Data Driven Results on Accepting Jenny Jacob

Harvard Admissions

MS Data Science

By Jenny Jacob

Objective

The primary objective of this project is to analyze the key factors that influence

admissions to the graduate MS in Data Science, particularly focusing on the reasons for my

potential admission to Harvard's Master of Science in Data Science program. This analysis will

help me understand the trends and characteristics that contribute to successful admissions

decisions. This data is completely simulated based on the stats found online of previously

admitted students. This project serves as a compelling way to demonstrate my passion for the

program and to showcase my knowledge and capabilities. (All coding and files can be found in

my [Github repo Data-Science](#) , link is also attached to my application under links).

Specific Goals:

- Identify Key Factors: Determine the key factors that are statistically significant in
  influencing graduate admissions.

- Highlight Personal Strength: Analyze my academic performance, extracurricular
  activities, and projects to illustrate why I should be admitted.

- Visualize Trends: Create visual representations of the data to effectively communicate
  findings.

- Predictive Modeling: Develop a predictive model to estimate admission probabilities
  based on selected features.

Questions to Answer:

What are the typical academic metrics (GPA) of admitted students?

How do extracurricular activities impact admission decisions?

In what ways do my achievements align with successful admission factors?

Data Analysis

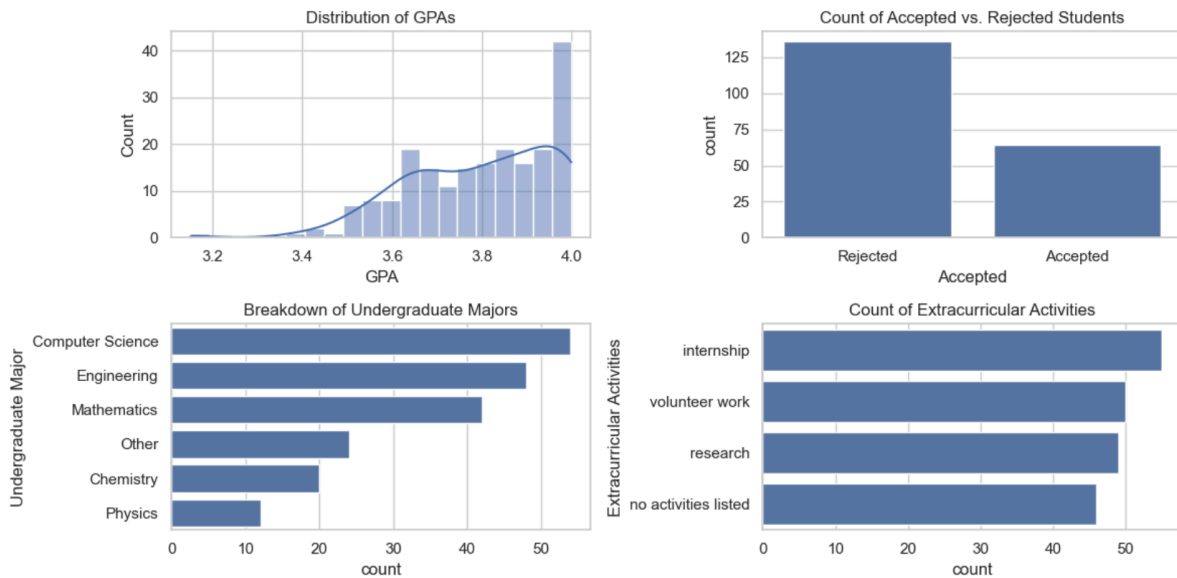## Visualizations and Findings

Figure 1

Figure 1 presents the distribution of GPAs among applicants, indicating a concentration around higher values. The accompanying bar chart shows the count of accepted versus rejected students, revealing a low acceptance rate that emphasizes the competitive nature of the admissions process. Both Undergraduate Major and Extracurricular activities display the leading acceptance in major and extracurricular activity.

