*USA University Recommendation System based on Student profile for UG/graduate studies*

**Rama Tejaswini Thotapalli (SJSU Id: 013785681)**

Under the guidance ofProfessor Shih Yu Chang, San Jose State University

**Abstract:**

In this paper, we present an applied research on

designing and developing a recommender system for graduate

admission seekers which can help them to choose graduate school

matching their entire academic profile. Here we have developed a

technique to transform relational database for students' all types

of relevant information into a universal database format using

academic data of successful students who have already got

opportunity to study abroad. After that we have developed an

algorithm for grad school recommender system which can

calculate similarity between training and test data set based on

weighted scores using mean squared deviation similarity metrics.

We have used K-nearest Neighbor algorithm for calculating top

N similar users for the test users and recommend Top K

universities to users from N similar users. Finally our proposed

recommender system will recommend list of universities to apply

for graduate admission to pursue higher study abroad with

funding.

In this paper, we present an applied research on

designing and developing a recommender system for graduate

admission seekers which can help them to choose graduate school

matching their entire academic profile. Here we have developed a

technique to transform relational database for students' all types

of relevant information into a universal database format using

academic data of successful students who have already got

opportunity to study abroad. After that we have developed an

algorithm for grad school recommender system which can

calculate similarity between training and test data set based on

weighted scores using mean squared deviation similarity metrics.

We have used K-nearest Neighbor algorithm for calculating top

N similar users for the test users and recommend Top K

universities to users from N similar users. Finally our proposed

recommender system will recommend list of universities to apply

for graduate admission to pursue higher study abroad with

funding.

In this paper, we present an applied research on

designing and developing a recommender system for graduate

admission seekers which can help them to choose graduate school

matching their entire academic profile. Here we have developed a

technique to transform relational database for students' all types

of relevant information into a universal database format using

academic data of successful students who have already got

opportunity to study abroad. After that we have developed an

algorithm for grad school recommender system which can

calculate similarity between training and test data set based on

weighted scores using mean squared deviation similarity metrics.

We have used K-nearest Neighbor algorithm for calculating top

N similar users for the test users and recommend Top K

universities to users from N similar users. Finally our proposed

recommender system will recommend list of universities to apply

for graduate admission to pursue higher study abroad with

funding.

In this paper, we present an applied research on

designing and developing a recommender system for graduate

admission seekers which can help them to choose graduate school

matching their entire academic profile. Here we have developed a

technique to transform relational database for students' all types

of relevant information into a universal database format using

academic data of successful students who have already got

opportunity to study abroad. After that we have developed an

algorithm for grad school recommender system which can

calculate similarity between training and test data set based on

weighted scores using mean squared deviation similarity metrics.

We have used K-nearest Neighbor algorithm for calculating top

N similar users for the test users and recommend Top K

universities to users from N similar users. Finally our proposed

recommender system will recommend list of universities to apply

for graduate admission to pursue higher study abroad with

funding.

In this paper, I have presented a recommender system for undergraduate & graduate admission seekers, which can help students to choose best graduate university matching their academic profile. Here I have used a different data mining techniques to transform database of students of relevant information into a universal database format using academic data of successful students who have already got opportunity to study abroad. After that I have developed a machine learning algorithm which can calculate similarity between training and test data based on weighted scores. I have used K-nearest Neighbor algorithm and feature-weighted algorithm for calculating top N similar users for the test users and recommend Top K universities to users from N similar users.

**Introduction:**

Many students who wants to pursue higher studies apply different universities with their academic profile as well as standardized test scores such as SAT, GRE, TOEFL, and IELTS. Institutions take in the students who are suitable candidates based on their academic profile, standardized test scores. But in this entire process university selection is the most crucial & tedious step for applying to graduate studies. Some of them succeed and get admission into their desired programs in desired universities, but some are not because of the academic level of colleges which they have applied. To resolve this problem of not getting admission because of applications, even though students have good academic profile, I have developed this recommendation system. In this project, the knowledge acquired from the database of successful applicants is used to predict the schools with various data mining techniques. This data will be modeled into machine learning algorithms to predict the universities and their acceptance rate for the given user academic details.

**Background & Objectives of Project:**

For an aspiring student who wants to apply for higher studies in other countries, university selection process is a challenging task as lot of different criteria need to consider during application process based on individual’s requirement. Some of them succeed and get admission into their desired programs in desired universities, but some are not because of the academic level of colleges, which they have applied.

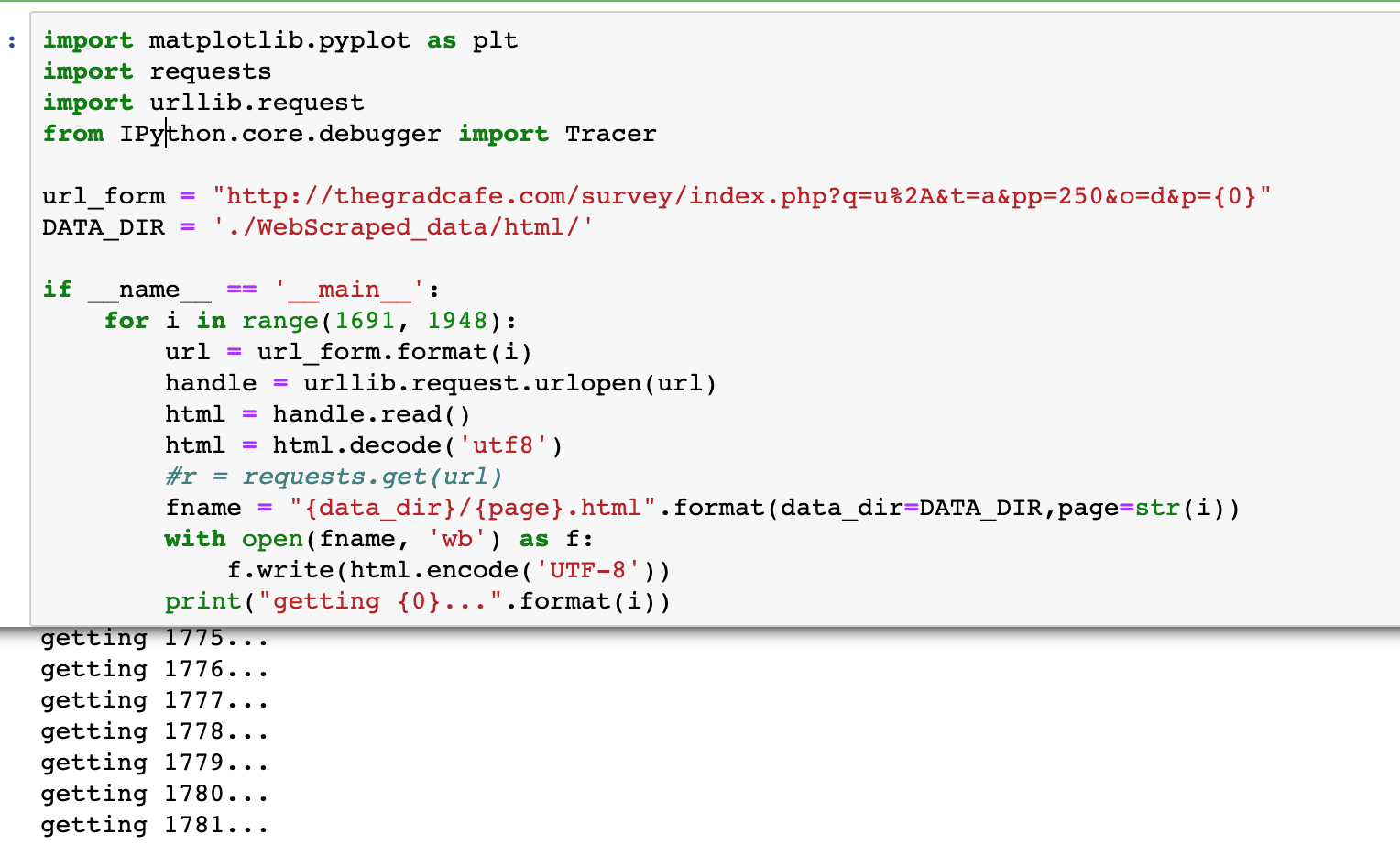
This problem can be addressed by modeling a recommender system based on various classification algorithms. In this project based on the student data set and the student profile who is looking for the admit, various models will be trained and a list of 10 best universities will be suggested such that it maximizes the chances of a student getting admit from that university list..

