EasyVisa Project

- **Business Presentation**

Contents

# Business Problem Overview and Solution Approach

- OFLC processes job certification applications for employers seeking to bring foreign workers into the United States and grants certifications in those cases where employers can demonstrate that there are not sufficient US workers available to perform the work at wages that meet or exceed the wage paid for the occupation in the area of intended employment.

- The process of reviewing every case is becoming a tedious task as the number of applicants is increasing every year.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 task at hand to analyze the data provided to facilitate the process of visa approvals by building a machine learning model and to 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 of 25480 employees and their employers.

- The data has 12 columns which includes information about employee's education, job experience, whether the employee requires training or not, continent,the number of employees in the company, year of establishment of the company, prevailing wage, etc.

- There were no missing and duplicate values in the dataset.

- All object data types were converted to category data types leading to a reduced memory space.

- The case_id column was dropped from the dataset since all the values in that column were unique.

- There were some negative values in the number of employees column which were replaced by their absolute values.

- There were quite a few outliers in the data, however, they were not treated since they were realistic values.
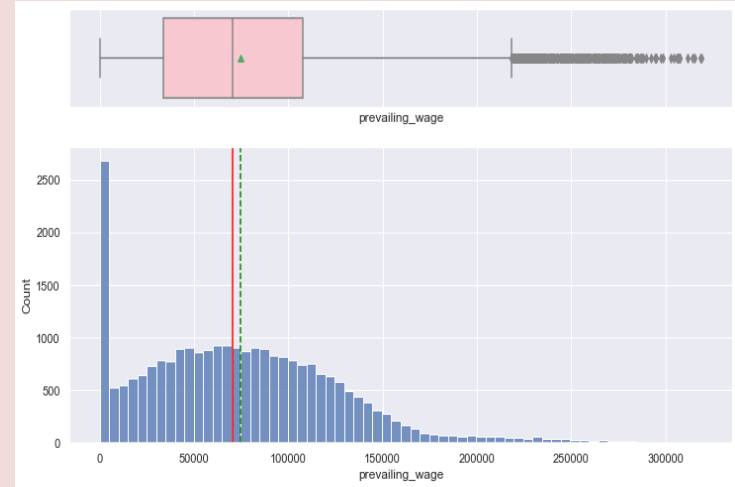
Exploratory Data Analysis (EDA)

No_of_employees
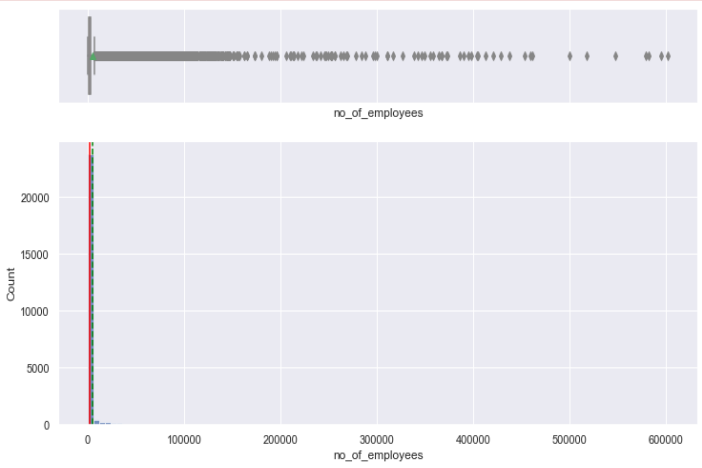
Prevailing_wage

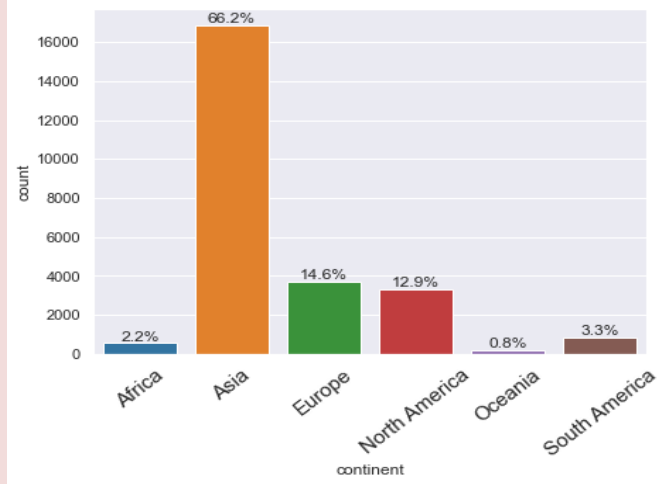- The distribution of the no_of_employees variable is highly right-skewed.
- There are outliers towards the right.

