

# DIIG Data Challenge

By Ayush Jain (adj51)



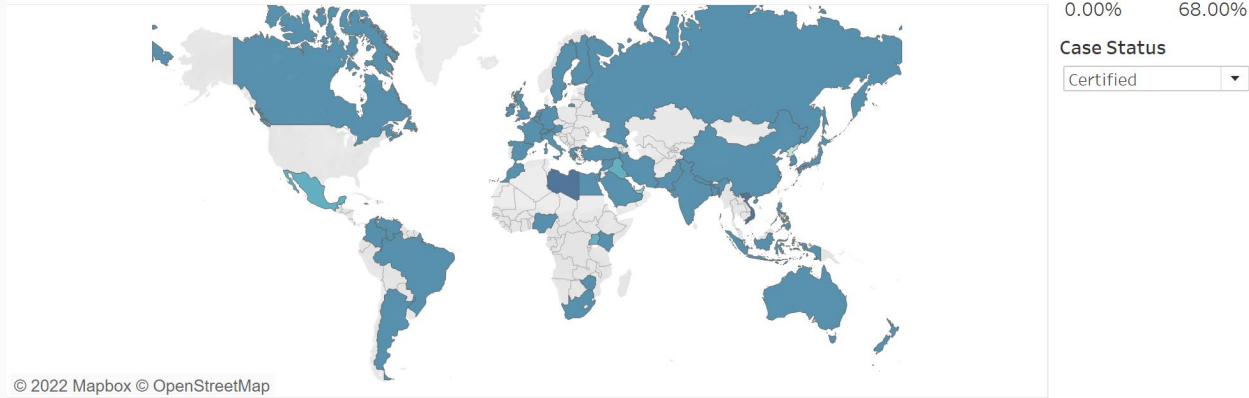
# Overview

- Dataset chosen: US Visa Applications
- Central question: What are the demographics of applicants who get accepted/denied for US visa applications and which variables play a role?
- Exploratory Data Analysis
- Findings
- Conclusions/Suggestions

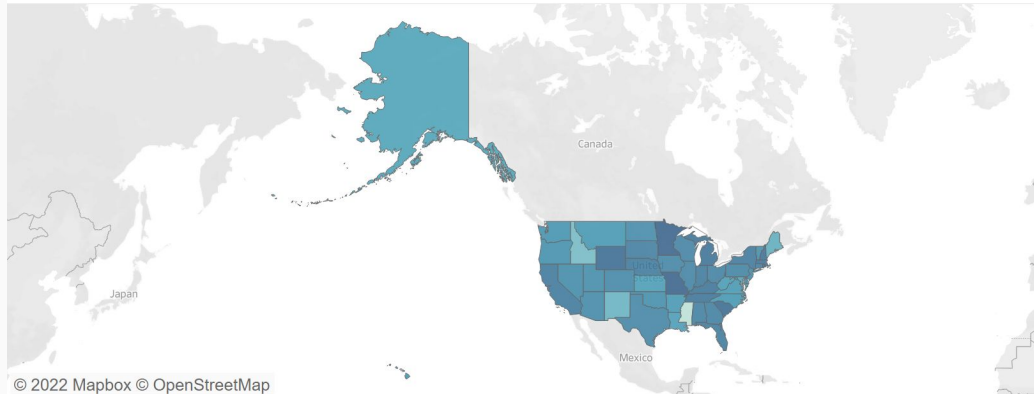
Central Question: What are the demographics of US visa applicants and what variables play a role in acceptance/denial?

