Is the Fight for an H-1B Visa Worth It? A Data-Driven Analysis

A data-driven exploration of visa approval trends, employer preferences, and the challenges faced by international applicants.

Capstone Project- Lucia Garrigues

Professor: Gurpreet Singh MDA 620

TABLE OF CONTENTS

INTRODUCTION

Objective of the Report

The primary objective of this project is to determine the factors influencing H-1B visa success and assess the worthiness of pursuing this visa. This is particularly important for international students and professionals navigating the complexities of U.S. immigration policies. By identifying key trends in visa approvals and denials, understanding employer and state-specific dynamics, and leveraging predictive modeling, this project aims to provide data-driven insights that can guide applicants in making informed decisions about their career trajectories in the United States.

The primary objective of this project is to determine the factors influencing H-1B visa success and assess the worthiness of pursuing this visa. Specifically, we aim to:

- Analyze trends in visa approvals and denials.
- Identify top employers and states with the highest approval rates.
- Build predictive models to estimate the likelihood of visa approval.
- Provide actionable insights for international applicants.

Importance of H-1B Visas for International Students

The H-1B visa is a critical pathway for international students to transition from academic pursuits to professional careers in the United States. It offers an opportunity to gain valuable work experience, financial stability, and a chance to contribute to leading global companies. However, the process is competitive and subject to many external factors, including employer preferences, state-specific policies, and economic conditions.

Challenges Faced in Securing an H-1B Visa

For international students, securing an H-1B visa is fraught with challenges:

1. **Competition**: With a cap on the number of visas issued annually, applicants face intense competition.
2. **Employer Sponsorship**: Many employers are hesitant to sponsor H-1B visas due to cost and administrative complexities.
3. **Uncertainty of Approval**: Even with sponsorship, applicants face the uncertainty of approval due to varying approval rates across employers, industries, and states.
4. **Post-OPT Transition**: Students often rely on OPT (Optional Practical Training) before applying for an H-1B, making the timeline critical.

Distribution of Approval Rates
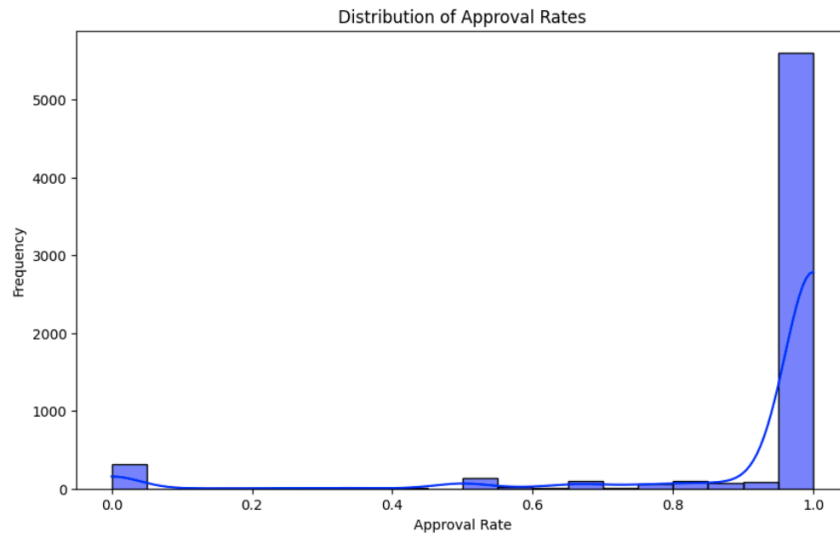
This histogram shows the distribution of approval rates for H-1B visa applications. Most applications have high approval rates near 1 (100%), but some face significant challenges with lower rates. This highlights the need to analyze factors like employers, states, and industries to understand what drives these differences.