- The distribution of prevailing_wage variable is right-skewed.
- There are outliers towards the right.

Exploratory Data Analysis (EDA)
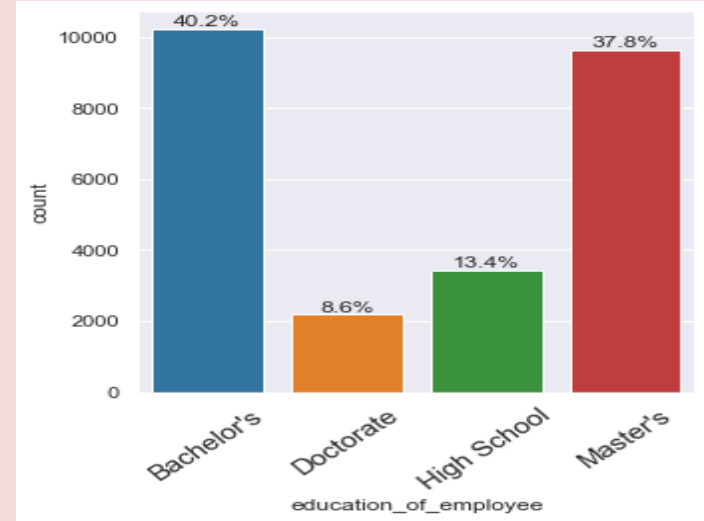
Continent

Education_of_Employee

- **66.2%** of candidates come from Asian continent followed by **14.6%** of candidates that come from the continent of Europe.
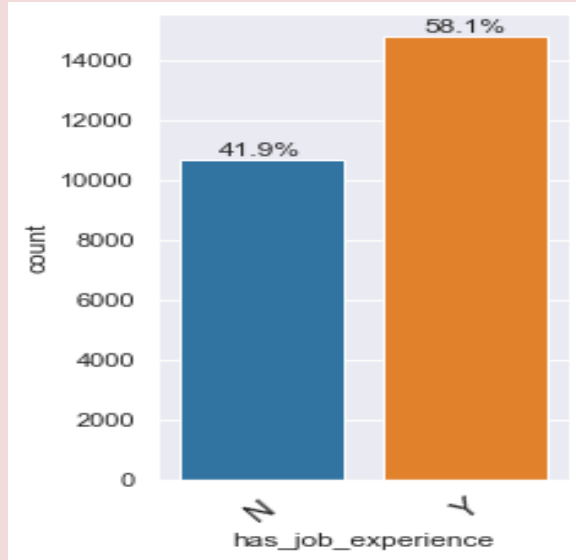- **0.8%** of candidates come from Oceania.

- **40.2%** of employees have a bachelor's degree, followed by **37.8%** having a master's degree.
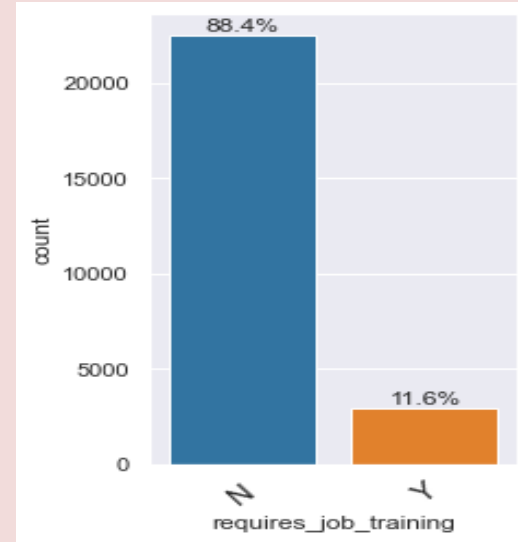- **8.6%** of employees have a doctorate degree.

