EasyVisa Presentation

Business Problem Overview and Solution Approach

Context

Business communities in the United States are facing high demand for human resources, but one of the constant challenges is identifying and attracting the right talent, which is perhaps the most important element in remaining competitive. Companies in the United States look for hard-working, talented, and qualified individuals both locally as well as abroad.

The Immigration and Nationality Act (INA) of the US permits foreign workers to come to the United States to work on either a temporary or permanent basis. The act also protects US workers against adverse impacts on their wages or working conditions by ensuring US employers' compliance with statutory requirements when they hire foreign workers to fill workforce shortages. The immigration programs are administered by the Office of Foreign Labor Certification (OFLC).

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.

Business Objective

Objective

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 the firm EasyVisa for data-driven solutions. You as a data scientist at EasyVisa have to 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 Description

Data Description

The data contains the different attributes of the employee and the employer. The detailed data dictionary is given below.

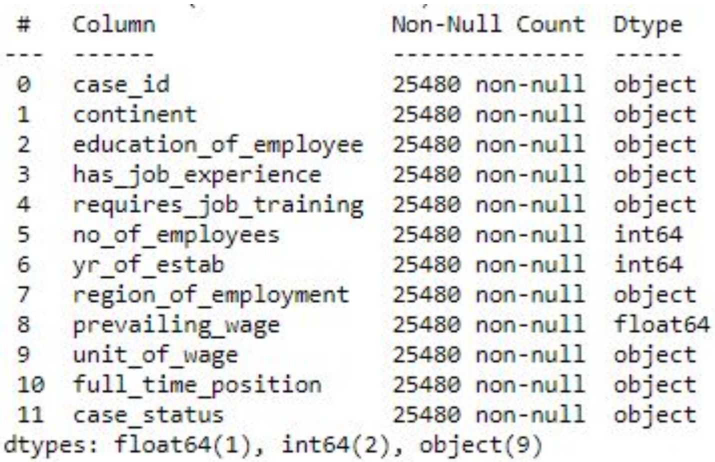
- case_id: ID of each visa application
- continent: Information of continent the employee
- education_of_employee: Information of education of the employee
- has_job_experience: Does the employee has any job experience? Y= Yes; N = No
- requires_job_training: Does the employee require any job training? Y = Yes; N = No
- no_of_employees: Number of employees in the employer's company
- yr_of_estab: Year in which the employer's company was established
- region_of_employment: Information of foreign worker's intended region of employment in the US.
- prevailing_wage: Average wage paid to similarly employed workers in a specific occupation in the area of intended employment. The purpose of the prevailing wage is to ensure that the foreign worker is not underpaid compared to other workers offering the same or similar service in the same area of employment.
- unit_of_wage: Unit of prevailing wage. Values include Hourly, Weekly, Monthly, and Yearly.
- full_time_position: Is the position of work full-time? Y = Full-Time Position; N = Part-Time Position
- case_status: Flag indicating if the Visa was certified or denied

Data Description Ctd.

- Data has no missing values

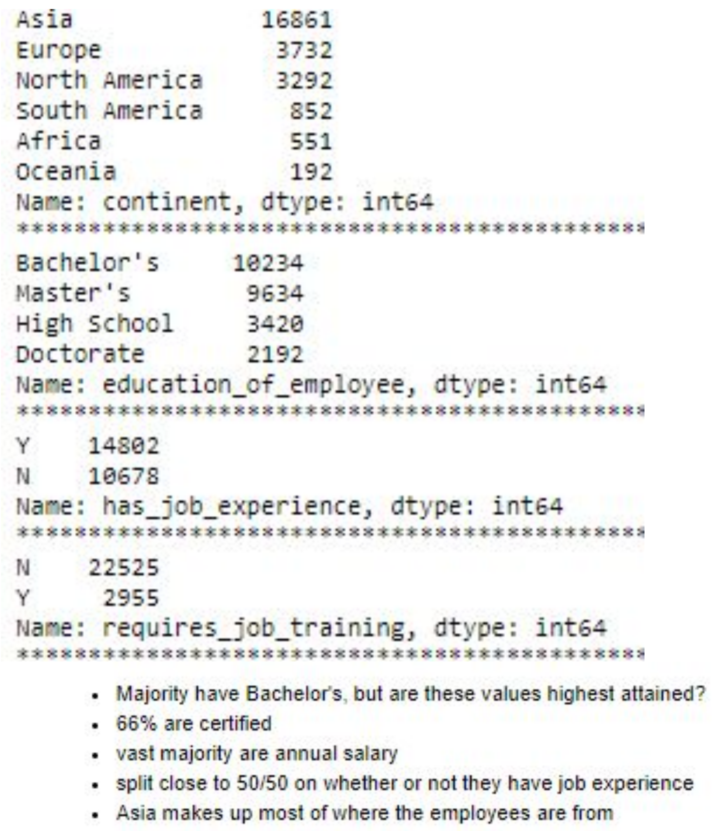
- There are 25480 observations and 12 rows in the dataset.

- Data is mostly in object/string format with a couple factors in numeric format



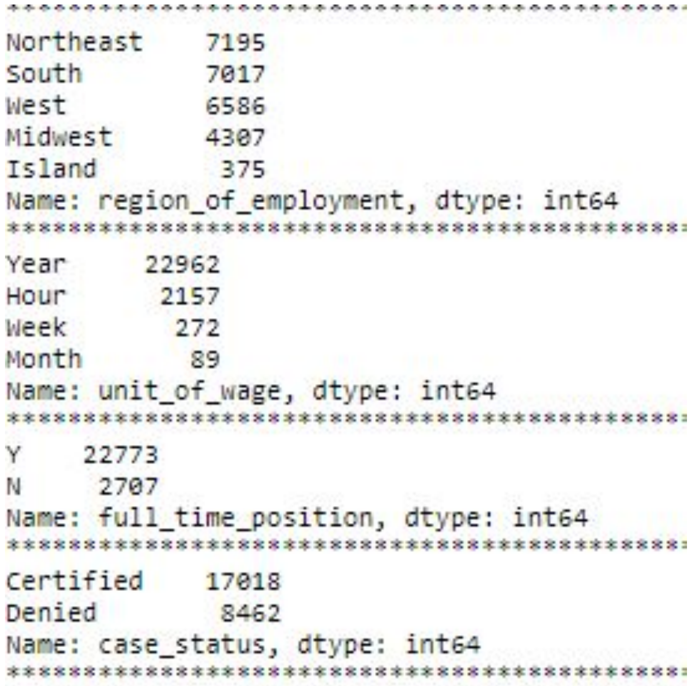
```
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   case_id                25480 non-null  object
 1   continent              25480 non-null  object
 2   education_of_employee  25480 non-null  object
 3   has_job_experience     25480 non-null  object
 4   requires_job_training  25480 non-null  object
 5   no_of_employees        25480 non-null  int64
 6   yr_of_estab            25480 non-null  int64
 7   region_of_employment   25480 non-null  object
 8   prevailing_wage        25480 non-null  float64
 9   unit_of_wage           25480 non-null  object
 10  full_time_position     25480 non-null  object
 11  case_status            25480 non-null  object
dtypes: float64(1), int64(2), object(9)
```

EDA: Univariate Analysis: Categorical Data



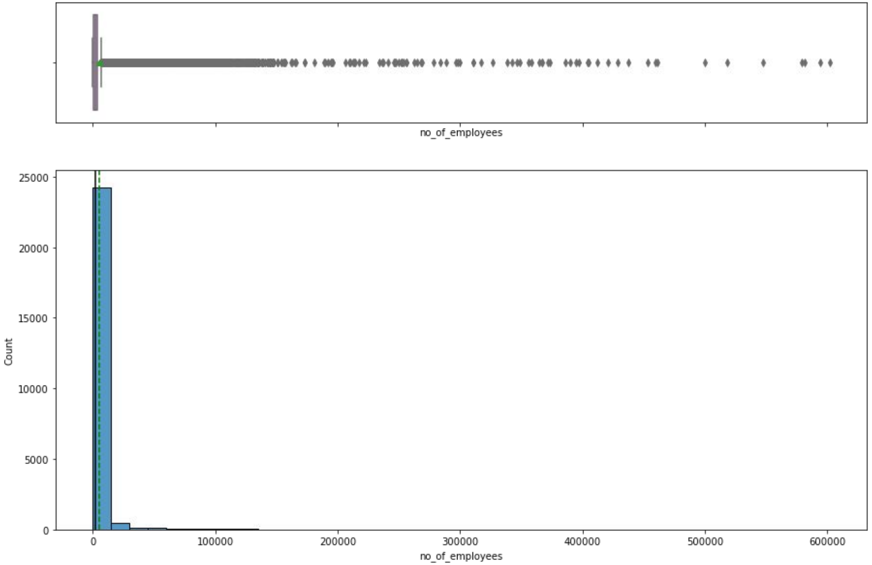
```
Asia             16861
Europe            3732
North America     3292
South America      852
Africa             551
Oceania            192
Name: continent, dtype: int64
*********************************************
Bachelor's     10234
Master's        9634
High School     3420
Doctorate       2192
Name: education_of_employee, dtype: int64
*********************************************
Y    14802
N    10678
Name: has_job_experience, dtype: int64
*********************************************
N    22525
Y     2955
Name: requires_job_training, dtype: int64
*********************************************
```