# Case acceptance

## Certified Rate by Country

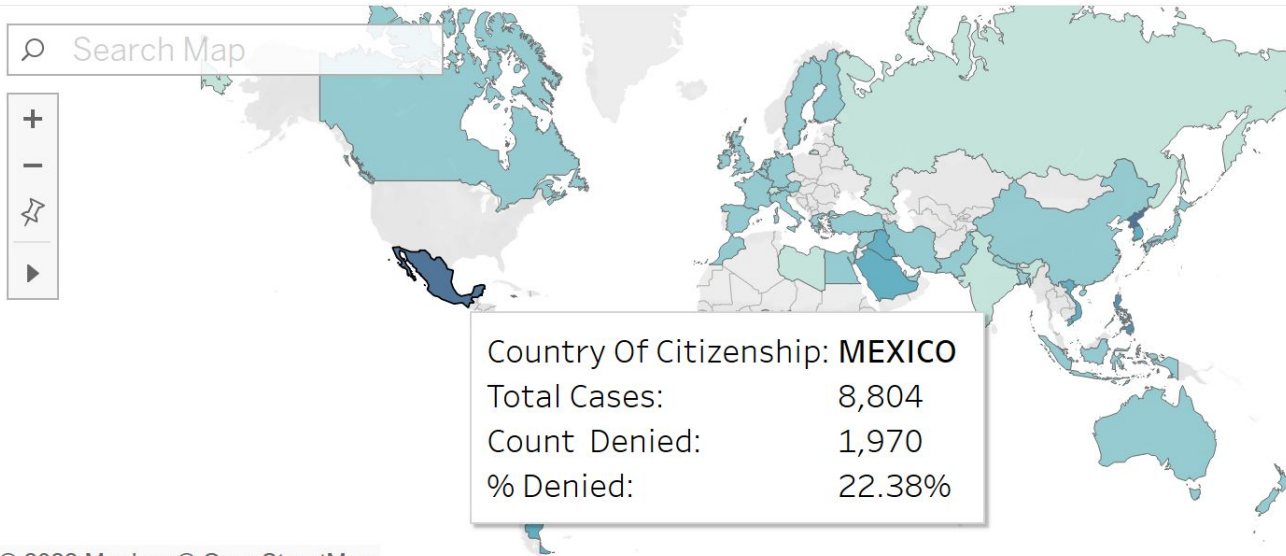


## Certified Rate By State of Employer



# Highest case rejection rate (by country)

## Denied