Easy Visa – Case Study

Jesus Torres

Committed to providing clients with a transparent, smooth, and client-focused immigration vetting process. No other company knows a career abroad like us.

CONTENT

- Business Problem Overview
- Data Overview
- Exploratory Data Analysis (EDA)
- Model Performance Summary
- Business Insights and Recommendations

Business Problem Overview & Solution Approach

- In FY 2016, the OFLC processed 775,979 employer applications for 1,699,957 positions for temporary and permanent labor certifications. This was a nine percent increase in the overall number of processed applications from the previous year. The process of reviewing every case is becoming a tedious task as the number of applicants is increasing every year.

- The increasing number of applicants every year calls for a Machine Learning based solution that can help in shortlisting the candidates having higher chances of VISA approval. OFLC has hired Easy Visa for data-driven solutions. A data scientist at Easy Visa must analyze the data provided and, with the help of a classification model:
  1. Facilitate the process of visa approvals.
  2. Recommend a suitable profile for the applicants for whom the visa should be certified or denied based on the drivers that significantly influence the case status.

Data Overview

- The data contains information about 25,480 VISA applications and their 12 features.

Observations	Features
25,480	12

- The features include case_id, continent, education_of_employee, has_job_experience, requires_job_training, no_of_employees, yr_of_estab, region_of_employment, prevailing_wage, unit_of_wage, full_time_position, and case_status.

- The feature prevailing wage although it contains some irregular values was not modified since additional information for unit of wage would be necessary to do so.
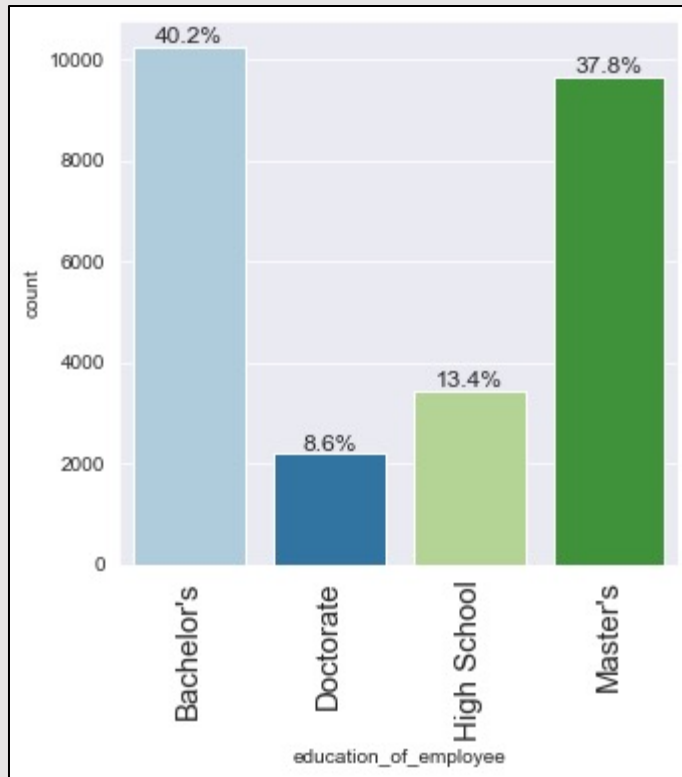
Exploratory Data Analysis (EDA)
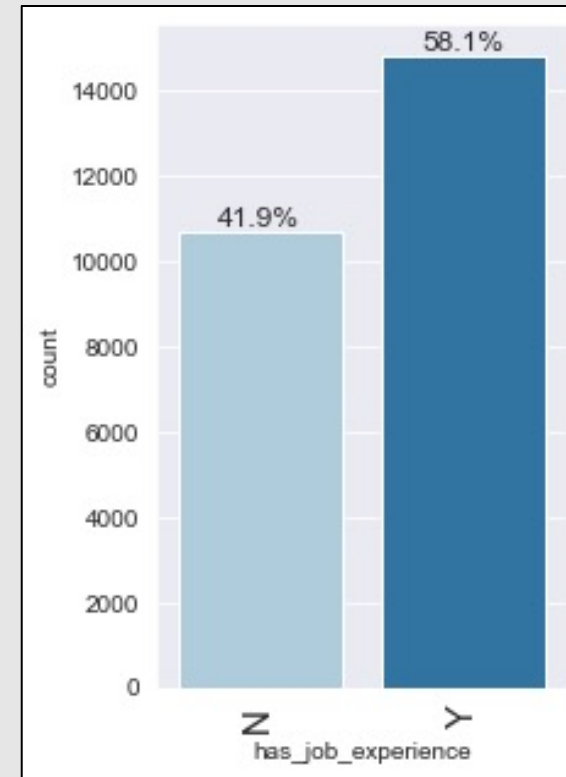
Univariate Analysis
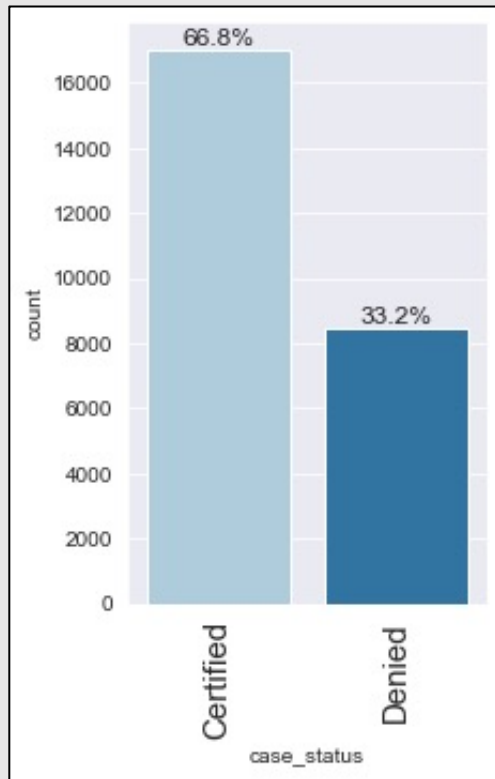
- 67% roughly two-thirds of all VISA applicants are certified.

