INFO 6105 Final Project

Analyzing University Data

University Recommendation Model

Tushar Vimalbhai Patel
002080292

Motivation

- Personal experience with lack of guidance during college applications
- Time-consuming research across multiple universities
- Need for a more efficient way to identify institutions matching preferences
- Desire to apply data science skills to real-world problem

Research Question

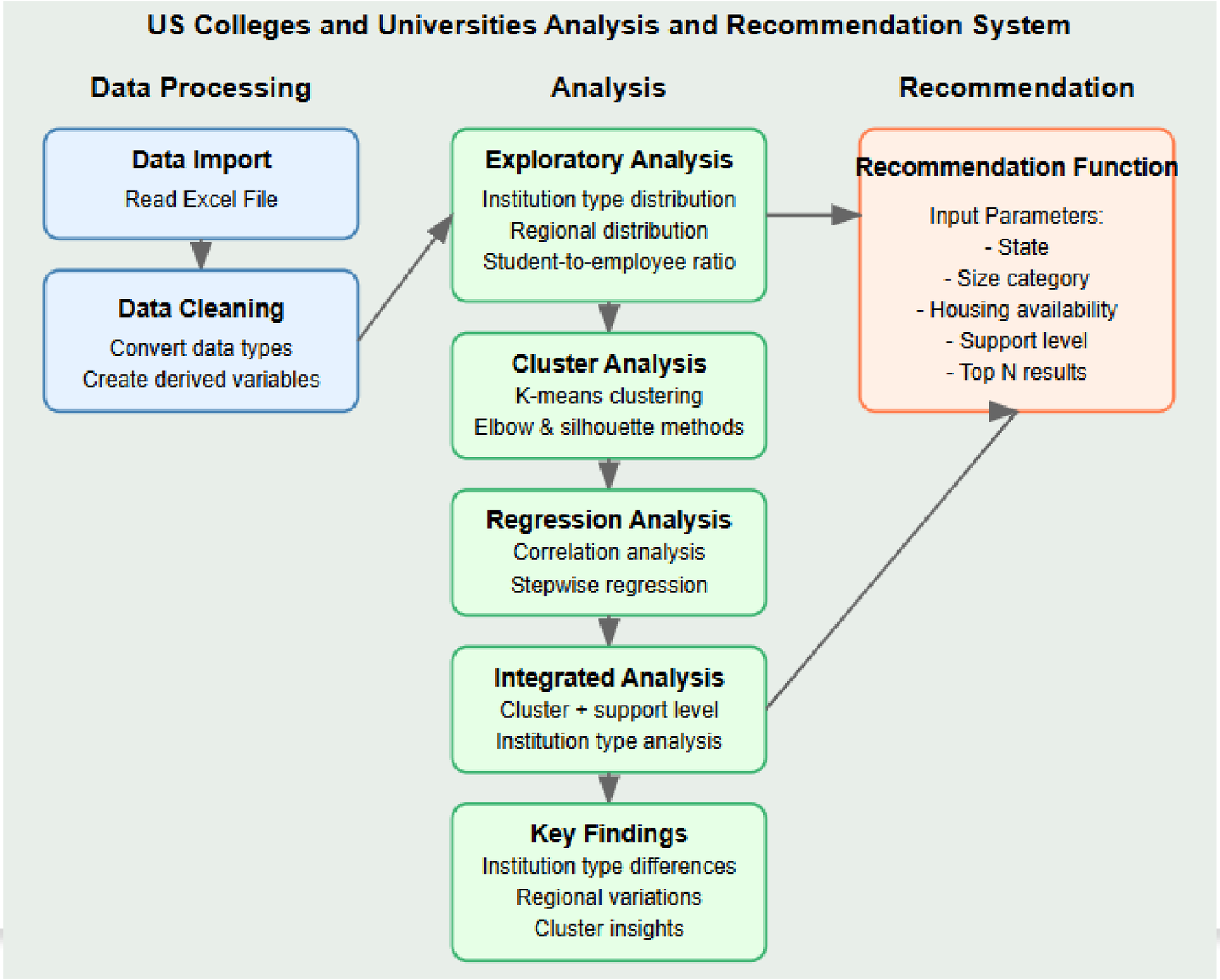
- How does the student-to-faculty ratio influence student enrollment size?
- How does geographic region affect enrollment size and cost of attendance?
- Can colleges be grouped into meaningful clusters based on shared characteristics like size, type, and location?