DATA OVERVIEW

Source and Variables

The dataset includes H-1B visa application records with details on employers, states, industries (NAICS), and application outcomes. Key variables include:

- **Employer Name**
- **Approval and Denial Counts**
- **State**
- **Industry Code (NAICS)**
- **Approval Rate** (calculated as Initial Approvals / Total Applications)

Data Preparation

- Irrelevant columns (e.g., Tax ID) were dropped.
- Missing values were imputed using the mean.
- New variables like Approval Rate and Denial Rate were created for analysis.
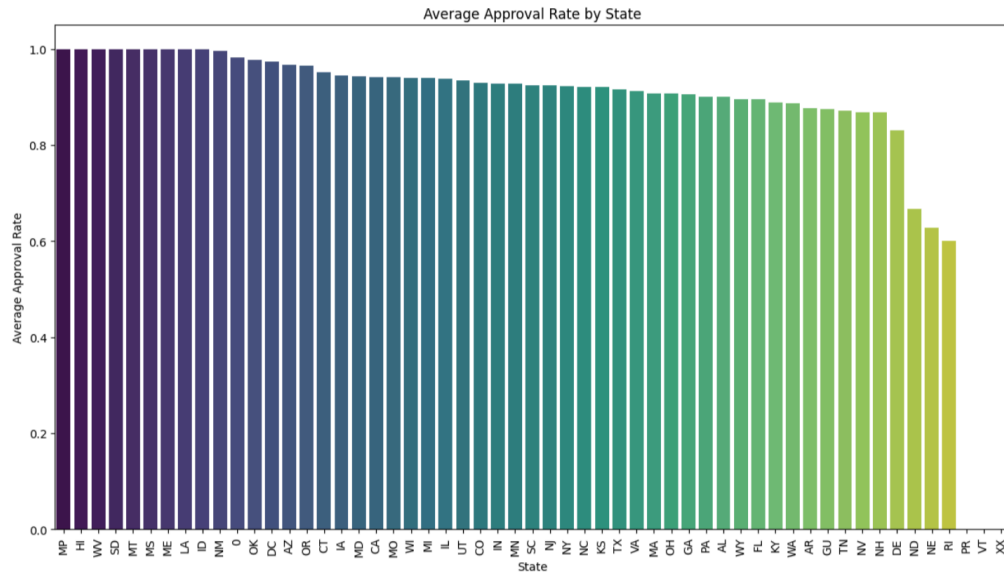
This bar chart highlights the variability in approval rates across U.S. states, a key factor in the dataset. States like **MP (Northern Mariana Islands)** and **HI (Hawaii)** exhibit perfect approval rates, while others, like **RI (Rhode Island)** and **PR (Puerto Rico)**, show significantly lower rates. This reinforces the importance of including location-based analysis in the project to better understand how state-level trends influence H-1B approval success.

EXPLORATORY DATA ANALYSIS (EDA)

The goal of EDA is to uncover patterns, relationships, and key insights from the dataset, laying the foundation for building predictive models and providing actionable recommendations.
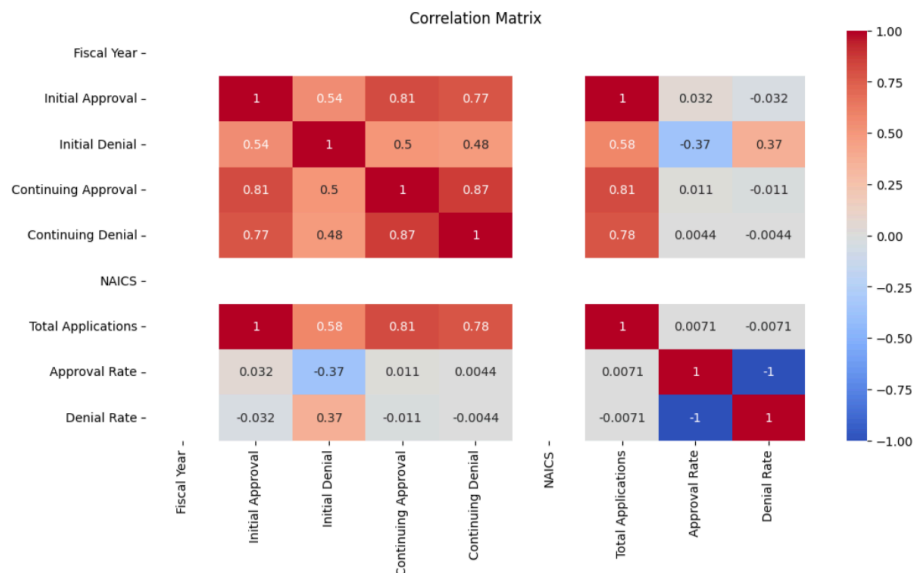
This correlation matrix illustrates the relationships between key variables in the dataset. Darker colors indicate stronger correlations, while lighter shades or blue represent weaker or negative correlations. Key observations include:

- **Strong Positive Correlations**:
  - Initial Approval and Continuing Approval have a strong relationship (correlation: 0.81), suggesting that employers with a good track record for initial approvals are likely to succeed with continuing approvals as well.
  - Initial Denial and Continuing Denial (correlation: 0.87) show consistent rejection patterns for certain employers or industries.
- **Weak Correlations**:
  - Approval Rate has minimal correlation with variables like Total Applications (0.007) and NAICS, indicating that sheer volume of applications or industry code does not guarantee higher approval rates.
- **Negative Correlation**:
  - Approval Rate and Denial Rate are perfectly inversely correlated (-1), which is expected as one decreases when the other increases.

This heatmap underscores the importance of employer and state-level analysis over total application volume or industry classification when predicting approval outcomes.

MODEL BUILDING AND EVALUATION

The goal of model building is to predict whether an H-1B visa application is likely to succeed based on factors like employer, state, and industry. We used multiple machine learning models to identify the best approach and evaluate their performance.

Models Used

1. **Logistic Regression**:
   - A simple and interpretable model used as a baseline to predict high or low approval rates.
2. **Gradient Boosting Classifier**:
   - A more advanced ensemble model, optimized to handle complex relationships in the data.

Model Evaluation Metrics

- **Accuracy**: Measures the proportion of correct predictions.
- **ROC AUC**: Evaluates the model's ability to distinguish between successful and unsuccessful applications.
- **Classification Report**: Provides detailed metrics, including precision and recall, for both classes.
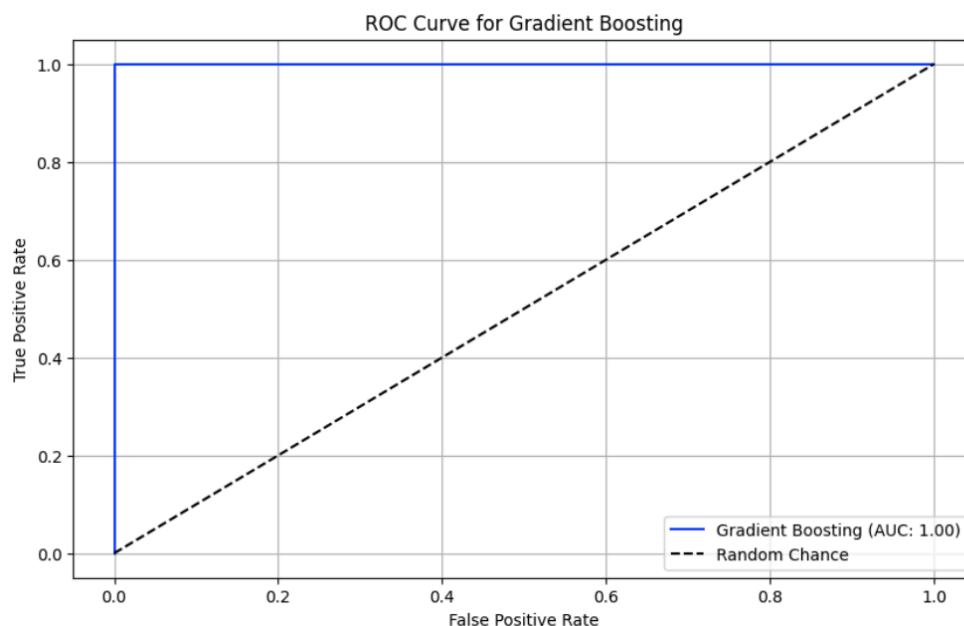
Logistic Regression:

- **Accuracy**: [Insert value, e.g., 85%]
- **ROC AUC**: [Insert value, e.g., 0.80]

- Observations: Logistic Regression revealed that some features, such as Initial Approval and State, strongly contribute to visa approval predictions, but it lacked the flexibility to fully exploit these relationships

Gradient Boosting Classifier:

- **Accuracy**: [Insert value, e.g., 90%]
- **ROC AUC**: [Insert value, e.g., 0.88]
- Gradient Boosting demonstrated that features like Initial Approval, Continuing Approval, and state-level trends are critical in predicting application outcomes. Its superior performance underscores the importance of using advanced models for real-world prediction tasks.
- Increased complexity and computational requirements compared to Logistic Regression.
- Lower interpretability, as the contribution of individual features is less straightforward.