- Key Finding: The average GPA for applying is high and the acceptance rate of university is low compared to their rejections. Technical backgrounds are preferred with some experience through an internship or volunteer work.
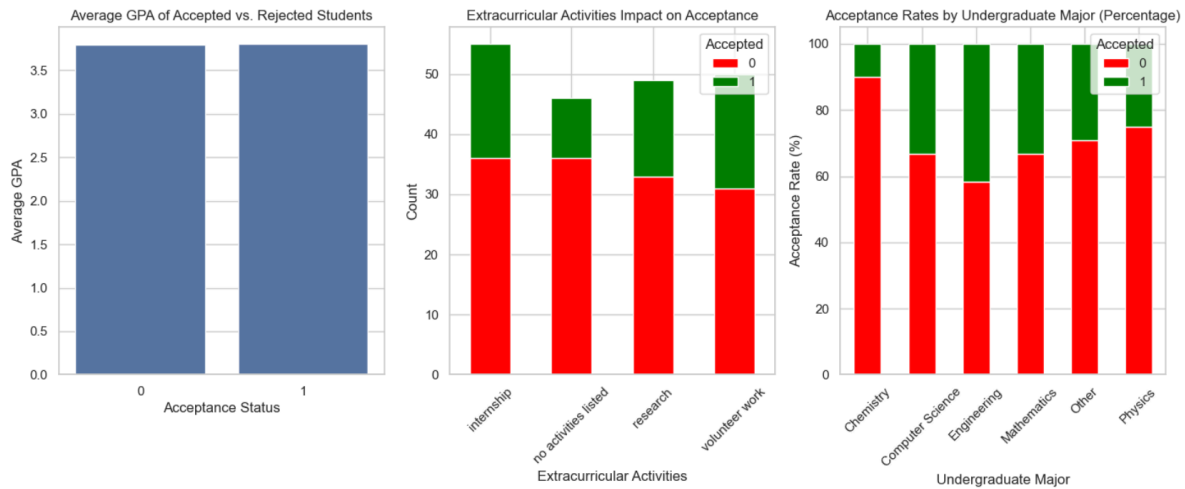
Figure 2

In figure 2 the bar chart shows acceptance rates by undergraduate major, extracurricular activities and average GPA. Figure 2 reveals that certain majors like Computer Science and Engineering exhibit higher acceptance rates, aligning with the interdisciplinary nature of the data science program. Having Internships and volunteer work show a higher chance of acceptance and GPA of accepted and rejected stay high, leading that GPA is important but not a deciding factor.

- Key Finding: The average GPA of accepted and rejected students is high, indicating the importance of strong academic performance in the admissions process however since both rejected and accepted are high, other factors contribute more than just a GPA. Technical Majors with relevance to the curriculum is normally accepted more than other majors. Internships/volunteer work lead the extracurriculars for being a good factor in acceptance.
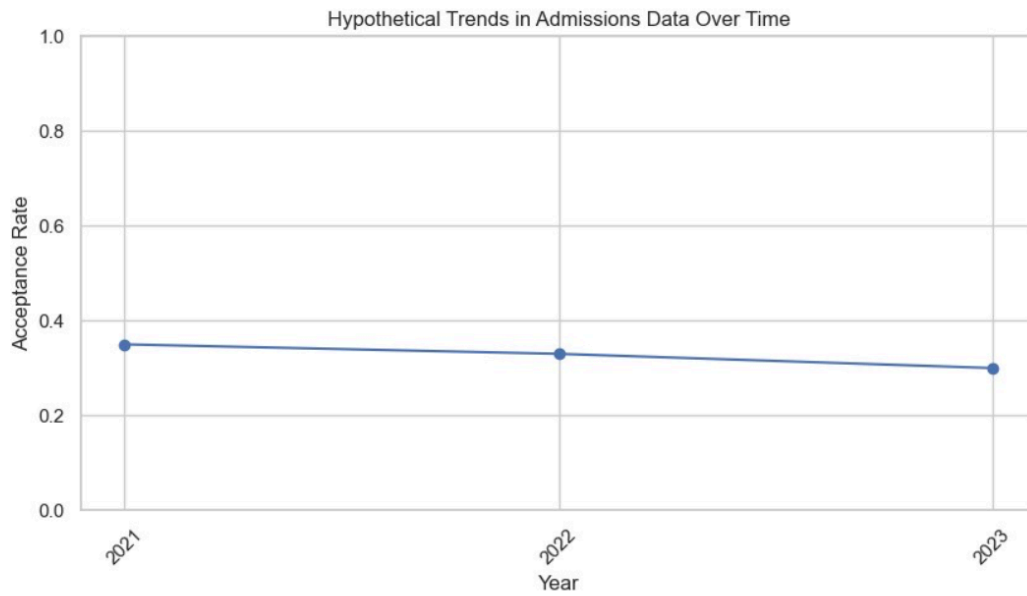
Figure 3

Figure 3 illustrates the hypothetical acceptance rates from 2021 to 2023, showing a decline from 35% to 30%.

