College Recommender System

by Kevin Trinh

Table of Contents

01

Problem Statement and Methods

02

Data Cleaning and EDA

03

Modeling and Evaluation

04

Conclusions and Application

Problem Statement

- High school can be a challenging time for both students and their parents.
- On top of balancing grades, standardized testing, and extracurricular activities, choosing which colleges to apply to can be another stressful task that high school students have to take on.
- To help take the stress off of picking colleges, a recommender system would be a very helpful tool for students and their parents.
- A college recommender system could filter the right colleges based on field of study, location preferences, and degree types that they are searching for.

Methods

- Collected data on U.S. colleges
- Cleaned data and selected specific features
- Calculated summary statistics
- Conducted exploratory data analysis and visualized initial findings
- Utilized models to cluster data
- Evaluated model and formed conclusions
- Developed recommender system for students and parents

Data Collection, Cleaning, and EDA

- Data was collected from the U.S. Department of Education's College Scorecard data files
- The dataset used was specifically from the 2022-2023 academic year
- Certain features of the dataset were selected and cleaned properly
- Exploratory Data Analysis was performed on data involving average SAT scores, admission rates, types of degrees awarded, etc.
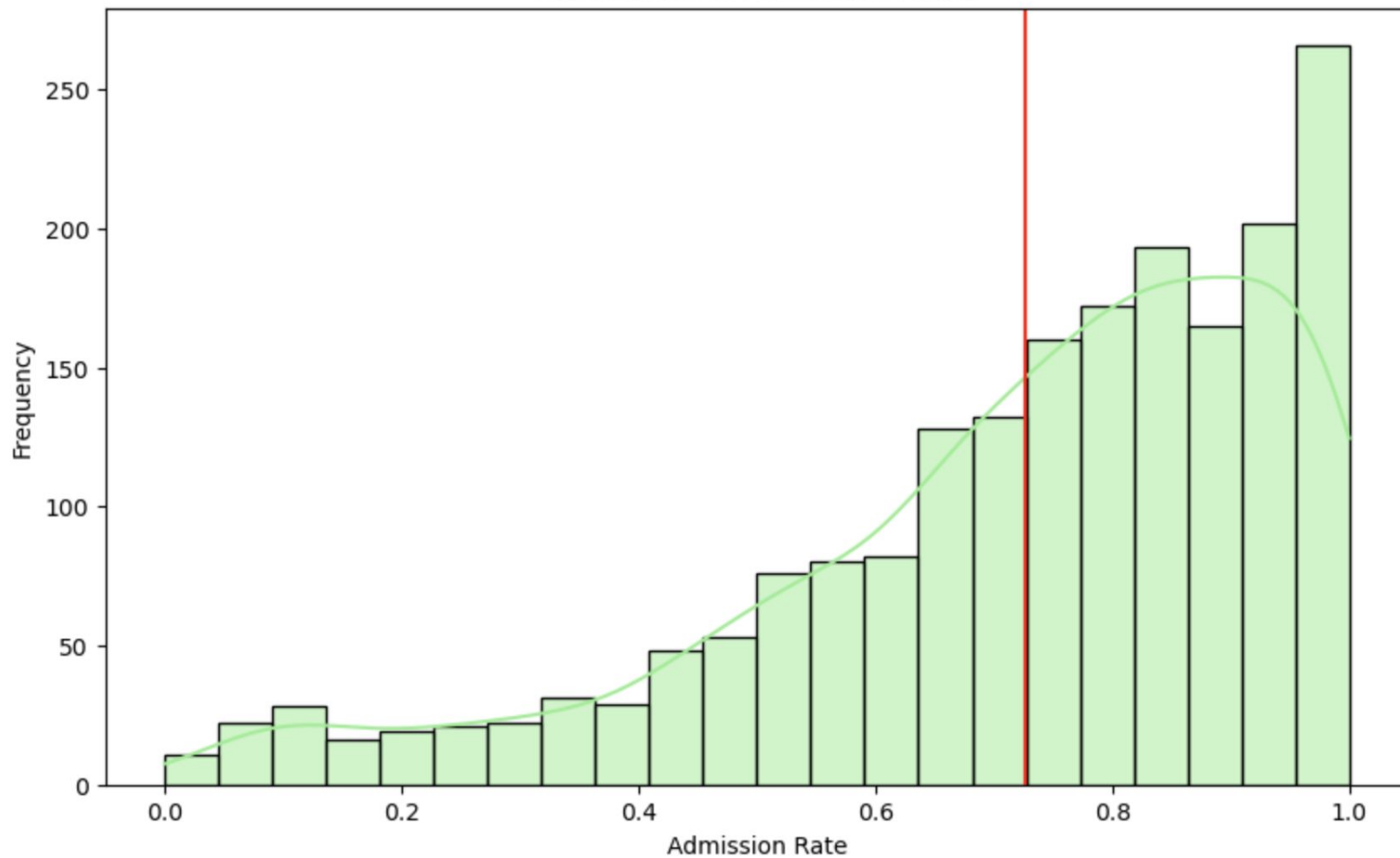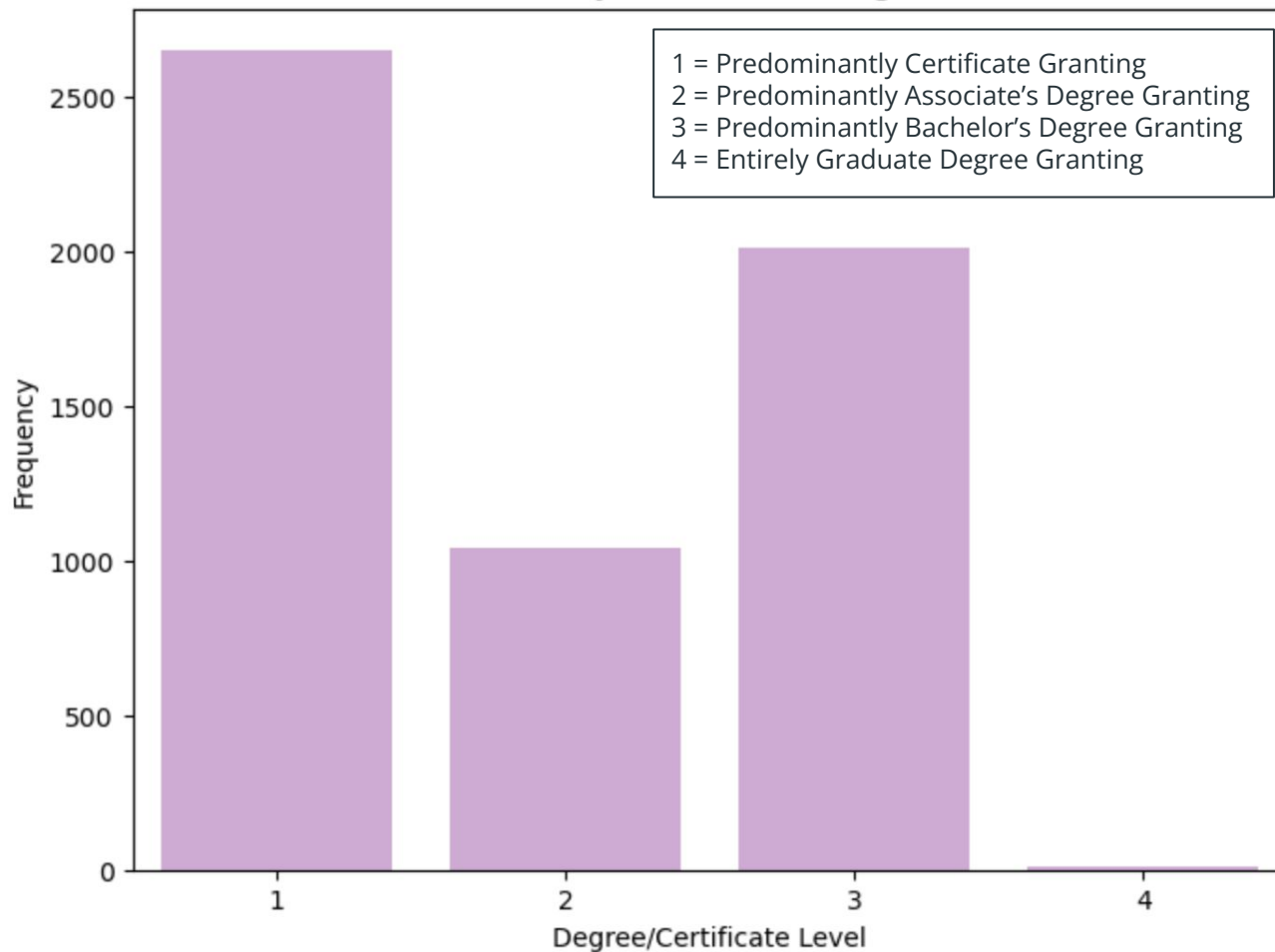
Initial Findings
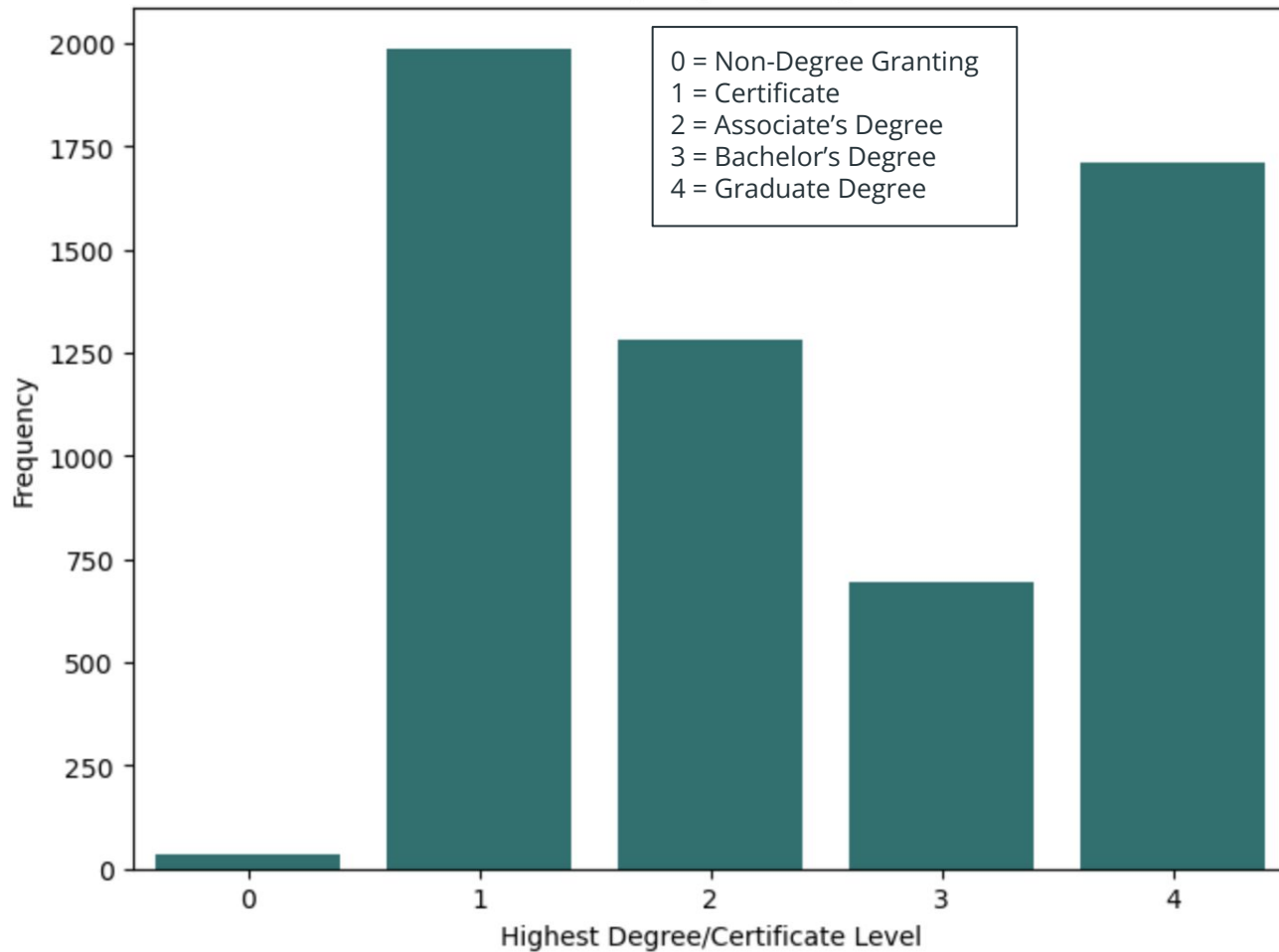
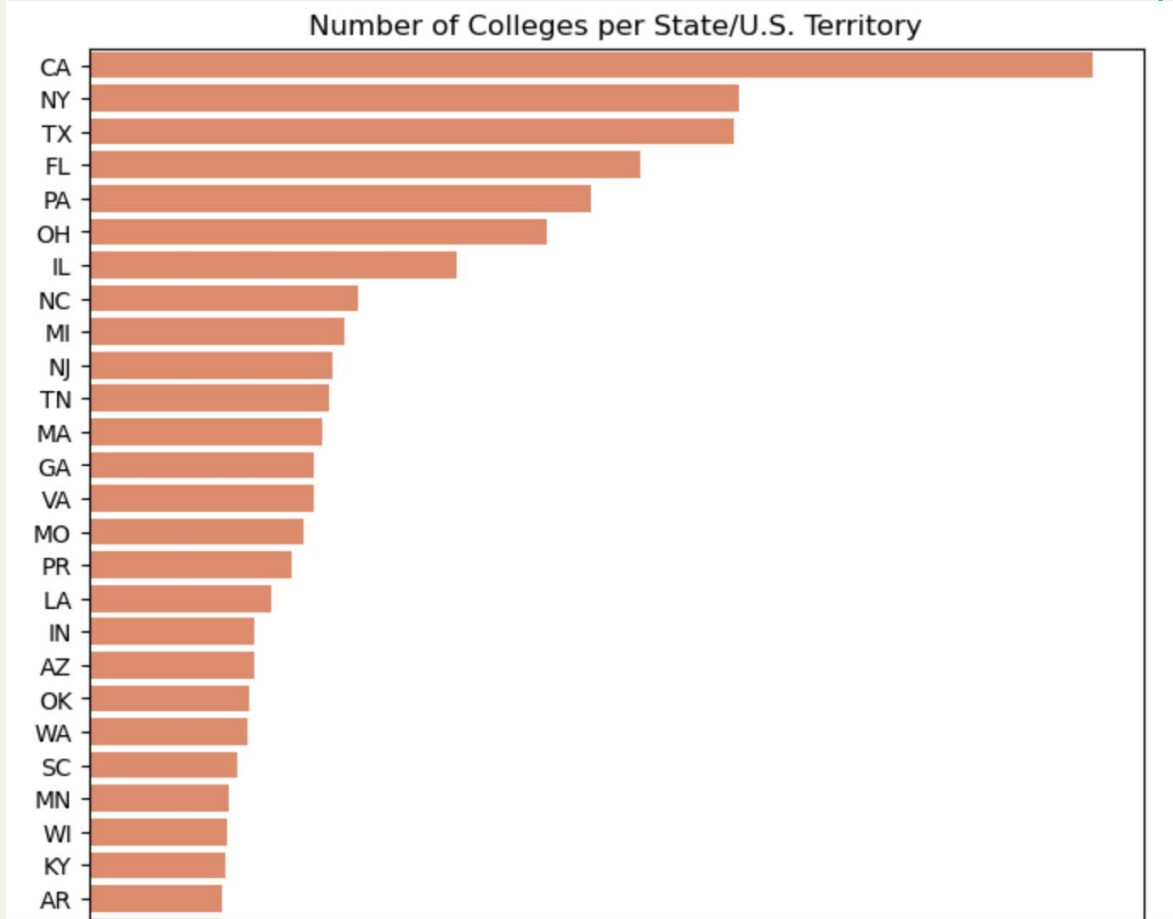Distribution of Average SAT Scores
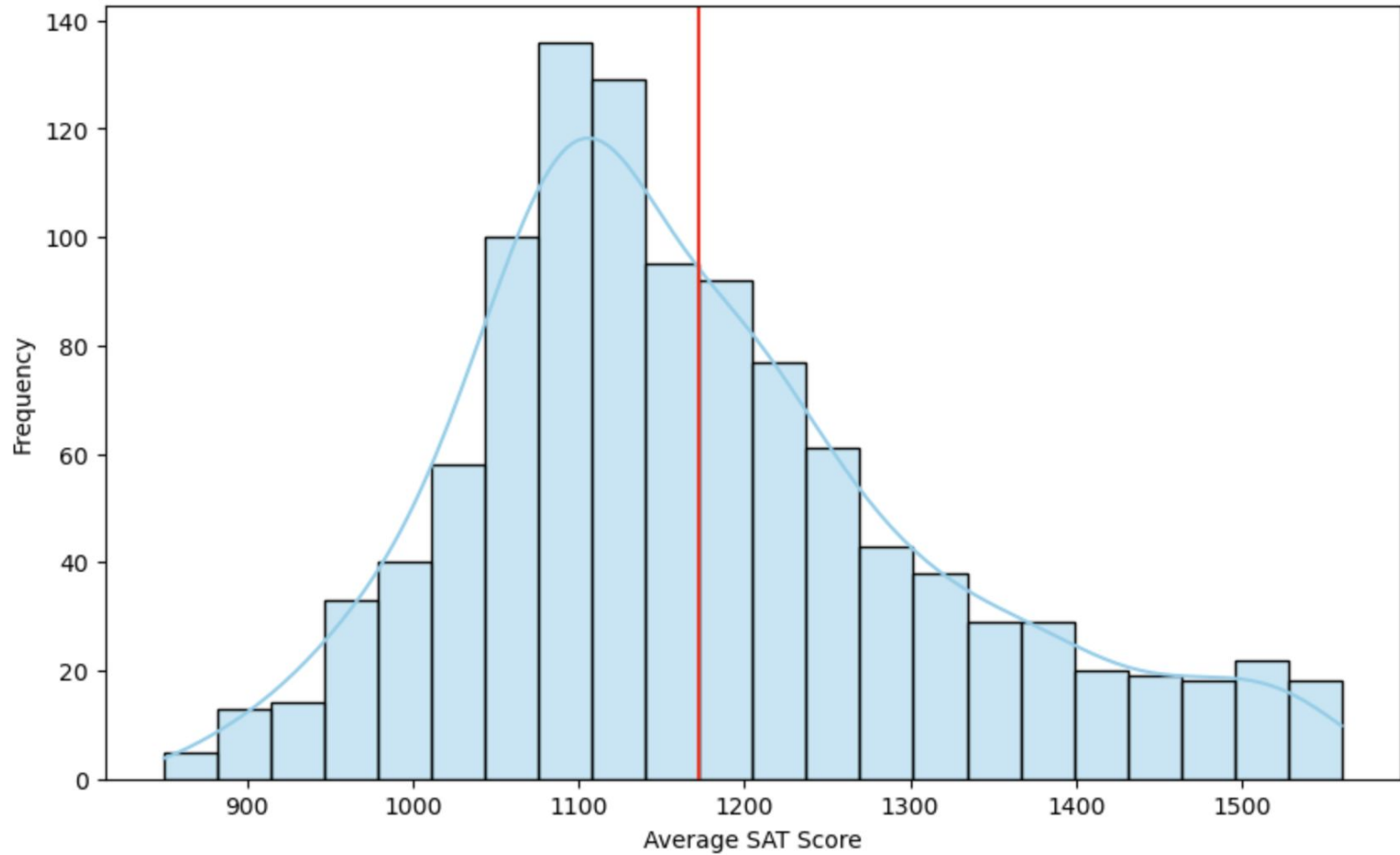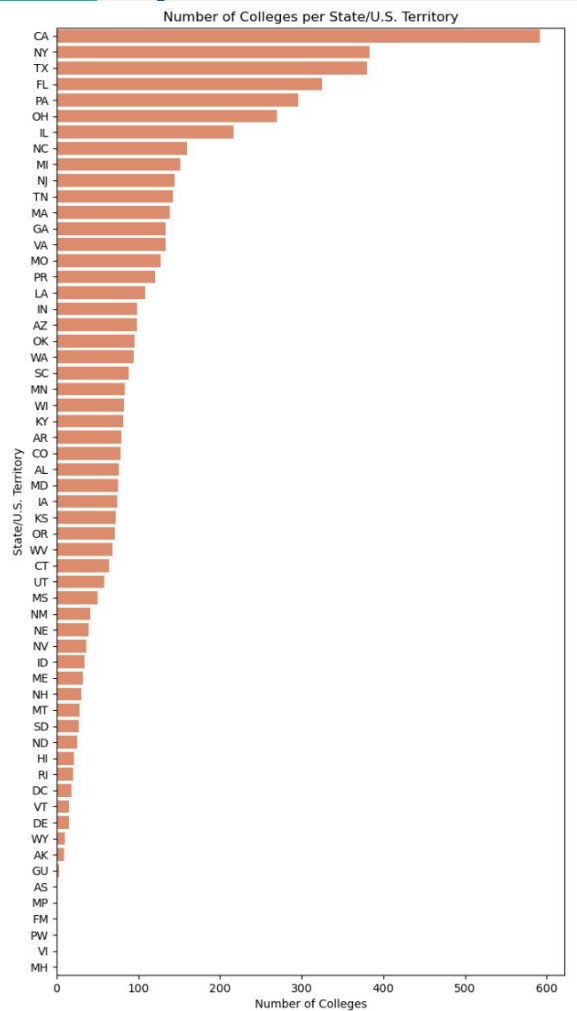
Distribution of Admission Rates

