHY OF INDIA",** submitted in partial fulfillment to Anna University as the project work of Bachelor Of Engineering (Computer Science and Engineering) Degree, is a record of original work done by us under the supervision and guidance of Mr.Sampath Kumar.S, M.Tech.,(Ph.D)., Assistant Professor, Sri Eshwar College Of Engineering, Coimbatore.

Place: Coimbatore

Date: [Manohar]

[Nikhil Srinivasan]

[Prawin R]

[Ranjith Kumar A]

Project Guided by

[Asst.prof Mr. Sampath Kumar, M.Tech., (Ph. D).,**]**

ABSTRACT

In India, there is a platform which shows the exact population, growth rate this platform helps to make plans and decisions using the population information. There may be many problems when it comes to manipulating the data and also projecting/predicting the future population and at a certain time like this where we can't get the data manually because of the Pandemic situation. To clear these problems and to get the data without placing anyone in harm's way we have developed a population Prediction algorithm using a machine learning. We have adopted a Linear Regression Model and also Polynomial Regression Model. Moreover, the population is difficult to predict because of some circumstances that can alter a location's population and displace them to different places. Events like Virus outbreaks, Natural calamities, Natality rate, Mortality rates and migration. Natural Calamities can alter population demographic very-fast, especially when it is a result of life-changing outcomes like earthquakes, tsunamis, volcanic eruptions. In this event, one area's demographic population is reduced while another place's population Grows. The results have shown the linear regression has higher percentage error when compared to Polynomial regression. We have also visualized the population data of the years 2001 and 2011 for a better understanding of what the future may hold for us.

Keywords – Population Prediction, Population Visualization, Machine Learning

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

An Essential part of most successful countries is that they have to be ready to face any challenges that the future may hold for them. This requires planning for both short as well as long term. The data with the help of which these planning's are done should be reliable as well as based on the present situation. However, the future will not always be a straight line, sometimes it may go up or sometimes even down but the data should always adjust along with the events so that it can be very much reliable. In our case the data of the population must very accurate in order for us to predict with future with low error rate's. The information obtained from our project may help the country prepare itself for a stable and positive population growth. The visualized data will help us understand how the population was distributed throughout the country and will help us get a clear understanding how the population ratio was for many number of things like education, working population, literates.

CHAPTER 2

SYSTEM SPECIFICATION

2.1 HARDWARE REQUIREMENTS:

System : Intel Dual Core

Screen Resolution : 1024x768 display

RAM : 4GB

2.2 SOFTWARE SPECIFICATIONS:

Operating System : Windows XP and above

Web server : Apache Web Server

Front End : HTML, CSS, PHP, JavaScript

Back End : MySQL server

Web Browser : Google Chrome, Mozilla Firefox

Compiler : Google Colab

CHAPTER 3

SOFTWARE DESCRIPTION

3.1 FRONT END

3.1.1 HTML

Hypertext Mark-up Language is the main mark-up language for creating web pages and other information that can be displayed in a web browser.

HTML is written in the form of HTML elements consisting of tags enclosed in angle brackets (like <html>), within the web page content. HTML elements form the building blocks of all websites. HTML allows images and objects to be embedded and can be used to create interactive forms. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items.

It can embed scripts written in languages such as JavaScript which affect the behaviour of HTML web pages. Web browsers can also refer to Cascading Style Sheets (CSS) to define the look and layout of text and other material. The W3C, maintainer of both the HTML and the CSS standards, encourages the use of CSS over explicit presentational HTML.

3.1.2 CASCADING STYLE SHEETS (CSS)

Cascading Style Sheets is a style sheet language used for describing the look and formatting of a document written in a mark-up language.

While most often used to style web pages and interfaces written in HTML and XHTML, the language can be applied to any kind of XML document, including plain XML, SVG and XUL.

CSS is a cornerstone specification of the web and almost all web pages use CSS style sheets to describe their presentation. CSS is designed primarily to enable the separation of document content from document presentation, including elements such as the layout, colours, and fonts.

CSS can also allow the same mark-up page to be presented in different styles for different rendering methods, such as on-screen, in print, by voice (when read out by a speech-based browser or screen reader) and on Braille-based, tactile devices. It can also be used to allow the web page to display differently depending on the screen size or device on which it is being viewed.

3.1.3 PHP

PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. PHP is now installed on more than 244 million websites and 2.1 million web servers.

PHP code is interpreted by a web server with a PHP processor module, which generates the resulting web page: PHP commands can be embedded directly into an HTML source document rather than calling an external file to process data. It has also evolved to include a command-line interface capability and can be used in standalone graphical applications.

PHP is free software released under the PHP License. PHP can be deployed on most web servers and also as a standalone shell on almost every operating system and platform, free of charge.

3.2 BACK END

3.2.1 MYSQL

MySQL is the world's most popular open source database software, with over 100 million copies of its software downloaded or distributed throughout its history. With its superior speed, reliability, and ease of use, MySQL has become the preferred choice for Web, Web 2.0, SaaS, ISV, Telecom companies and forward-thinking corporate IT Managers because it eliminates the major problems associated with downtime, maintenance and administration for modern, online applications.

Many of the world's largest and fastest-growing organizations use MySQL to save time and money powering their high-volume Web sites, critical business System, and packaged software — including industry leaders such as Yahoo!, Alcatel-Lucent, Google, Nokia, YouTube, Wikipedia, and Booking.com.

The flagship MySQL offering is MySQL Enterprise, a comprehensive set of production-tested software, proactive monitoring tools, and premium support services available in an affordable annual subscription.

MySQL is a key part of LAMP (Linux, Apache, MySQL, PHP / Perl / Python), the fast-growing open source enterprise software stack. More and more companies are using LAMP as an alternative to expensive proprietary software stacks because of its lower cost and freedom from platform lock-in.

3.3 SOFTWARE FEATURES

3.3.1 XAMPP

XAMPP stands for Cross-Platform(X), Apache (A), MySQL (M), PHP(P) and Perl(P). It is a simple, lightweight Apache distribution that makes it extremely easy for developers to create a local web server for testing purposes.

XAMPP has four primary components. These are:

- Apache
- MySQL
- PHP

Apache → Default web server application

MySQL→ Database Management System

PHP → Server-Side programming Language

phpMyAdmin → Admin tool for working with MySQL

3.3.2 XAMPP CONTROL PANEL

The XAMPP control panel gives a complete control over all installed XAMPP components.

3.3.3 APACHE 2.2

The Apache HTTP Server commonly referred to as Apache, is a web server application notable for playing a key role in the initial growth of the World Wide Web.

Originally based on the NCSA HTTPd server, development of Apache began in early 1995 after work on the NCSA code installed.

Apache quickly overtook NCSA HTTPd as the dominant HTTP server, and has remained the most popular HTTP server in use since April 1996. In 2009, it became the first web server software to serve more than 100 million websites.

Apache is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation. Most commonly used on a Unix-like system, the software is available for a wide variety of operating systems, including Unix, FreeBSD, Linux, Solaris, Novell NetWare, OSX, Microsoft Windows, OS/2, TPF, OpenVMS and eComStation. Released under the Apache License, Apache is open-source software.

Apache supports a variety of features, many implemented as compiled Modules which extend the core functionality. These can range from server-side programming language support to authentication schemes. Some common language interfaces support Perl, Python, Tcl and PHP.

Virtual hosting allows one Apache installation to serve many different websites. For example, one machine with one Apache installation could simultaneously serve www.example.com, www.example.org, test47.test-server.example.edu, etc.

Apache features configurable error messages, DBMS-based authentication databases, and content negotiation. It is also supported by several graphical user interfaces (GUIs).

It supports password authentication and digital certificate authentication. Because the source code is freely available, anyone can adapt the server for specific needs, and there is a large public library of Apache add-on.