Univariate Analysis

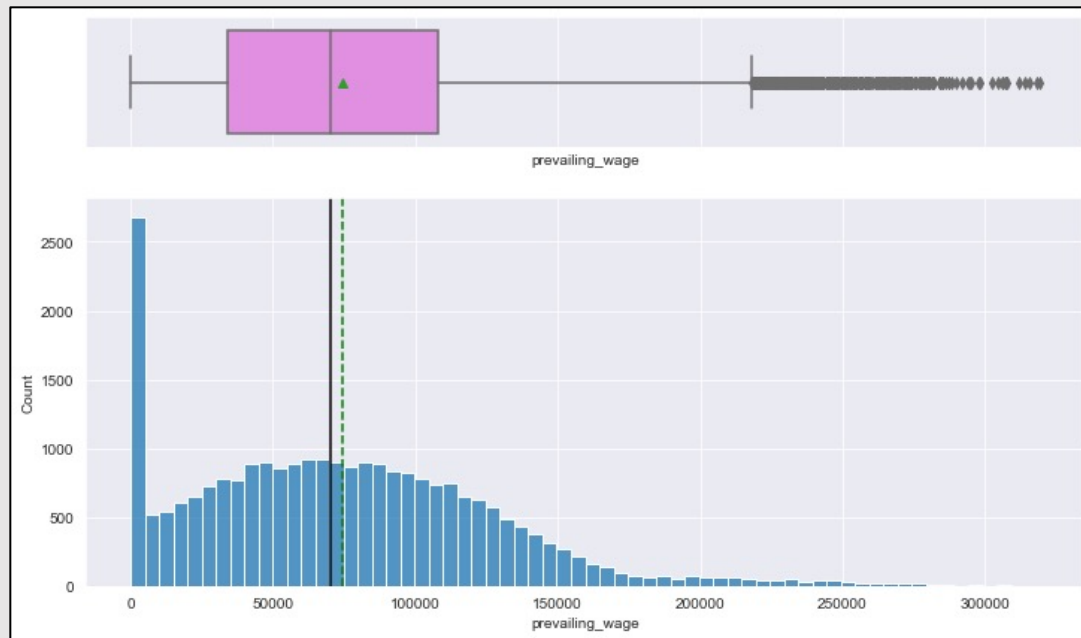- Most VISA applicants, 40%, hold a bachelor's degree.

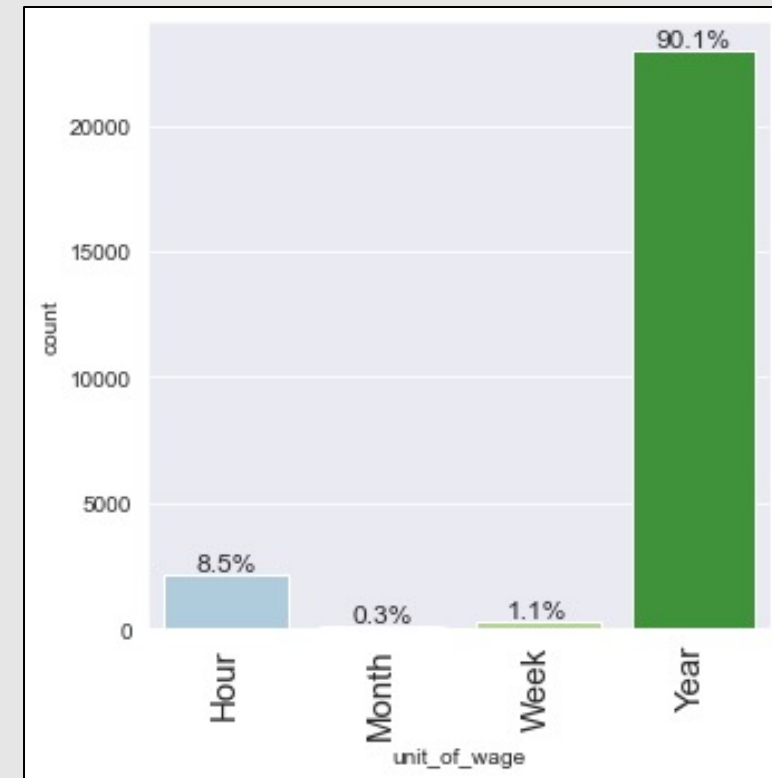- About 58% of VISA applicants have previous work experience.

Univariate Analysis

- Prevailing wage follows a right-skewed distribution with a median at about 70,828 USD.
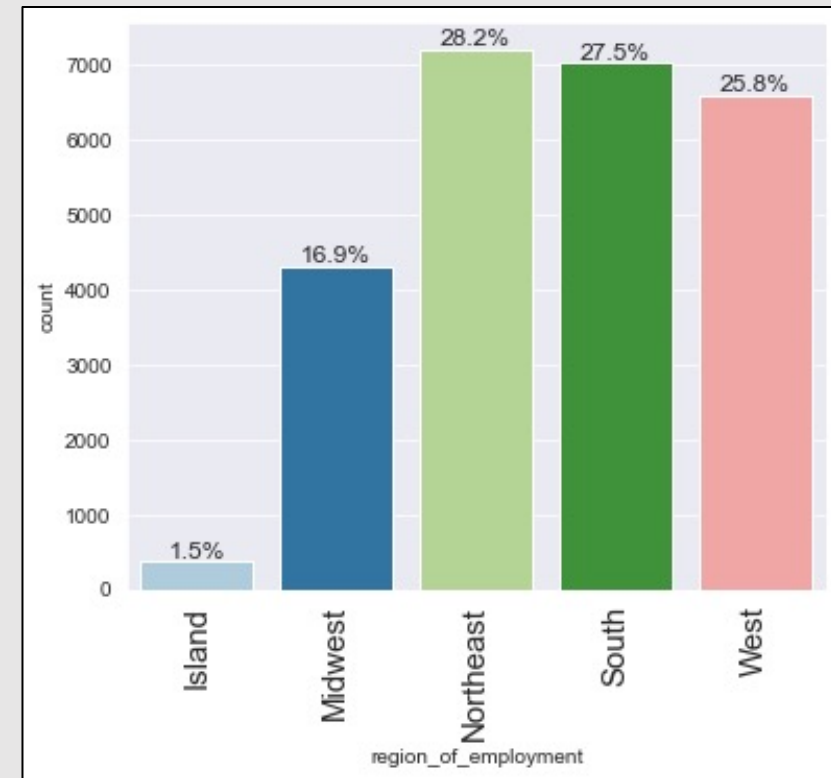
- The majority, 90%, of VISA applicants show a yearly unit wage.