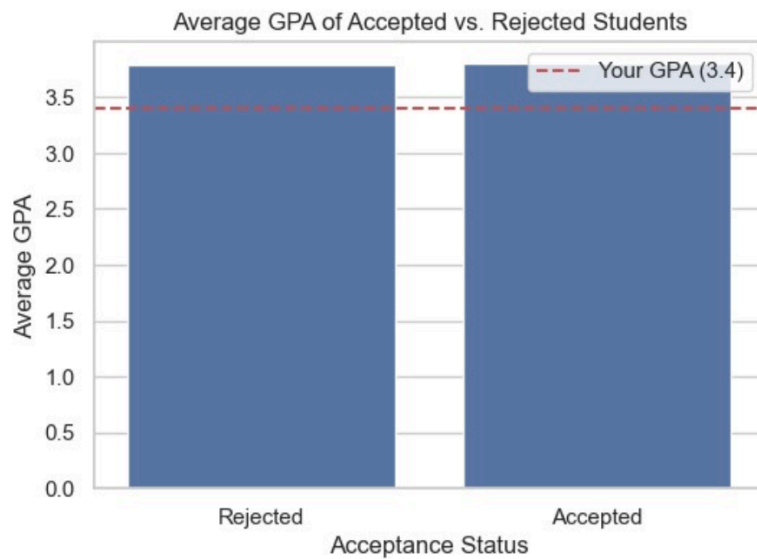- Key Finding: Highlights the increasing competitiveness of the admissions process over the years.

Figure 4

Figure 4 shows the average GPA of accepted and rejected students is similar, reinforcing the critical role of academic performance in the admissions criteria. My GPA of 3.4 is indicated as below the average for accepted students, suggesting an area for improvement.
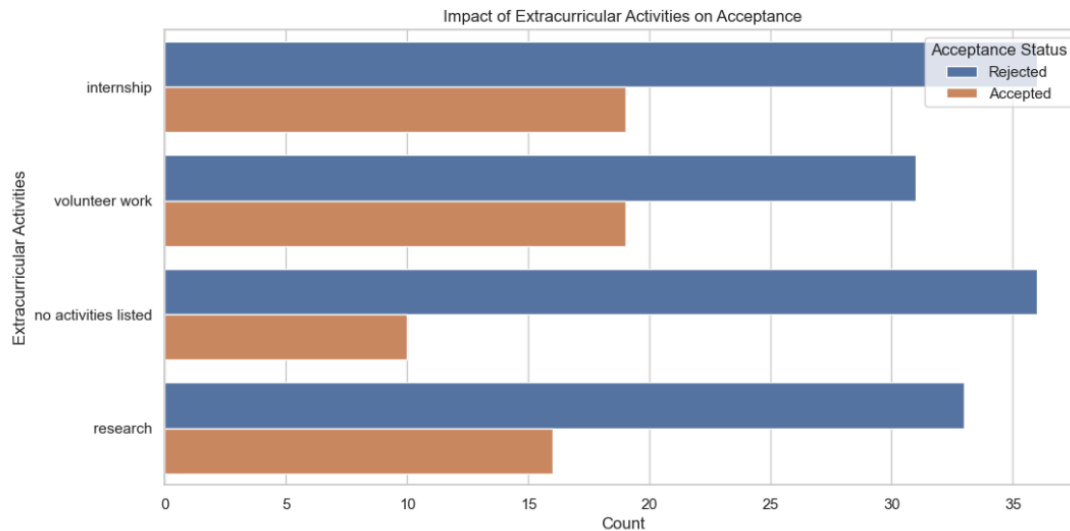
- Key Finding: My GPA should be higher.