3.4.0 Colab

Colaboratory, or 'Colab' for short, allows you to write and execute Python in your browser, with

Zero configuration required

Free access to GPUs

Easy sharing

Whether you're a student, a data scientist or an AI researcher, Colab can make your work easier. Watch Introduction to Colab to find out more, or just get started below!

Getting started

The document that you are reading is not a static web page, but an interactive environment called a Colab notebook that lets you write and execute code.

For example, here is a code cell with a short Python script that computes a value, stores it in a variable and prints the result:

[ ]

```
seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day
```

86400

To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut 'Command/Ctrl+Enter'. To edit the code, just click the cell and start editing.

Variables that you define in one cell can later be used in other cells:

[ ]

```
seconds_in_a_week = 7 * seconds_in_a_day
seconds_in_a_week
```

604800

Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, LaTeX and more. When you create your own

Colab notebooks, they are stored in your Google Drive account. You can easily share your Colab notebooks with co-workers or friends, allowing them to comment on your notebooks or even edit them. To find out more, see Overview of Colab. To create a new Colab notebook you can use the File menu above, or use the following link: Create a new Colab notebook.

Colab notebooks are Jupyter notebooks that are hosted by Colab. To find out more about the Jupyter project, see jupyter.org.

3.4.1 Data science

With Colab you can harness the full power of popular Python libraries to analyse and visualise data. The code cell below uses numpy to generate some random data, and uses matplotlib to visualise it. To edit the code, just click the cell and start editing.

[1]
0s

```
import numpy as np
from matplotlib import pyplot as plt

ys = 200 + np.random.randn(100)
x = [x for x in range(len(ys))]

plt.plot(x, ys, '-')
plt.fill_between(x, ys, 195, where=(ys > 195), facecolor='g', alpha=0.6)

plt.title("Sample Visualization")
```

```
plt.show()
```

You can import your own data into Colab notebooks from your Google Drive account, including from spreadsheets, as well as from GitHub and many other sources. To find out more about importing data, and how Colab can be used for data science, see the links below under Working with data.

3.4.2 Machine learning

With Colab you can import an image dataset, train an image classifier on it, and evaluate the model, all in just a few lines of code. Colab notebooks execute code on Google's cloud servers, meaning you can leverage the power of Google hardware, including GPUs and TPUs, regardless of the power of your machine. All you need is a browser.

Colab is used extensively in the machine learning community with applications including:

Getting started with TensorFlow
Developing and training neural networks
Experimenting with TPUs
Disseminating AI research
Creating tutorials

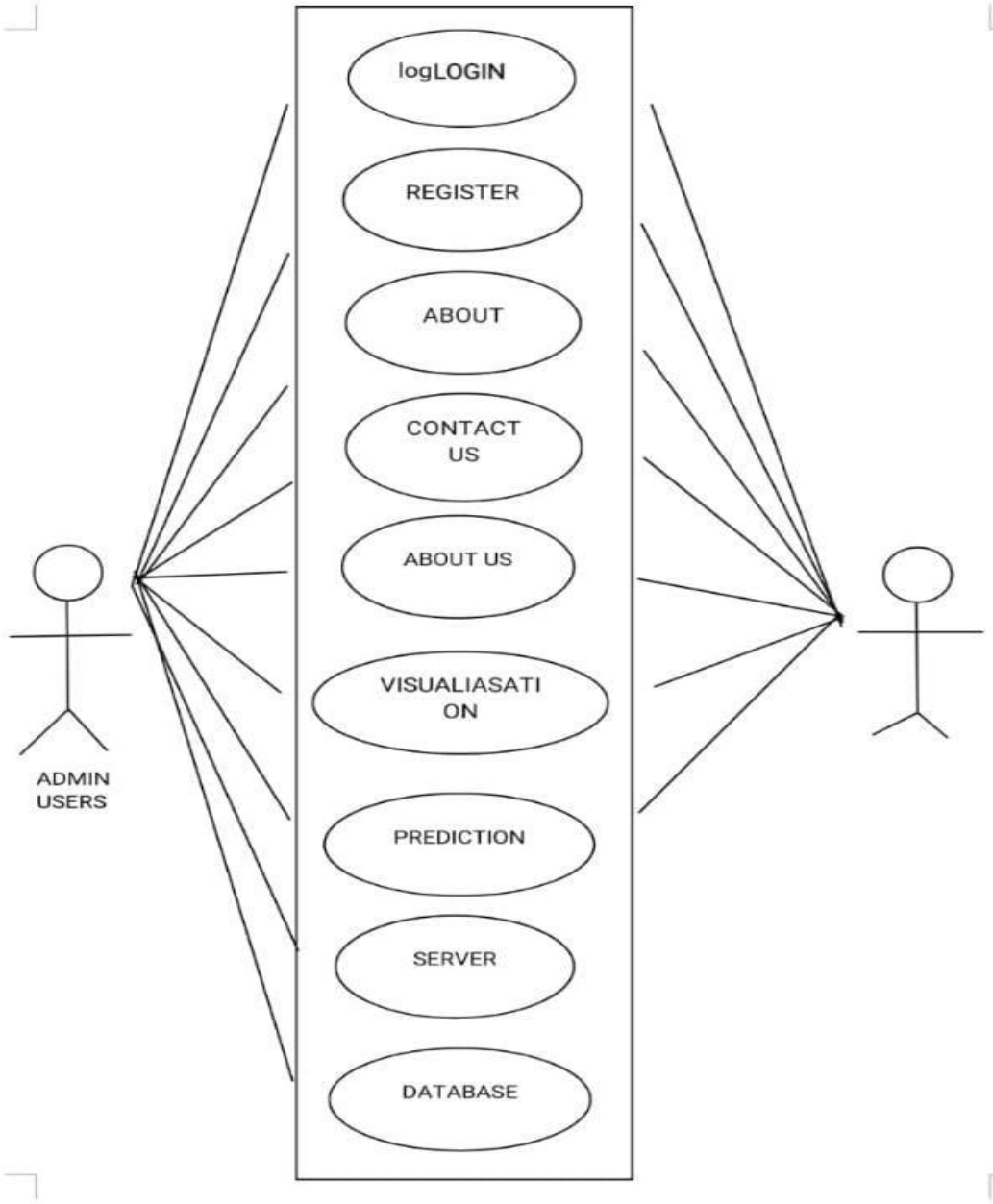
CHAPTER 4

PROJECT DESCRIPTION

Predicting the population trend of humans is a very complicated process. There many numbers of uncertainties with a demography of a country. Many traditional approaches can be used in predicting a countries population but these methods are not suitable for this purpose because these methods assume many things like the population growth will only be linear and will not take into account the factors such as Natality rates and Mortality rates. These things could lead to a prediction where there are many errors. Secondly population growth is difficult to predict because some events like earthquakes, tsunamis, volcanic eruptions could alter a person's location in a short period of time. Migration can rapidly alter a country's demography especially when it is caused by events which cause high mortality rates like war. In case of events like war the population of that particular place will decrease while the population of another place will increase due to the exodus. Some model like polynomial regression models will have higher accuracy when compared to models like linear regression models because they consider the factors of demography while the latter does not. Optimizing the polynomial model will give us a prediction with high accuracy rate.

CHAPTER 5

SYSTEM DESIGN



s

CHAPTER 6

SYSTEM IMPLEMENTATION

We have adopted a Polynomial regression model**. Polynomial regression is a special case of linear regression where we fit a polynomial equation on the data with a curvilinear relationship between the target variable and the independent variables.**

In a curvilinear relationship, the value of the target variable changes in a non-uniform manner with respect to the predictor (s).