**Dataset:**

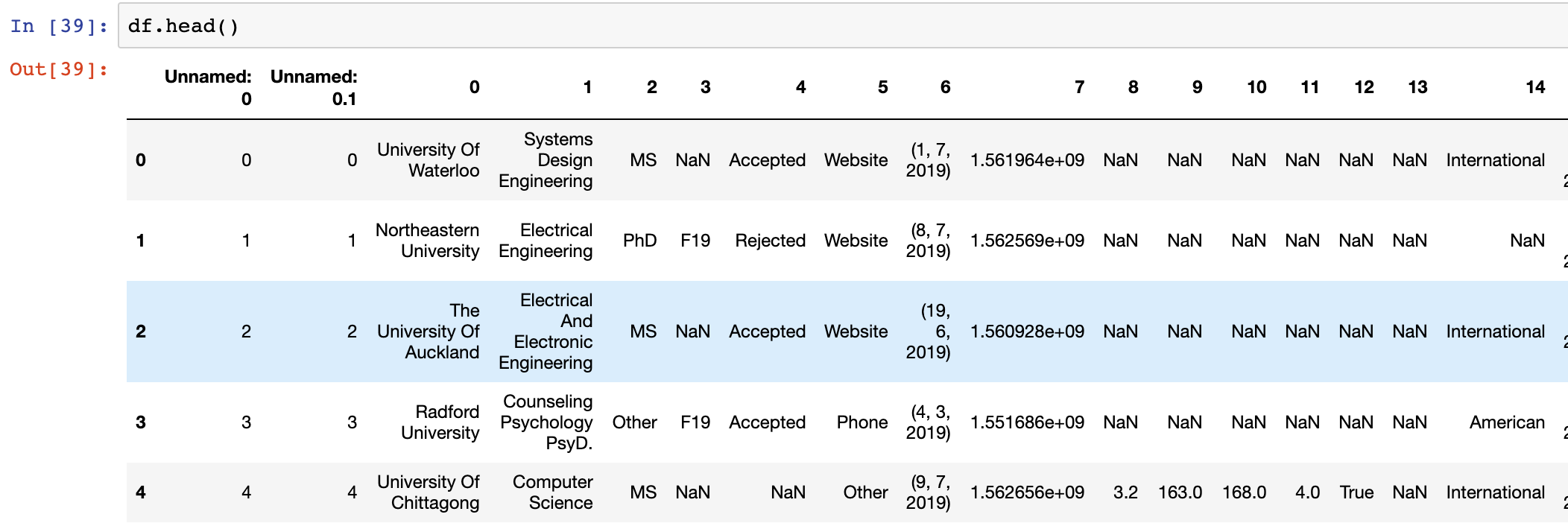
The first step in building any recommendation system is the identification of the data set. In order to build the classification model for the recommender system, this data has to be organized with appropriate labels. This core data for the application process is not readily available on the internet for direct consumption. However, this whole approach is based on making maximum use of the available information. The graduate student data was scraped from the following websites [www.thegradcafe.com](http://www.thegradcafe.com), and the Undergraduate university student data was scraped from <https://collegescorecard.ed.gov/data/>.

Graduate Student Dataset:

For Graduate Student data, we scraped [www.thegradcafe.com](http://www.thegradcafe.com) website. About 271807 rows of raw student data was obtained as a result of web scraping. Each sample corresponds to the profile of a student. We have got 1949 html pages of the data and need to change it into CSV files.



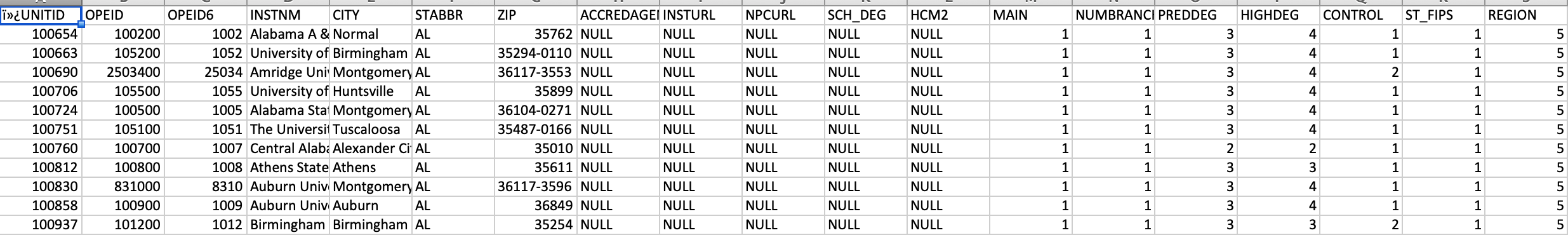
After scraping the final data will look like



The list of attributes are made as dataset for pre-process cleansing. For graduate students the dataset consists of University Name, Major, Degree, Season, Decision, Decision Method, Decision Date, Undergraduate GPA, Is New GRE Verbal, GRE Quant,   GRE Writing, Status,   Postdate Comments, Research Experience, Recommendations and Undergraduate GPA. For Under graduate students dataset consists of Student profile and SAT scores.

Under Graduate student dataset:

Under Graduate student data is taken from the College rank score card website <https://collegescorecard.ed.gov/data/>. The data before cleaning looked like below.



Data Preprocessing:

In order to use the obtained data for our analysis, we need to do the preprocessing and cleansing, as there are lots of anomalies in the dataset. For this we use pandas and numpy frameworks.

Cleansing the data was done by

* Removing the irrelevant columns by using the drop column feature
* Filling the null values with the appropriate value or deleting the row containing null values.
* Removing the spaces in the data and reducing the size of the dataset.

In our graduate dataset, The GRE scores were also cleansed since they contained scores of both old and new versions of the examination. Similarly the GPA scores available were based on different point systems, so all the GPA scores were uniformly scaled to 4 point scale by using normalize functions.