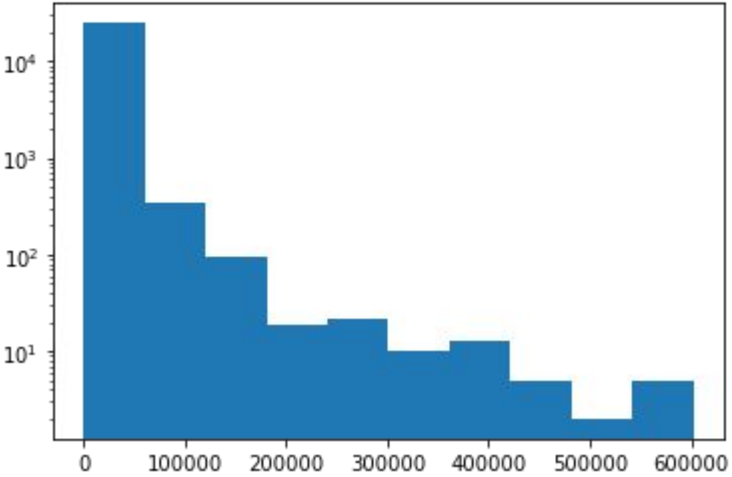
- Majority have Bachelor's, but are these values highest attained?
- 66% are certified
- vast majority are annual salary
- split close to 50/50 on whether or not they have job experience
- Asia makes up most of where the employees are from



```
************************************************
Northeast    7195
South        7017
West         6586
Midwest      4307
Island        375
Name: region_of_employment, dtype: int64
************************************************
Year     22962
Hour      2157
Week       272
Month       89
Name: unit_of_wage, dtype: int64
************************************************
Y    22773
N     2707
Name: full_time_position, dtype: int64
************************************************
Certified    17018
Denied        8462
Name: case_status, dtype: int64
************************************************
```

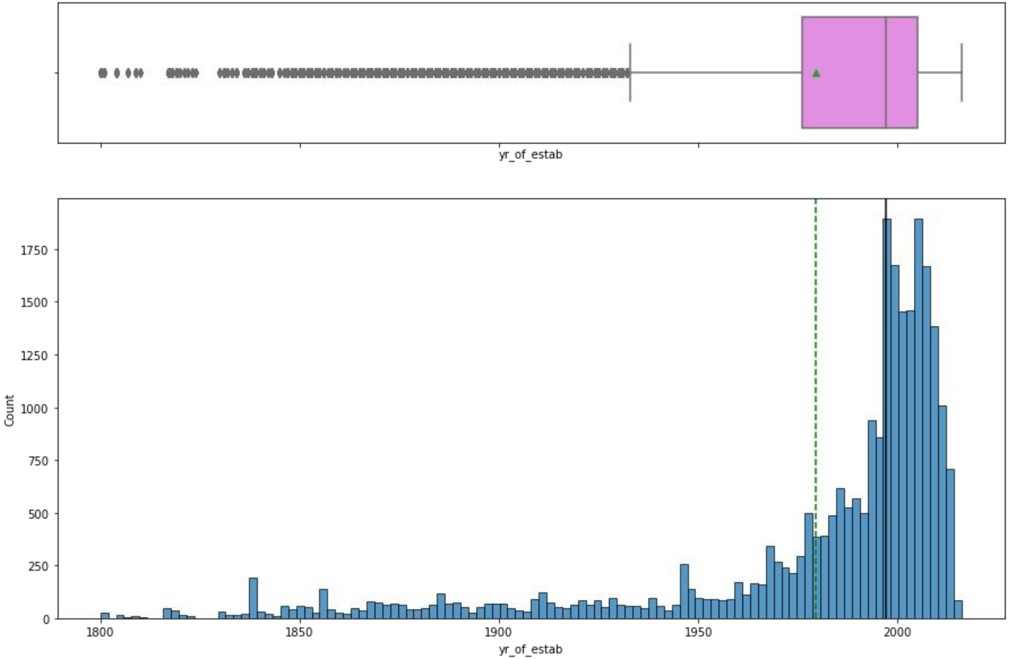
EDA: Univariate Analysis: Number of Employees



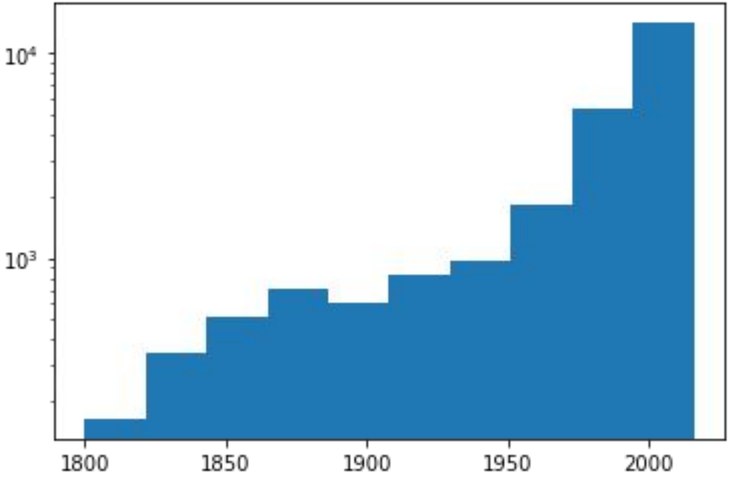


- Very heavily right skewed, with average company size of ~4,000
- 33 values of negative employee numbers, which will be removed from the data set as there can not be negative employee numbers.
-

EDA: Univariate Analysis: Year Company was Established

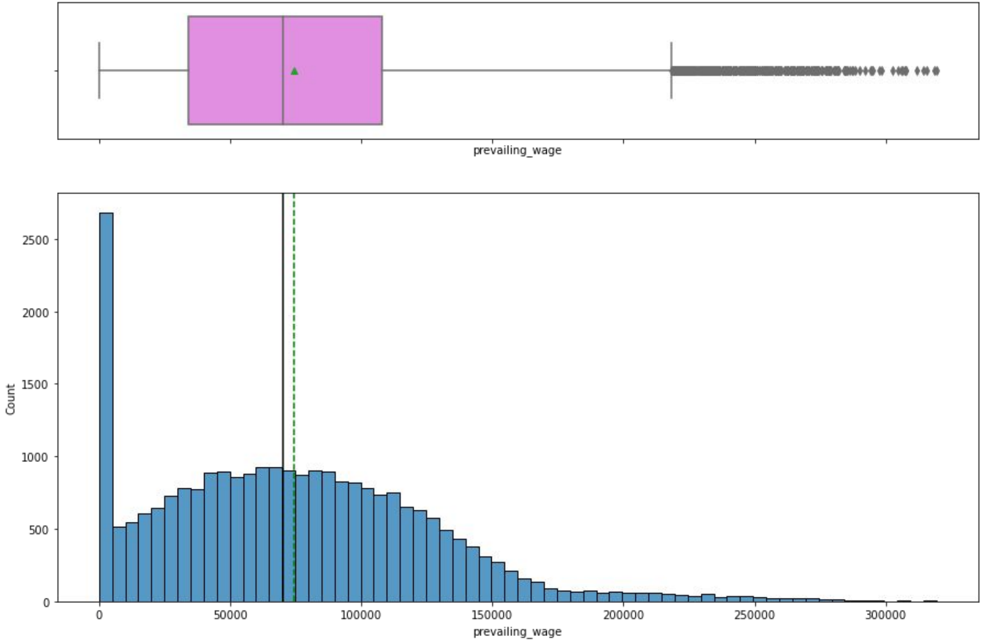


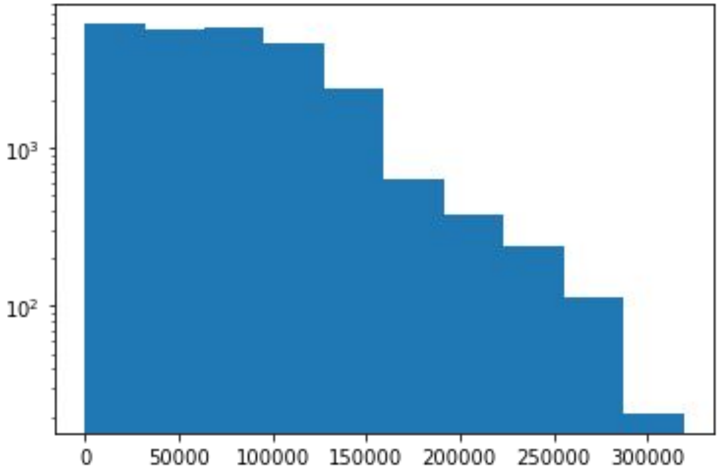
Log Scale

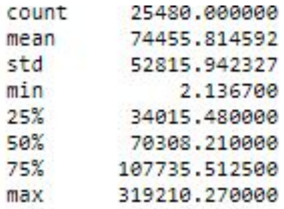


- Another heavily skewed graph, this time to the left.
- Most of the companies were established in the 1990s and beyond, with a large number in 2000's and beyond, possible tech/housing/MFG boom.