Univariate Analysis

- Roughly two-thirds, 66.2%, of VISA applicants represent the Asian continent.

- The Northeast, South, and West regions of employment represent ~80% of all VISA applicants.
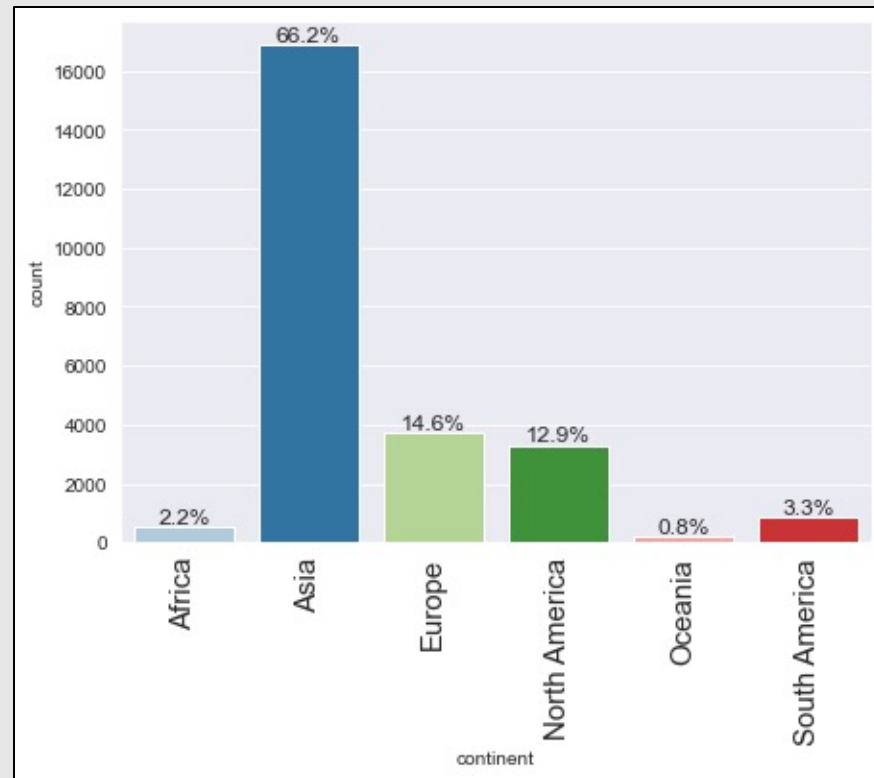
Bivariate Analysis

• Our target variable is the case status of a VISA applicant.

• Positively correlated features include:

1. Employee education

2. Job experience

3. Region of employment

4. Unit of wage/salary

5. Continent

Bivariate Analysis
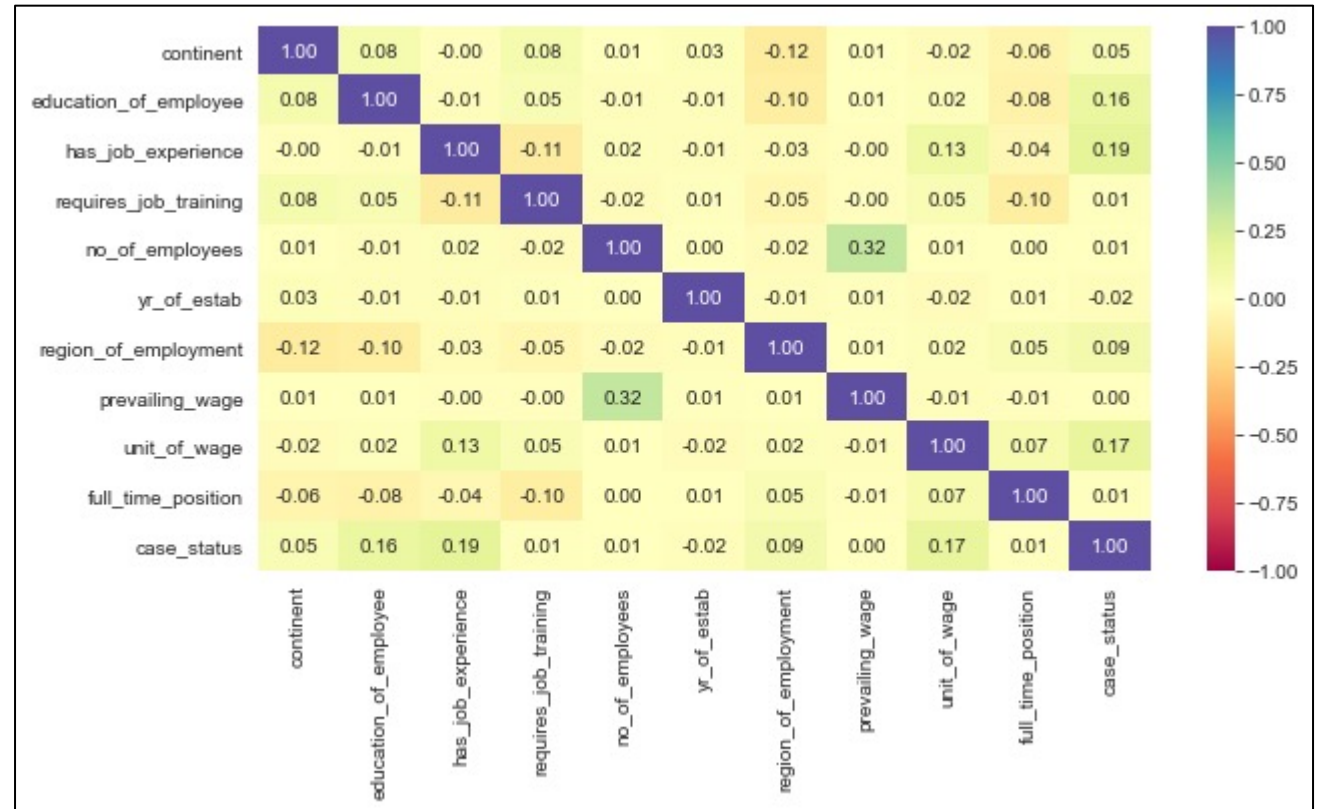
• For one unit increase in education level the likelihood of the VISA applicant getting certified increases too.