Exploratory Data Analysis (EDA)

Has_job_experience

- **58.1%** of employees have a job experience.
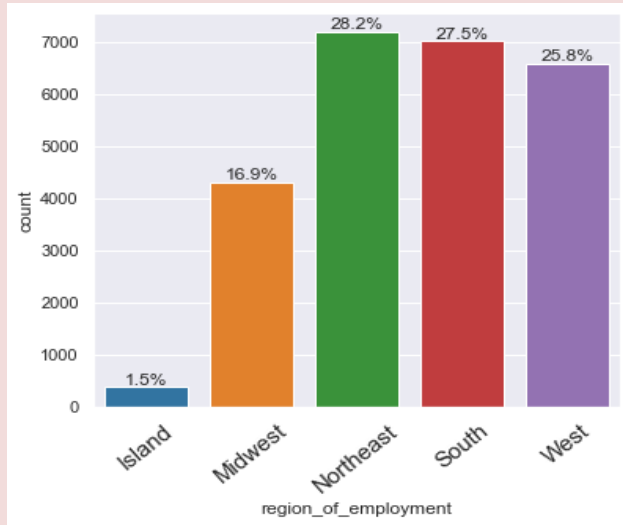- **41.9%** of employees **do not** have a job experience.

Requires_job_training

- **88.4%** of candidates **do not** require a job training
- **11.6%** of candidates require a job training.

Exploratory Data Analysis (EDA)

Region_of_employment

Full_time_position

- **Northeast, South,** and **West regions** have **28.2%,27.5%,25. 8%** of candidates respectfully.
- The Island region have only **1.5%** of the candidates.
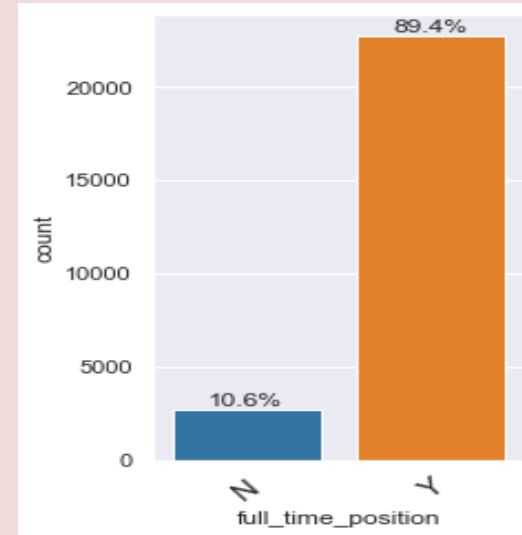
- **89.4%** of candidates go in for the full time work position.
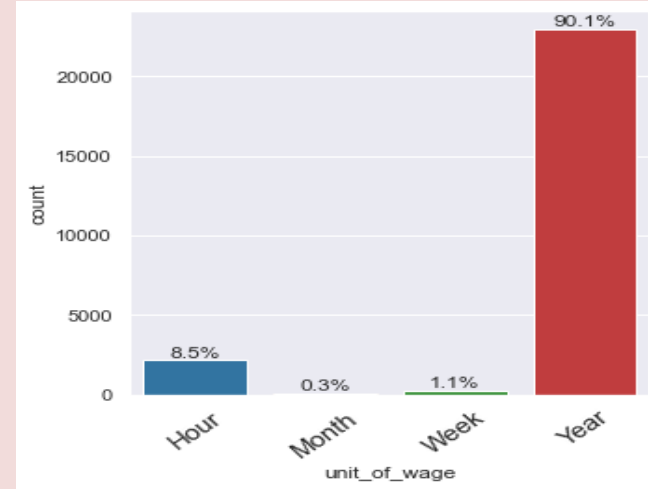- **10.6%** of candidates go in for the part time work position.