EDA: Univariate Analysis: Prevailing Wage



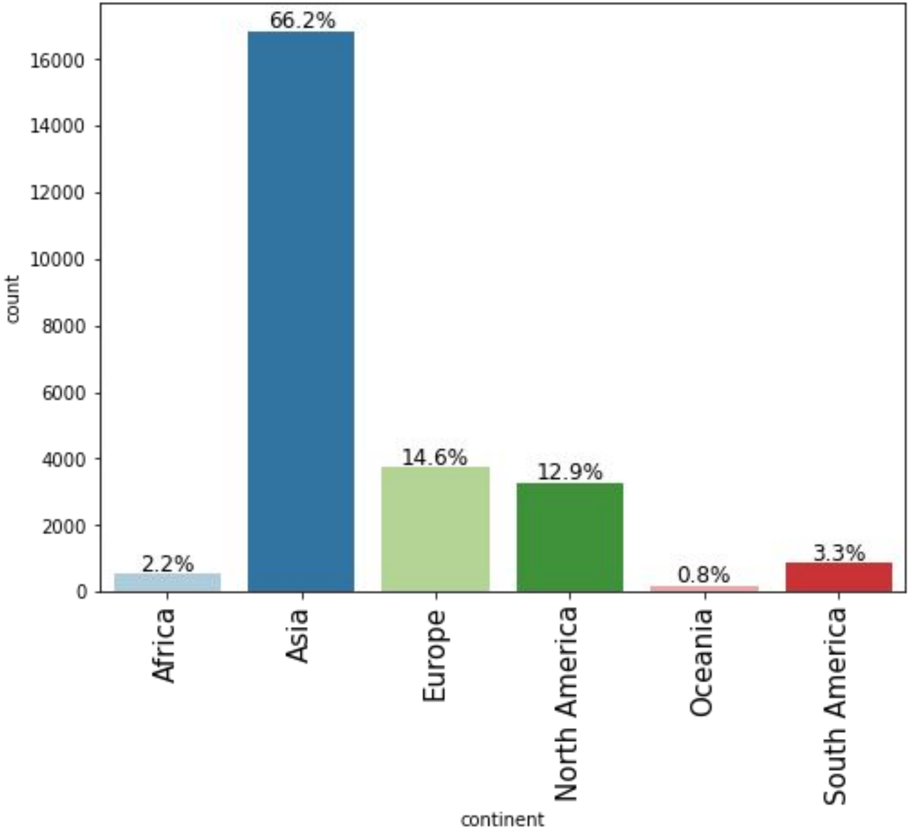




```
count      25480.000000
mean       74455.814592
std        52815.942327
min            2.136700
25%        34015.480000
50%        70308.210000
75%       107735.512500
max       319210.270000
```

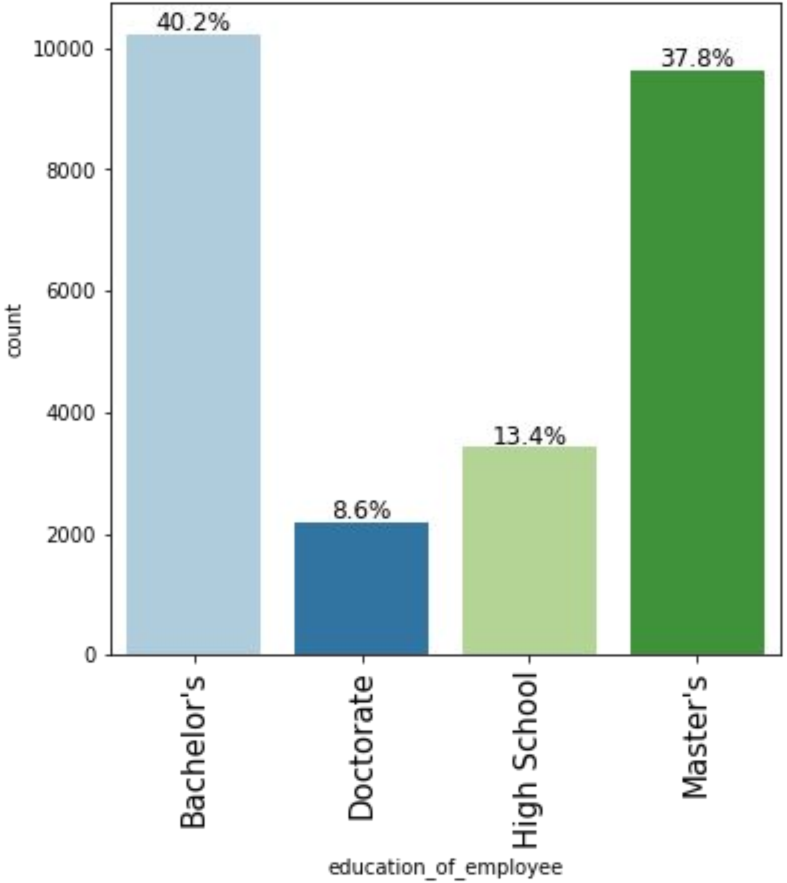
- Wage is heavily right skewed with wages ranging from $2-319,210, with an average of 75k.

EDA: Univariate Analysis: Continent



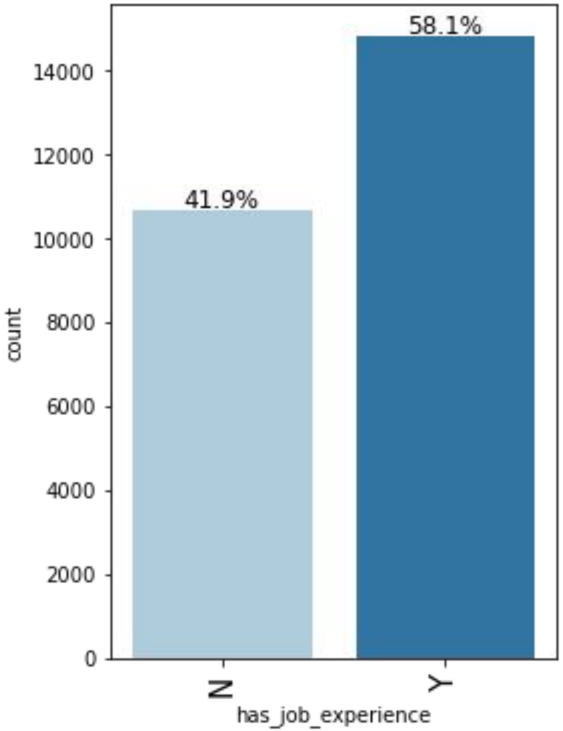
- 66.2% of visa applicants are from Asia
- Oceania has the lowest number of applicants

EDA: Univariate Analysis: Education



- Over 75% of the applicants have secondary education degrees

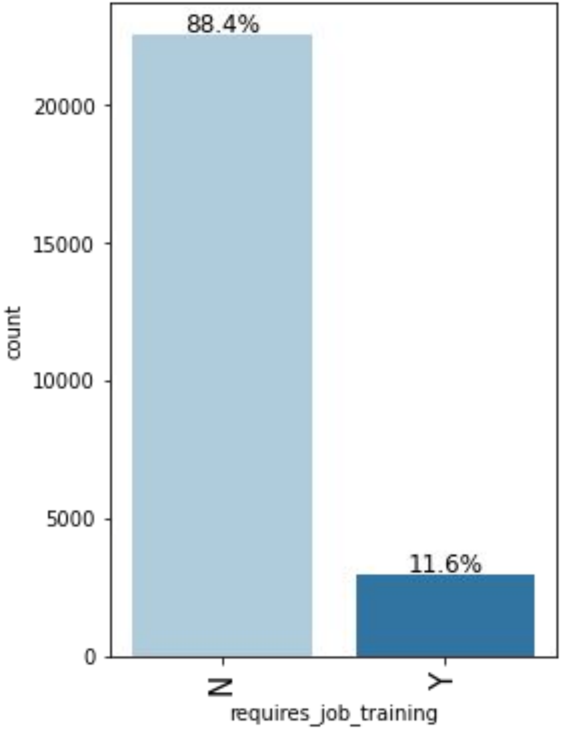
EDA: Univariate Analysis: Job Experience