TOOLS AND LIBRARIES

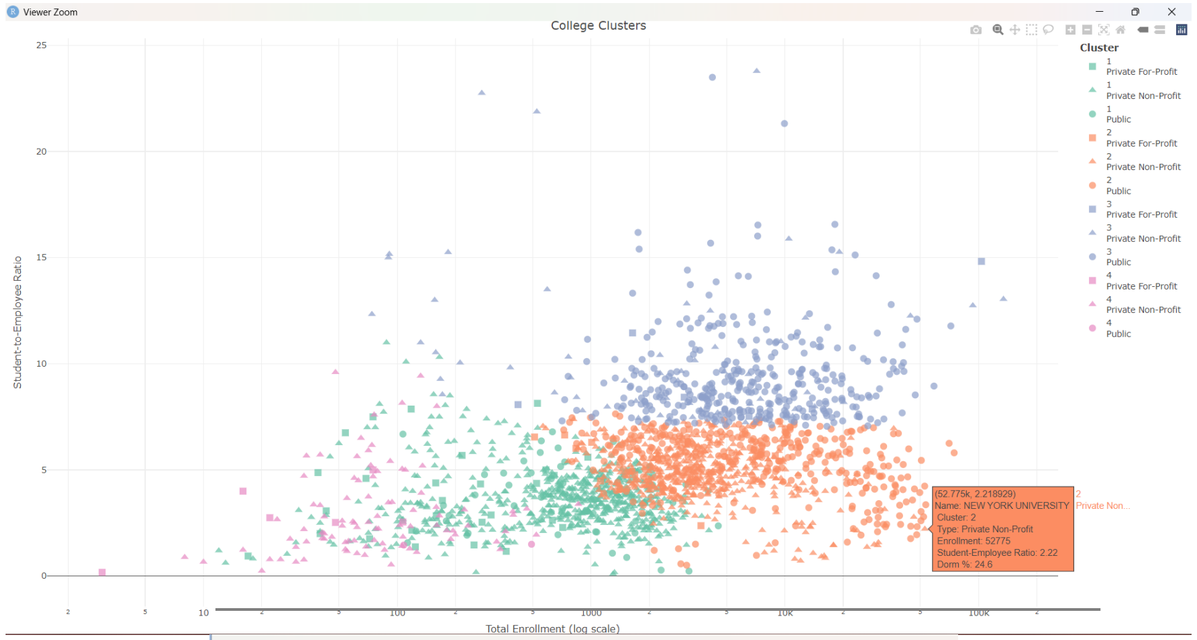
R Libraries Used:

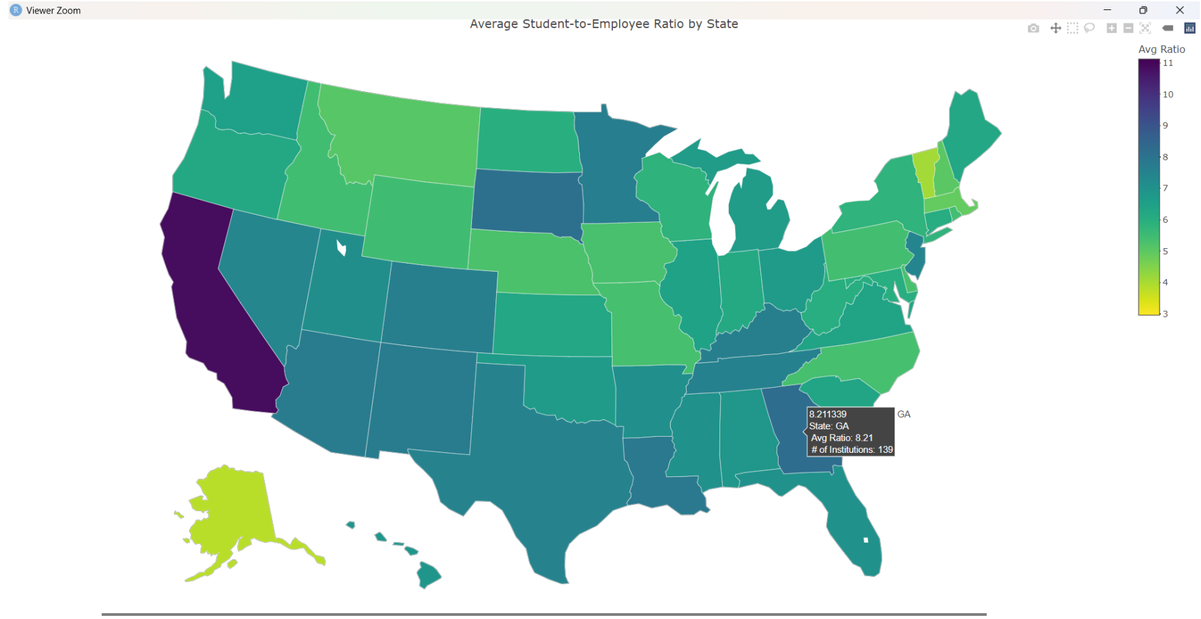
- **Data Manipulation:** tidyverse, readxl - for data import, transformation, and piping
- **Statical Modeling:** , car, leaps, caret - regression diagnostics, feature selection, cross-validation
- **Visualization:** plotly, viridis, ggplot2, corrplot - static + interactive + heatmaps
- Clusturing and Analysis: Cluster, factoextra - for K-means, silhouette, and cluster visuals
- **Geographic Analysis:** maps - to show enrollment/ratios across US states
- **Interactive Application:** shiny, DT - to create an interactive recommendation system UI