Exploratory Data Analysis (EDA)

Case_status
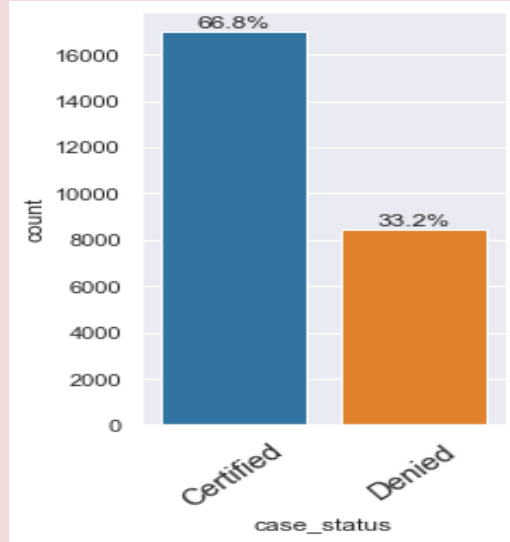
- **66.8%** of visas was certified.
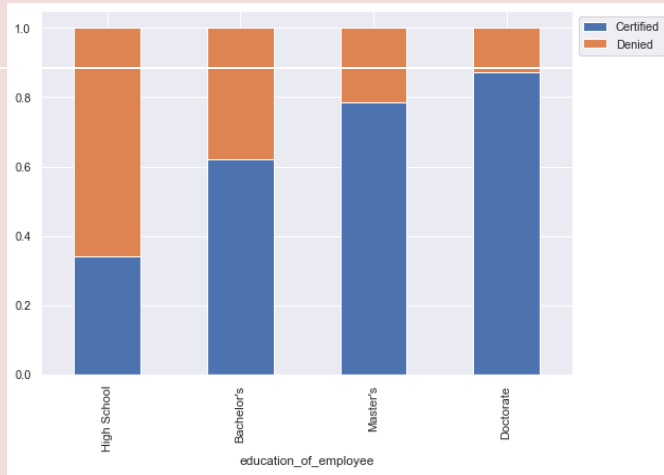- **33.2%** of visas denied.

Unit_of_wage

- **90.1%** is yearly unit wage.This makes sense as large values of prevailing wage was observed.
- **8.5% ,1.1%, and 0.3%** are **hourly,weekly and monthly** unit wage respectfully**.**

Exploratory Data Analysis(EDA)

Case_status vs Education_of_employee

- **Applicants with high school education are more likely to be denied a visa.**

- Most applicants (**about 60%**) with Bachelor's degree were certified a visa.

- Most applicants (**about 80%**) with Master's degree were certified a visa.

- **Applicants with a doctorate education are more likely to be certified visa. About 90%** of the visa applications got certified for the applicants with Doctorate degree..

- **Applicants from Europe and Africa more likely of getting certified for a visa.** Around 80% of applicants from Europe and about 75% of applicants from Africa were certified a visa.

- **Asia has the third-highest percentage (Around 60%) of visa certification and has the highest number of applications.**
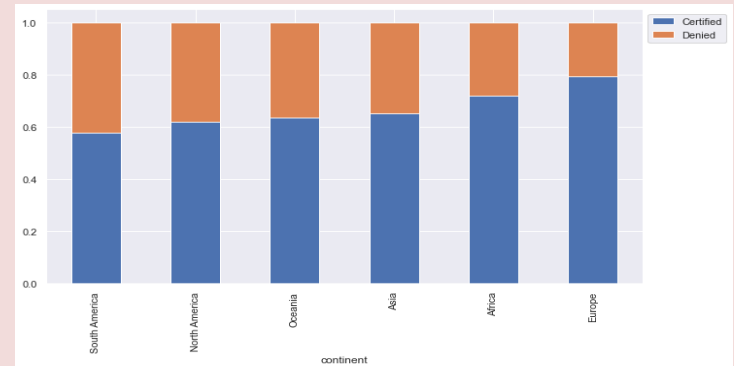
Case_status vs Continent

Exploratory Data Analysis(EDA)

- **Employees with work experience are more likely to be certified a visa** as compared to employees without work experience.
- Around 80% of the applications were certified for the applicants who have some job experience as compared to the applicants who do not have any job experience.
- Applicants without job experiences saw only about 60% of the visa applications getting certified.