*x* normalized = (*x* – *x* minimum) / (*x* maximum – *x* minimum)

Where x is the value of the GPA

Exploratory data Analysis:

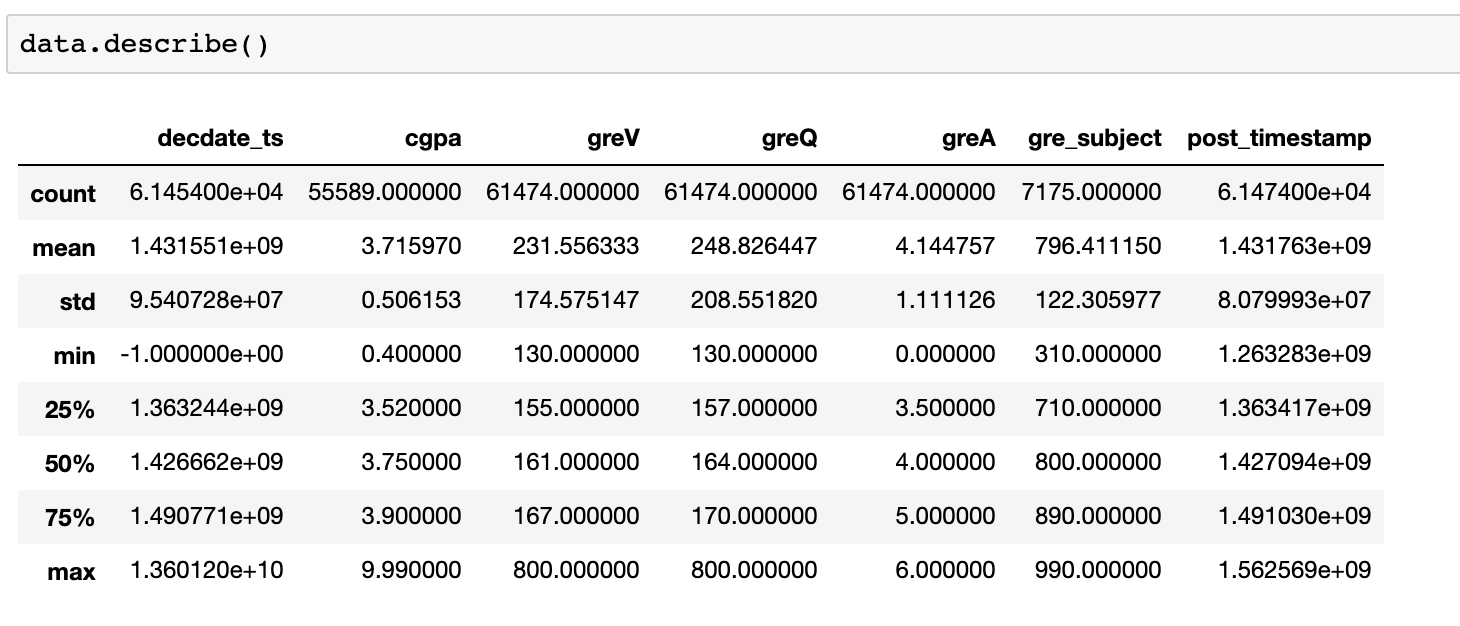
Exploratory data Analysis is a technique, which employs a variety of techniques. It consists of various techniques like below.

1.Plotting raw data

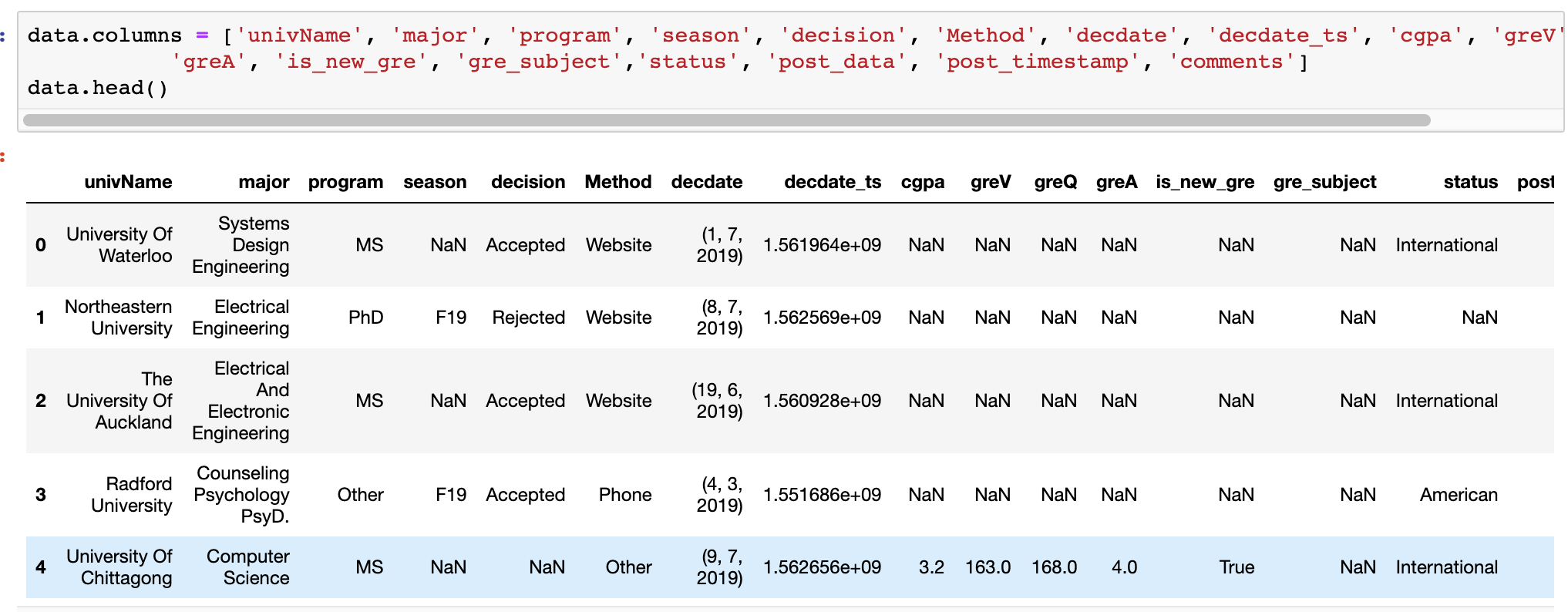
2. Plotting simple statistics such as mean , standard deviation, etc.