• VISA applicants with a doctorate hold the highest certification rate at about 87%.

• Followed by:
  • Master's – 79%
  • Bachelor's – 62%
  • High School – 36%
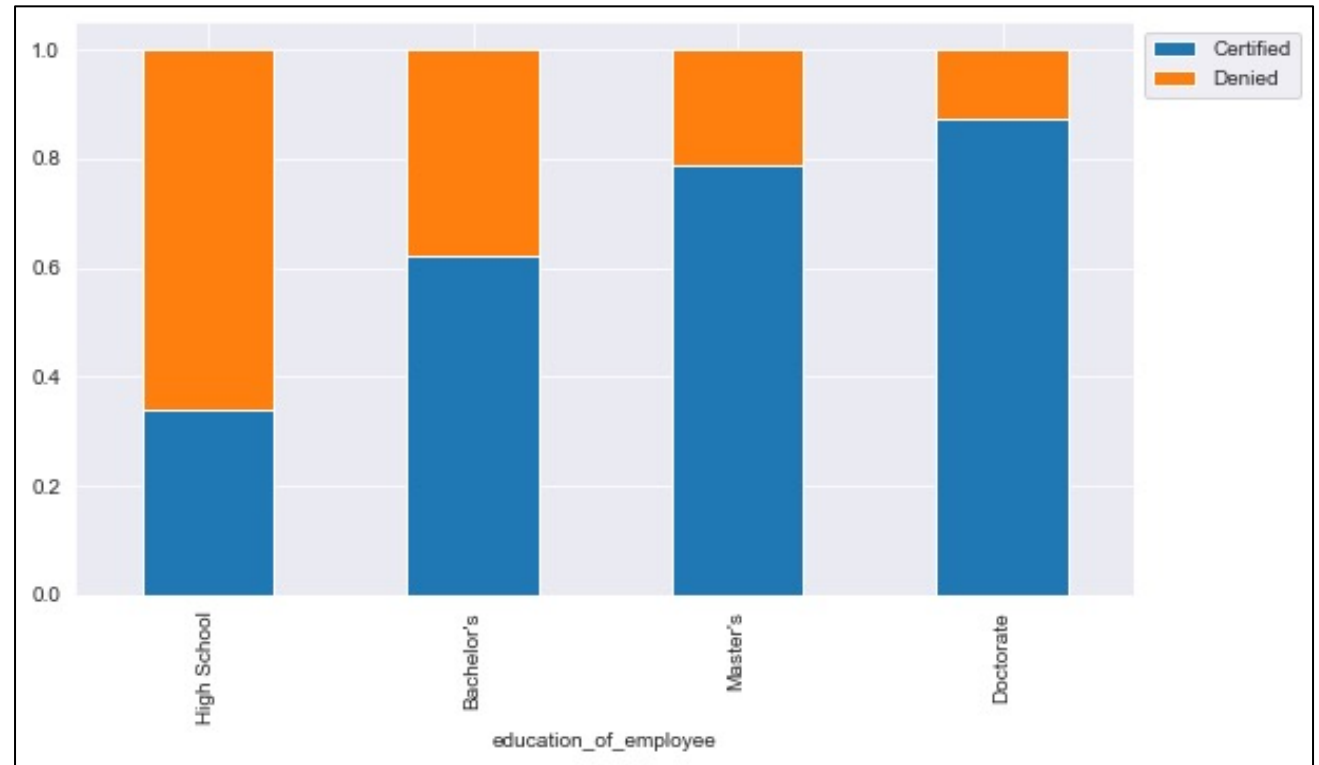
Bivariate Analysis

• VISA applicants who do have previous work experience show a higher certification rate of about 77%.

• VISA applicants with no previous work experience have a certification rate of about 58%.

• VISA candidates with previous work experience almost show a 20% increase in getting certified/approved.

Bivariate Analysis

• For one unit increase in the region of employment the likelihood of the VISA applicant getting certified increases too.

• The Midwest region holds the highest VISA application certification rate at about 77%.

• Followed by:
  • South – 70%
  • Northeast – 62%
  • West – 61%
  • Island – 60%