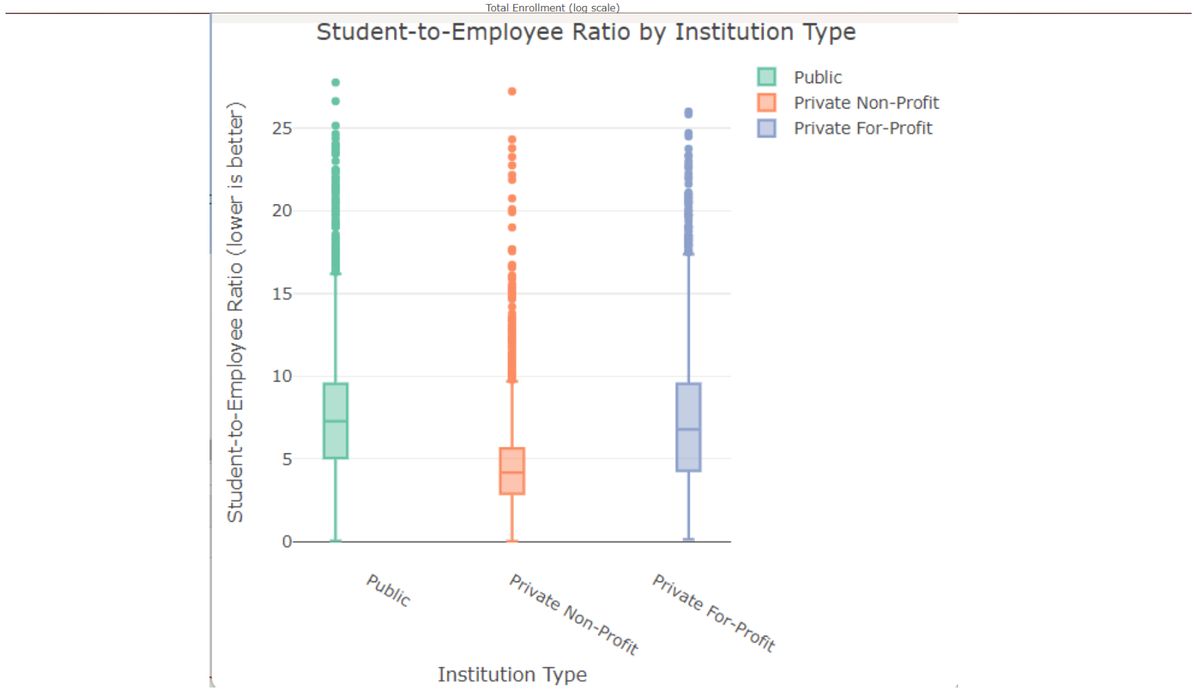
Models and Analysis Used

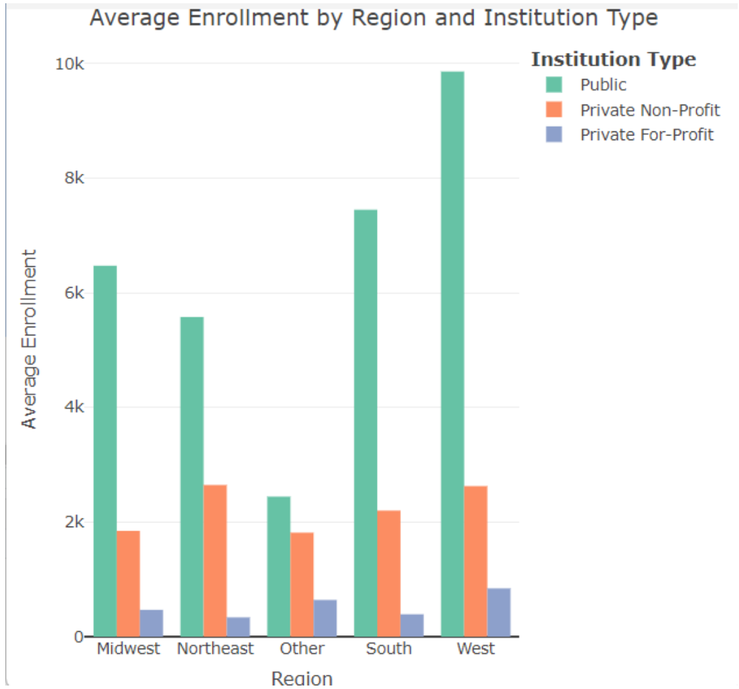


Exploratory Data Analysis And Cluster Analysis

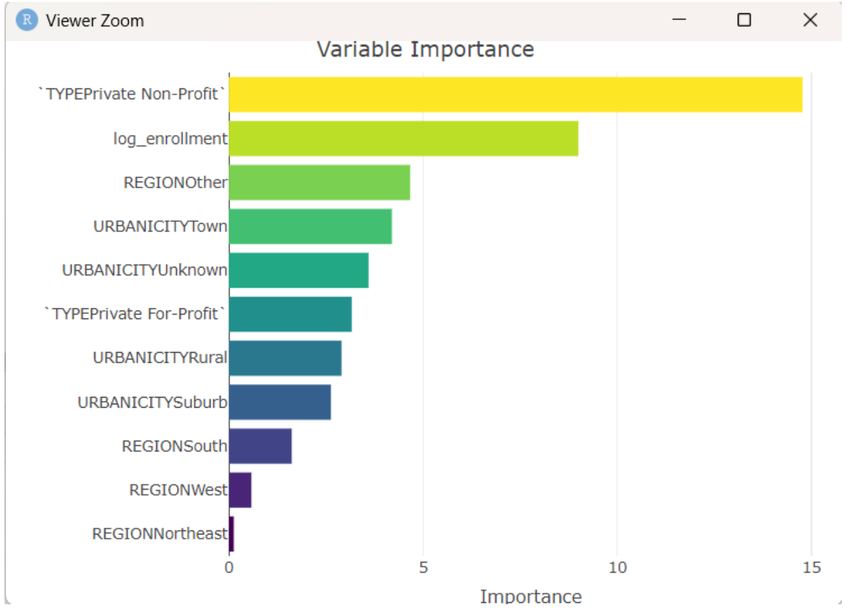


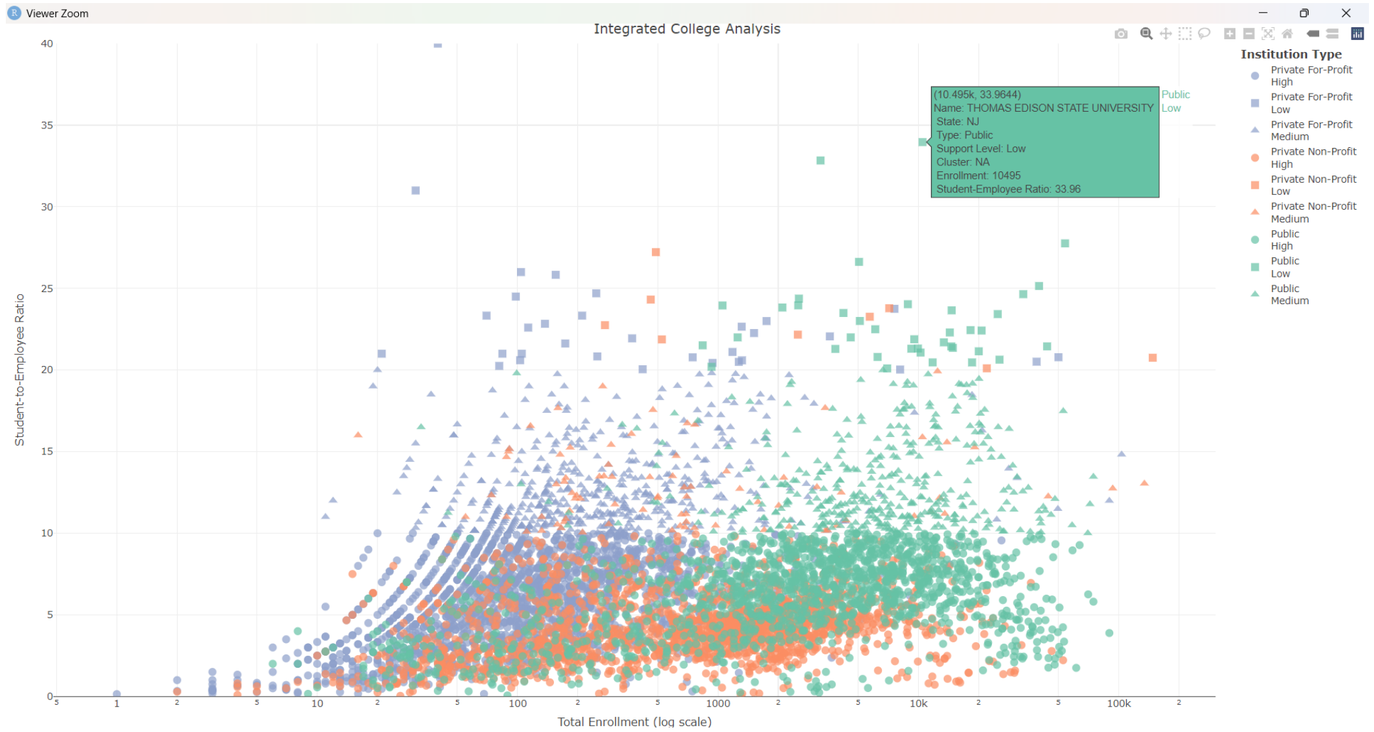


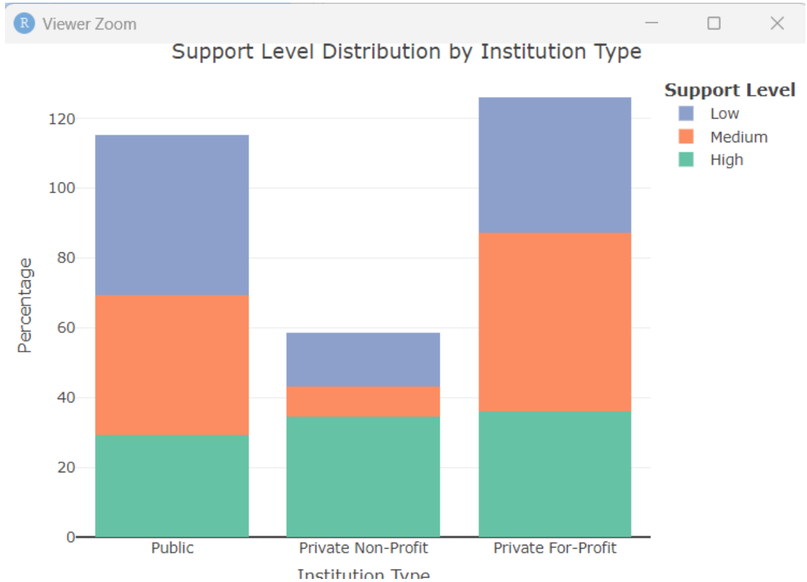




Regression Analysis and Integrated Analysis







University Recommendation System

Location (State):

CA

Enrollment Size:

- [x] Small (< 5,000)
- [ ] Medium (5,000-15,000)
- [ ] Large (> 15,000)

Housing Available:

- (•) Either
- ( ) Yes
- ( ) No

Support Level:

- (•) Any
- ( ) High
- ( ) Medium
- ( ) Low

Region:

- [ ] Northeast
- [ ] Midwest
- [ ] South
- [x] West

Institution Type:

- [x] Public
- [ ] Private Non-Profit
- [ ] Private For-Profit
- [ ] Other

Number of Recommendations:

10