3. Positioning such plots, so as to maximize our natural pattern recognition.

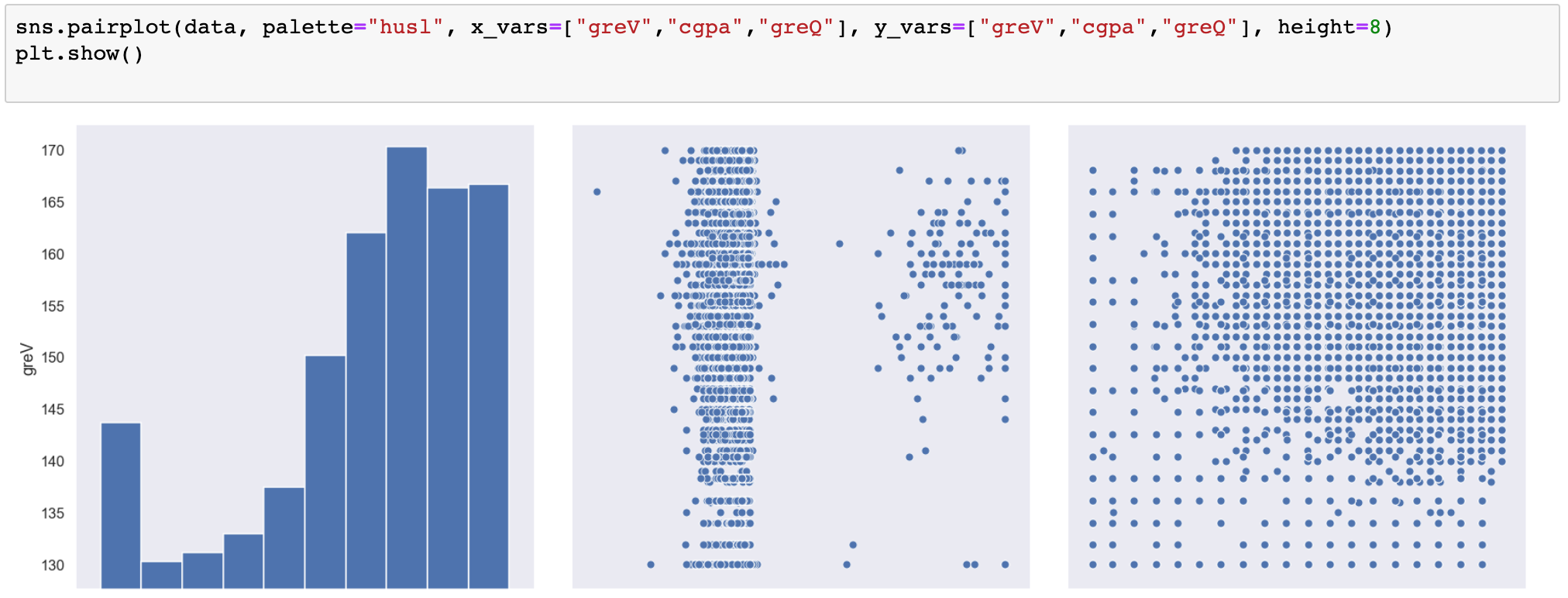
This is the description of the data.

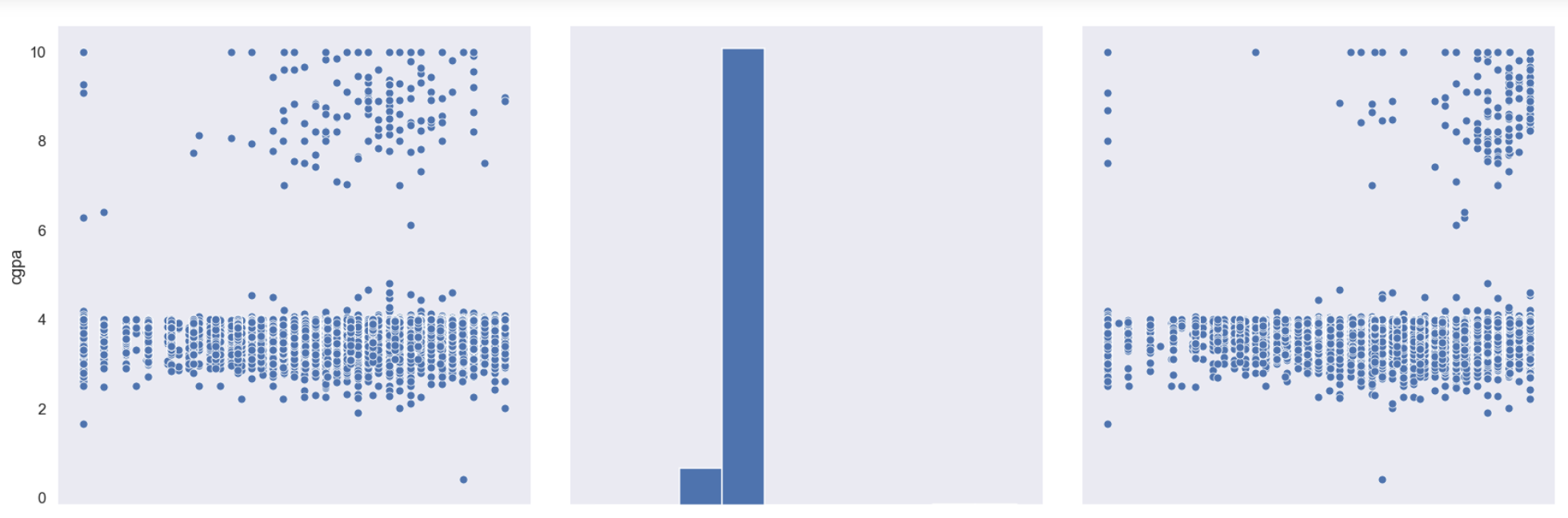


The Processed data will look like below.



Exploratory data analysis: pair plots of GRE verbal, GRE Quantitative and GPA.



**Undergraduate Data:**

In Undergraduate data, we have taken below few rows of data like Institution name, city, tuition Fees, Sat Score, Admission rate, Debt and Men Ratio.

INSTNM', 'CITY', 'STABBR' ,'TUITIONFEE\_OUT', 'SAT\_AVG\_ALL', 'ADM\_RATE\_ALL', 'DEBT\_MDN\_SUPP', 'UGDS\_MEN'])

This data will be used for training the model and test data as SAT score and Maximum tuition fees.

**Analysis and Methodology of project:**

Here I have used Knowledge based recommendation System where User inputs are taken into account and compare with the training data.

For Graduate University Recommendation I have used Case based knowledge recommendation as it will take the User inputs and compare with trained data.

For Undergraduate Recommendation System, I have used Constraint based Knowledge recommendation system where user inputs taken into account as constraints and based on the constraints I compared with trained data.

I used two different models like K-Nearest Neighbors for Graduate data and Feature weighted algorithms for Undergraduate data.

# K Nearest Neighbor:

# In KNN, the trained data is compared with test data and distances are calculated using Euclidean distance. It then classifies an instance by finding its nearest neighbors and recommend the top n nearest neighbor universities.

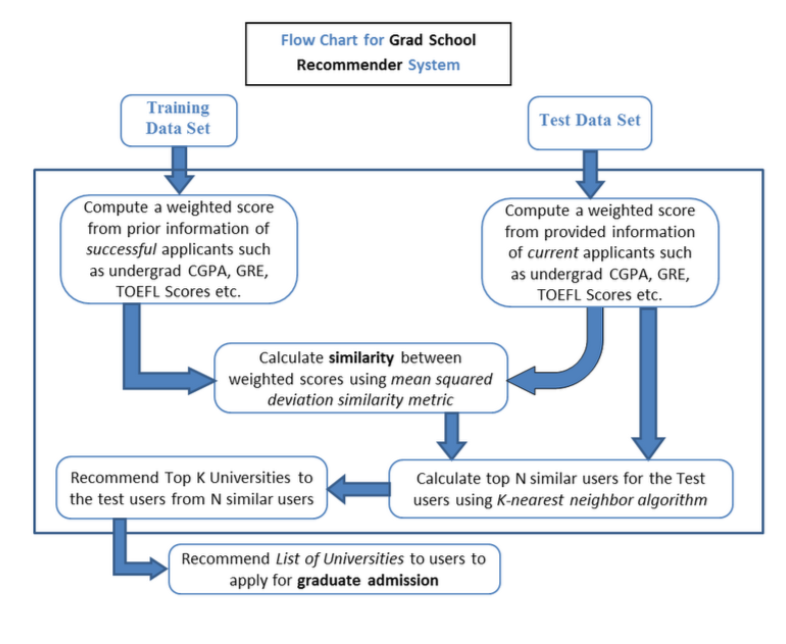
# Algorithm is stated as below.