Rate by Country



% of Cases Effected

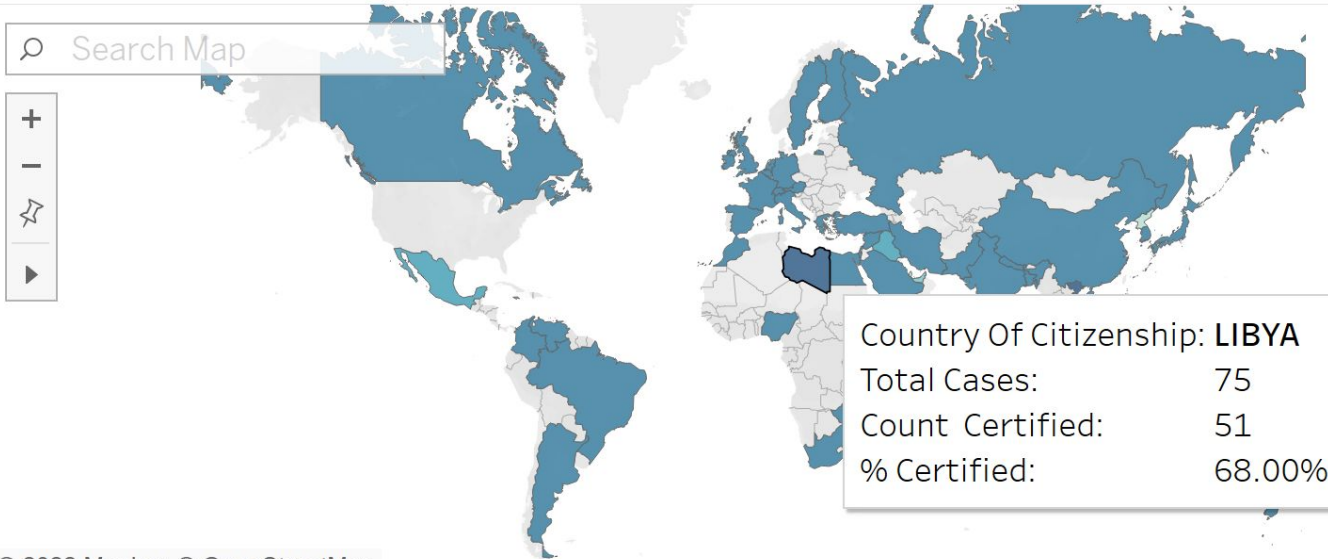
0.00% 25.00%

Case Status

Denied

# Highest case acceptance rate (by country)

## Certified Rate by