Find Universities

Results | Statistics

Recommended Universities

Show 10 entries

Search:

NAME	STATE	REGION	TYPE	TOT_ENROLL	SIZE_CATEGORY	HAS_HOUSING	STUDENT_EMP_RATIO	SUPPORT_LEVEL	cluster
NAPA VALLEY COLLEGE	CA	West	Public	4931	Small (< 5,000)	No	9.4	High (<=10)	
COMPTON COLLEGE	CA	West	Public	4612	Small (< 5,000)	No	10.4	Medium (>10)	
WOODLAND COMMUNITY COLLEGE	CA	West	Public	4598	Small (< 5,000)	No	22.0	Low (>20)	
COLLEGE OF MARIN	CA	West	Public	4509	Small (< 5,000)	No	9.3	High (<=10)	
GAVILAN COLLEGE	CA	West	Public	4494	Small (< 5,000)	No	11.3	Medium (>10)	
WEST HILLS COLLEGE-COALINGA	CA	West	Public	4229	Small (< 5,000)	Yes	23.5	Low (>20)	3
PORTERVILLE COLLEGE	CA	West	Public	3964	Small (< 5,000)	No	14.7	Medium (>10)	
WEST HILLS COLLEGE-LEMOORE	CA	West	Public	3932	Small (< 5,000)	No	16.0	Medium (>10)	
COLLEGE OF THE REDWOODS	CA	West	Public	3891	Small (< 5,000)	Yes	8.8	High (<=10)	3
PALO VERDE COLLEGE	CA	West	Public	3854	Small (< 5,000)	No	21.3	Low (>20)	

Showing 1 to 10 of 10 entries

Previous 1 Next

/08

Key Findings

- Institution type is a strong predictor of student-employee ratio and housing availability
- Four distinct institutional clusters emerged from the analysis
- Geographic location significantly impacts institutional characteristics
- Student support levels vary substantially by institution type and size

Conclusion

This project demonstrates how data analysis can transform the college selection process, making it more efficient and personalized for prospective students, addressing the challenges I faced during my own application experience.

Thank You