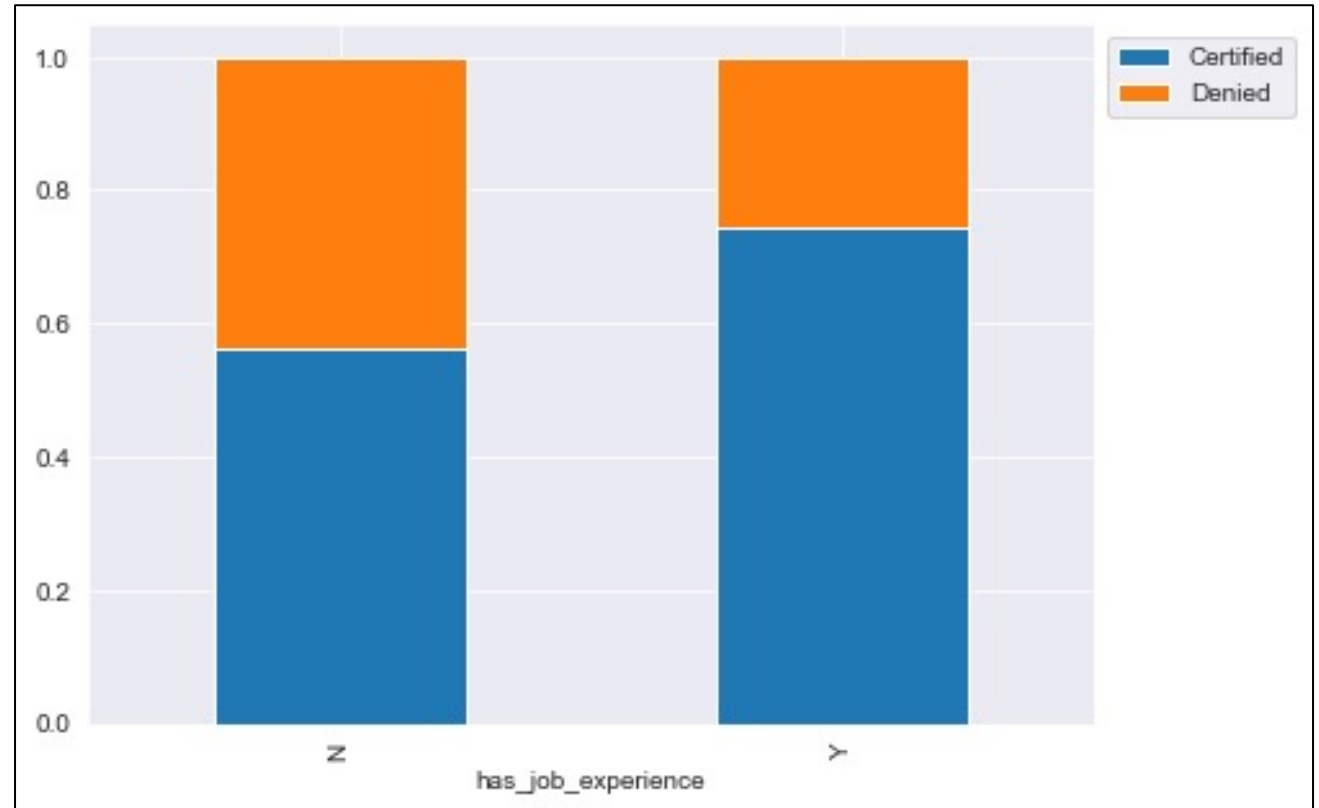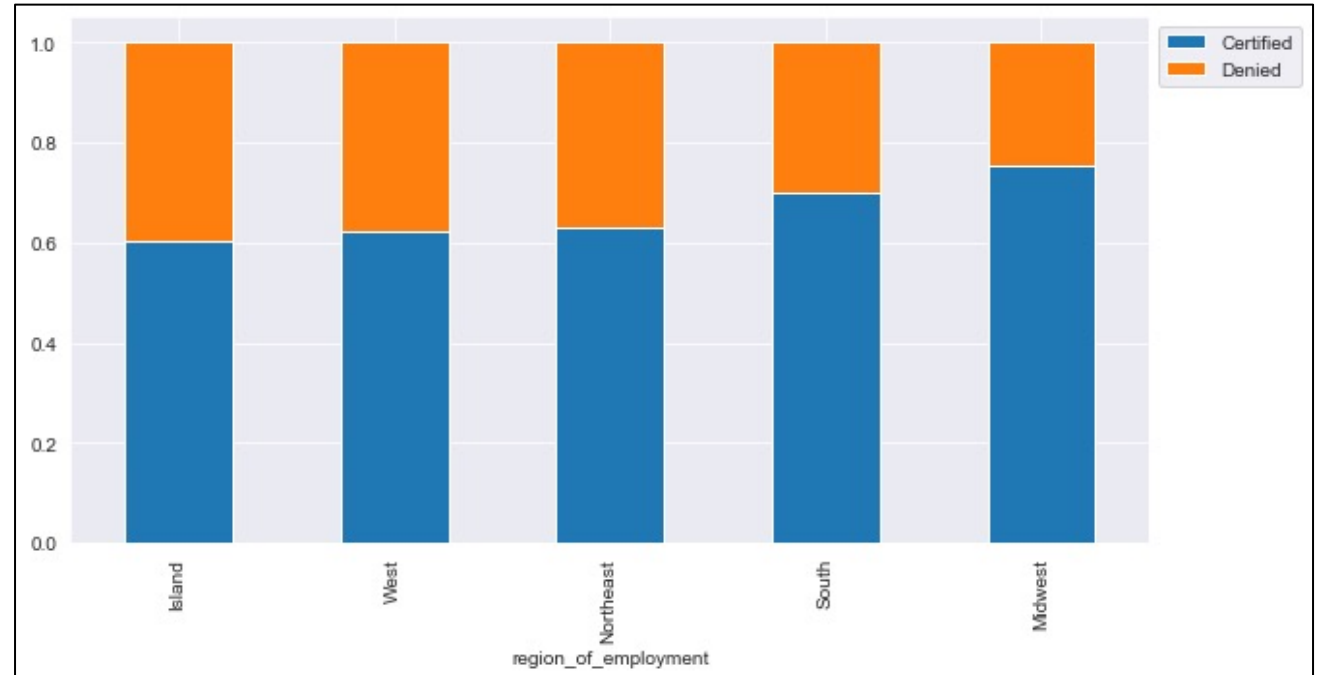# Bivariate Analysis

• For one unit increase in the unit of wage the likelihood of the VISA applicant getting certified increases too.

• A yearly unit of wage shows the highest likelihood of VISA application certification at about 70%.