Country



% of Cases Effected

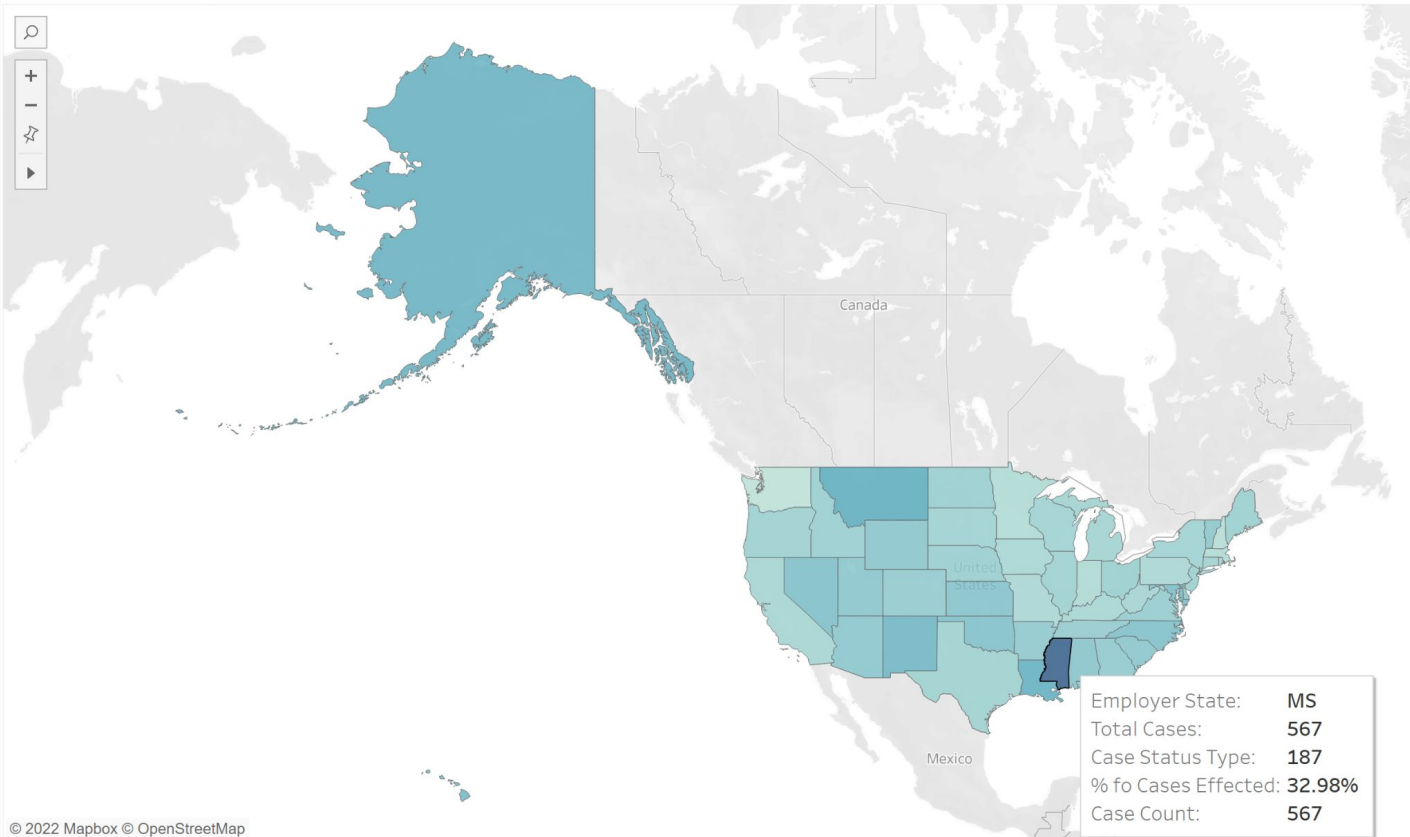


Case Status

Certified ▼

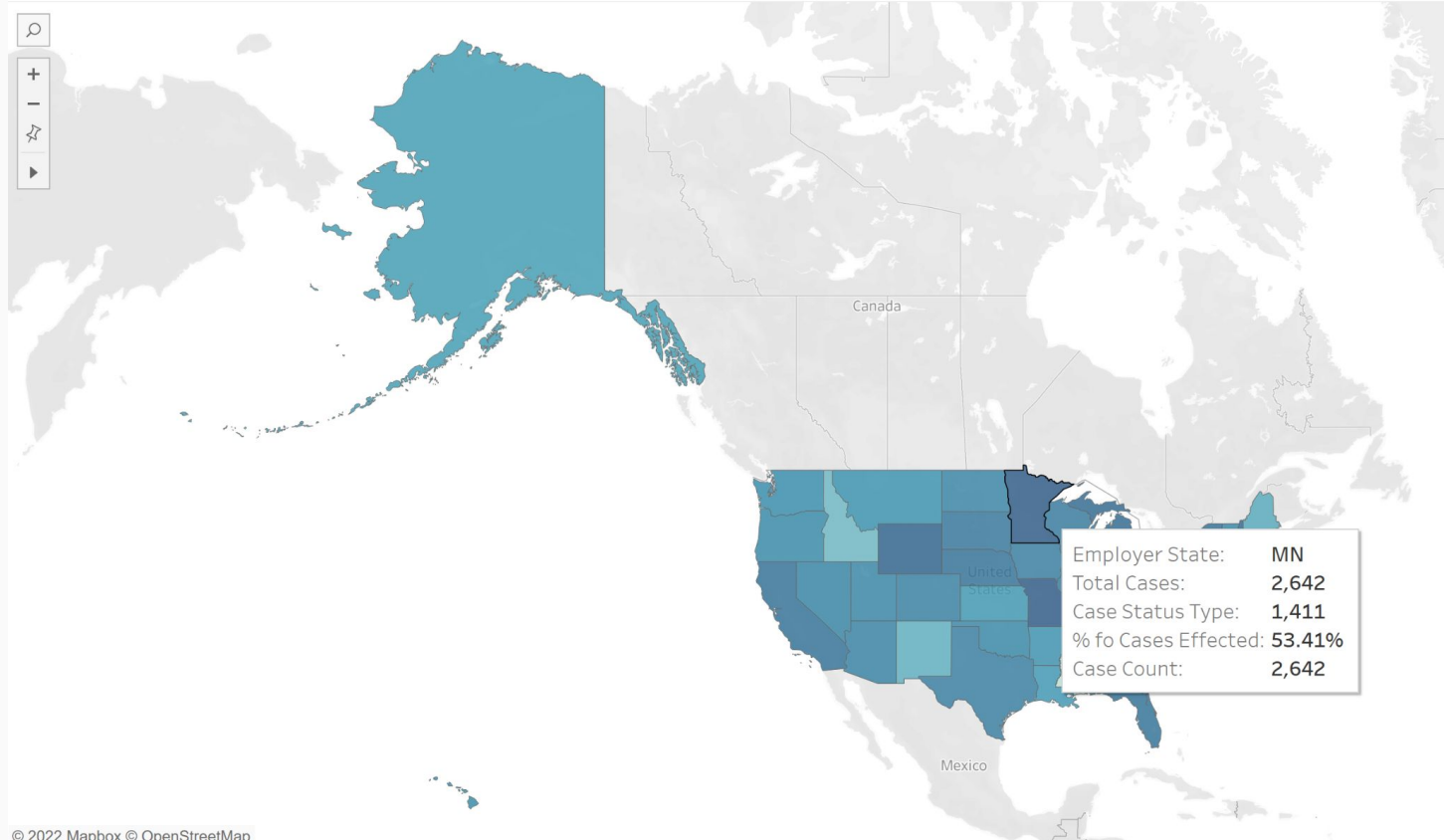
# Highest case rejection rate (by state)