Case_status vs Region_of_employment
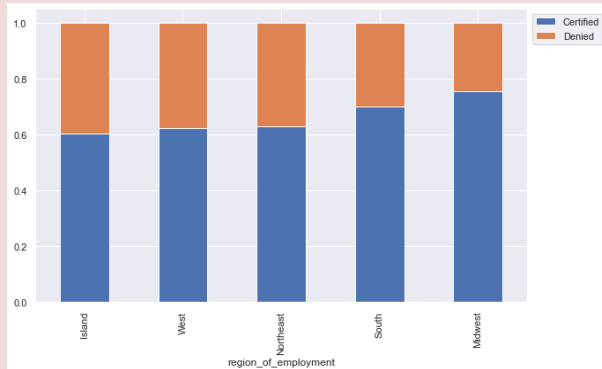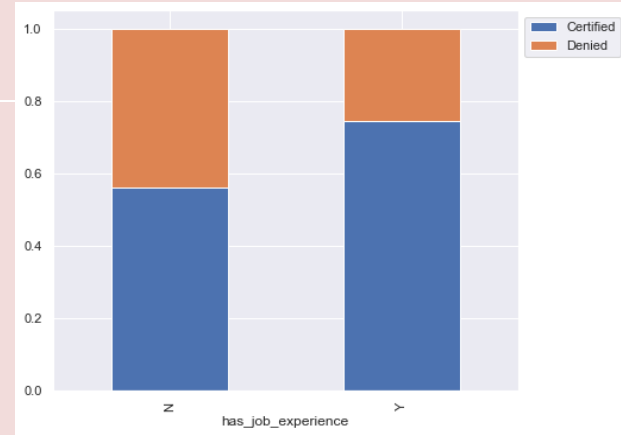
- Applicants intending to be employed at Midwest region are more likely to be certified a visa followed by the South region.

- There is no significant difference of Island, West and Northeast regions with respect to visa status.

Exploratory Data Analysis(EDA)

Case_status vs Unit_of_wage
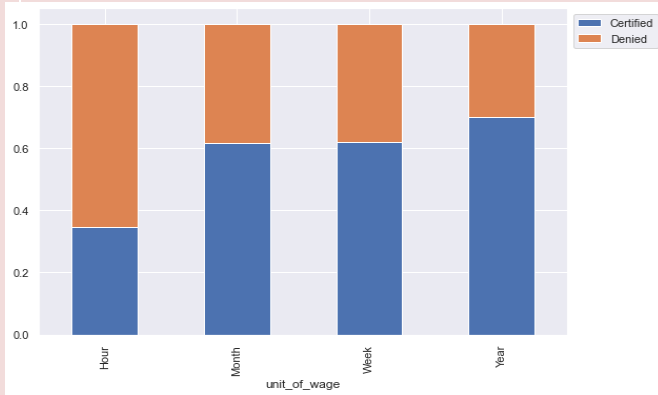
- **The yearly pay unit is most likely to be certified a visa as compared to the other pay units.** About 75% of applications of applicants who have a yearly unit wage were certified for visa.

- **The hourly pay unit is more likely to be denied a visa** .Only about 35% of the applications of applicants who have an hourly unit of wage were certified a visa.

- The monthly and weekly pay units are almost at same level (around 60%) of being certified for a visa.

- Candidates/applicants going for full time or part-time work positions have almost the same chance (around 70%) of getting certified for a visa.
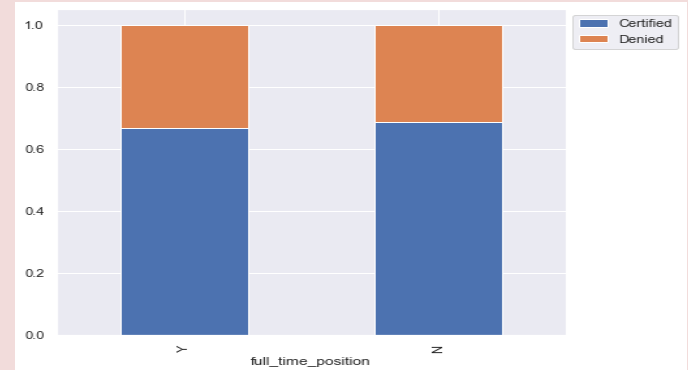
Case_status vs Full_time_position¶

Exploratory Data Analysis(EDA)

