

Title: SOIS student admission demography

|  |  |
| --- | --- |
| **Sachin B K.** | **171046023** |
| **Sumanth S.** | **171046020** |

**Course: ME- Big Data and Data Analytics**

**Guide: Deepak Rao**

**Contents**

1. Abstract

2. Introduction

3. Specifications

4. Design

5. Analysis

6. Results

7. Scope for further work

8. Bibliography

**ABSTRACT**

The student admission demographic information is used for “predictive analytics”. The details of the student data like Name, Date of birth, Age, Address, Marks, Ranking and B.E or B-Tec college name collected. The college rank students based on this data. By analyzing these data of students and creating the graphical format. It enables decision makers to see analytics presented visually. So, they can grasp difficult concepts.

**INTRODUCTION**

The age of big data is now coming. But the traditional data analytics may not be able to handle such large quantities of data. The question that arises now is, how to develop a high-performance platform to efficiently analyze the big data and how to develop an appropriate model. To deeply discuss this issue, this project begins with a data collection, data analysis, followed by data visualization.

We gathered the required data set from the office of the SOIS department. For the analysis and visualization, we clean the collected dataset into the requirements. We took some of the major questions based on the dataset given and we analyzed it. For the visualization part we plotted some pie and bar graph on datasets of students and can predict from which states more students are coming, gender and the results of students.

**SPECIFICATION**

The hardware and software requirements of the project are as follows:

**Software Requirement:**

1. Python:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

**Python is Interpreted** − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

**Python is Interactive** − you can sit at a Python prompt and interact with the interpreter directly to write your programs.

**Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

**Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

**2. Pandas:**

Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

**3. Matplotlib (Plotting):**

Matplotlib is a Python 2D plotting library which produces publication quality figures in variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and [I Python](http://ipython.org/) shell, the Jupiter notebook, web application servers, and four graphical user interface toolkits

**DESIGN**

Cleaning and processing of data

Data Visualization

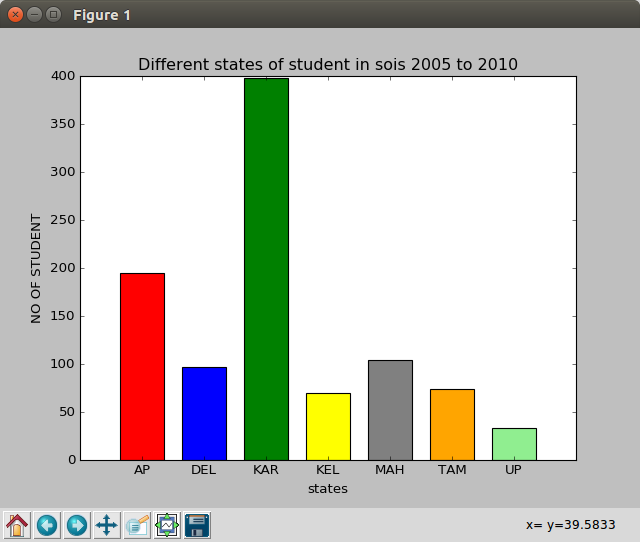
Data Analysis

We gathered the required data set from the office of the SOIS department. For the analysis and visualization, we clean the collected dataset. We took some of the major questions based on the dataset given and we analyzed it. And for the visualization part we plotted some pie and bar graph by using python packages like Numpy, Pandas and Matplotlib can predict from which states more students are coming, gender and the results of students

**MODULES**

**STATES**:

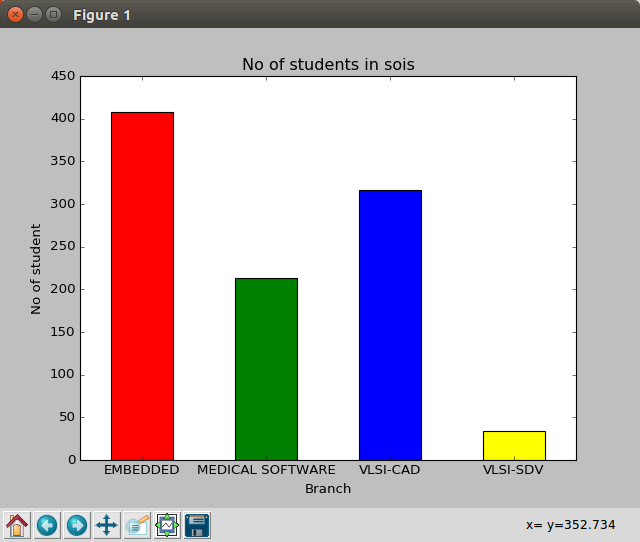
The number of students coming from the different states in year 2005 to 2010



In this analysis maximum people are coming in KAR(Karnataka) and AP(Andrapradesh), MAH(Maharashtra), DEL(Delhi), TAM(Tamilnadu), KEL(Kerla), UP (Uttar Pradesh).