- Over half of the applicants have job experience (58.1%)

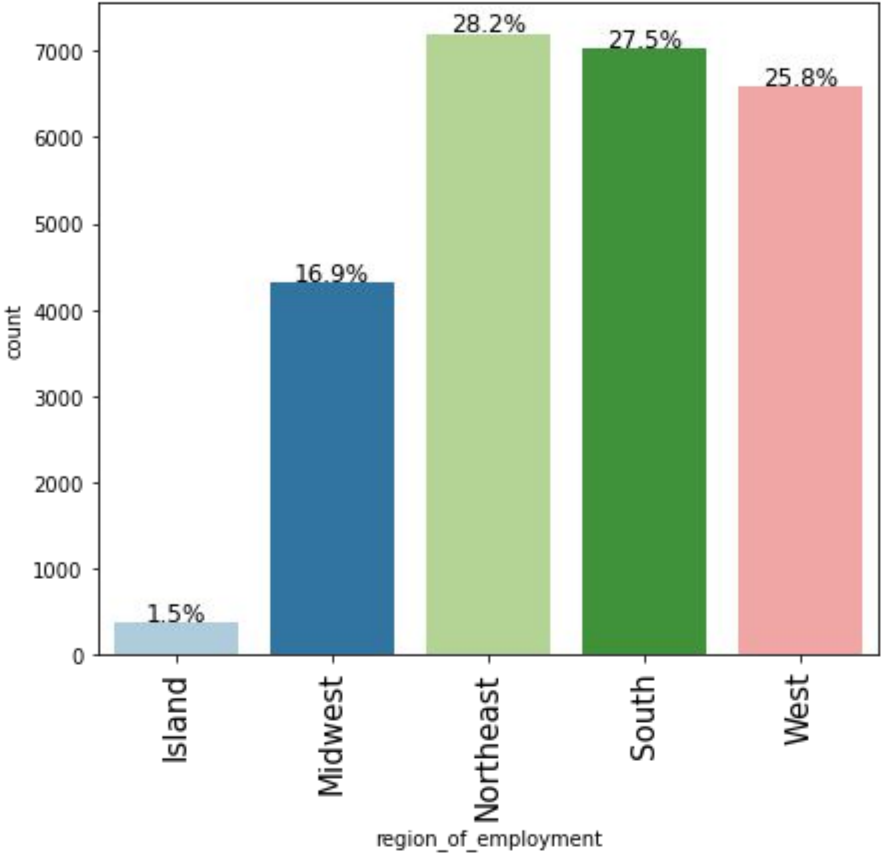
- 41.9% do not have job experience.

EDA: Univariate Analysis: Requires Job Training



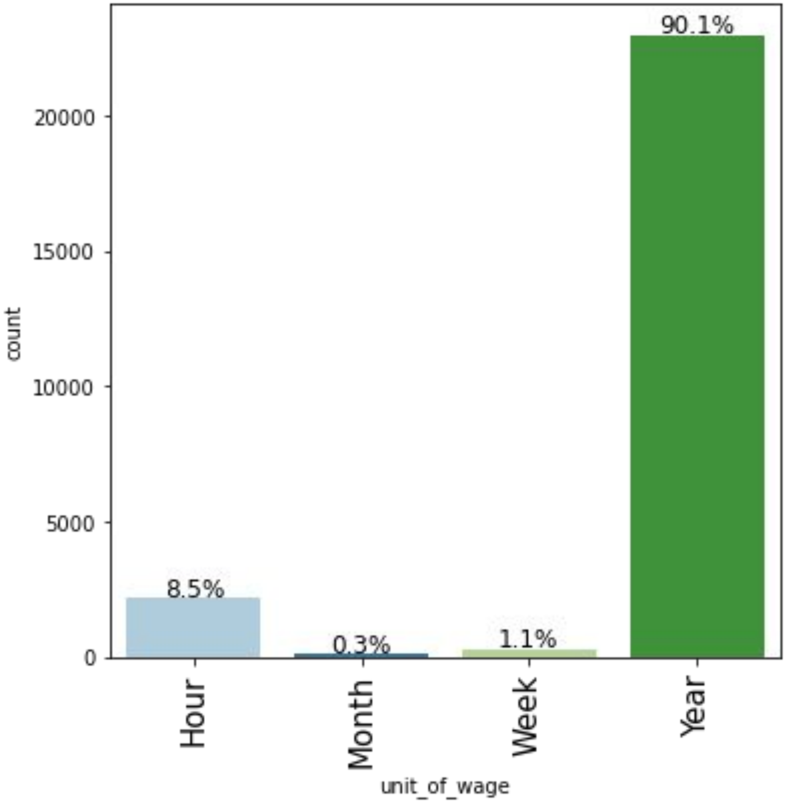
- 88.4% do not require on the job training.
- 11.6% do require on the job training.

EDA: Univariate Analysis: Region of Employment



- Over 75% of applicants are wanting to work in the Northeast, South, and Western United States.
- Island has the lowest amount of applicants, as it is probably mostly hospitality and agriculture.
- Midwest only have 16.9%, which could be due to coastal regions accessibility.

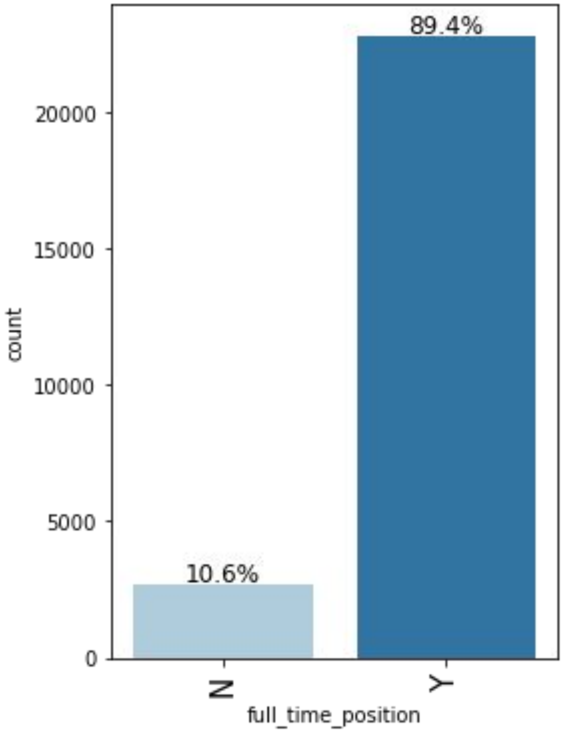
EDA: Univariate Analysis: Unit of Wage



- 90.1% of applicants have a yearly salary.

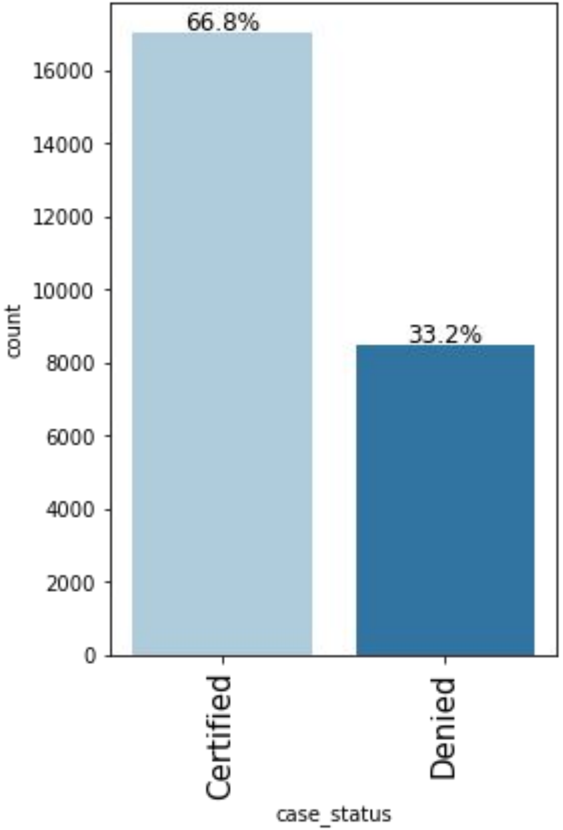
- Next highest is 8.5% at an hourly wage.
- Week and Month are the lowest units of wage as those are probably rather seasonal/temporary positions.