Denied Rate By State of Employer



# Highest case acceptance rate (by state)

Certified Rate By State of Employer





# 86,682

Largest raw number of cases processed by a state

Associated acceptance rate: 50.18%

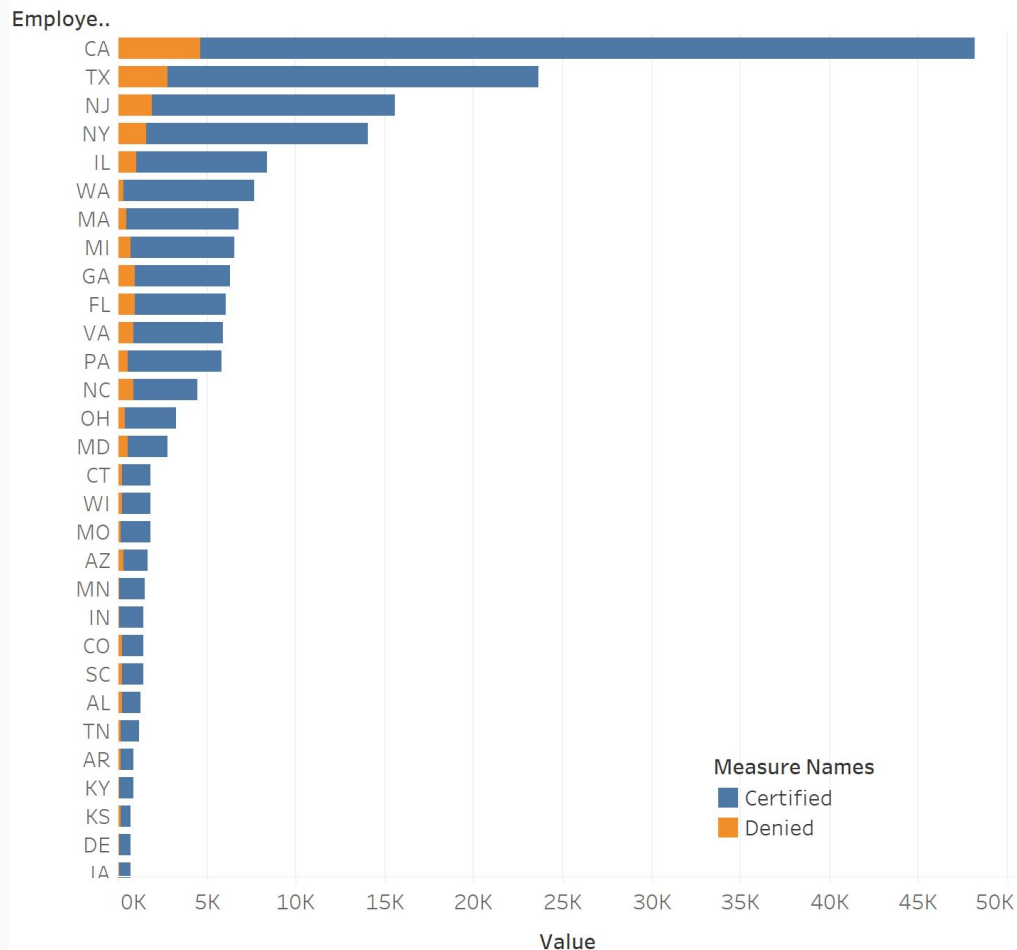
# Acceptance/Denial by state in numbers

California overwhelmingly  
beats other states with  
almost 50K cases

Unsurprising given that  
California attracts a  
diverse and multicultural  
crowd