• Followed by:
  • Weekly – 61%
  • Monthly – 61%
  • Hourly – 37%

Bivariate Analysis

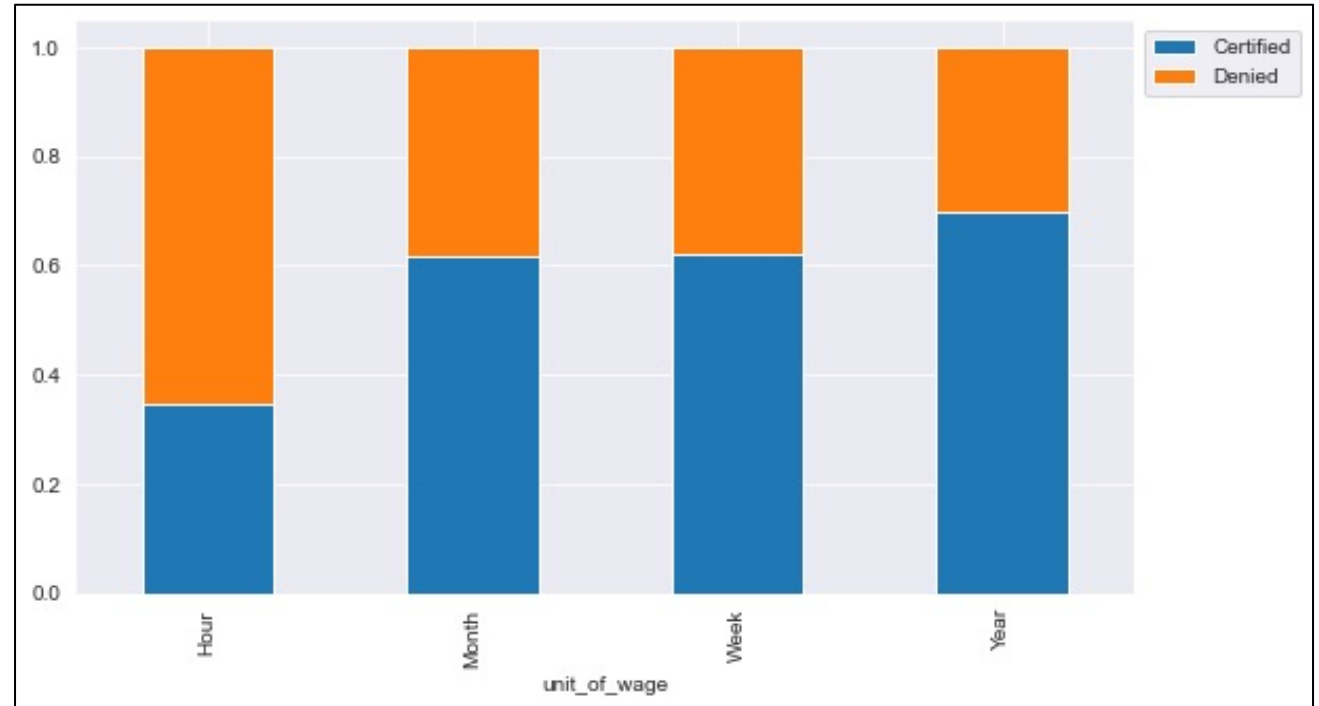• For one unit increase in continent the likelihood of the VISA applicant getting certified increases too.

• Certified VISA applicants are primarily from Europe with a certification rate of about 80%.

• Followed by:
  • Africa – 76%
  • Asia – 63%
  • Oceania – 62%
  • North America – 61%
  • South America – 58%
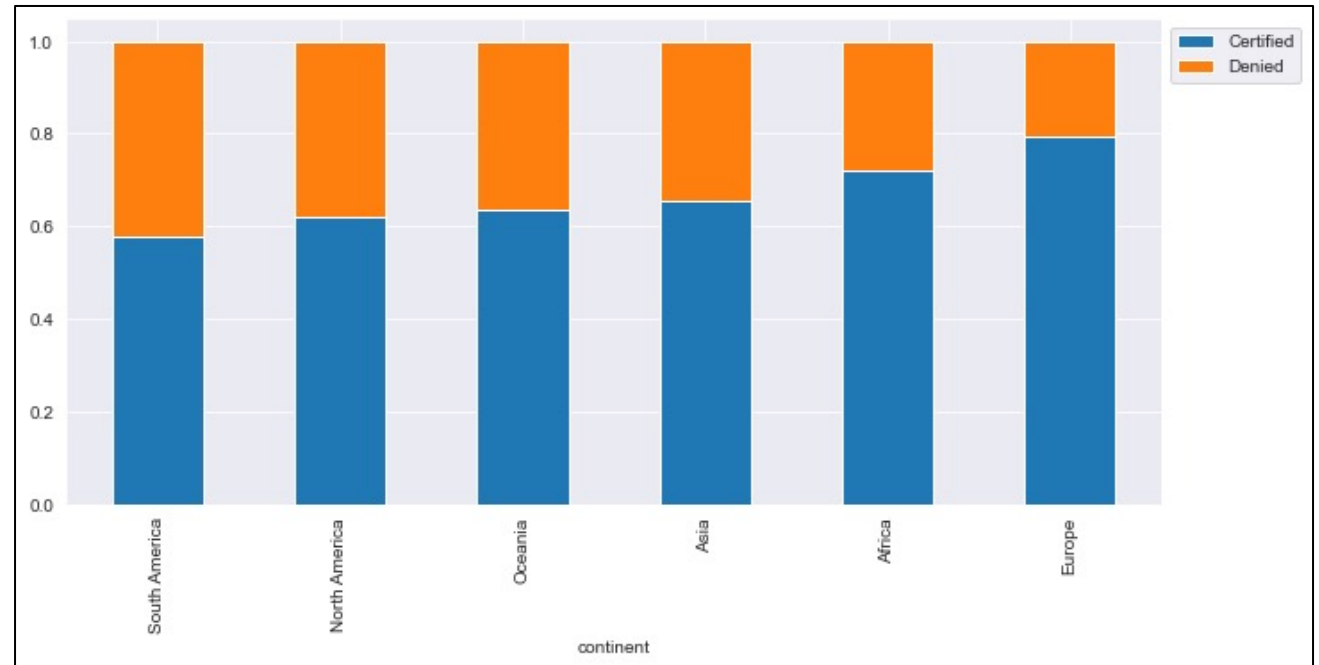
Bivariate Analysis

- VISA applicants with a higher prevailing wage are more likely to get certified.