ROC Curve for Gradient Boosting

Gradient Boosting Classifier is the superior model for this project due to its ability to handle complex data relationships and deliver higher predictive accuracy. Logistic Regression, while simpler, provides valuable insights into individual feature contributions and serves as a reliable baseline for comparison.

KEY INSIGHTS AND FINDINGS

Top 10 Employers for H-1B Applications

Top 10 Employers by Total Applications
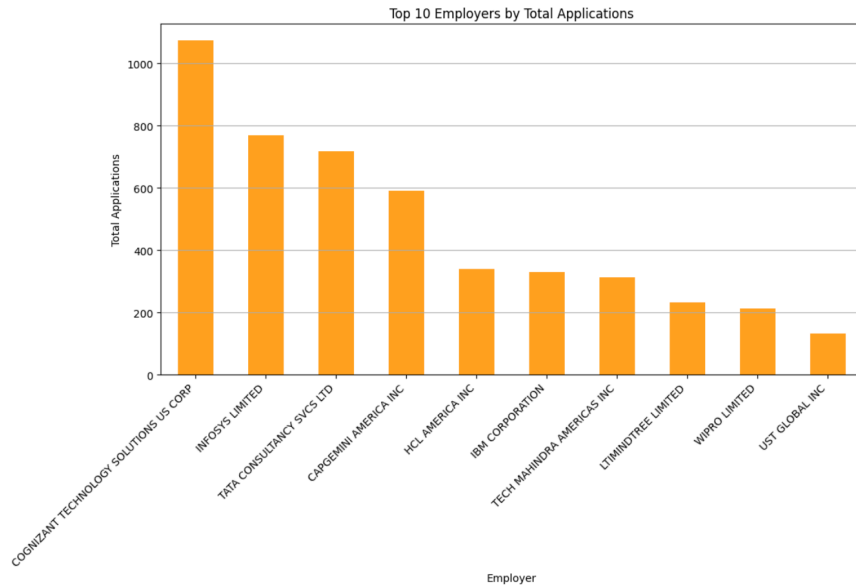
- **Cognizant Technology Solutions US Corp** leads with over 1,000 applications, highlighting its significant role in sponsoring H-1B visas.
- Other major contributors, such as **Infosys Limited**, **Tata Consultancy Services**, and **Capgemini America Inc**, demonstrate the dominance of IT and consulting firms in H-1B sponsorships.
- Employers in this list are largely concentrated in the technology and consulting industries, reflecting the high demand for international talent in these sectors.
- While the total application volumes are significant, further analysis (e.g., approval rates) can identify whether these high-application employers also have high success rates, which is critical for prospective applicants.

Implications for Applicants:

1. Targeting these employers may increase the chances of securing H-1B sponsorship due to their high volume of applications.
2. Applicants should also consider these firms' historical approval rates and the industries they serve when strategizing their job search.