Input: undergraduate university, department, CGPA, GRE Scores of User

1. Initialize the value of k
2. For getting recommendation, iterate from 1 to number of trained data
3. Calculate distance between test data and each row in the trained data.
4. Sort the distances in ascending order
5. Get top k rows and recommend to the user

Output: highly recommended N Outgoing University analyzing Universal Table of Previous Successful Students

**Flowchart of the graduate recommendation System:**



**Training data for the KNN algorithm:**



**KNN algorithm code:**

# Macintosh HD:Users:Sagar:Desktop:Screen Shot 2019-08-06 at 11.30.34 PM.png

# Macintosh HD:Users:Sagar:Desktop:Screen Shot 2019-08-06 at 11.30.49 PM.png

# Feature weighted algorithm for Undergraduate universities:

# The weightage of all the features are taken and find the similarity score. Based on the similarity score, the universities with highest similarities will be recommended to student. Suppose w1, w2 are weights and f1 and f2 are features the similarity is calculated by formula Similarity

# score = w1\* f1+w2\*(1-f2)

# Algorithm is stated as below.

Input: SAT Score and Maximum tuition fees of User

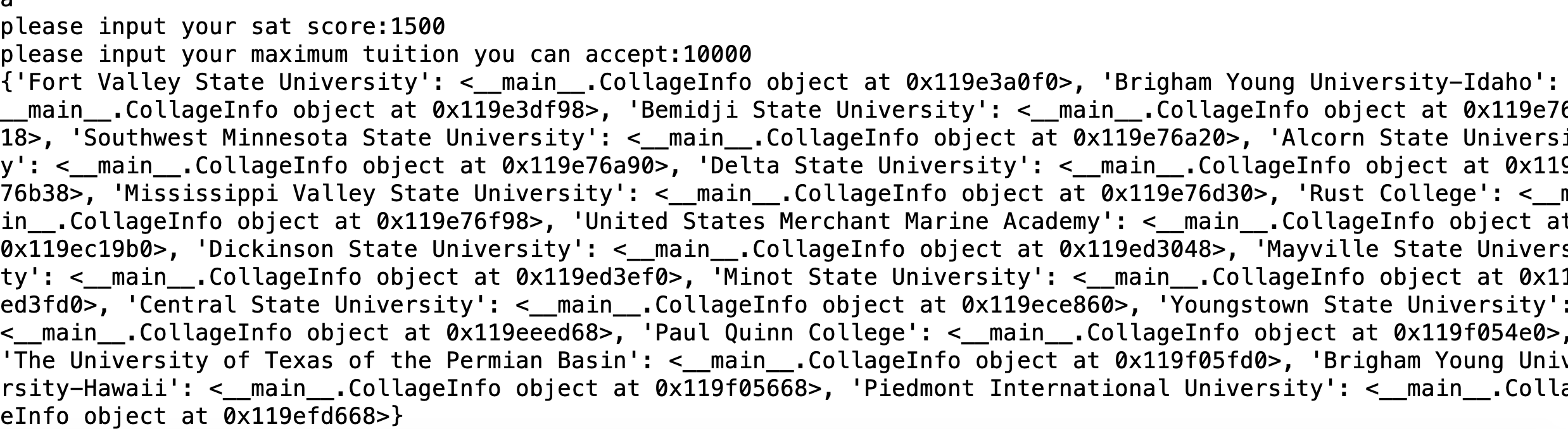
1. For getting recommendation, iterate from 1 to number of trained data
2. Find the rows in the training data similar to the user provided SAT score and max tuition fees.
3. Calculate the weightage of both the attributes and calculate the score as acceptance rate
4. Sort the distances in ascending order
5. Get top k rows and recommend to the user

Output: Top 5 Recommended Universities

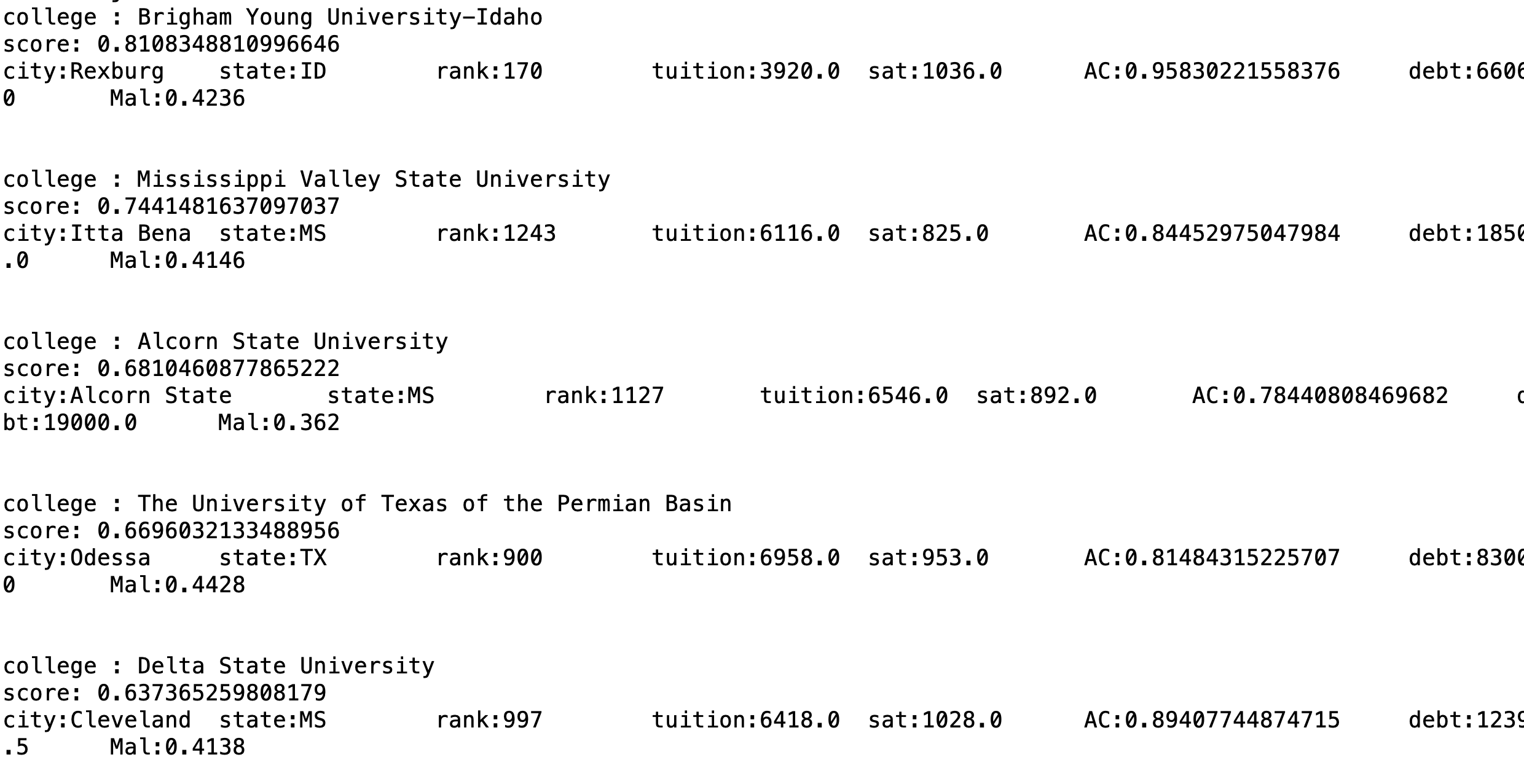
**Results of the models:**

For Undergraduate Universities recommendation, weighted algorithm outputs:

