

What College is Best for Me?

Vidhyalakshimi Sreenivasan, Aman Chauhan,
Sal Camassa, Path Nagori, and Wyatt Maxey





Project Recap

- College search is often overwhelming and we have to maximize the student input and factor in preferences to narrow down college choices
- Display a finite collection of colleges for the stakeholders to choose from based on their selection criteria - student profile, financial state, and location



Problem Statement Level

<i>Questions for Problem Statement</i>	<i>Attributes</i>
Which universities can student afford?	Cost, Aid, Earnings
What is the student's profile?	Academics, School Environment
Is the student likely to be accepted?	Academics, Admissions
What is the school's profile?	Academics, Earnings, Completion, Cost, Admissions, Rankings, Location, Student Body



Which Universities can Student Afford?

<i>Questions for Problem Statement</i>	<i>Attributes</i>
Cost	Average cost of attendance (Living expenses), Tuition and Fees
Aid	Federal Loans, Federal Grants, Typical Monthly Loan Payment, Cohort Default Rate, Repayment Rate
Earnings	Average earnings, Median Earnings, Earnings of former students



What is the Student's Profile?

<i>Questions for Problem Statement</i>	<i>Attributes</i>
Academics	SAT Scores, ACT score, Degree desired
School Environment	Student Body size, Greek life, Diversity, Religious affiliation
Budget	Private/public, In-State/Out-State, Income bracket

What is the School Profile?



<i>Questions for Problem Statement</i>	<i>Attributes</i>
Academics	Programs offered
Earnings	Average earnings, median earnings
Cost	Tuition and fees, Average net price
Admissions	Admission rate, Midpoint ACT Score, Midpoint SAT Score
Rankings	Program rankings
Location	State, Main/Branch, Online-only
Student Body	Number of undergrad, Demographics, Part time/Full time, Percent over 25 years old



Question

WHICH COLLEGE IS
BEST FOR “ME”?



Methodology

- Gather university data
 - College Scorecard
 - Google Places
 - NCDC Climate Data Online for Temperature Stats
 - US Department of Education Campus Safety and Security
- Organize university data
 - Selecting subset of College ScoreCard Data
 - Joining temperature and crime rate with main dataset using zipcode.
- Gather User (student) Data through Survey
- Custom Weighted K-Nearest Neighbors
- Present results to user
 - Table
 - D3.js map
- Allow user to sort results via TOPSIS for best fit
 - User defined weights
 - Nightlife, crime rate, and weather



Data cleaning - University data



Selected a subset of CollegeScorecard Data.

Filtered the data based on Undergrad Colleges

Filled NaN values with context specific means.

Crime Data was merged.

Weather Data was merged.



Data cleaning - Crime rate

FBI Crime data Explorer:

Contains only State level crime rates.

US Department of Education Campus Safety and Security:

- Contains crime information for different crime categories - Murder, burglary, etc. for each zip code.
- For missing crime info for each zipcode, the average crime for the corresponding state is filled in.



Data cleaning - Weather



NCDC Climate Data Online provides API to collect Temperature data from their Global Summary of The Month dataset.

For each zipcode, average temperature for Spring, Summer, Fall and Winter was fetched from each Weather Station in the nearby region.

If the zipcode became too specific, the FIPS associated with that zipcode was used to fetch data from Weather Stations in the FIPS region.

If no data could be obtained even then, these missing values were filled with mean temperature values from the Weather station in that state.

These columns were then appended to the cleaned dataset.



Survey Construction



- Survey JS library used to construct user survey
- Survey markup maintained in a JSON file, output from survey in JSON format
- Questions in radio, dropdown, input and sliders
- Questions intuitively formed for more personal and less logical feel



Architecture

Dataset → Filtering Rows based on user input →

Filtering columns based on user input → Data Standardisation →

PCA → Apply KNN (weighted) → Push UNITID of Top 10 →

TOPSIS preference from user → Rank the College based on
TOPSIS.



Data Sources

<https://collegescorecard.ed.gov/data/documentation/>

<https://ope.ed.gov/campussafety/#/datafile/list>

<https://www.ncdc.noaa.gov/cdo-web/>

<https://developers.google.com/places/>



Demo

<https://bestcollege.herokuapp.com/bestcollege/index>

<https://tinyurl.com/ybatdrmj>