Other notable states:  
Texas, New Jersey, New  
York

Number of Cases Certified vs Denied by Employer



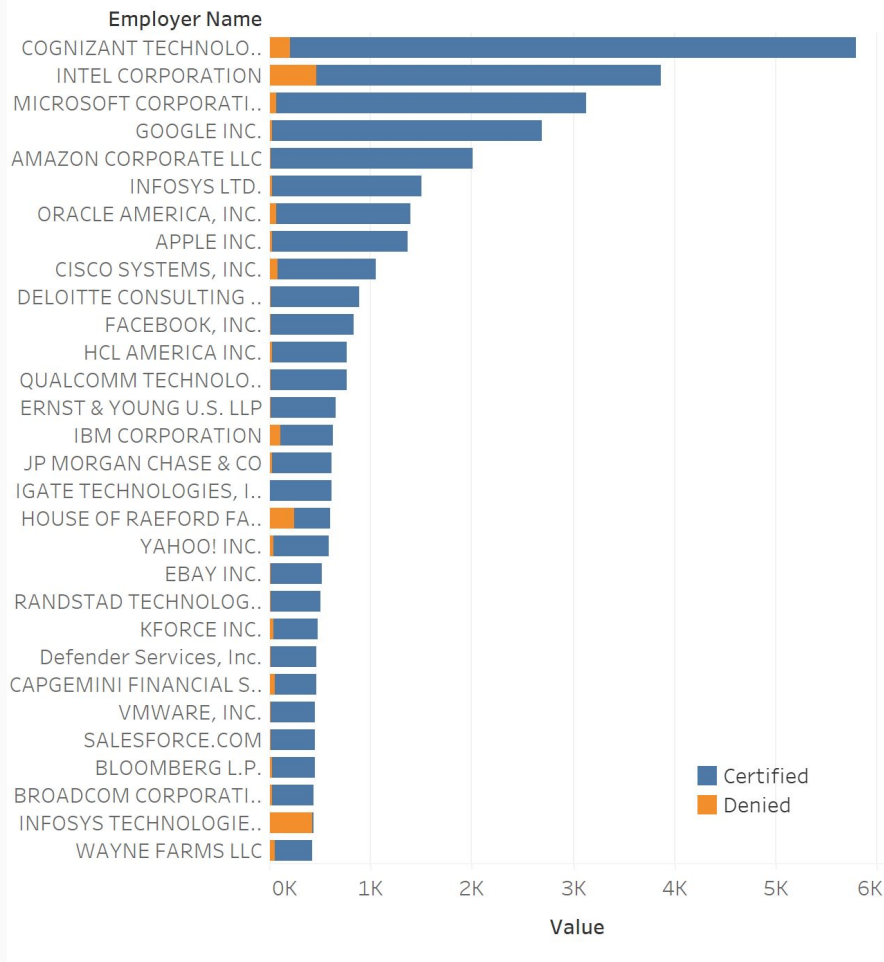
# Filtering by employer

Cognizant and Intel topped the charts with highest number of accepted applicants

This was further filtered by visa type to see what type of visas ranked the highest

