

The value of international students to the United States. Probability of getting a non-immigrant visa.



Zinaida Dvoskina | Kirill Ilin
Johnathan Conley | Cindy Ye Fung

~~X~~



Business Question

- Northeastern University is among the top US colleges for international students under the F-1 visa, and our MSBA cohort represents 16 foreign countries
- International students are beneficial to the US economy and create thousands of workplaces
- The job market is challenging

X

Business Question



- Recently the immigration issues have become very political
- Seeking to understand if various factors such as the passport that the student holds, local location of the applicant, type of visa the student is seeking, the political party in office, wage rate, job title, etc. determines the likelihood of an international student of obtaining a working visa.



DATA SOURCES

Publicly Available data from
governmental websites

U.S. Department of Labor

USCIS



Data Sources

- USCIS has data on number of visas issued per
 - country
 - category
 - year



Data Sources

- U.S. Department of Labor has Performance Data
- Information about employment-based immigration applications
- Dataset is cumulative for the fiscal year
- Records provided by employers such as an applicant's received dates, decision dates, the most recent date a case determination decision was issued, etc.



Data Cleaning

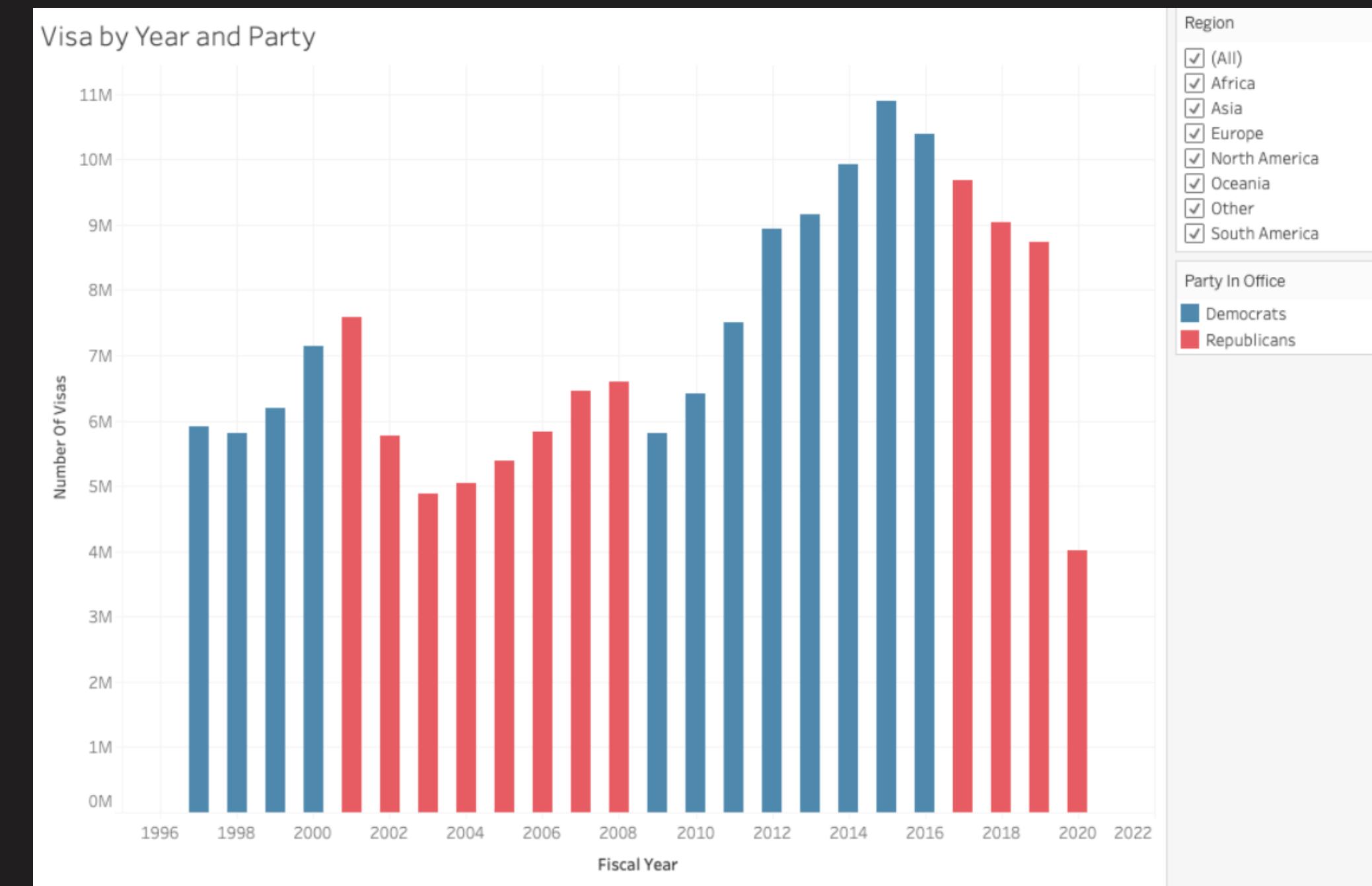
- "Status" column where we only include rows with "Certified" and "Denied" results.
- Eliminate the rows with a "withdrawn" status and changed "Certified-withdrawn" status to "Certified"
- Added the "pw_Fixed" column to standardize the prevailing wage rate to be yearly
 - Calculated by multiplying the rows with hourly values by 2087
- Eliminating columns with high number of BLANK fields or that didn't provide valuable information to our analysis