Distribution of Institutions by Predominant Degree/Certificate Awarded

1 = Predominantly Certificate Granting
2 = Predominantly Associate's Degree Granting
3 = Predominantly Bachelor's Degree Granting
4 = Entirely Graduate Degree Granting

Distribution of Institutions by Highest Degree/Certificate Awarded

0 = Non-Degree Granting
1 = Certificate
2 = Associate's Degree
3 = Bachelor's Degree
4 = Graduate Degree

Number of Colleges per State/U.S. Territory

Modeling and Evaluation

- Clustering was done using KMeans and DBSCAN
- KMeans Silhouette Score 0.41526884799239544
- DBSCAN Silhouette Score: 0.4057589009879435

Silhouette Score for Number of Clusters
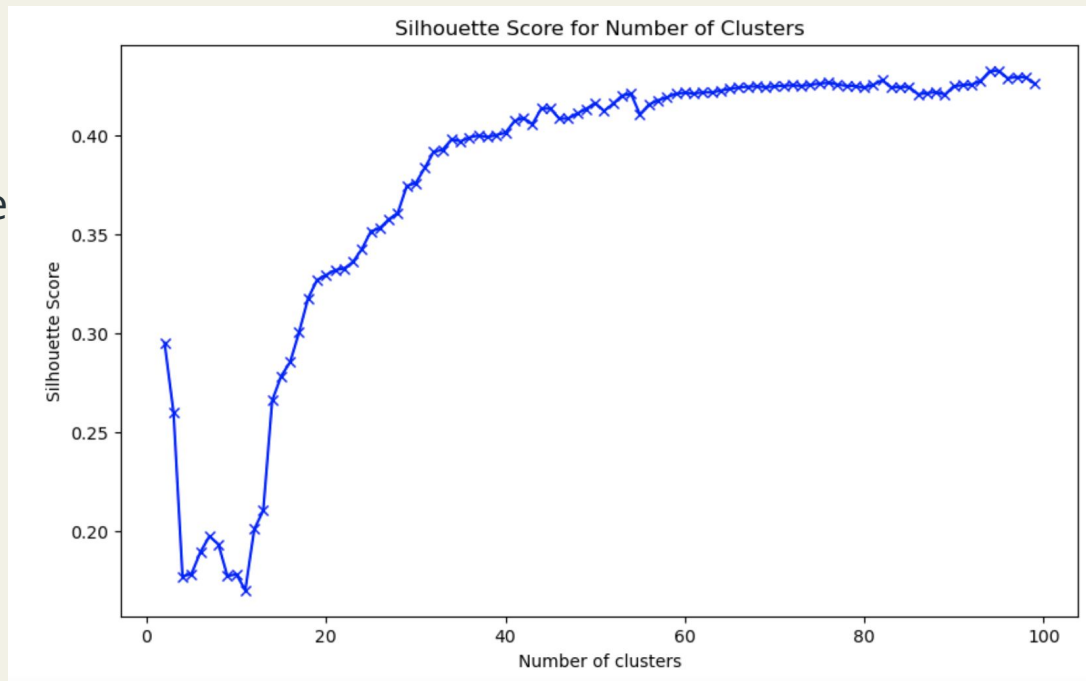
Conclusions and Recommendations

- Clustering worked well for some fields of study, but not all
- Silhouette Score is not very good even after using two methods of clustering
- College Recommender System was created for clustering and value sorting

- Recommend using value sorting for fields of study as clustering is not always the most accurate with this particular dataset

Limitations and Future Studies

Limitations

- Data was only taken from the 2022-2023 academic year

- Specific features, namely the percentage of degrees awarded in a particular field of study, were selected for clustering

- Current college recommender system primarily filters based on field of study, location, and predominant degree type offered
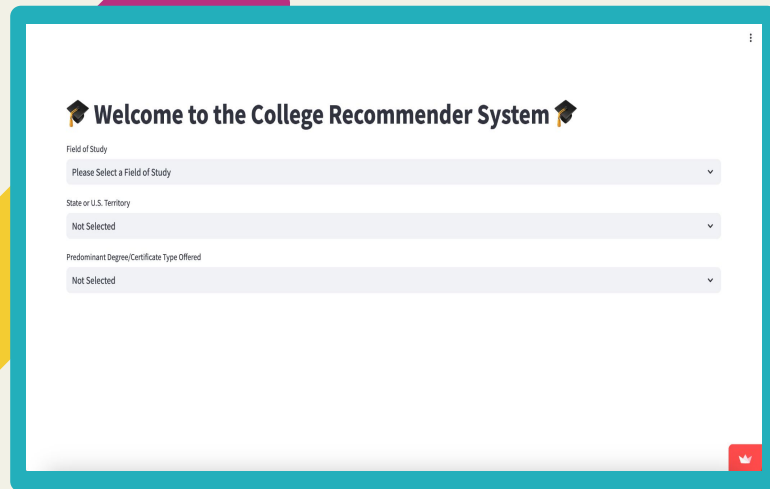
Future Studies

- Other datasets from the past decade or so should be used

- Other features in dataset could be used to improve clustering and silhouette score

- College recommender system can be augmented by additional filtering options such as cost, type of institution, or graduation rate

Which college should I go to?

Streamlit
Application

(Click link above)

Sources

- U.S. Department of Education:
  - College Scorecard data files: https://collegescoreca rd.ed.gov/data
- Data Dictionary:
  - data.yaml
- Datasets:
  - MERGED2022_23_PP.csv
  - data_cleaned.csv
- Streamlit.io

Thank You!

Client Testimonials

Peter W.

Thanks to Kevin, I was able to pick the best college for my chosen field of study!

Emma H.

Kevin's College Recommender System is very useful and helped my child pick colleges close to home!

Barry C.

Wow! The College Recommender System is super cool and allowed me to narrow down my college choices.