Input of the data for SAT Score and Maximum Tuition Fees:



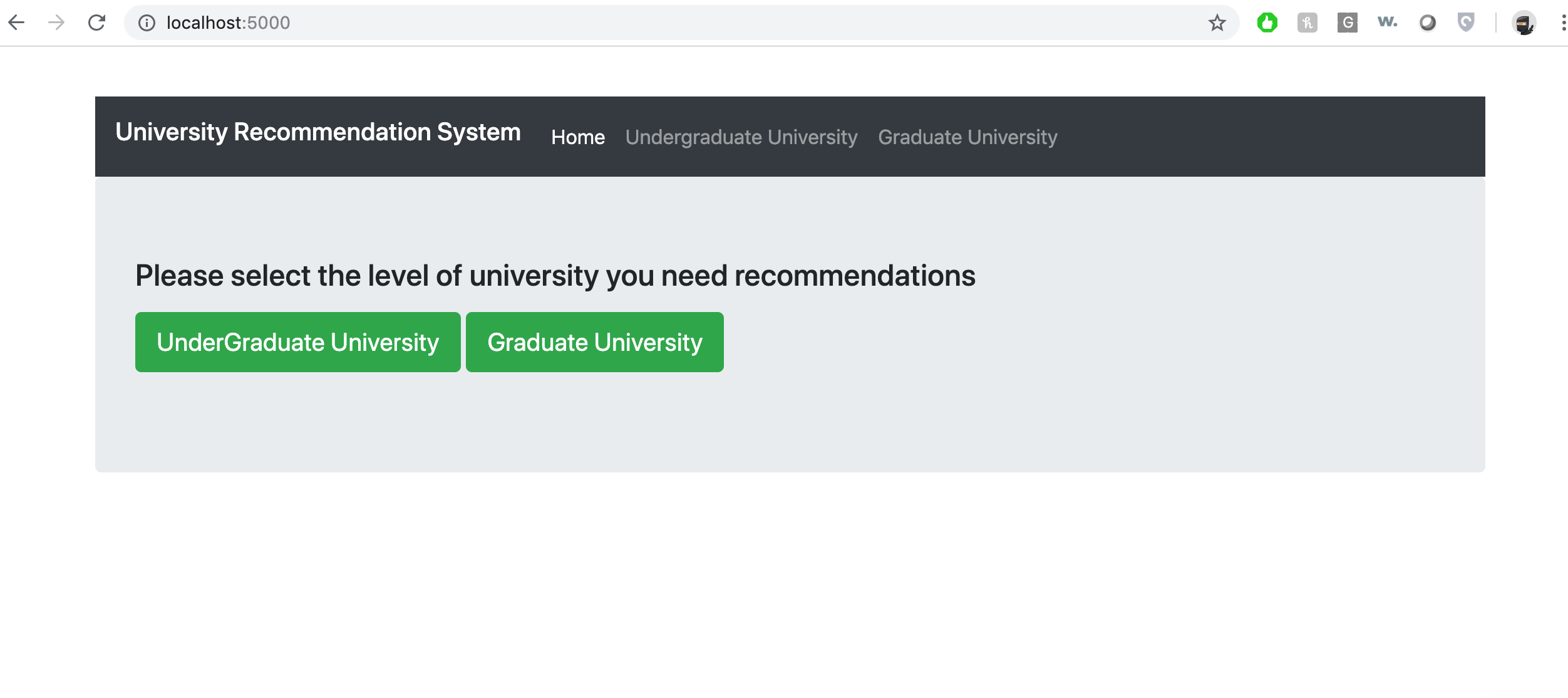
Output:



**Web Application:**

I have developed a web application for University Recommendation System. Below are the Screenshots for this.

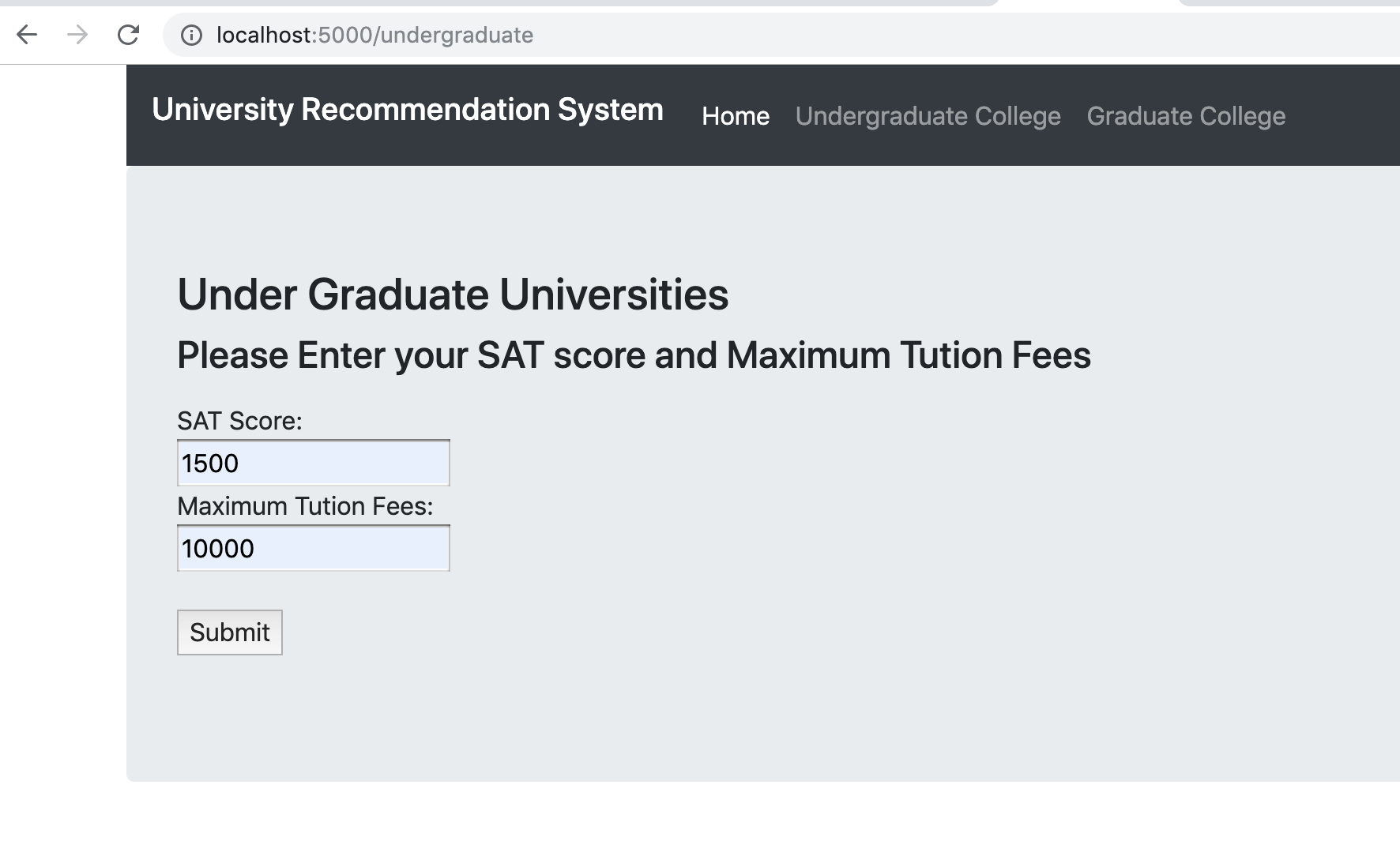
Home Page: This is the home page which has links to Under graduate and graduate Recommendation system