In Linear Regression, with a single predictor, we have the following equation:

$$Y = \theta_0 + \theta_1 x$$

where,

Y is the target,

x is the predictor,

***θ*0** is the bias,

and ***θ*1** is the weight in the regression equation. This linear equation can be used to represent a linear relationship. But, in polynomial regression, we have a polynomial equation of degree ***n*** represented as:

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \ldots + \theta_n x^n$$

Here:

θ**0** is the bias,

θ**1**, θ**2**, **...**, θ**n** are the weights in the equation of the polynomial regression,

and ***n*** is the degree of the polynomial

The number of higher-order terms increases with the increasing value of ***n,*** and hence the equation becomes more complicated.

We have used three different datasets from Kaggle for both visualization and prediction. For visualization we have used the population data of the years 2001 and 2011 which includes data such as population of male, population of female, population of male literates, population of female literates, population of people who have received primary, middle and secondary education, total number of graduates and population of working class for each state of India. For the Prediction of population, we have used a dataset which has the actual population of India from the year 1950 to 2011, using this dataset we have predicted the population of the Indian country from 2012 to 2030. The Visualization and prediction were done using data science and machine learning through python programming language.

CHAPTER 7

CONCLUSION & FUTURE ENHANCEMENT

CONCLUSION

By the year 2030 Indian population would be 1,55,63,79,282 i.e. One Hundred Crore Fifty-Five Crore Sixty-Three Lakh Seventy-Nine Thousand Two hundred eighty-Two. Our Country should be ready for this population and we should be able to produce much needed jobs and increase medical facilities and also school facilities. This data would be really helpful our government to be ready for the Future.

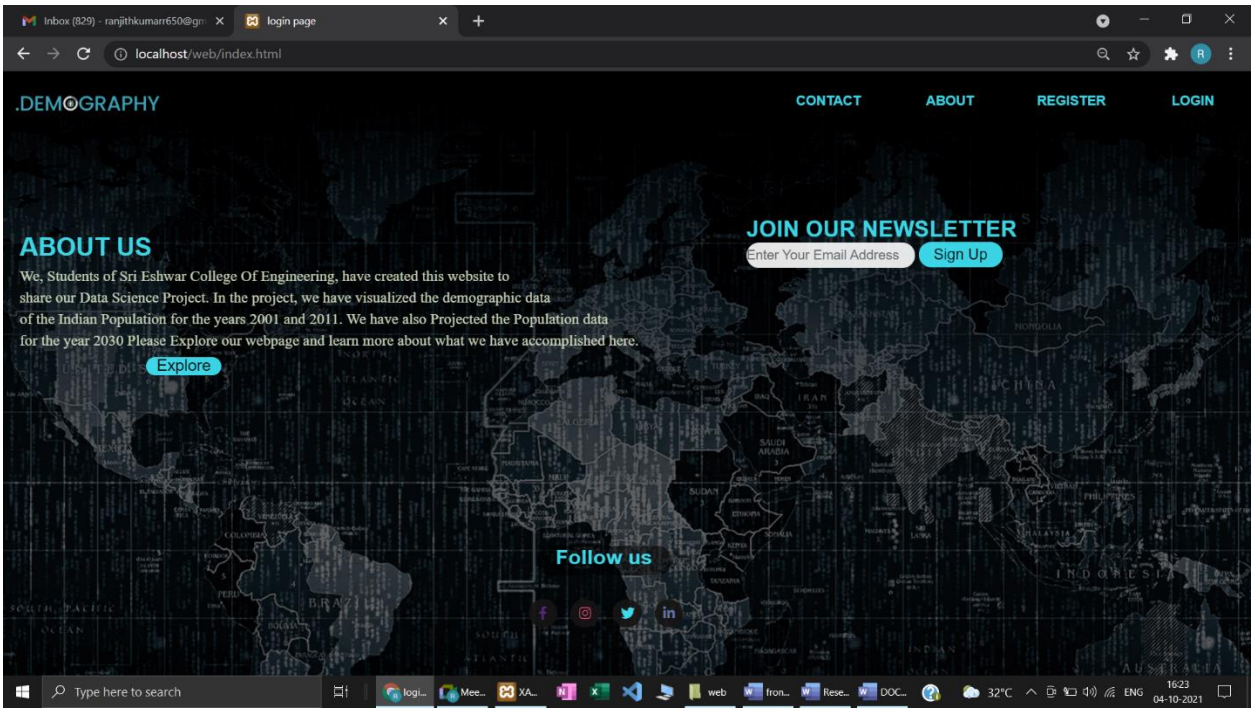
FUTURE ENHANCEMENT

- It is possible to give monthly growth rate population data.
- And also having plan of prediciting **2050**.