H-1B and F-1 visas were the top 2

Number of Cases Certified vs Denied by Employer



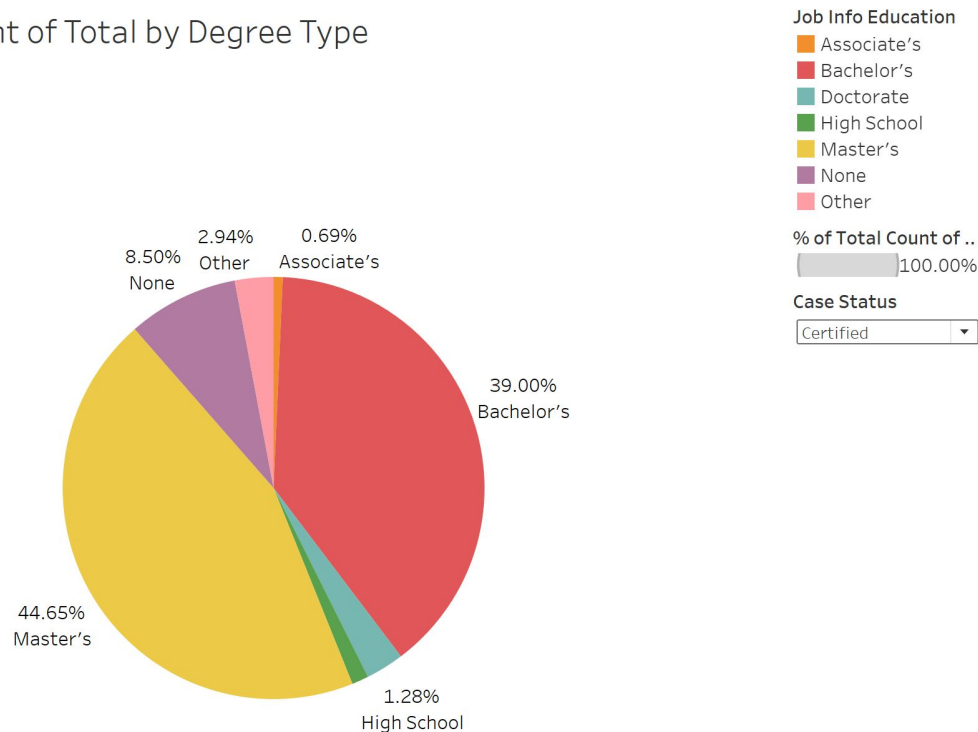
# Qualifications

Applicants with a Master's or Bachelor's degree had the highest acceptance rate

Note that 'None' has been used for missing data rather than no qualifications

**\*\* Change "Case status" filter to the right to see changes\*\***

Certified Percent of Total by Degree Type



## Qualifications

Although high school students should be higher ideally, this data set mostly has entries for work visas (not F-1, J-1 etc) so it makes sense that higher education qualifications have a higher acceptance rate

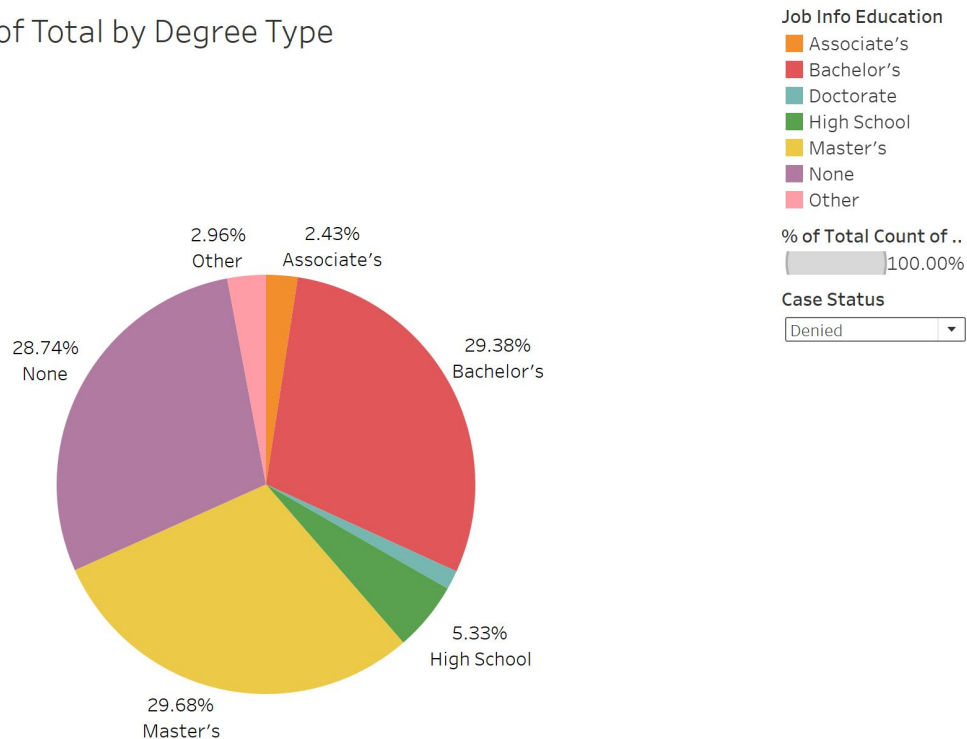
class_of_admission	n	percent
H-1B	265411	76.42631%
	20529	5.91142%
L-1	18013	5.18693%
F-1	13402	3.85917%
Not in USA	7995	2.30220%
TN	4244	1.22208%
E-2	4017	1.15671%
Parolee	2628	0.75674%
B-2	2524	0.72680%
EWI	1281	0.36887%
F-2	895	0.25772%
E-3	873	0.25138%
J-1	783	0.22547%
H-4	690	0.19869%
O-1	672	0.19351%
E-1	513	0.14772%
L-2	494	0.14225%