| Column Name | Description |
|--------------------------------|-----------------------------------------------------------------------------------------------|
| Year | Fiscal Year |
| Status | Status associated with the last significant event or decision: "Certified", Denied". |
| LCA_CASE_SUBMIT | Date and time the application was submitted |
| DECISION_DATE | Date on which the last significant event or determination was issued by OFLC. |
| LCA_CASE_EMPLOYMENT_START_DATE | Beginning date of employment |
| LCA_CASE_EMPLOYMENT_END_DATE | Ending date of employment |
| LCA_CASE_EMPLOYER_NAME | Employer's name |
| LCA_CASE_SOC_CODE | The Standard Occupational Classification code which classifies workers by occupational groups |
| LCA_CASE_SOC_NAME | Title of the SOC occupational group |
| LCA_CASE_JOB_TITLE | Job title |
| LCA_CASE_WAGE_RATE_FROM | Employer's proposed wage rate |
| WORK_LOCATION_CITY1 | City of the job opening |
| WORK_LOCATION_STATE1 | State of the job opening |
| PW_Fixed | Hourly Prevailing wage rate |

Visualizations

Visa by Year and Party

- No strong correlation
- No obvious trend



Visualizations

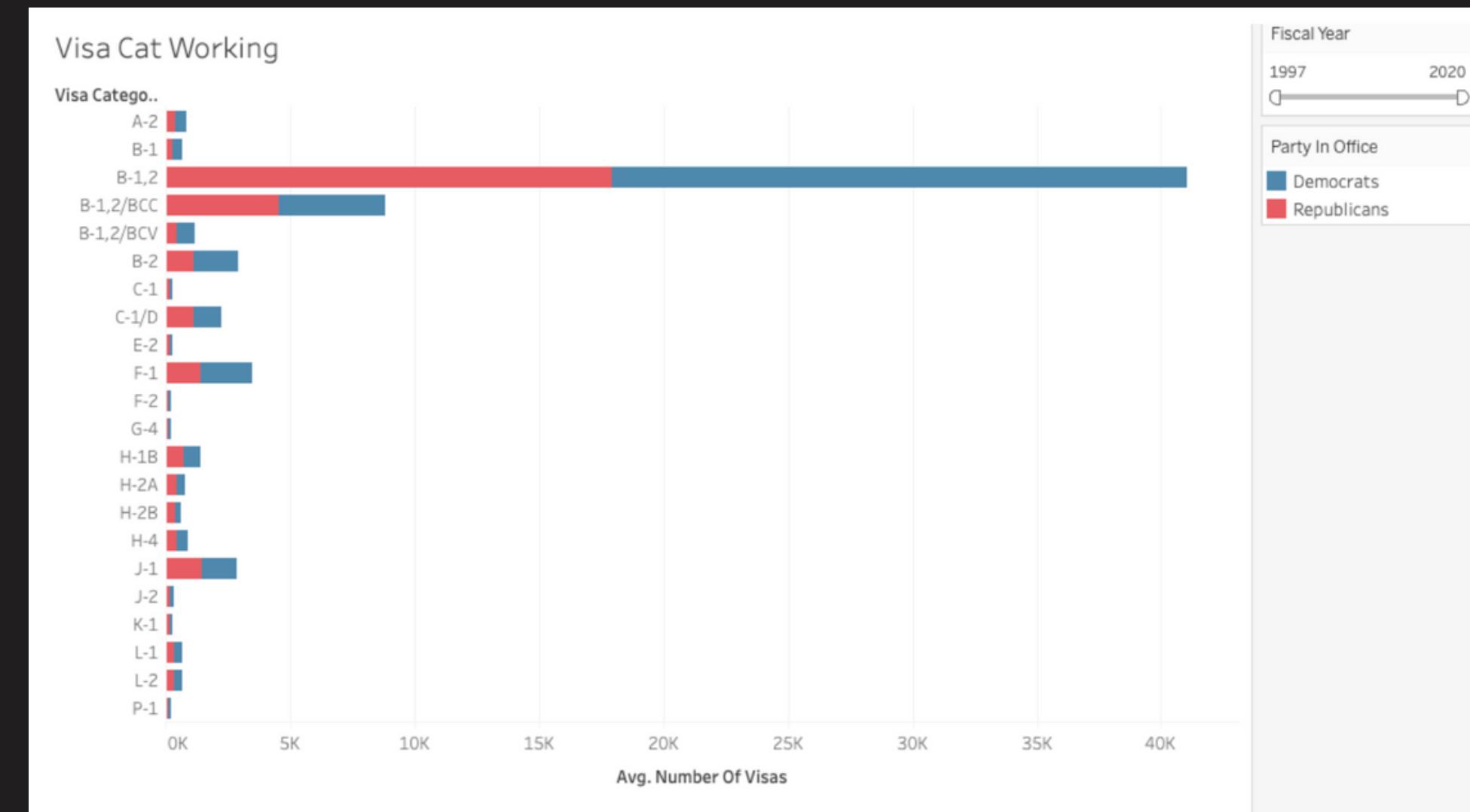
Visa Time Lapse

- "Geo map that can be filtered by year and country to show number of visas



Visualizations

- Visa Category WORKING
 - BAR graph shows average number of visas for each visa category
 - B-1: persons who want to enter the US temporarily for business
 - B-2: for tourism
 - B-1/B-2 dominates: for a combination of both purposes
- Distinction between Democrats and Republicans party
- Not showing a significant difference



Key 2020 Stats

Job Title

| | |
|-----------------------------------------|--------------|
| Software Engineer | 21092 |
| Software Developer | 15669 |
| Senior Systems Analyst JC60 | 9334 |
| Manager JC50 | 7834 |
| SOFTWARE DEVELOPER | 6810 |
| SOFTWARE ENGINEER | 6327 |
| Senior Software Engineer | 5836 |
| Assistant Professor | 4030 |
| SOFTWARE DEVELOPMENT ENGINEER II | 3372 |
| System Analyst JC65 | 2831 |