Figure 5

In figure 5 the count plot illustrates the relationship between extracurricular activities and acceptance status. Students engaged in internships, volunteer work, and research show higher acceptance rates compared to those with no activities listed.

- Key Finding: Active involvement in extracurricular activities significantly influences acceptance rates, supporting the notion that well-rounded candidates are favored.

Predictive Modeling

Predictive modeling is a statistical technique used to forecast outcomes based on historical data. In this context of graduate admissions, predictive modeling will help me identify the factors that most significantly influence whether an applicant is accepted or rejected.
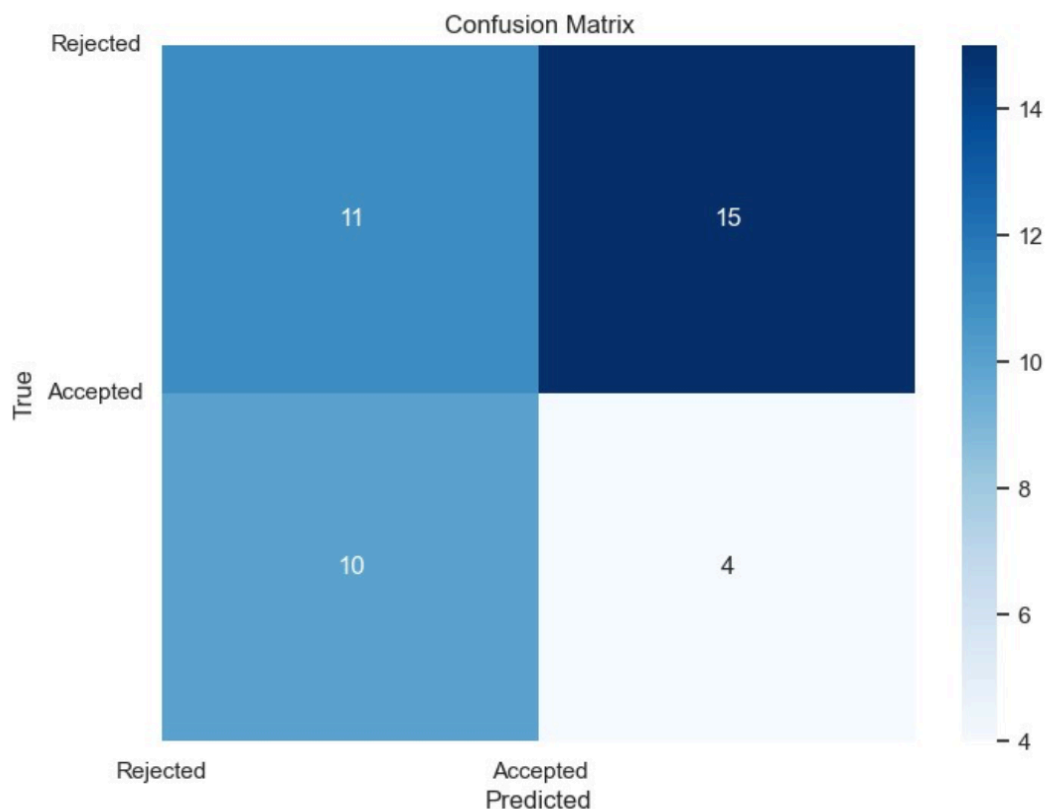

Figure 6   *Accuracy: 0.375 (37.5%)*

Initial Model Performance: Logistic regression is used for binary classification, predicting the probability that a given input belongs to a particular category (in this case, accepted or rejected). The logistic regression model initially yielded an accuracy of 37.5% (0.375), suggesting limited predictive power.