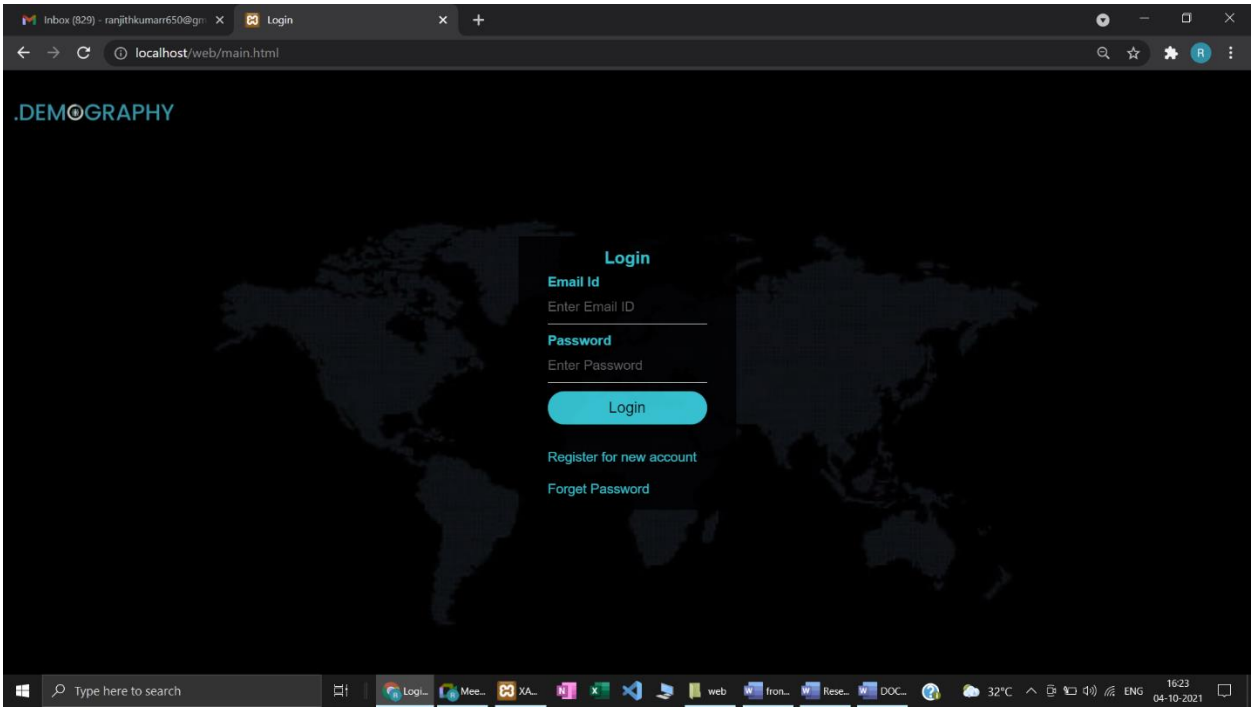
CHAPTER 8

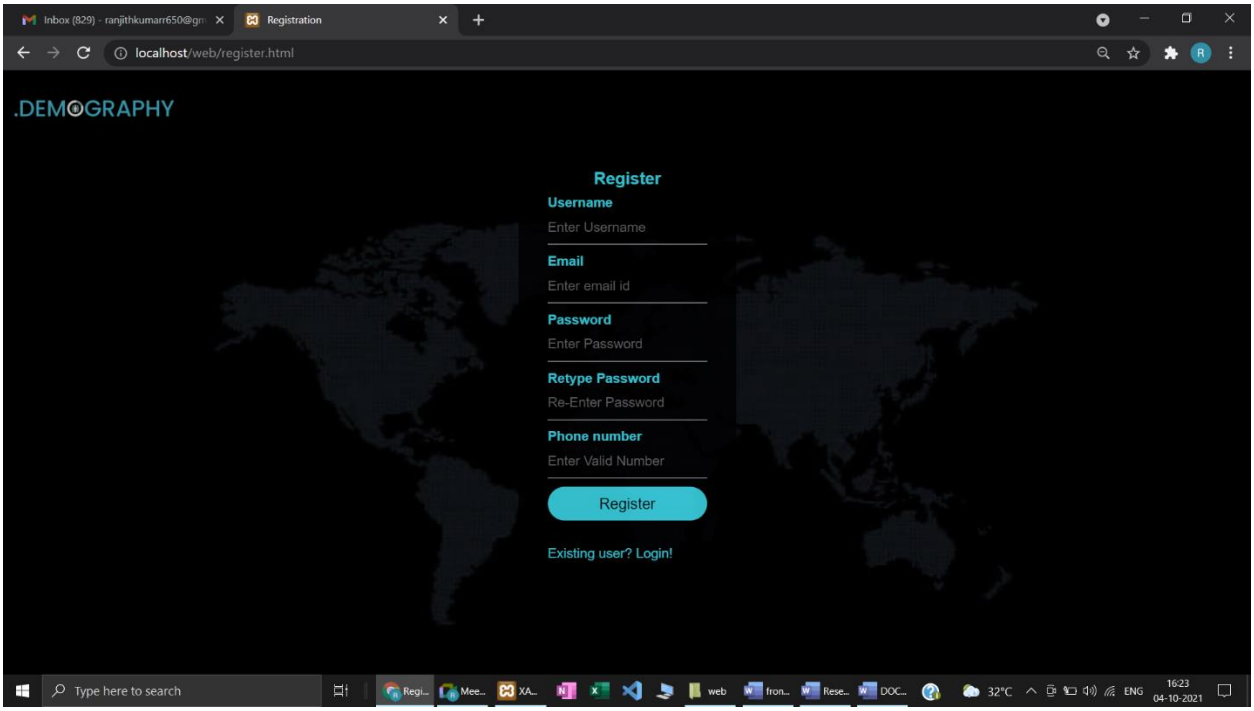
APPENDIX

Index

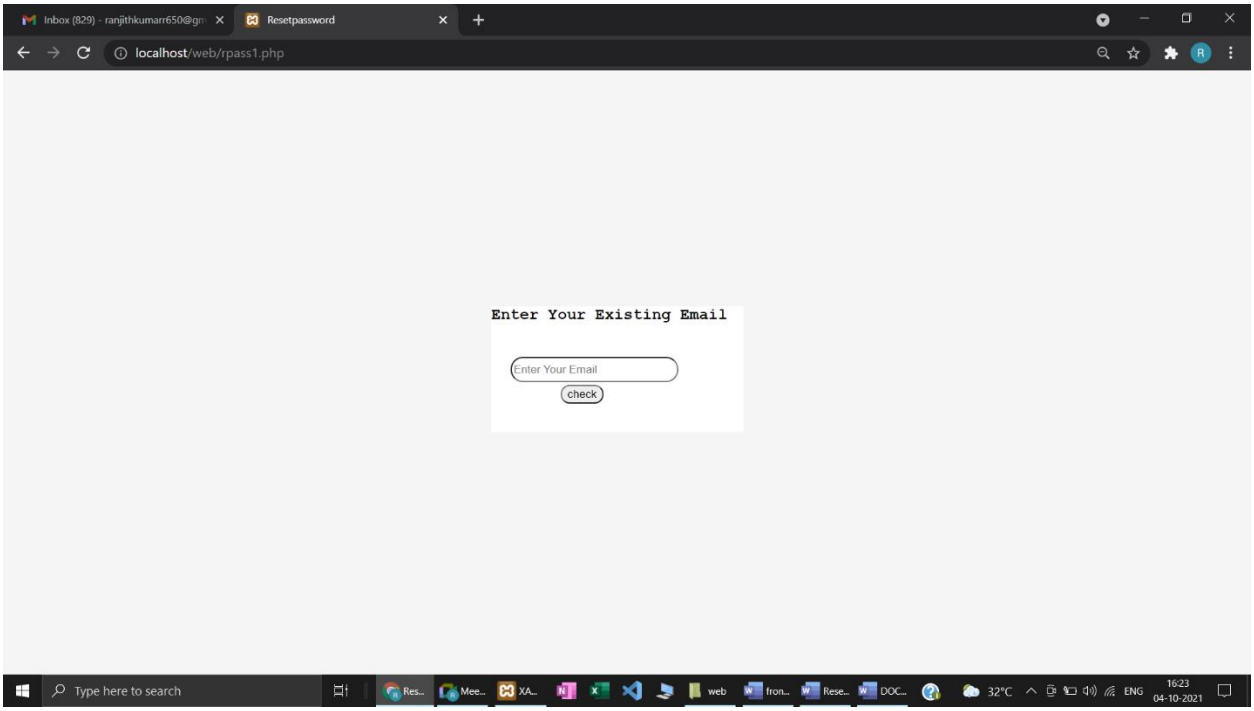


Login and Registration page

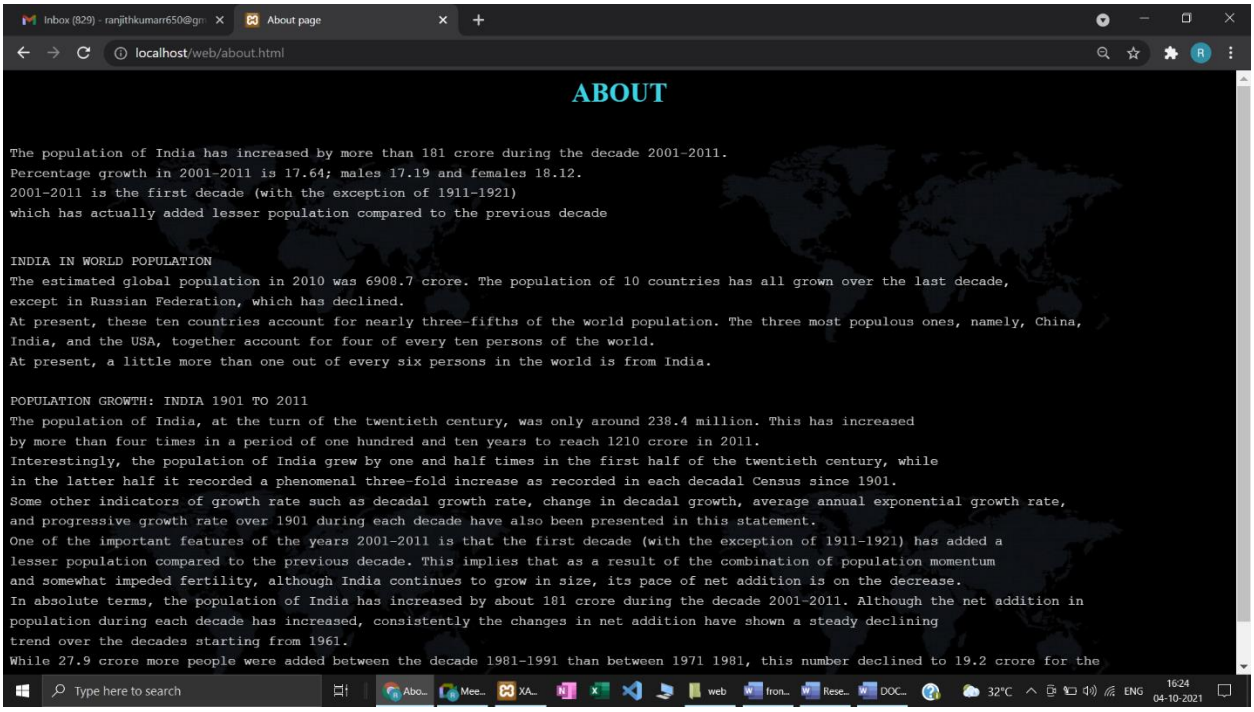


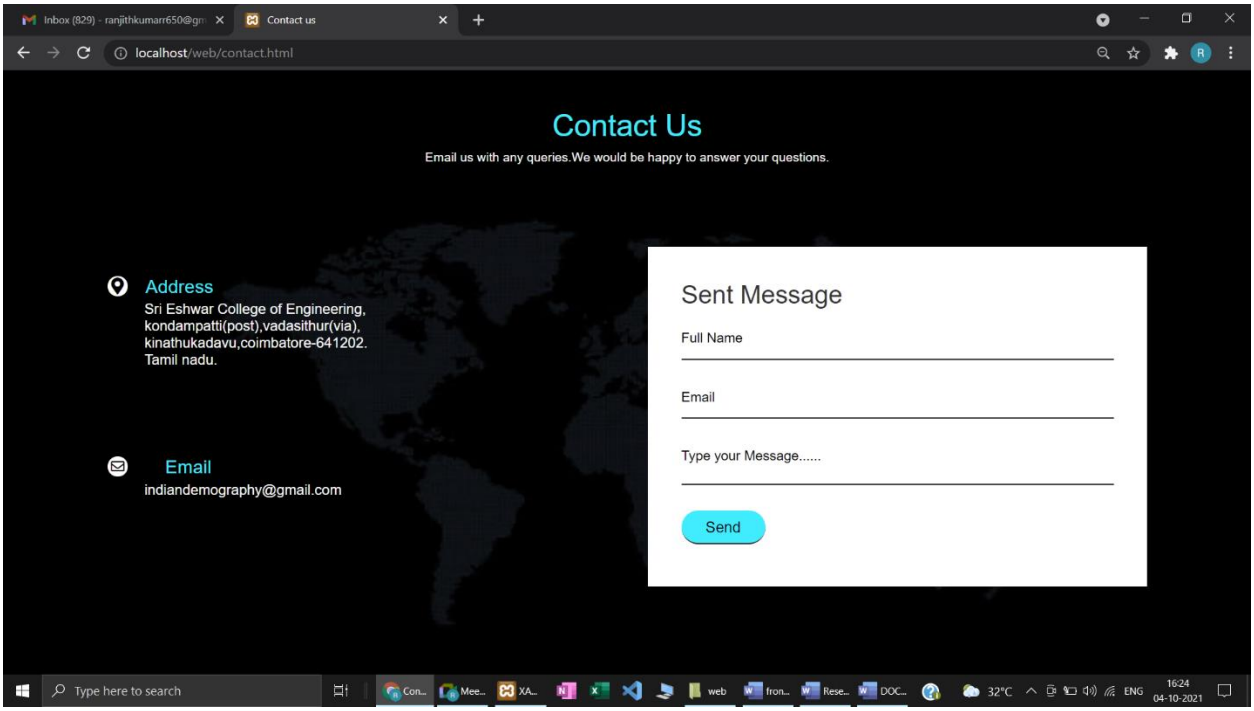


Reset Password page

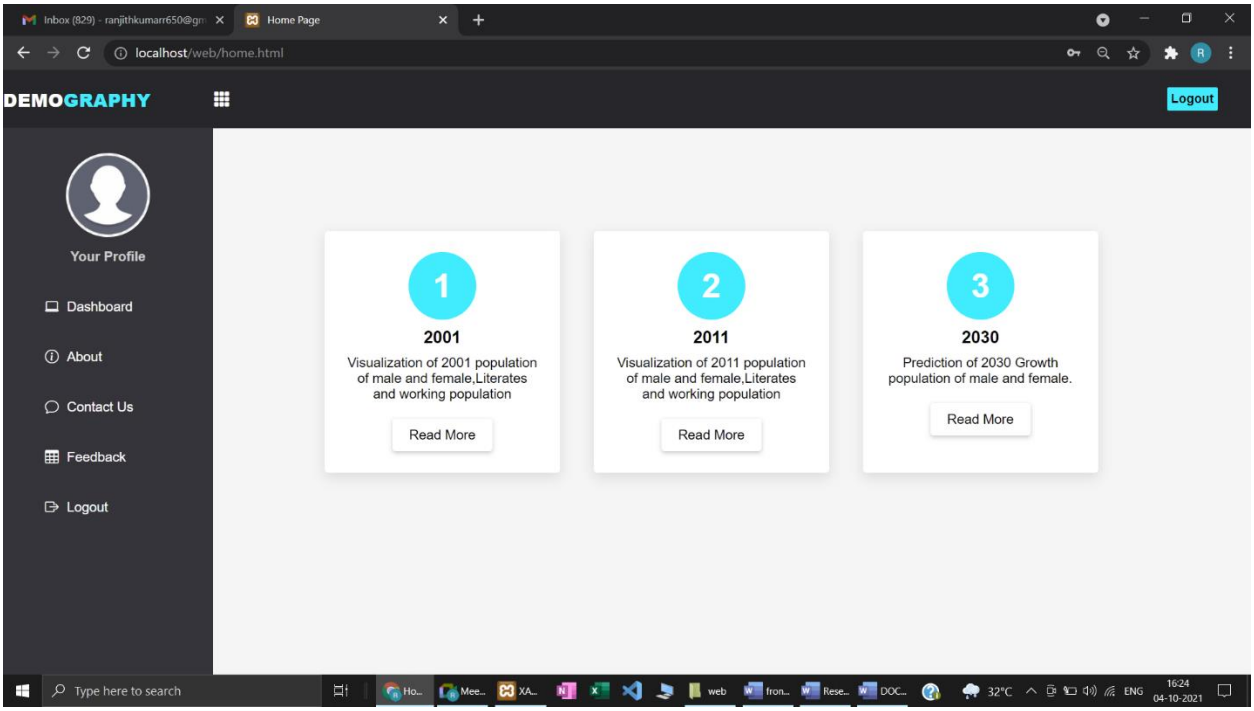