Case_status vs Prevailing_wage

Case_status vs No_of_employees
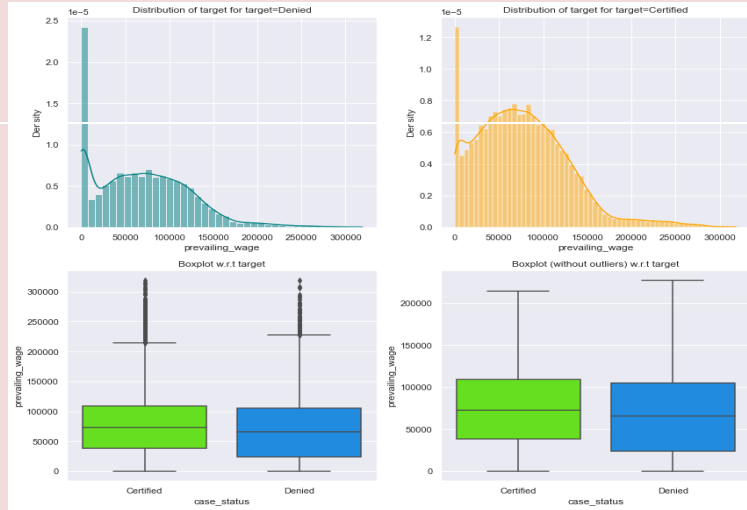
- The average and median prevailing wage of employees that were certified a visa is slightly greater than the average and median prevailing wage of those that were denied a visa.
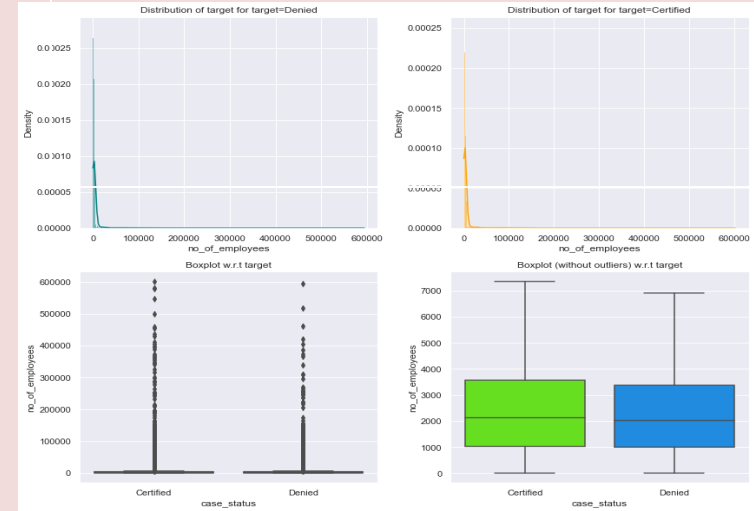
- Employer's company with a larger number of employees has a greater chance for its applicants to be certified a visa.

Exploratory Data Analysis(EDA)

Region_of_employment vs Prevailing wage
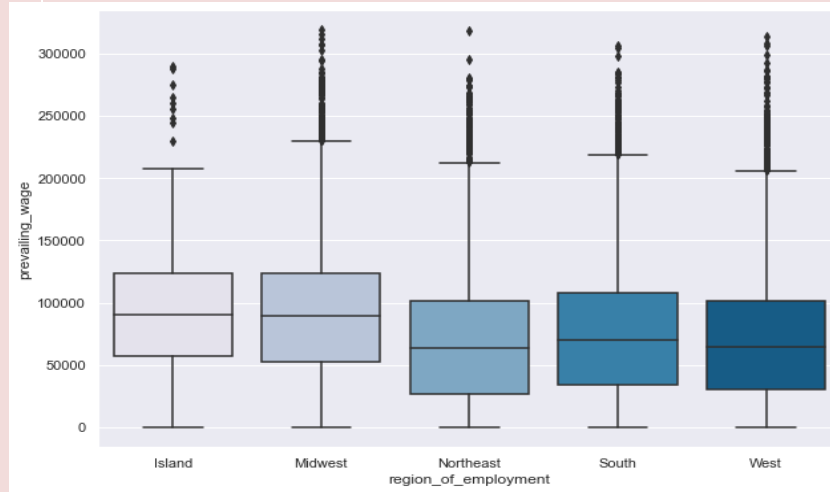
- **Island and Midwest regions of the US have slighter higher average and median prevailing wage than other regions.**
- The distribution of prevailing wage is quite similar across West,**Northeast,and South regions.But Northeast region have the lowest average prevailing wage.**

Model Performance Summary

- We want to predict whether the visa application will get certified or not using the information provided.