EDA: Univariate Analysis: Full Time Positions



- 89.4% of employees are applying for full time positions

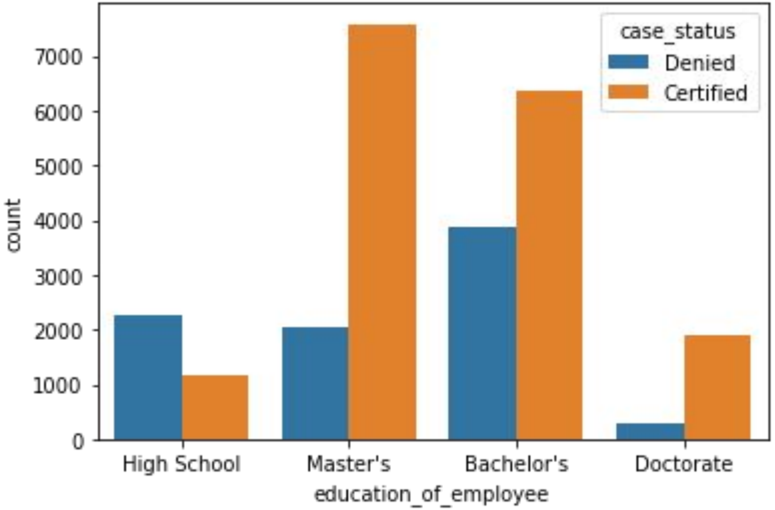
EDA: Univariate Analysis: Case Status



- 66.8% of applicants get a work Visa.

- 33.2% of applicants do not get their Visa approved

Bivariate Analysis: Education with Case Status

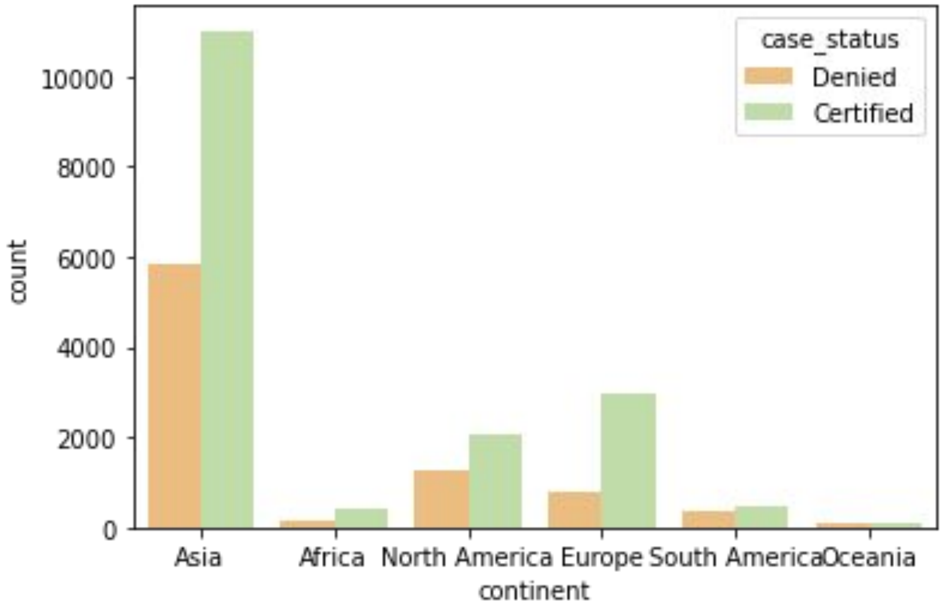


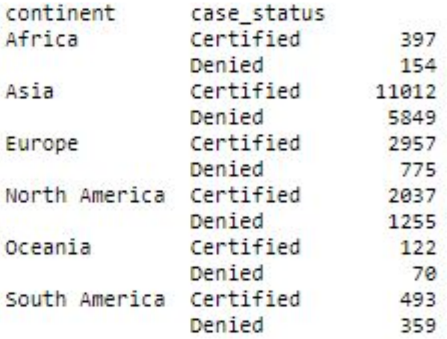
- Higher education does increase the likelihood of getting a work Visa.

- Better educated workforce is a better investment for a company and makes sponsoring an employee worth the time and cost.

- Majority of workers have a Master’s or a Bachelor’s degree.

Bivariate Analysis: Continents with Case Status

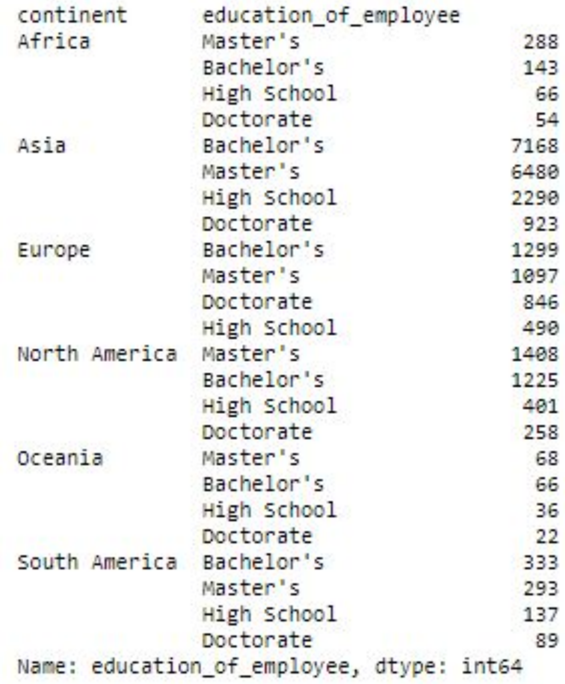




```
continent      case_status
Africa         Certified         397
               Denied            154
Asia           Certified       11012
               Denied           5849
Europe         Certified        2957
               Denied            775
North America  Certified        2037
               Denied           1255
Oceania        Certified         122
               Denied             70
South America  Certified         493
               Denied            359
```

- Most of the applicants are from Asia, then Europe and North America.
- Oceania has the least amount applying for a work Visa.
- All of the continents have a higher certified rate than denied rate.

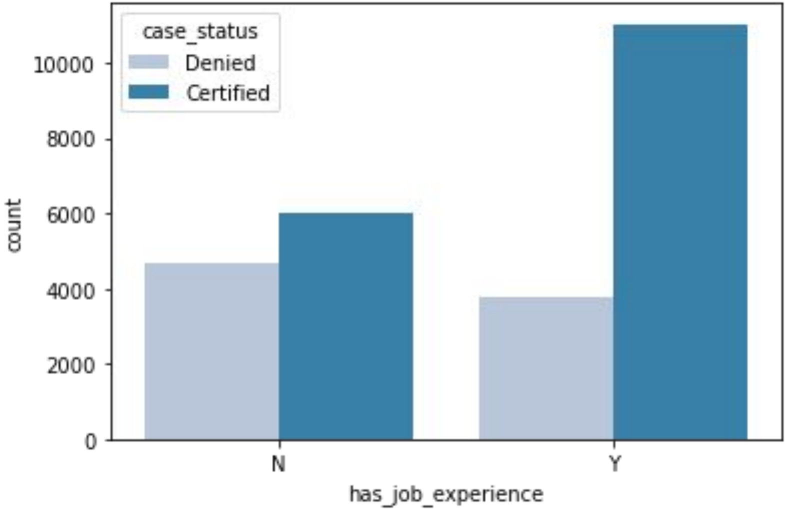
Bivariate Analysis: Continent with Education of Employee

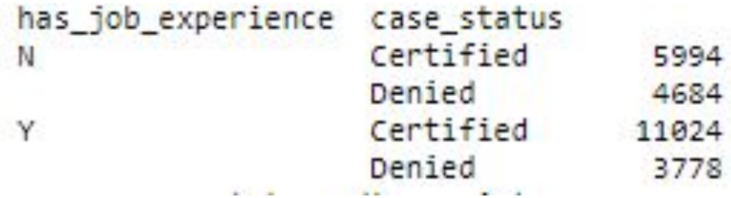


```
continent      education_of_employee
Africa         Master's                 288
               Bachelor's               143
               High School               66
               Doctorate                 54
Asia           Bachelor's              7168
               Master's                6480
               High School             2290
               Doctorate                923
Europe         Bachelor's              1299
               Master's                1097
               Doctorate                846
               High School              490
North America  Master's                1408
               Bachelor's              1225
               High School              401
               Doctorate                258
Oceania        Master's                  68
               Bachelor's                66
               High School               36
               Doctorate                 22
South America  Bachelor's               333
               Master's                 293
               High School              137
               Doctorate                 89
Name: education_of_employee, dtype: int64
```

- European employees have the lowest number of employees with at least High School education.
- Africa, Asia, North American, Oceania, and South America have Doctorate level education as the lowest amount.
- Africa, Oceania, and North America have Master’s Degrees as the highest prevalence.
- The data is missing information on whether or not this is the highest level obtained by the employee or not.