About and Contact us page

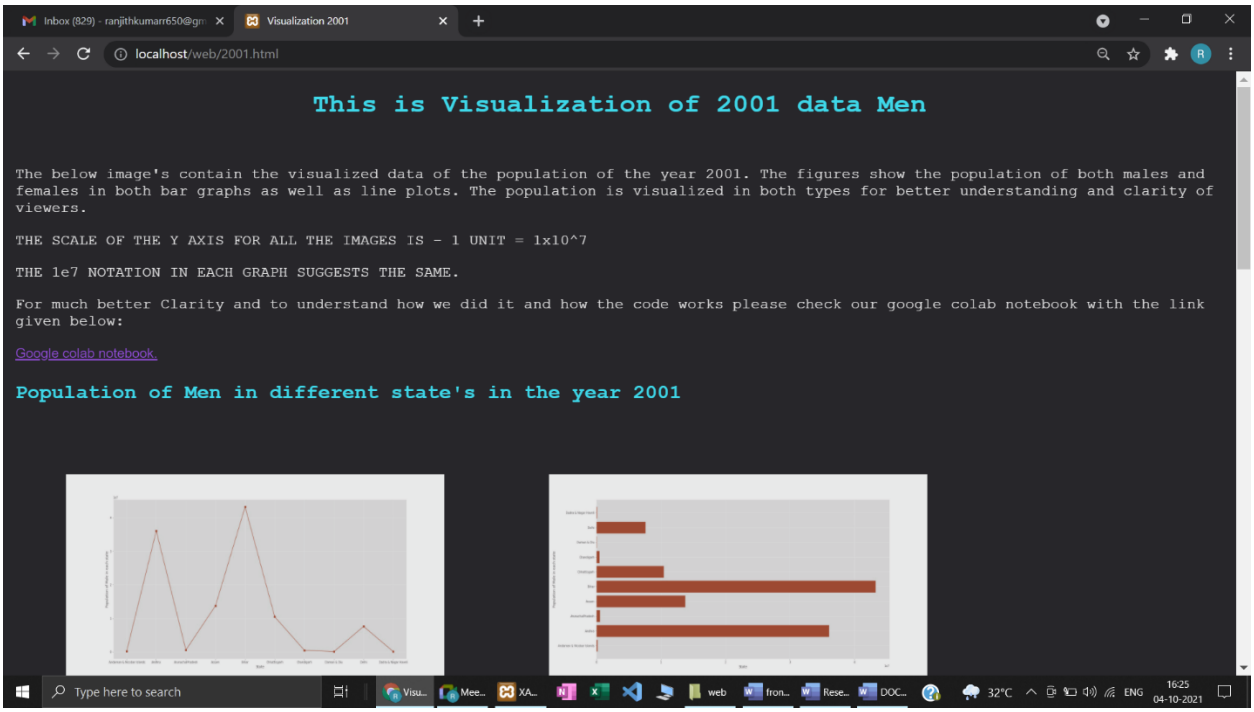


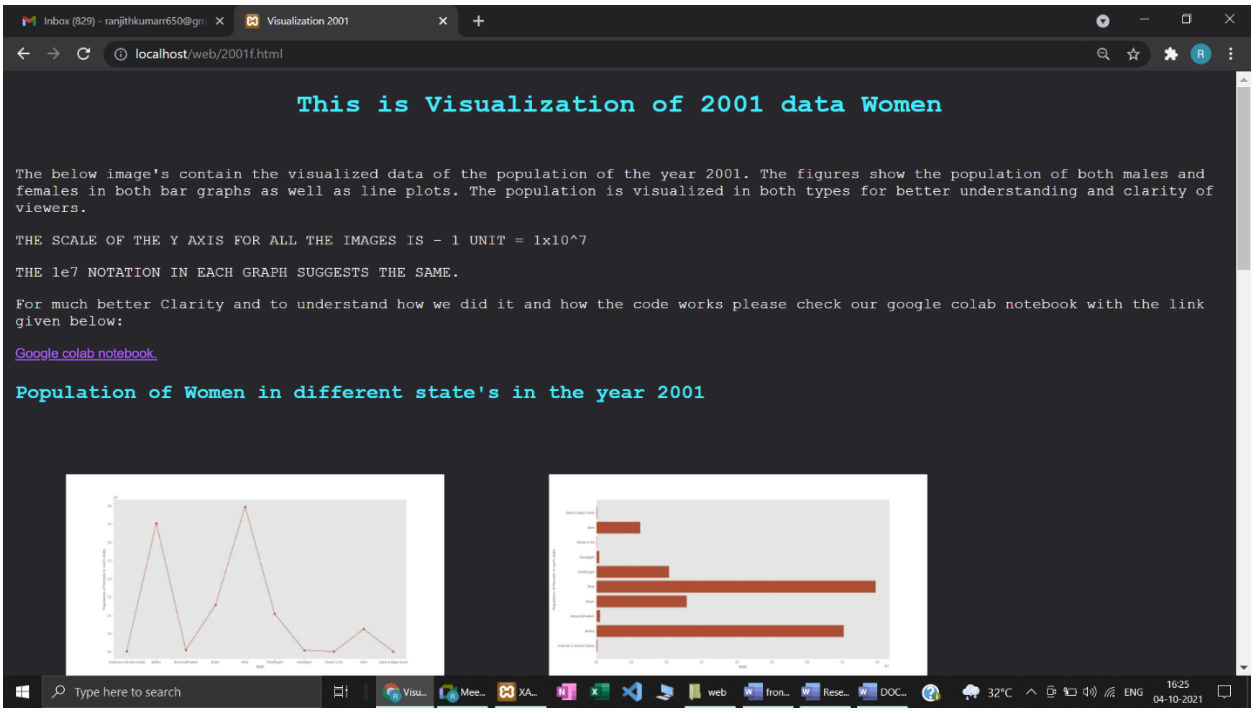


Home page

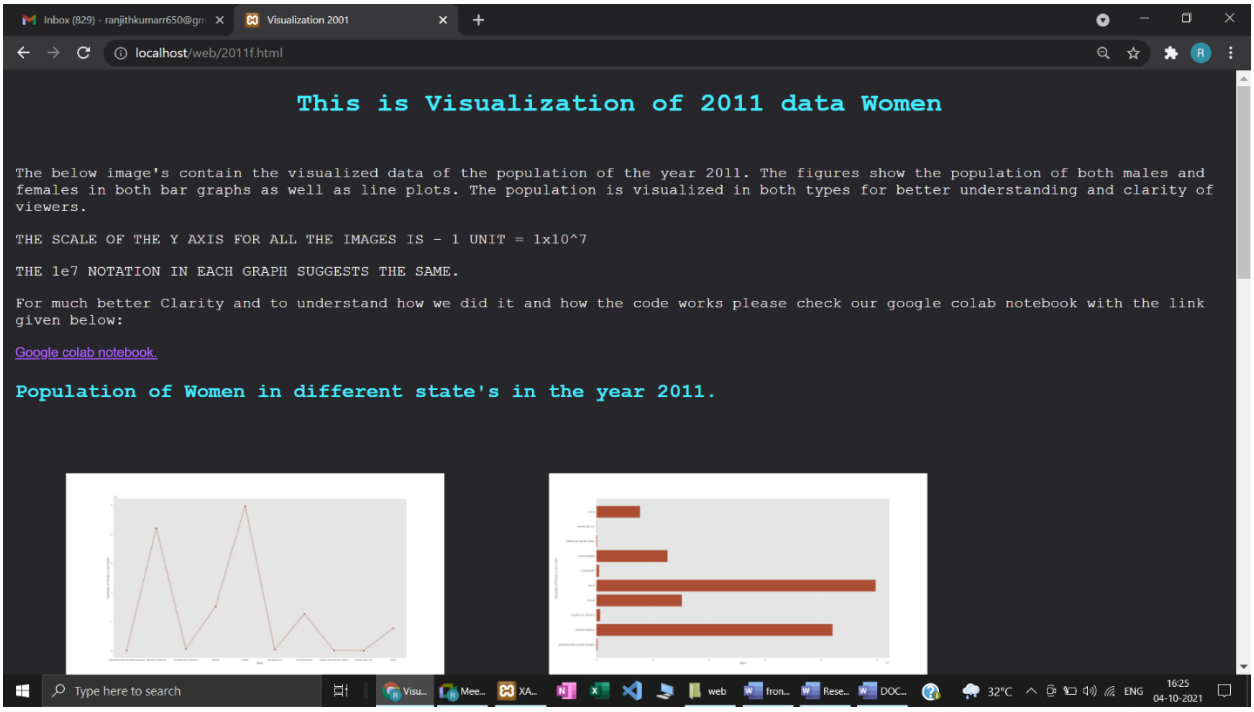

Inbox (829) - ranjithkumarr650@gm
Home Page
localhost/web/home.html
DEMOGRAPHY
Logout
Your Profile
Dashboard
About
Contact Us
Feedback
Logout
1
2001
Visualization of 2001 population of male and female,Literates and working population
Read More
2
2011
Visualization of 2011 population of male and female,Literates and working population
Read More
3
2030
Prediction of 2030 Growth population of male and female.
Read More
Type here to search
32°C
ENG
16:24
04-10-2021