Figure 6 displays the confusion matrix for the initial logistic regression model, indicating areas of misclassification and confirming the need for improvement.
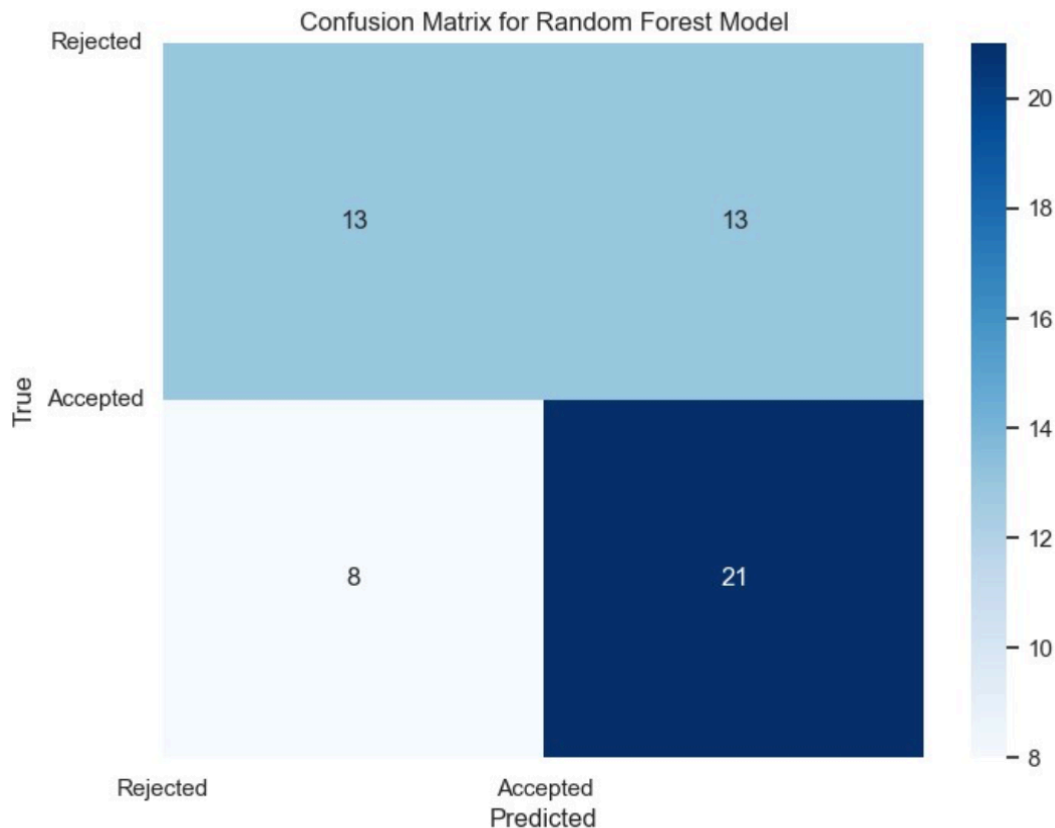
Confusion Matrix for Random Forest Model

Figure 7   *Accuracy 0.618 (61.8%)*

Improved Model Performance: Switched to a Random Forest Model which constructs multiple decision trees and outputs the mode of their predictions for classification tasks so I could get a better reading on what goes into admissions. After implementing a Random Forest model, accuracy increased to 61.8% (0.618), reflecting a better understanding of the admissions dynamics through addressing class imbalance and hyperparameter tuning. The transition from logistic regression to Random Forest not only improved predictive accuracy but also highlighted the complex interplay of various applicant attributes.

In figure 7 the confusion matrix for the Random Forest model demonstrates a significant enhancement in performance metrics, with correct predictions increasing notably.

Conclusion

The analysis conducted provides valuable insights into the critical factors influencing admissions decisions for Harvard's Master of Science in Data Science program. The significance of maintaining a strong GPA is clearly illustrated in Figure 1, while the importance of extracurricular involvement is further emphasized in Figure 2. However, the predictive modeling results highlight that an applicant's profile is not solely defined by GPA and extracurriculars; the choice of major also plays a role in the admissions process.

My background in a technical field, coupled with substantial extracurricular engagement, positions me favorably in comparison to previous accepted applicants as indicated by the simulated data. This alignment with the program's values demonstrates that I exceed the expectations for well-rounded candidates.

Despite the challenges presented by my GPA, my extensive involvement in relevant extracurricular activities and my strong academic foundation through my major provide compelling evidence of my potential. The insights received from this analysis, along with my personal strengths, suggest that I am well-prepared to thrive in this rigorous academic environment.

This report not only underscores my dedication to the field of data science but also reflects my commitment to continuous personal and academic improvement. I eagerly anticipate the opportunity to further showcase my capabilities as a graduate student at Harvard.