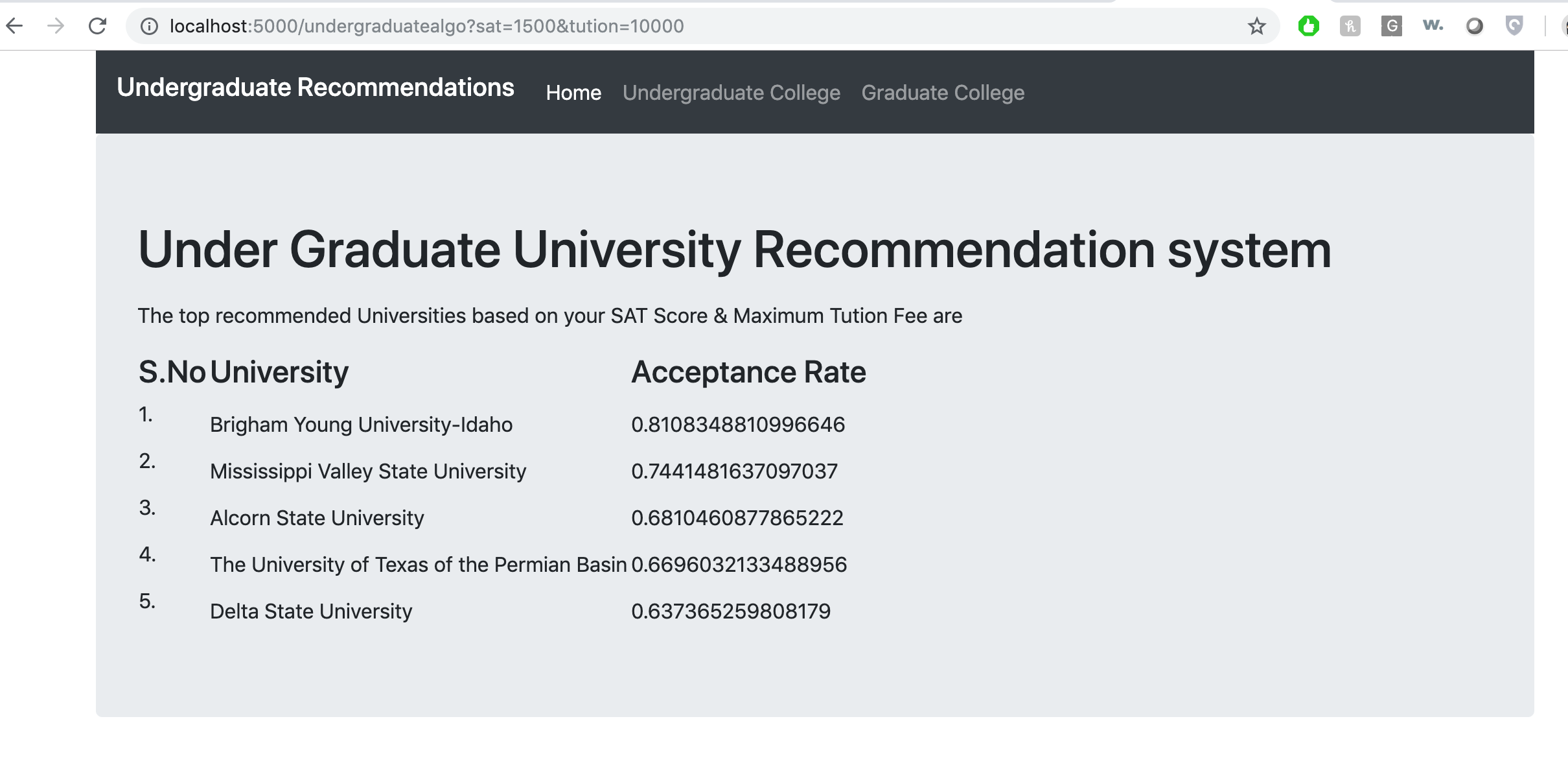
2. Once the user clicks on the Undergraduate College, he/she will land on this page where they can provide the SAT score and Maximum Tuition fees for their college recommendation.



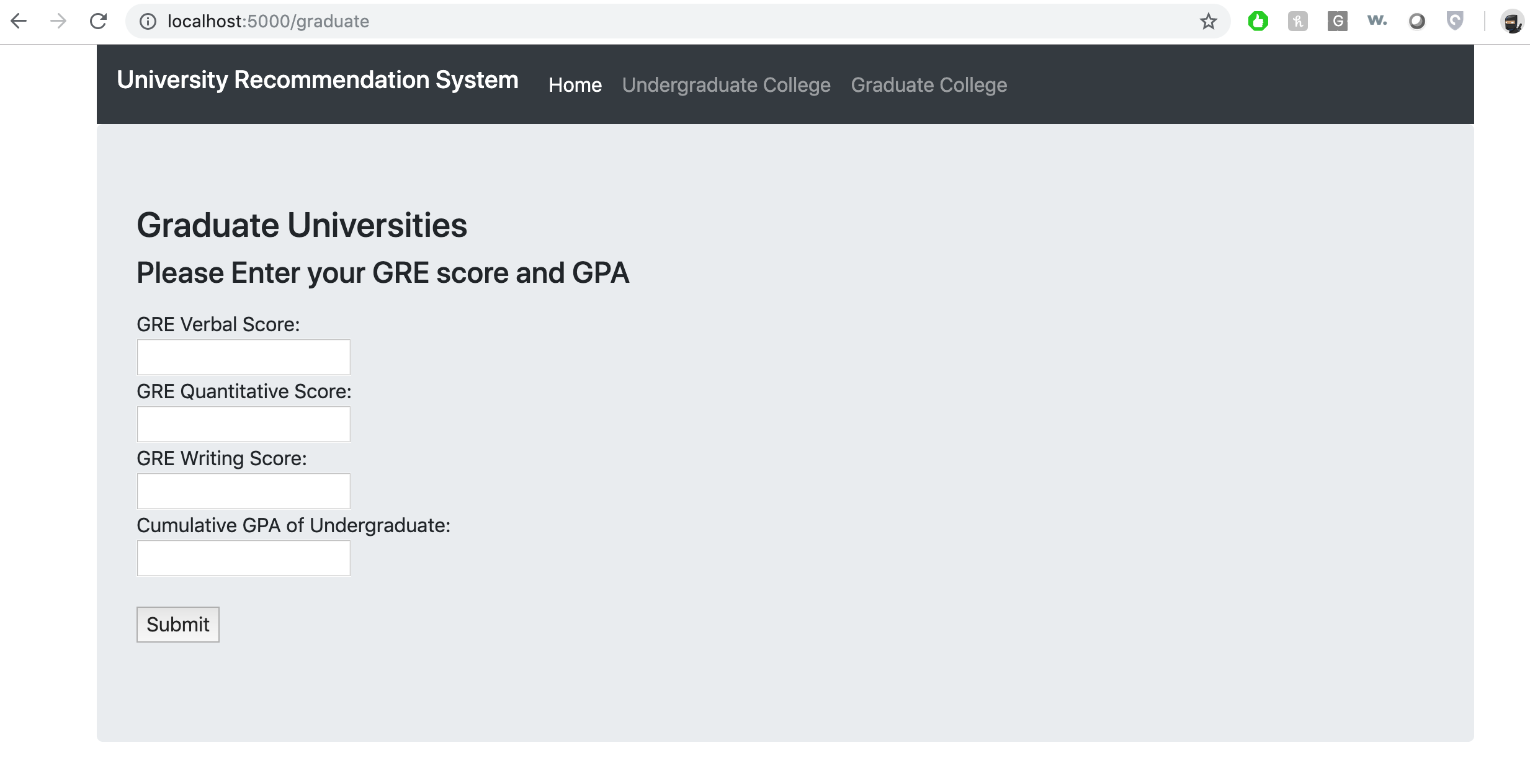
3. If user provides SAT Score as 1500 and Max tuition fees as 10000, Then the output of University names he gets is in the below screenshot.