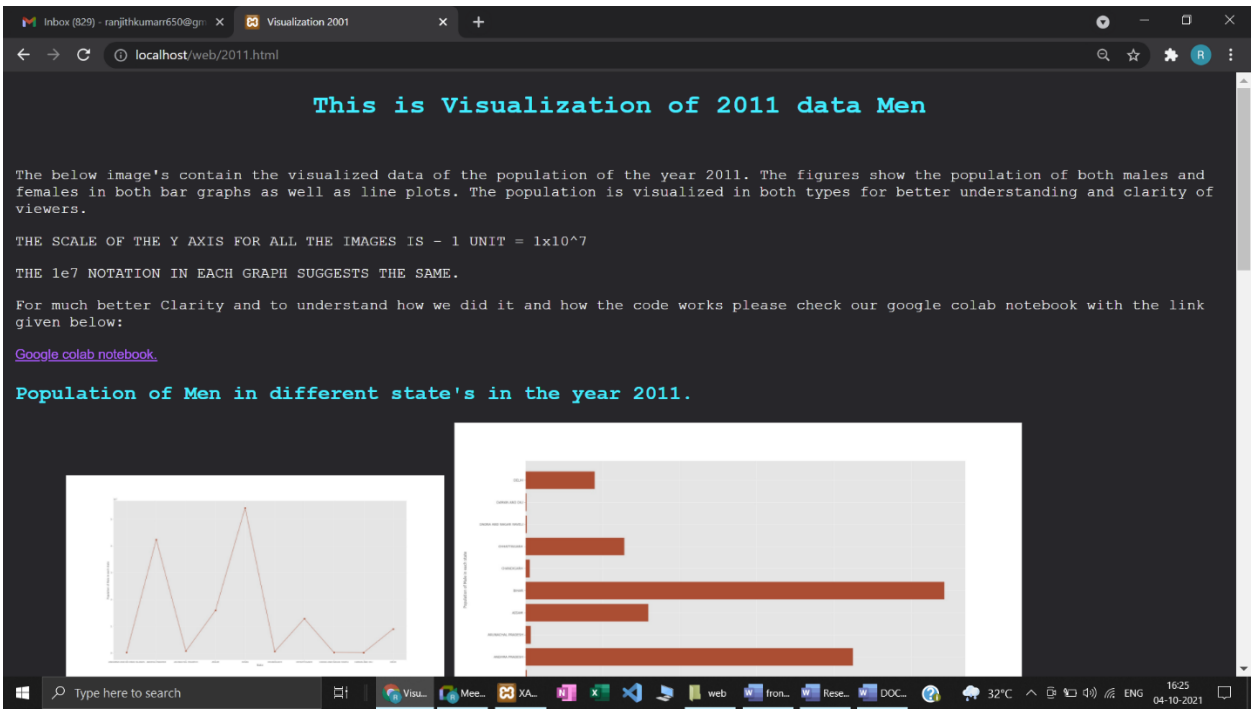
Visualization of 2001 male and female population page