Company

| | |
|-----------------------------------------------|--------------|
| COGNIZANT TECHNOLOGY SOLUTIONS US CORP | 28625 |
| INFOSYS LIMITED | 8906 |
| Ernst & Young U.S. LLP | 8846 |
| TATA CONSULTANCY SERVICES LIMITED | 8748 |
| Microsoft Corporation | 7792 |
| AMAZON.COM SERVICES LLC | 7375 |
| Google LLC | 6514 |
| Accenture LLP | 5713 |
| Deloitte Consulting LLP | 5487 |
| CAPGEMINI AMERICA INC | 5335 |

SOC Code

| | |
|---------------------------------------------------------|---------------|
| Software Developers, Applications | 183111 |
| Computer Systems Analysts | 48455 |
| Software Developers, Systems Software | 28754 |
| Computer Systems Engineers/Architects | 16576 |
| Software Quality Assurance Engineers and Testers | 14332 |
| Information Technology Project Managers | 12056 |
| Computer Programmers | 11955 |
| Computer and Information Systems Managers | 11022 |
| Mechanical Engineers | 9839 |
| Operations Research Analysts | 9676 |

City

| | |
|----------------------|--------------|
| New York | 24462 |
| San Francisco | 11561 |
| Chicago | 9127 |
| SEATTLE | 8060 |
| San Jose | 7733 |
| Sunnyvale | 7633 |
| Atlanta | 7581 |
| Austin | 7086 |
| Redmond | 6998 |
| Houston | 6980 |

State

| | |
|-----------|---------------|
| CA | 114003 |
| TX | 57956 |
| NY | 45918 |
| WA | 32092 |
| NJ | 30715 |
| IL | 27196 |
| MA | 22449 |
| NC | 20451 |
| PA | 19391 |
| GA | 19262 |



Challenges

- .7% denial when excluding withdrawn
 - Large amounts of data (too much)
 - Speed
 - Equipment
 - Method
 - Constant changing immigration laws
 - Changing data files



Future Work

- Model for software engineers
 - Make up the most applications
 - However, have the most denials
 - Historical Dashboard
 - maintain for student use



Data Processing

- Prevailing wage
 - Year - 1
 - Hour - 2087
 - Month - 12
 - Week - 51 (Assumes some holidays)
- Case status
 - Certified (Target)
 - Denied
 - Certified withdrawn
 - withdrawn
- Standard Scaler

Analysis

This classification matrix evaluates results of the KNN model that uses following variables as predictors:

- Received month
- Agent representing employer
- Annual wage rate
- Annual prevailing wage
- PW wage level
- H-1B dependent status
- Support H1B status

| | | precision | recall | f1-score | support |
|--|--------------|-----------|--------|----------|---------|
| | 0 | 0.41 | 0.07 | 0.11 | 1172 |
| | 1 | 0.99 | 1.00 | 1.00 | 168730 |
| | accuracy | | | 0.99 | 169902 |
| | macro avg | 0.70 | 0.53 | 0.55 | 169902 |
| | weighted avg | 0.99 | 0.99 | 0.99 | 169902 |

Analysis

| | | precision | recall | f1-score | support |
|--|--------------|-----------|--------|----------|---------|
| | 0 | 0.41 | 0.07 | 0.11 | 1172 |
| | 1 | 0.99 | 1.00 | 1.00 | 168730 |
| | accuracy | | | 0.99 | 169902 |
| | macro avg | 0.70 | 0.53 | 0.55 | 169902 |
| | weighted avg | 0.99 | 0.99 | 0.99 | 169902 |

Classification Report shows that model performs very well predicting positive outcome for visa approval, however it's accuracy is a lot smaller when predicting negative result. Overall model performs very well with f-1 score being 0,99.

Recommendations/ Key Findings

- The political party in office is not a great variable to determine the likelihood of an international student obtaining a professional working visa as there's no correlation
- To improve our prediction, including other variables that capture macro and micro factors is necessary
- Models with different predictors show very high accuracy in predicting positive result of visa application, however they perform poorly in predicting negative results.
- This can be due to low amount of data on visa denial cases that do not allow to increase the number of successful negative predictions

To The Success of
International
and Domestic
Students!

CLASS OF 2021!