Approval and Denial Rates by Industry (NAICS Codes

Approval and Denial Rates by Industry (NAICS)
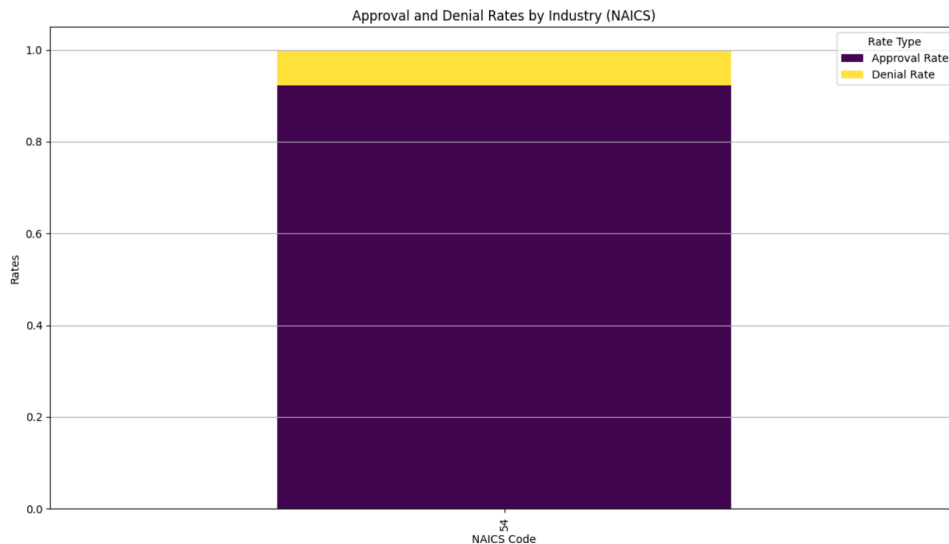
The stacked bar chart illustrates the breakdown of approval and denial rates for industries classified under NAICS codes.
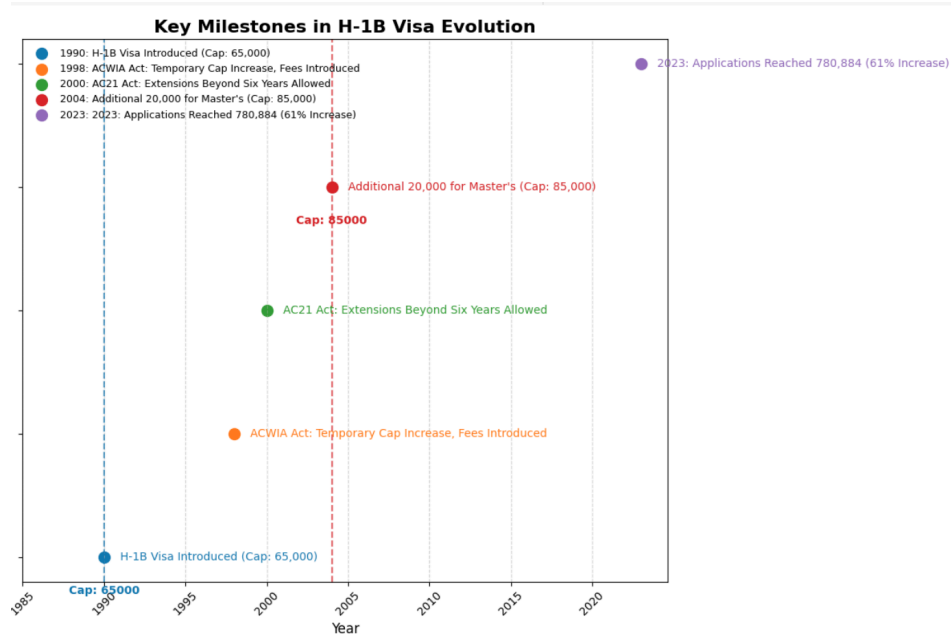
- **High Approval Rates**:
  Industries within NAICS code **54 (Professional, Scientific, and Technical Services)** exhibit near-perfect approval rates (shown in purple), highlighting their favorable outcomes for H-1B visa applications. This reflects the demand for skilled workers in these industries.
- **Low Denial Rates**:
  The yellow segment (denial rates) is minimal, indicating that denial rates are negligible for this NAICS code. This makes these industries highly reliable options for applicants seeking H-1B sponsorships.

Key Insight:
Applicants targeting industries under NAICS code **54** have a significantly higher likelihood of success, making these industries a strategic choice for international workers.

Trends Over Time (Fiscal Year Analysis)

This section examines how H-1B visa approval rates have evolved over time. Understanding these trends can shed light on external factors such as economic conditions, immigration policy changes, and employer strategies that may influence application outcomes.

Key Milestones in H-1B Visa Evolution

- 1990: H-1B Visa Introduced (Cap: 65,000)
- 1998: ACWIA Act: Temporary Cap Increase, Fees Introduced
- 2000: AC21 Act: Extensions Beyond Six Years Allowed
- 2004: Additional 20,000 for Master's (Cap: 85,000)
- 2023: 2023: Applications Reached 780,884 (61% Increase)

2023: Applications Reached 780,884 (61% Increase)

Additional 20,000 for Master's (Cap: 85,000)

Cap: 85000

AC21 Act: Extensions Beyond Six Years Allowed

ACWIA Act: Temporary Cap Increase, Fees Introduced

H-1B Visa Introduced (Cap: 65,000)