- Inter-Quartile Range

- 25%: 38,375 USD

- 50%: 72,486 USD

- 75%: 108,879 USD
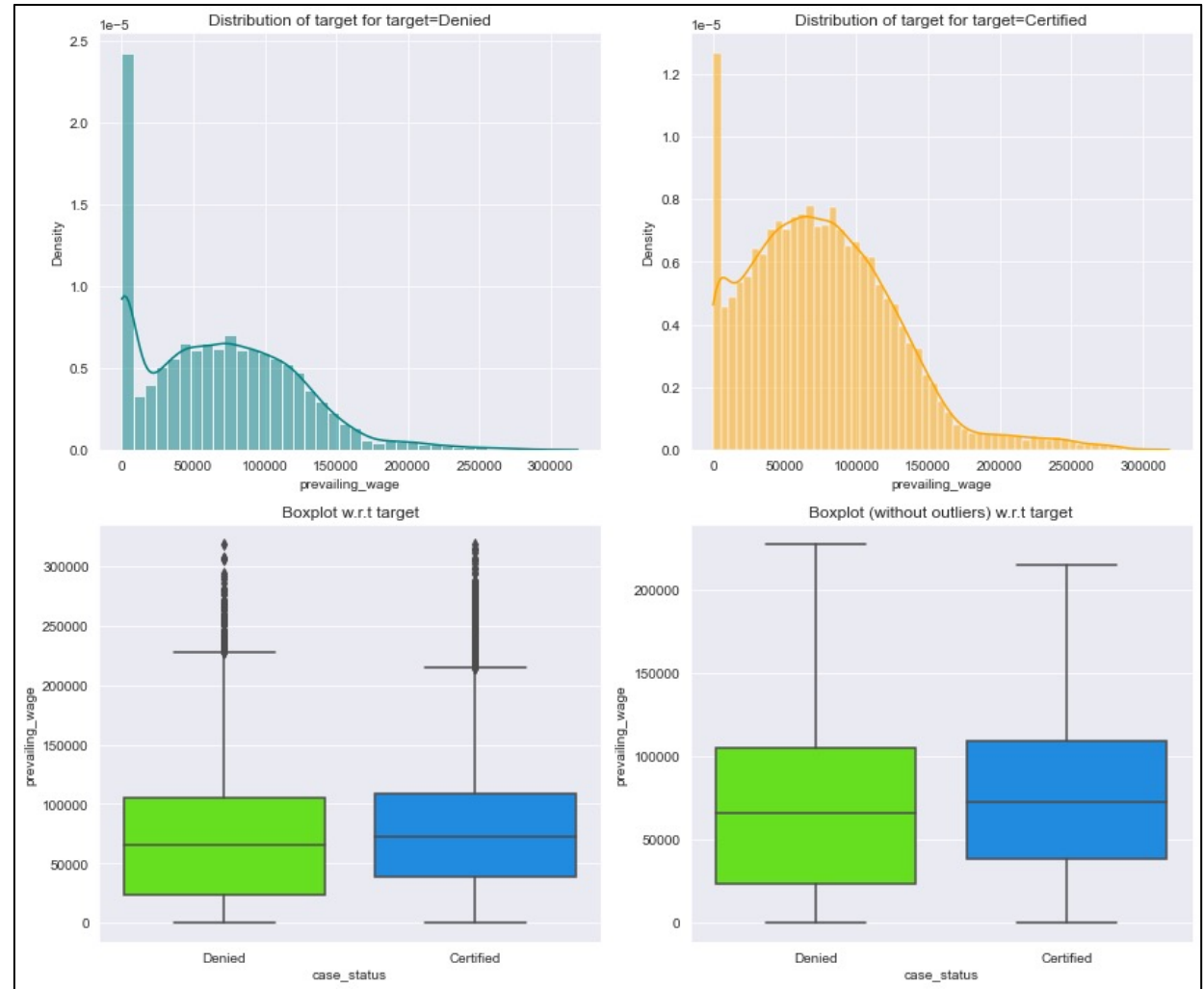
Model Performance Summary

- We want to predict whether a VISA applicant will be certified or denied based on the features provided to us.

- We used the binary version of "case_status" to build robust models. Also, we used the training data to check the performance on the testing data to understand the predictive power of our model.

- Which case is more important?
    - If a visa is certified when it had to be denied a wrong employee will get the job position while US citizens will miss the opportunity to work on that position.
    - If a visa is denied when it had to be certified the U.S. will lose a suitable human resource that can contribute to the economy.

EASY VISA
easy as simple as that

Model Performance Summary

	Accuracy	Recall	Precision	F1
Training Model Performances				
Decision Tree	0.712548	0.931923	0.720067	0.812411
Tuned Decision Tree	0.712548	0.931923	0.720067	0.812411
Bagging Classifier	0.985367	0.986317	0.991729	0.989016
Tuned Bagging Classifier	0.995851	0.999916	0.993909	0.996904
Random Forest	0.999944	0.999916	1.000000	0.999958
Tuned Random Forest	0.735815	0.923613	0.743195	0.823639
Adaboost Classifier	0.738058	0.887434	0.760411	0.819027
Tuned Adaboost Classifier	0.719163	0.781415	0.794690	0.787997
Gradient Boost Classifier	0.759419	0.882901	0.784106	0.830576
Tuned Gradient Boost Classifier	0.754597	0.879040	0.781026	0.827140
XGBoost Classifier	0.838248	0.931000	0.843166	0.884908
XGBoost Classifier Tuned	0.769399	0.885000	0.793542	0.836779
Stacking Classifier	0.755719	0.881978	0.780725	0.828268

Model Performance Summary

	Accuracy	Recall	Precision	F1
Testing Model Performances				
Decision Tree	0.706567	0.930852	0.715447	0.809058
Tuned Decision Tree	0.706567	0.930852	0.715447	0.809058
Bagging Classifier	0.692177	0.766112	0.771400	0.768747
Tuned Bagging Classifier	0.726975	0.897356	0.745605	0.814472
Random Forest	0.719126	0.830167	0.768032	0.797891
Tuned Random Forest	0.727499	0.923408	0.735873	0.819043
Adaboost Classifier	0.732993	0.885015	0.756530	0.815744
Tuned Adaboost Classifier	0.716641	0.781587	0.791510	0.786517
Gradient Boost Classifier	0.744636	0.873262	0.773555	0.820390
Tuned Gradient Boost Classifier	0.742543	0.872478	0.771790	0.819051
XGBoost Classifier	0.735348	0.863271	0.768842	0.813325
XGBoost Classifier Tuned	0.745290	0.871107	0.775279	0.820404
Stacking Classifier	0.743066	0.874241	0.771478	0.819651

Tuned Random Forest – Best Performing Model

- According to the tuned random forest model:

Testing Model Performance			
	Accuracy	Recall	Precision	F1
Tuned Random Forest	0.727499	0.923408	0.735873	0.819043

- A high recall score of 0.923 indicates that the model is returning most of all positive results.