# Qualifications

Although the denial rate is high for applicants with Master's or Bachelor's qualifications, it is hard to make conclusions because a lot of data is simply missing

**\*\* Change "Case status" filter to the right to see changes\*\***

Denied Percent of Total by Degree Type



# Decision Year vs Application Year

Decision Year	Case Received Year											
	Null	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
2011	10,152											
2012	21,394											
2013	37,761											
2014	55,212			7	15	53	24	245	2,033	14,480		
2015	1	1	3	55	127	141	273	373	11,145	42,095	35,848	
2016			3	3	38	153	211	224	393	2,148	45,335	67,331

# Final conclusions

What factors played a significant  
role?





# Conclusions

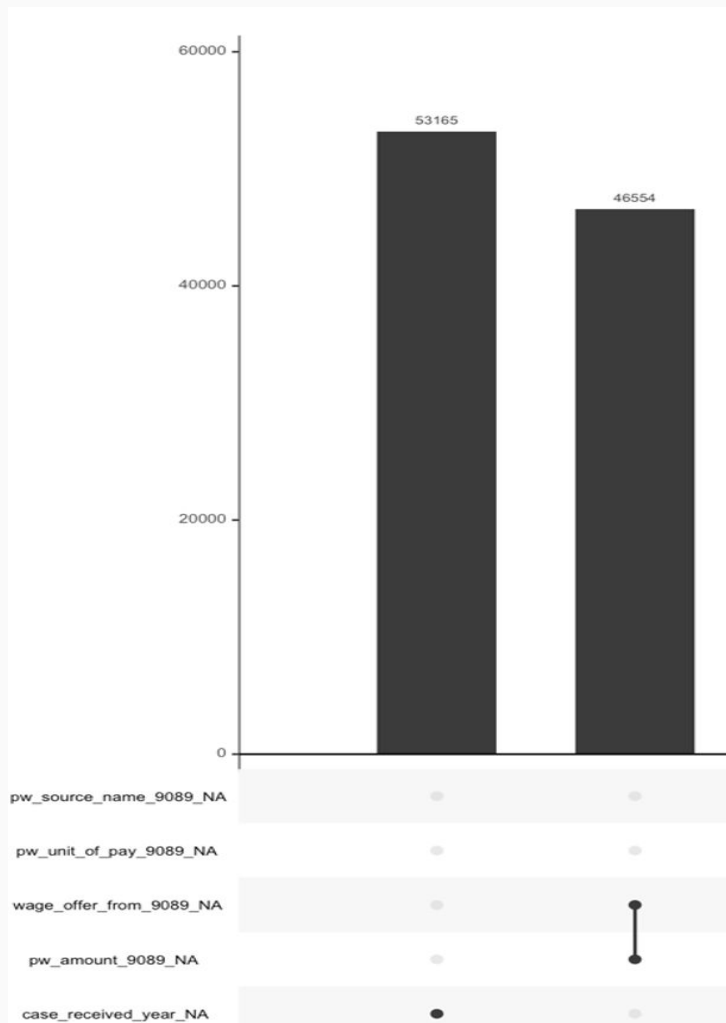
- Highest denial rate for Mexican applicants
- Highest number of applicants and acceptances for California, Texas, Jersey, NY. Minnesota marginally higher by %
- Cognizant and Intel had most cases + acceptances, Intel lower % rate
- Higher education qualifications more likely to be accepted
- Heavy backlog in case processing but still high volume

# Limitations

The data had a lot of missing values.

52,165 values were missing from the case\_received year.

Wage offer from 9089 along with pw\_amount 9089 had 46,554 values missing



# Thank you!

*Questions?*