- We will use F1 Score as the metric for evaluation of the model because

  - 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 of the country.
  - F1 score will help us to minimize both false positives and false negatives.

- We will use **balanced class** weights so that model focuses equally on both classes.
- We will build different models - DecisionTreeClassifier, RandomForestClassifier, BaggingClassifier, AdaBoostClassifier, GradientBoostingClassifier, XGBClassifier, and StackingClassifier

- **We will also perform hyperparameter tuning for these models and evaluate their performance using different metrics and confusion matrix.**

Model Performance Summary – Training Performance

| Model/ Metric	Decision Tree	Tuned Decision Tree	Random Forest	Random Forest Tuned	Bagging Classifier	Bagging Classifier Tuned	Adaboost Classifier	Adabosst Classifier Tuned	Gradient Boost Classifier	Gradient Boost Classifier Tuned	XGBoost Classifier	XGBoost Classifier Tuned	Stacking Classifier
Accuracy	1	0.712547656	1	0.760372281	0.985254541	0.996187486	0.73822606	0.71899529	0.758802422	0.764016596	0.838753084	0.765474322	0.762222471
Recall	1	0.931923109	1	0.794762025	0.987072946	0.999916058	0.88718207	0.781247377	0.883740452	0.882649207	0.931419458	0.881641904	0.878703937
Precision	1	0.720067454	1	0.838098661	0.990815639	0.994406879	0.760688067	0.794587211	0.783042023	0.789058982	0.843481566	0.791126845	0.78920386
F1	1	0.812410816	1	0.815855235	0.988940751	0.997153859	0.819080095	0.787860831	0.830349397	0.833234288	0.885272026	0.833935448	0.831552608

Model Performance Summary – Testing Performance

Model/ Metrics	Decision Tree	Tuned Decision Tree	Random Forest Estimator	Random Forest Tuned	Bagging Classifier	Bagging Estimator Tuned	Adaboost Classifier	Adabosst Classifier Tuned	Gradient Boost Classifier	Gradient Boost Classifier Tuned	XGBoost Classifier	XGBoost Classifier Tuned	Stacking Classifier
Accuracy	0.65489272 6	0.70656724 2	0.72736787	0.71899529	0.69453165 9	0.72475143 9	0.73430141 3	0.71650968 1	0.74476713 8	0.74345892 2	0.73325484	0.74515960 2	0.73979591 8
Recall	0.7475024 49	0.93085210 6	0.84720861 9	0.76376101 9	0.77668952	0.89500489 7	0.88501469 1	0.78139079 3	0.87600391 8	0.87130264 4	0.86072478	0.86953966 7	0.86464250 7
Precision	0.7388189 74	0.71544715 4	0.76834251 2	0.80541210 5	0.76841085 3	0.74450057	0.75779939 6	0.79146825 4	0.77236614 9	0.77329624 5	0.76791331 7	0.77591330 2	0.77275910 4
F1	0.7431353 46	0.80905763 2	0.80585056 8	0.78403378 2	0.77252800 8	0.81284469	0.81648143 1	0.78639724	0.82092703 1	0.81937920 2	0.81167451 7	0.82006281 2	0.81612277