- A high precision score of 0.736 indicates the model is returning accurate results.

- Therefore, the F1 score of 0.819 indicates the perfect balance between recall and precision.

Tuned Random Forest – Feature Importance

- The top three important features to predict the likelihood of a VISA applicant getting certified include:
  - High school / higher education
  - Job experience – Yes
  - Prevailing wage

Feature Importances
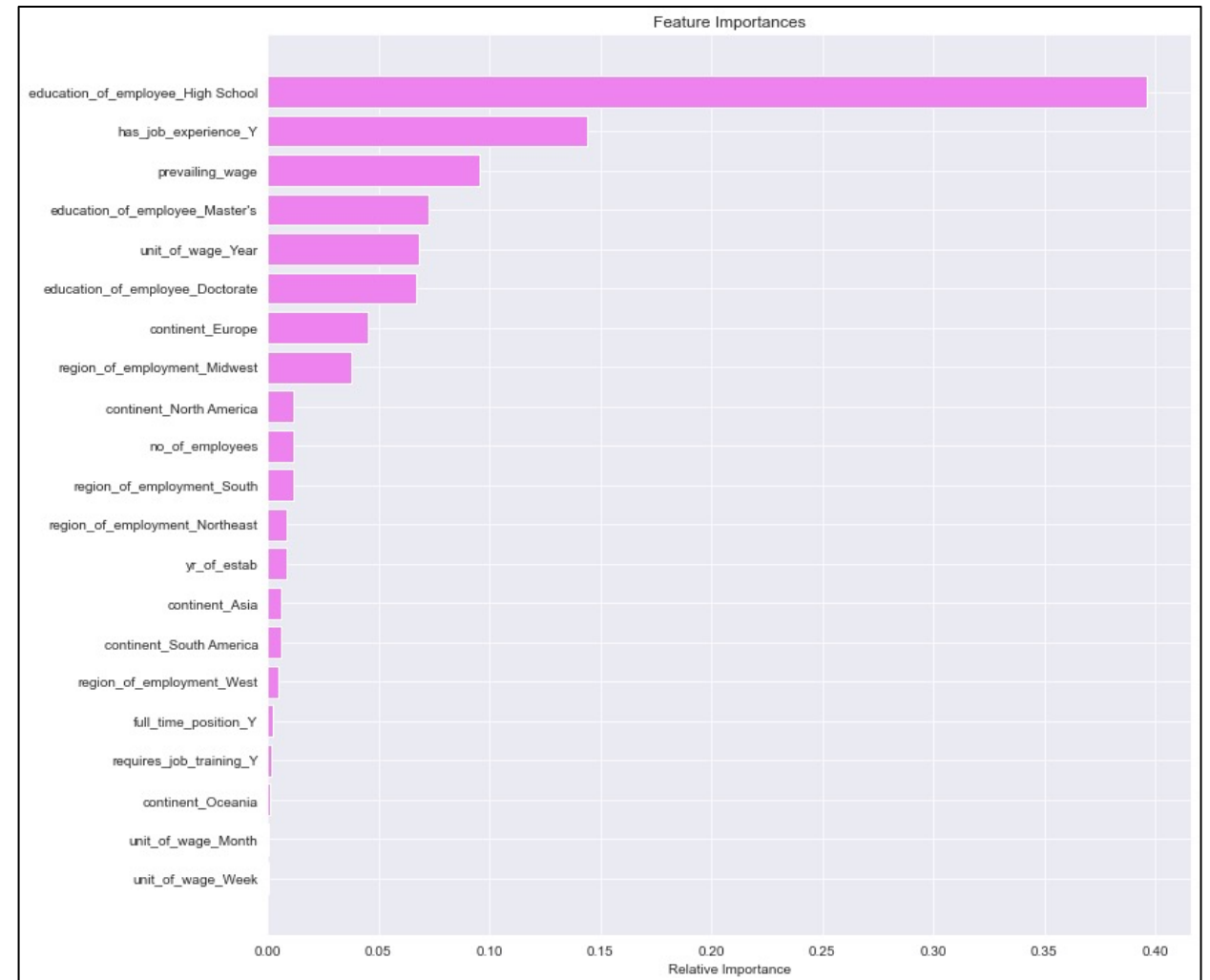
Final Thoughts on Model Performances

- The best performing classification model is the Tuned Random Forest.

- Furthermore, the model finds high school / higher education, job experience – yes, and prevailing wage as the top three features for VISA application certification.

IHG®
HOTELS & RESORTS

Business Insights

- Factors that drive visa certification: higher education, previous work experience, and their prevailing wage

- A visa applicant with a higher education is more likely to get certified.
  - Bachelor's – 26% increase
  - Master's – 17% increase
  - Doctorate – 8% increase

- Visa applicants with previous work experience almost show a 20% increase in getting certified.

- VISA applicants with a higher prevailing wage are more likely to get certified.

- Europe holds the highest visa certification rate at 80%.

Business Recommendations

Appropriate Profiles for VISA Certification vs Denial

- VISA Certification
  - A visa applicant with previous work experience and a higher education is more likely (80%) to get their visa certified.
  - A visa applicant who has a higher education and no previous work experience but whose intended prevailing wage is greater than 970 USD and of the European continent is more likely (75%) to get their visa certified.

- VISA Denial
  - A visa applicant who has a higher education but no previous work experience and whose intended prevailing wage is less than 970 USD demonstrates a higher likelihood (82%) of being denied a visa.
  - A visa applicant whose highest education level is high school demonstrates a higher likelihood (80%) of being denied a visa.

Profiles that require further evaluation

- Visa applicants who have a higher education but no previous work experience and whose intended prevailing wage is less than 970 USD:
  - Top unit of wage: Hourly rate
  - Hourly unit of wage does only show a 37% certification rate

- Visa applicants whose highest education level is high school:
  - Certification rate: ~20%
  - Top region of employment: South

  - Denial rate: ~80%
  - Top region of employment: Northeast

IHG®

HOTELS & RESORTS