Once user submits he will get these recommendations as below.



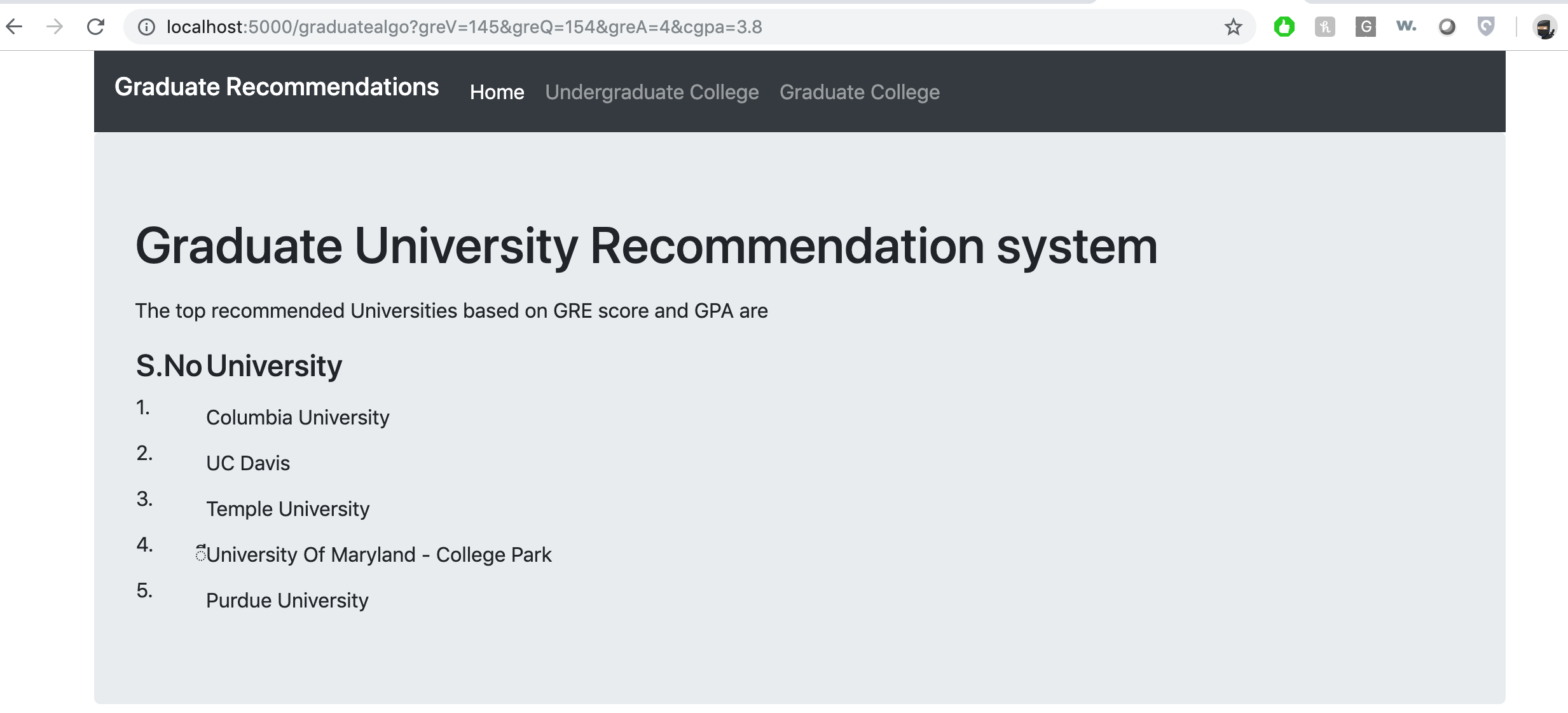
4. To Get the Graduate University recommendation list, User need to provide his/her GRE scores and GPA of the undergraduate University.



1. If user provides, GRE score and CGPA as below and click on submit, the recommendation system will provide output as below.



1. This is the output for the Graduate Recommendation System.



Demo Video:

**Conclusion:**

This project helps students in the decision making of the universities in which they apply. The data of the previous successful applicants can be taken into account. The data from the academic records of applicants is very important for the admission seekers in foreign. In this research, I have developed a technique of using those academic records of successful applicants for making school recommender system, which can help the current admission seekers. At first, I calculate similarity between training and test data set based on weighted scores. The weighted scores are calculated from prior information of successful applicants such as undergrad CGPA, GRE, TOEFL Scores and all other relevant records found in the universal database. I have used K-nearest Neighbor algorithm for graduate universities and feature weighted algorithm for Undergraduate Universities in order to calculate top N similar users and then recommend top K universities to the users. Our proposed recommender system will recommend list of universities to applicants trying to pursue higher study abroad and assist them to apply for graduate admission in appropriate universities.

**References:**

1. <https://ieeexplore.ieee.org/document/7760053>
2. <https://www.semanticscholar.org/paper/Recommender-System-for-Graduate-Studies-in-USA-Suresh/22924fda3f293f80a3f62f32799c08d0b81a9b20>
3. <https://www.researchgate.net/publication/311758642_Graduate_school_recommender_system_Assisting_admission_seekers_to_apply_for_graduate_studies_in_appropriate_graduate_schools>
4. <http://jmcauley.ucsd.edu/cse258/projects/fa15/026.pdf>
5. <https://devpost.com/software/graduate-school-recommendation-system>
6. <http://paper.ijcsns.org/07_book/201801/20180111.pdf>