Feature Importance

- **The top 3 important features** to look for while certifying a visa are :-

  - **Education_of_employee _highschool**

  - **Job experience(has_job_exp erience_Y)**

  - **Prevailing Wage.**

Feature Importances
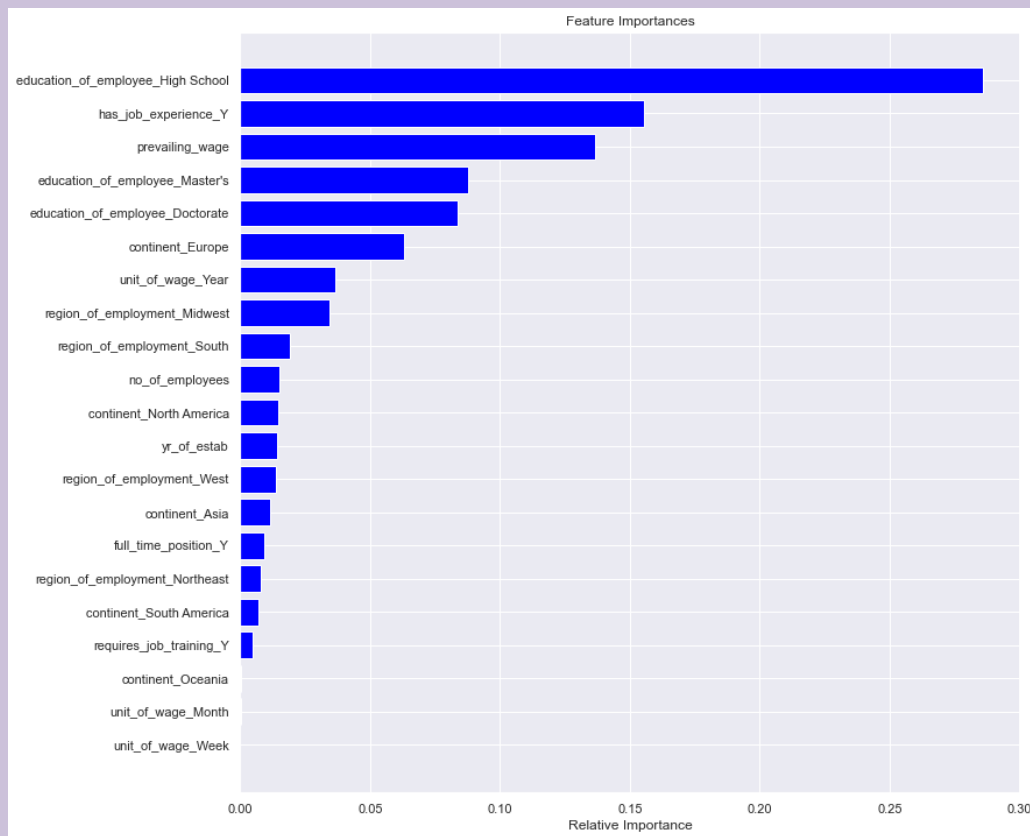
Conclusion

Based on the analysis done, the suitable profile of applicants for whom the visa status can be certified :

Primary information to look at:

- **Education Level** - Should have an educational certificate above high school - at least a Bachelor's degree. Those with Master's and doctorate have a higher chance.
- **Job Experience** - Should have some job experience.
- **Prevailing Wage** - Should have a considerable high yearly prevailing wage - that is prevailing wage from 80000 USD and above.

Secondary information to look at:

- **Continent** -
  Ideally the nationality and ethnicity of an applicant shouldn't matter to work in a country but it has been observed that applicants from Europe have higher chances of getting a certified visa and
  applicants South America, North America, and Oceania have higher chances of visa applications getting denied.
- **Region of Employement** - Applicants intending to be employed in the Midwest region and the South region of US have higher chances of being certified a visa.
- **Applicants that go in for full time work positions.**

- In order to facilitate the process of visa approvals, EasyVisa should target more applicants that have a master's or doctorate degree with work experience and are applying for full time work positions as those with high school certificates have very high chances of being denied a visa.
- EasyVisa should also target employers seeking employees to work in the Midwest and South Regions of the US as these regions were shown to have higher chances of their applicants to be certified for a visa.
- EasyVisa should further investigate applications for the Island region though having a considerable high prevailing wage had higher chances of being denied a visa.
- EasyVisa should also target applicants coming from Europe and Africa as these profiles have a higher chance to be certified for a visa.
- Applicants coming from South America should be thoroughly and keenly reviewed and made sure they have very strong profile(that is have at least master's degree, many years of work experience, going in for a full time position with a quite high yearly prevailing wage,etc).For applicants from this continent have a higher chance of being denied a visa.
- Applications from companies with larger number of employees should be targeted as their applicants are more likely to be certified for a visa.

Business Recommendations :