Bivariate Analysis: Work Experience with Case Status

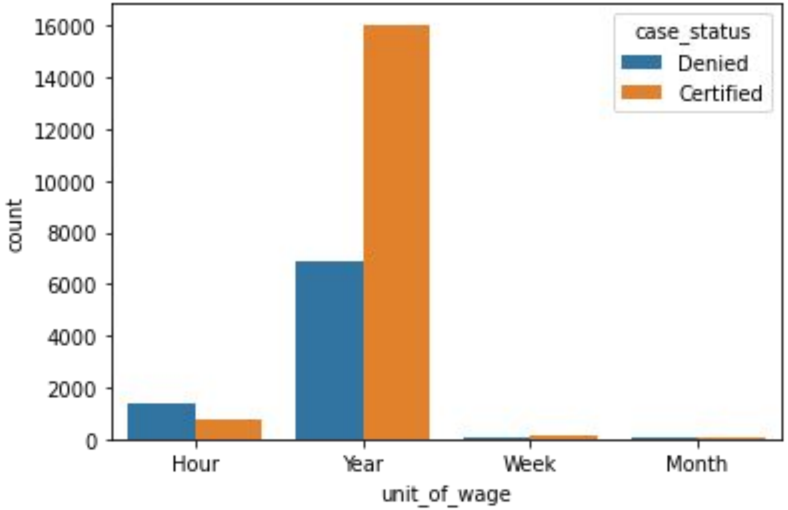


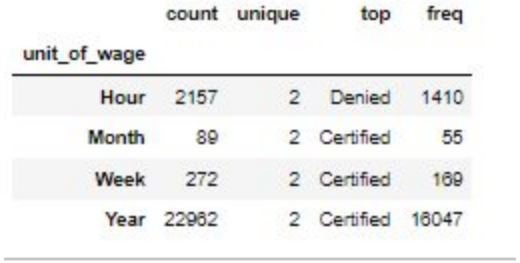


```
has_job_experience  case_status
N                   Certified       5994
                    Denied          4684
Y                   Certified      11024
                    Denied          3778
```

- Having work experience increases the chance of getting certified.
- But, even having no job experience there is a greater chance of getting certified. Is job experience referring to any experience or experience in the job they are applying for?

Bivariate Analysis: Pay Interval with Case Status

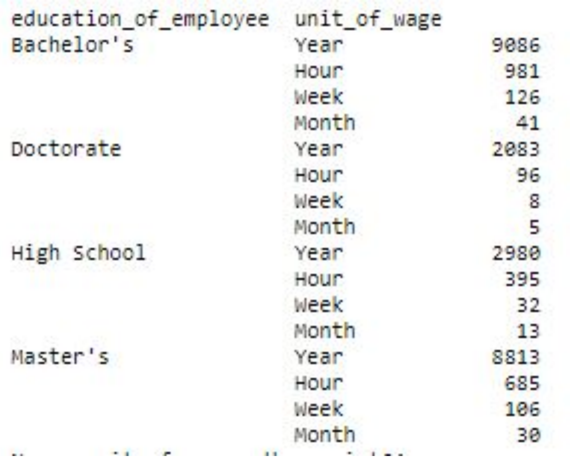




	count	unique	top	freq
unit_of_wage				
Hour	2157	2	Denied	1410
Month	89	2	Certified	55
Week	272	2	Certified	169
Year	22962	2	Certified	16047

- Applicants with yearly pay have the greatest chance of getting certified.
- Hourly waged applicants have the least chance of getting certified
- Majority of applicants have a yearly unit of wage.

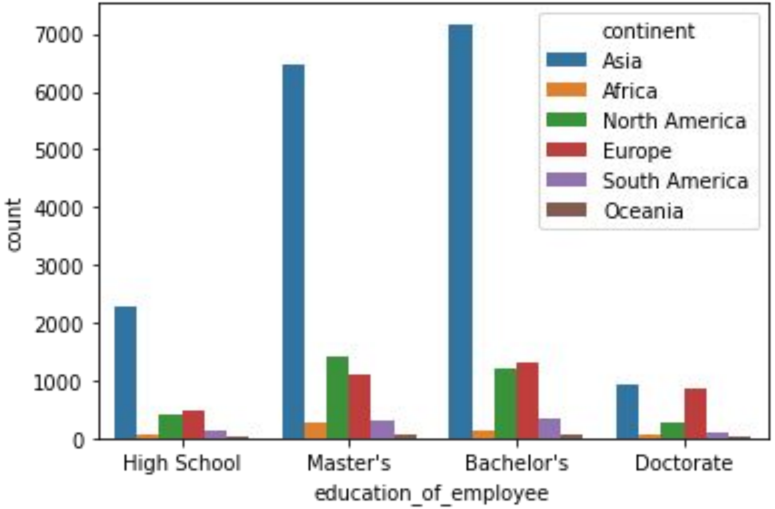
Bivariate Analysis: Education with Unit of Wage

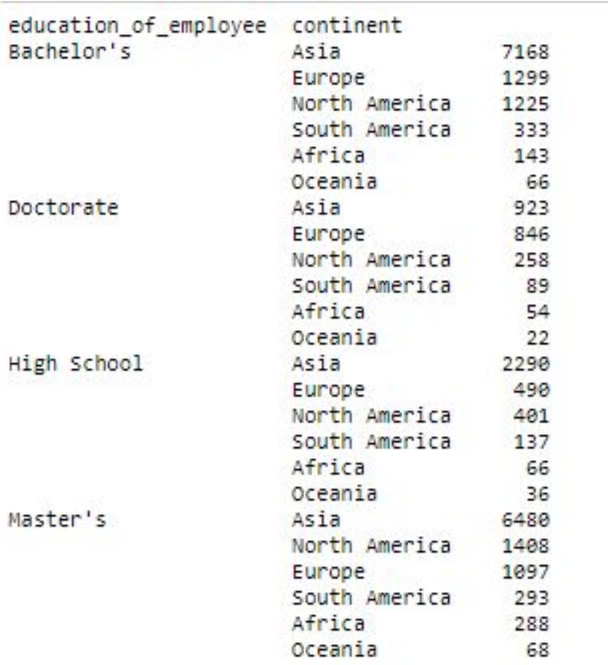


```
education_of_employee  unit_of_wage
Bachelor's             Year            9086
                       Hour             981
                       Week             126
                       Month             41
Doctorate              Year            2083
                       Hour              96
                       Week               8
                       Month              5
High School            Year            2980
                       Hour             395
                       Week              32
                       Month             13
Master's               Year            8813
                       Hour             685
                       Week             106
                       Month             30
```

- All levels of education have yearly unit of wage as the highest count.
- All levels of education follow the same trend in the units of wage with year being the most prevalent, followed by hour, week and then month.
-

Bivariate Analysis: Continent and Education

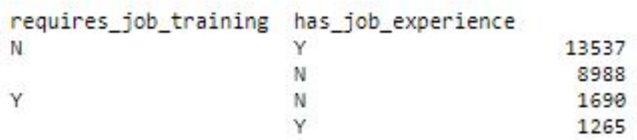




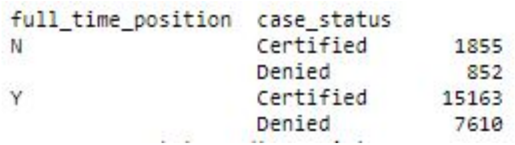
```
education_of_employee  continent
Bachelor's             Asia             7168
                       Europe           1299
                       North America    1225
                       South America     333
                       Africa            143
                       Oceania            66
Doctorate              Asia              923
                       Europe            846
                       North America     258
                       South America      89
                       Africa             54
                       Oceania            22
High School            Asia             2290
                       Europe            490
                       North America     401
                       South America     137
                       Africa             66
                       Oceania            36
Master's               Asia             6480
                       North America    1408
                       Europe           1097
                       South America     293
                       Africa            288
                       Oceania            68
```

- Asia has the highest counts across all of the levels of education.
- Bachelor’s is the most prevalent education level, followed by Master’s.
- Africa and Oceania have the lowest numbers represented, which is probably due to them having lower populations than other regions.