**BRANCH:**

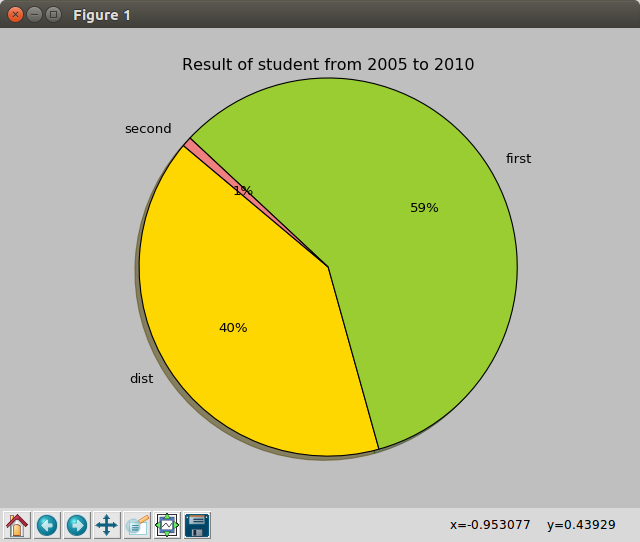
The number of students coming from the different states in year 2005 to 2010



2015The numbers of students in a branch like Embedded System, Medical software, VLSI-CAD, VLSI-SDV

**RESULT:**

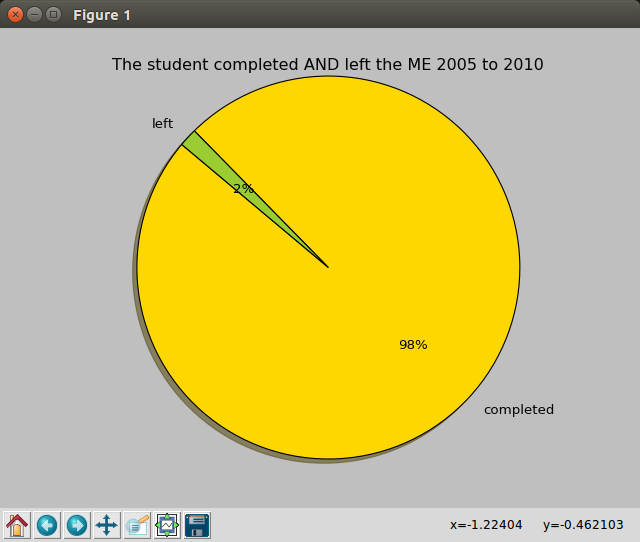
The number of students coming from the different states in year 2005 to 2010



In the sois the student results are be like First class and Second class and Distinction. First class = 59%, Second class =1%, Distinction = 40%

**STATUS:**

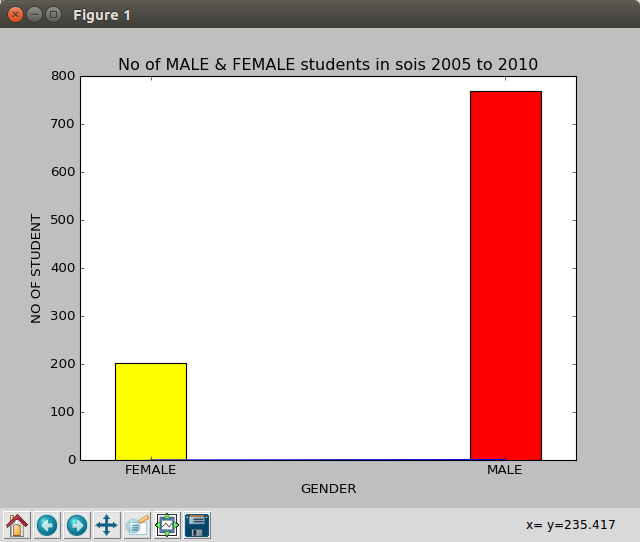
The number of students coming from the different states in year 2005 to 2010



In the 2005 to 2010 the total percentage of students are completed and left college. Completed = 98%, Left = 2%

**GENDER:**

The number of students coming from the different states in year 2005 to 2010



The number of Male and Female students joined in the year 2005 to 2010. The red color indicates strength of the Male students and yellow color indicates the strength of the female students

**ANALYSIS**

We Analyzed the SOIS student admission demography dataset by putting up some of the questions

1. Display the strength of students states from 2005-2010?

2. Display the results of the student of all the branches from 2005-2010?

3. Display the gender of students in percentage of all the branches from 2005-2010?

4. Display the status of students in percentage of all the branches from 2005-2010?

5. Display the strength students branch from 2005-2010?

We Visualized the SOIS student admission demography dataset by putting up some of the questions

1. Show the Visualization for the strength of students states from 2005-2010

2. Show the Visualization for results of the student of all the branches from 2005-2010

3. Show the Visualization for gender of the student of all the branches from 2005-2010

4. Show the Visualization for status of the student of all the branches from 2005-2010

5. Show the Visualization for strength students branch from 2005-2010

**RESULTS**

We got output for Analysis and Visualization that we had done on student’s admission demography dataset. On the basis on Analysis and Visualization, we can predict how that took place in SOIS department since past 5 years. Can predict from which states more students are coming, gender and the results of students.

**SCOPE OF FUTURE WORK**

The scope of the future work in sois student admission demography provided with the district address and engineering marks of a students in dataset graph can be plotted for it. which inference form which part of the district more students are going to come and clearly visualize the marks of the student before joining to the sois and marks increased after joining the sois.

**BIBLIOGRAPHY**

* Ten Simple Rules for Better Figures by Nicolas P. Rougier, Michael Droettboom, Philip E. Bourne, <https://hal.inria.fr/hal-01063732>
* A survey of powerful visualization techniques, from the obvious to the obscure by Jeffrey Heer, Michael Bostock, and Vadim Og ievetsky
* Interactive Data Visualization for the web by Scott Murray.