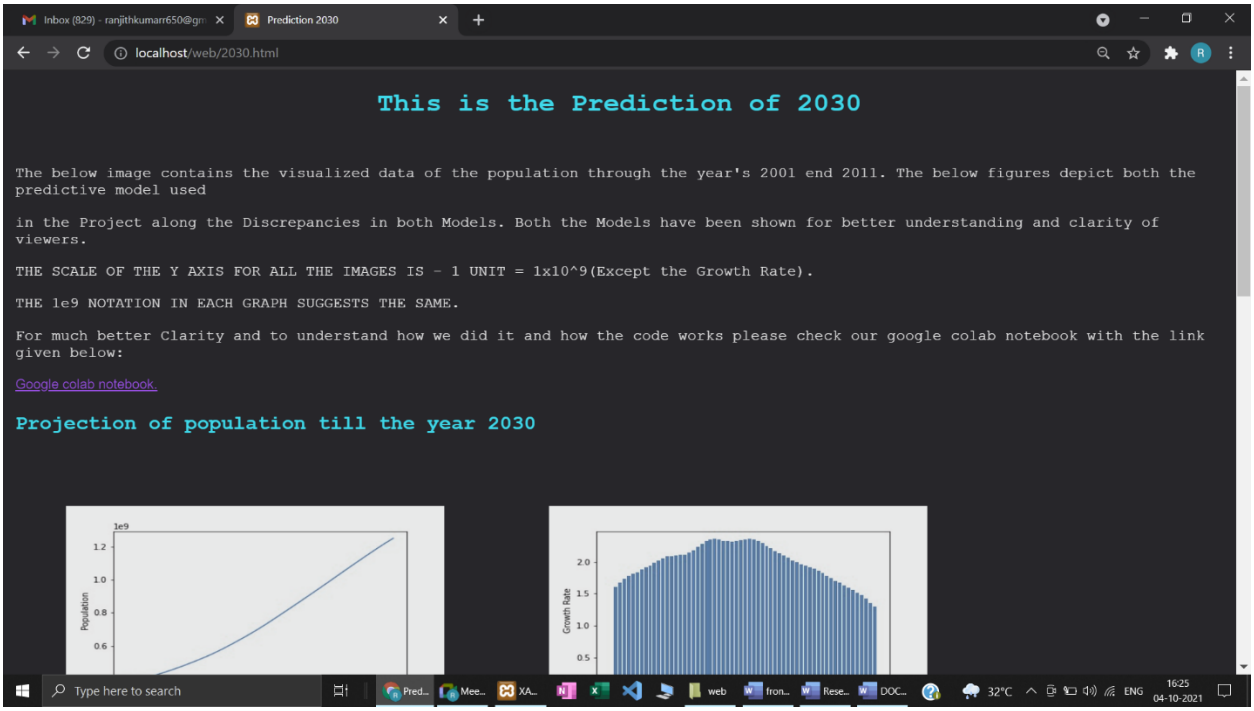


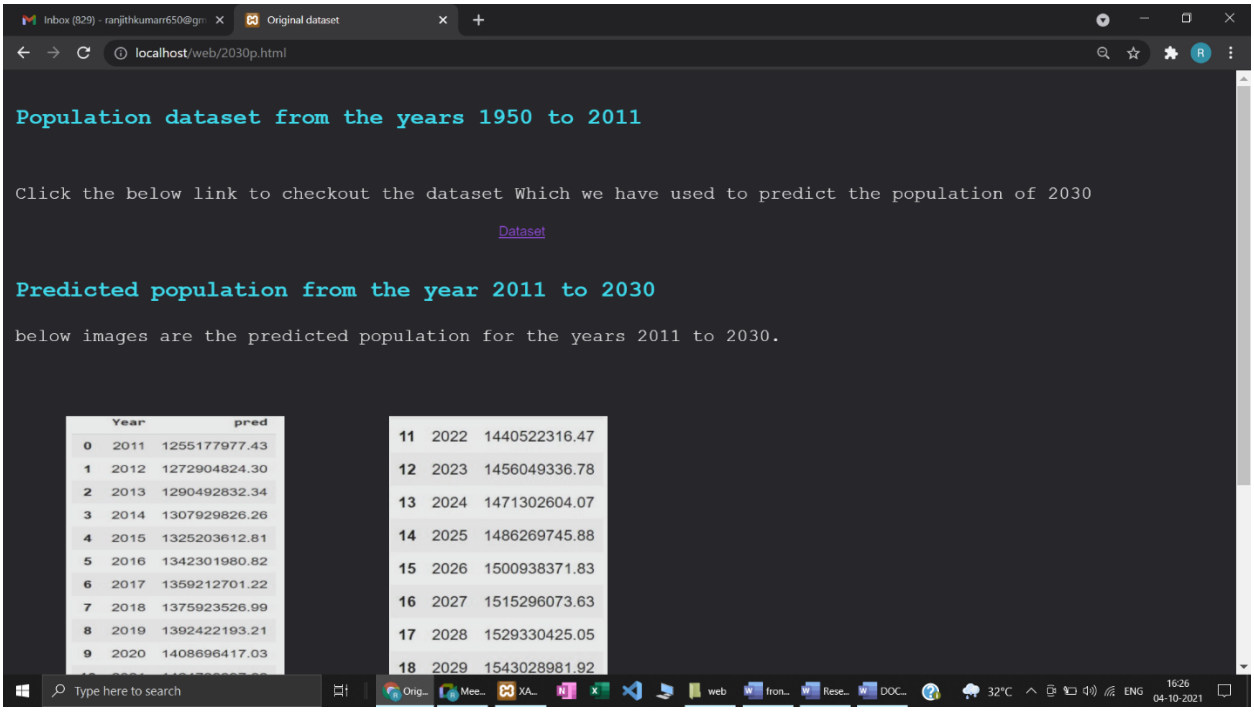
Visualization of 2011 male and female population page



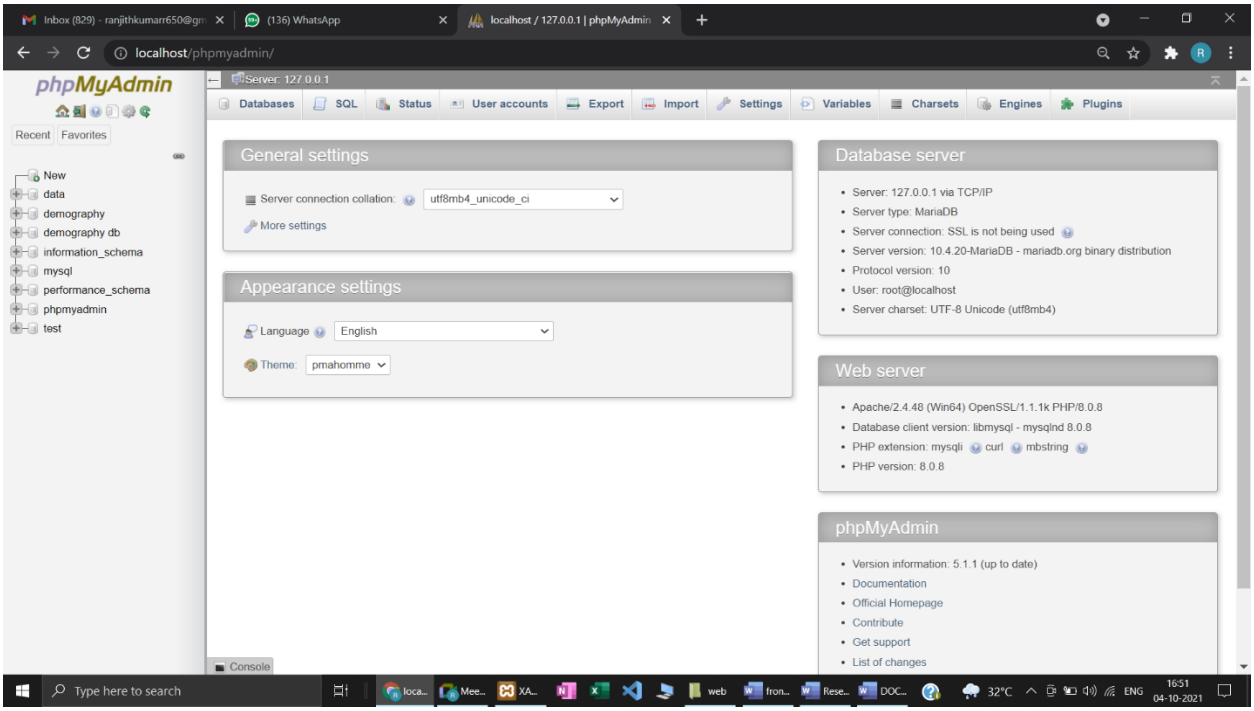


Prediction of 2030 page

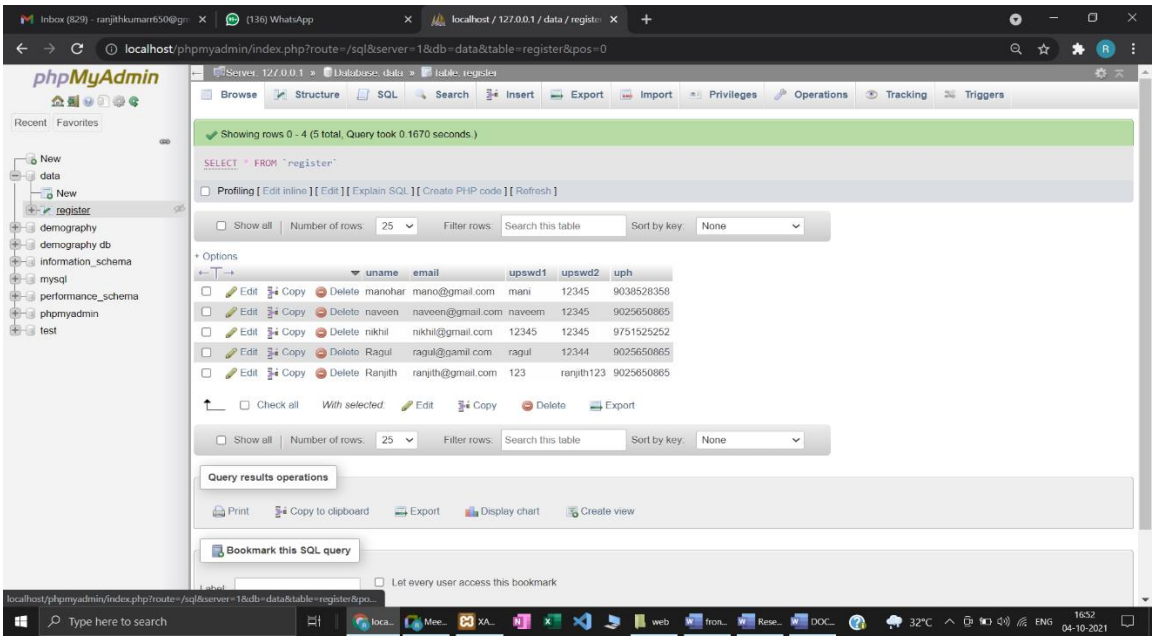




Xampp server



Xampp database



CHAPTER 10

REFERENCES

- https://www.kaggle.com/bazuka/census2001
- https://www.kaggle.com/danofer/india-census
- https://www.kaggle.com/sansuthi/indian-population
- https://learn.datacamp.com/courses/machine-learning-for-time-series-data-in-python
- https://learn.datacamp.com/courses/introduction-to-data-science-in-python
- https://learn.datacamp.com/courses/introduction-to-data-visualization-with-matplotlib
- **N. Keilman, D. Pham, and A. Hetland. Why population forecasts should be probabilistic – illustrated by the case of Norway. Demographic Research, 6, (2002) 409–453.**