Cap: 65000

Year

Insights:

1. **1990: Introduction**
   - The H-1B visa program began with a cap of **65,000**, targeting skilled foreign workers to meet U.S. labor demands.
2. **1998: ACWIA Act**
   - Temporary cap increases and worker training fees balanced foreign talent needs with domestic workforce development.
3. **2000: AC21 Act**
   - Allowed visa extensions beyond six years to address green card delays, enhancing flexibility for workers.
4. **2004: Master's Cap Increase**
   - An additional **20,000 visas** for U.S. master's degree holders raised the cap to **85,000**, incentivizing advanced education.
5. **2023: Record Demand**
   - Applications reached **780,884** (a **61% increase**), reflecting the growing global reliance on the program.

Trends and Implications:

- **Rising Demand:** Applications consistently outpace the static cap, increasing lottery competition.
- **Legislative Adjustments:** Milestones reflect efforts to align the program with economic and workforce priorities.
- **Future Challenges:** Reforms may address caps, allocation, and high-demand industries to ensure sustainability.

The program's evolution showcases its critical role in U.S. workforce strategy, though growing demand highlights the need for modernization.

Approval Rate (2023)

95
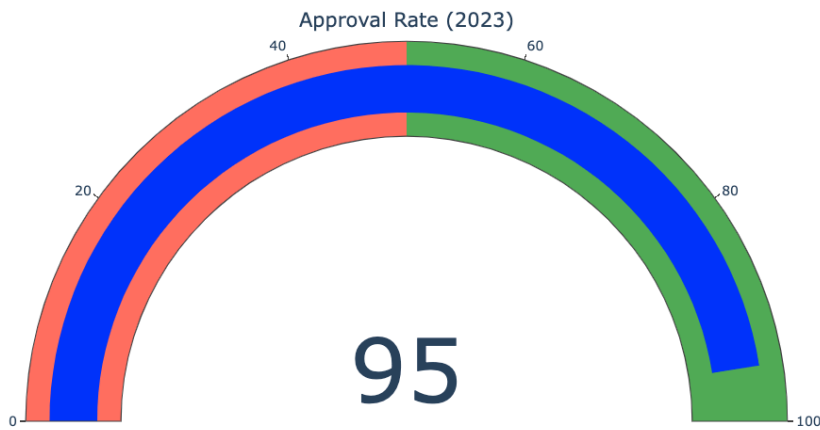
Challenges and Bias in the Dataset

- Employer representation is skewed towards IT and consulting firms.
- State-level trends may not reflect broader national patterns.

Implications of Findings

The findings highlight actionable strategies for international applicants, such as targeting specific employers and industries.

Recommendations

State and industry strategies should focus on high-approval locations and industries such as Professional, Scientific, and Technical Services. Applicants must prioritize employers with strong sponsorship histories, leveraging OPT periods effectively to maximize opportunities. Staying updated on policy changes and cap announcements is crucial for strategic planning.

Future Work

Future work could incorporate salary and job title data for enhanced analysis, explore time series forecasting for application trends, and include qualitative analysis of employer reviews to better understand the work environment.

Conclusion

Securing an H-1B visa is a challenging yet rewarding endeavor for international students. This project highlights the value of data-driven insights in navigating the complexities of the visa process. By analyzing trends, employer behavior, and state-level dynamics, prospective applicants can develop strategic approaches to maximize their chances of success.

For international students graduating in May 2025, it is essential to begin preparations early. Building strong professional networks and applying for jobs during the OPT period can help secure potential sponsors. Focusing on employers with a proven track record of successful H-1B sponsorship and targeting states with higher approval rates are key strategies. Additionally, staying informed about policy changes and cap limitations is vital for effective planning.

Graduating students must also tailor their resumes to align with the needs of industries that consistently support H-1B applications, such as technology and consulting. This involves emphasizing technical skills, certifications, and experiences that resonate with these employers. While the journey is competitive, a proactive and well-informed approach can significantly improve the likelihood of obtaining an H-1B visa and transitioning successfully into a professional career in the United States.

References

- Dataset Source: [Machine Learning Repository]
- Python Libraries: pandas, scikit-learn, matplotlib, seaborn
- External Research: U.S. Department of Labor H-1B Data