Bivariate Analysis: Job with Case Status



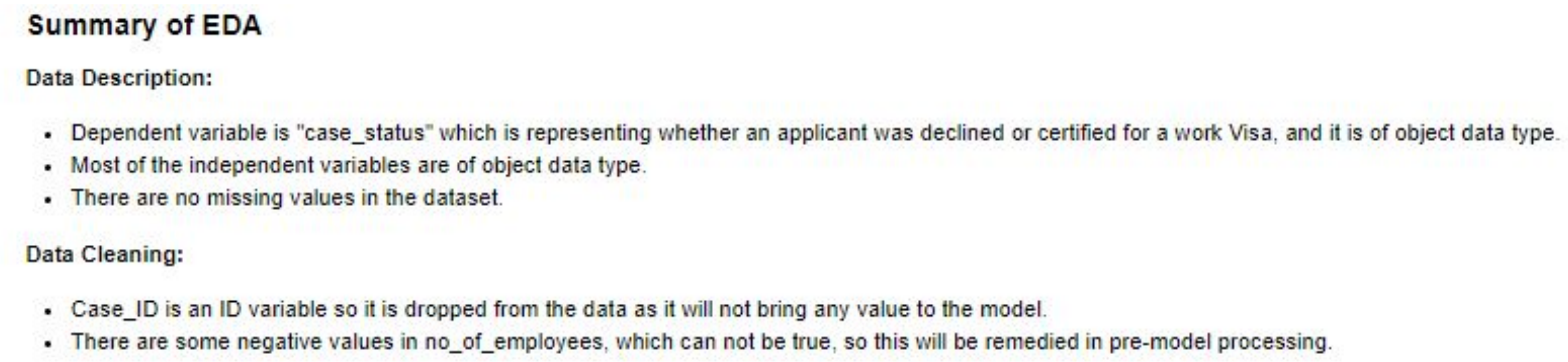
```
requires_job_training  has_job_experience
N                      Y                     13537
                       N                      8988
Y                      N                      1690
                       Y                      1265
```



```
full_time_position  case_status
N                   Certified       1855
                    Denied           852
Y                   Certified      15163
                    Denied          7610
```

- Those with job experience do not require as much on the job experience, which can help benefit companies as training cost will be reduced and make sponsoring the employee more profitable especially if they are educated.
- Applying for full time work far exceeds part time work counts, but both still have around a 50% chance of getting certified.

EDA: Summary



Summary of EDA

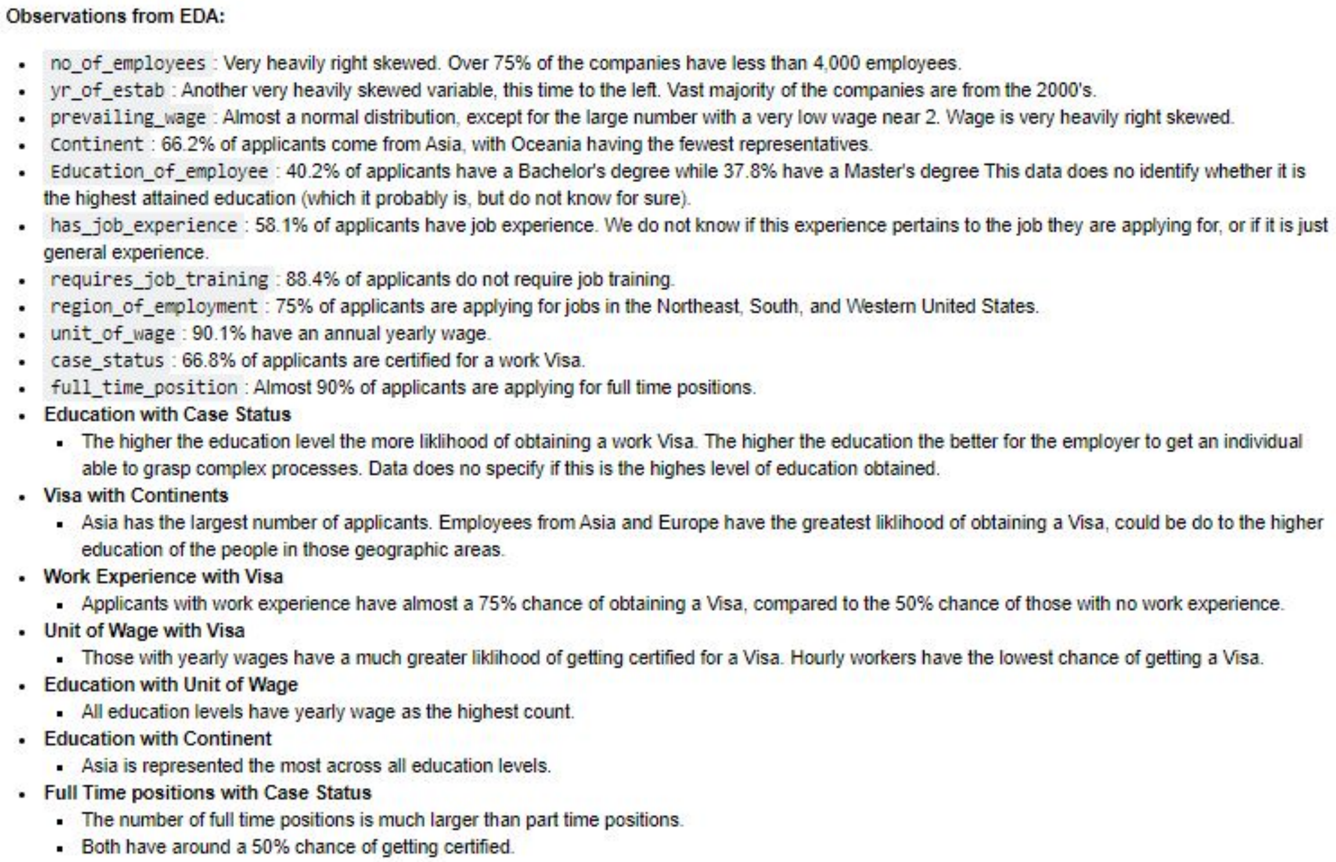
Data Description:

- Dependent variable is "case_status" which is representing whether an applicant was declined or certified for a work Visa, and it is of object data type.
- Most of the independent variables are of object data type.
- There are no missing values in the dataset.

Data Cleaning:

- Case_ID is an ID variable so it is dropped from the data as it will not bring any value to the model.
- There are some negative values in no_of_employees, which can not be true, so this will be remedied in pre-model processing.

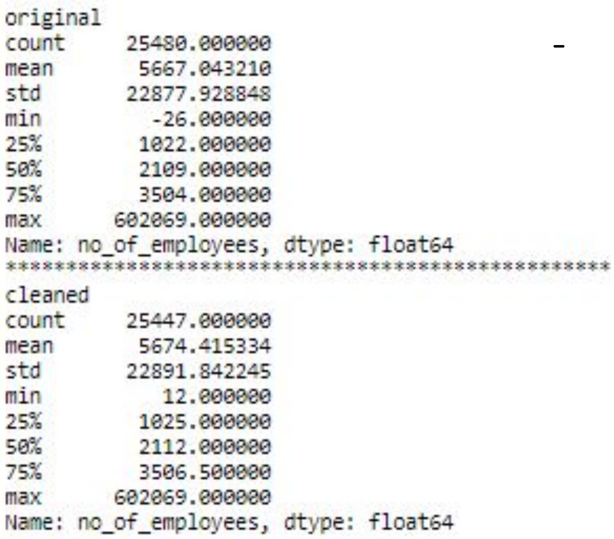
EDA: Summary



Observations from EDA:

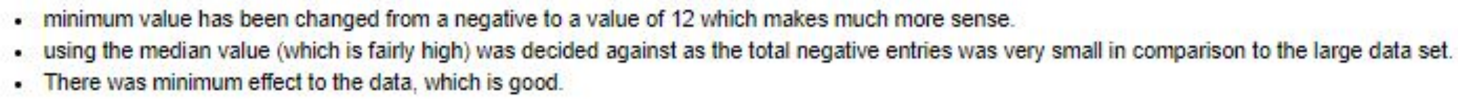
- `no_of_employees` : Very heavily right skewed. Over 75% of the companies have less than 4,000 employees.
- `yr_of_estab` : Another very heavily skewed variable, this time to the left. Vast majority of the companies are from the 2000's.
- `prevailing_wage` : Almost a normal distribution, except for the large number with a very low wage near 2. Wage is very heavily right skewed.
- `continent` : 66.2% of applicants come from Asia, with Oceania having the fewest representatives.
- `Education_of_employee` : 40.2% of applicants have a Bachelor's degree while 37.8% have a Master's degree This data does no identify whether it is the highest attained education (which it probably is, but do not know for sure).
- `has_job_experience` : 58.1% of applicants have job experience. We do not know if this experience pertains to the job they are applying for, or if it is just general experience.
- `requires_job_training` : 88.4% of applicants do not require job training.
- `region_of_employment` : 75% of applicants are applying for jobs in the Northeast, South, and Western United States.
- `unit_of_wage` : 90.1% have an annual yearly wage.
- `case_status` : 66.8% of applicants are certified for a work Visa.
- `full_time_position` : Almost 90% of applicants are applying for full time positions.
- **Education with Case Status**
  - The higher the education level the more liklihood of obtaining a work Visa. The higher the education the better for the employer to get an individual able to grasp complex processes. Data does no specify if this is the highes level of education obtained.
- **Visa with Continents**
  - Asia has the largest number of applicants. Employees from Asia and Europe have the greatest liklihood of obtaining a Visa, could be do to the higher education of the people in those geographic areas.
- **Work Experience with Visa**
  - Applicants with work experience have almost a 75% chance of obtaining a Visa, compared to the 50% chance of those with no work experience.
- **Unit of Wage with Visa**
  - Those with yearly wages have a much greater liklihood of getting certified for a Visa. Hourly workers have the lowest chance of getting a Visa.
- **Education with Unit of Wage**
  - All education levels have yearly wage as the highest count.
- **Education with Continent**
  - Asia is represented the most across all education levels.
- **Full Time positions with Case Status**
  - The number of full time positions is much larger than part time positions.
  - Both have around a 50% chance of getting certified.

Data Preprocessing



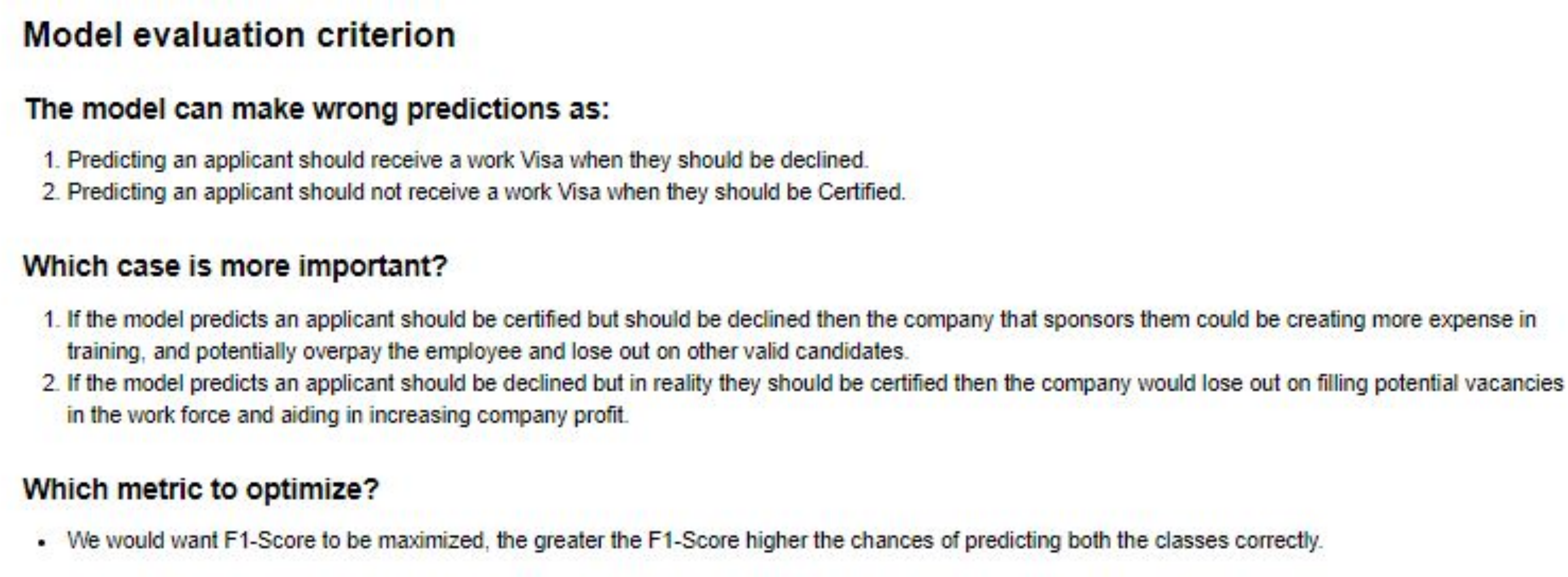
```
original
count     25480.000000
mean       5667.043210
std       22877.928848
min         -26.000000
25%        1022.000000
50%        2109.000000
75%        3504.000000
max      602069.000000
Name: no_of_employees, dtype: float64
**************************************************
cleaned
count     25447.000000
mean       5674.415334
std       22891.842245
min          12.000000
25%        1025.000000
50%        2112.000000
75%        3506.500000
max      602069.000000
Name: no_of_employees, dtype: float64
```

- Number of employees will be adjusted to remove the value below zero.



- minimum value has been changed from a negative to a value of 12 which makes much more sense.
- using the median value (which is fairly high) was decided against as the total negative entries was very small in comparison to the large data set.
- There was minimum effect to the data, which is good.

Model Evaluation Criteria:



Model evaluation criterion

The model can make wrong predictions as:

1. Predicting an applicant should receive a work Visa when they should be declined.
2. Predicting an applicant should not receive a work Visa when they should be Certified.

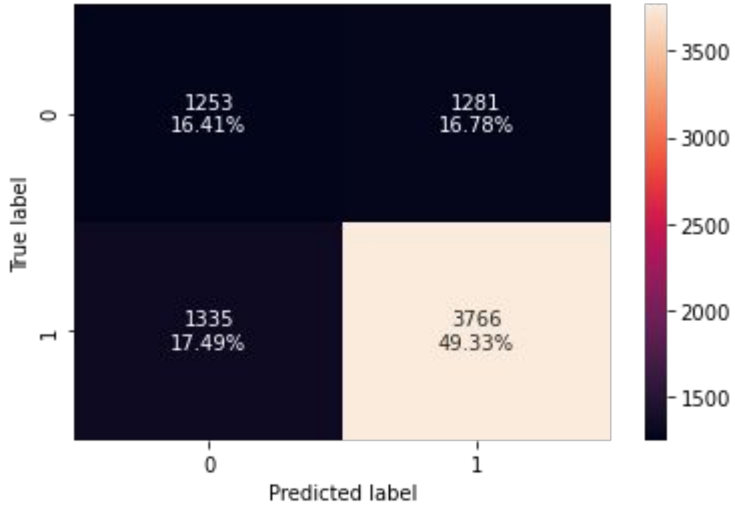
Which case is more important?

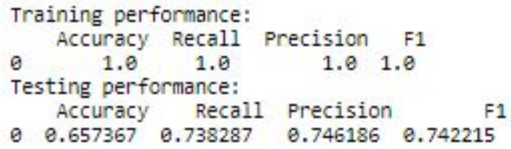
1. If the model predicts an applicant should be certified but should be declined then the company that sponsors them could be creating more expense in training, and potentially overpay the employee and lose out on other valid candidates.
2. If the model predicts an applicant should be declined but in reality they should be certified then the company would lose out on filling potential vacancies in the work force and aiding in increasing company profit.

Which metric to optimize?

- We would want F1-Score to be maximized, the greater the F1-Score higher the chances of predicting both the classes correctly.

Decision Tree:

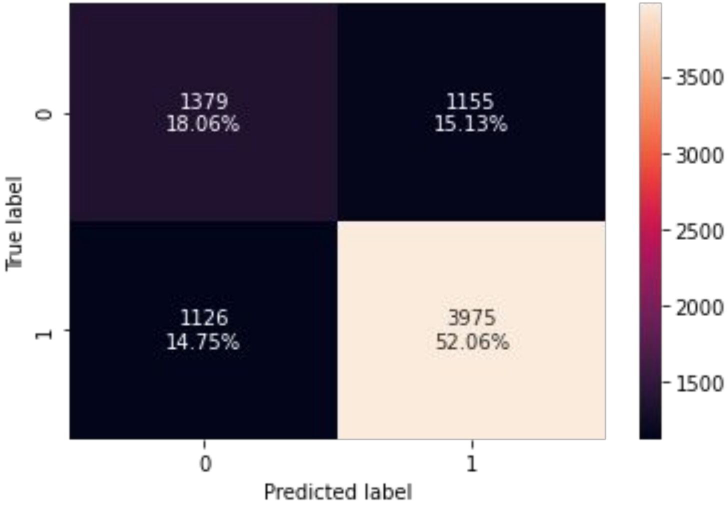


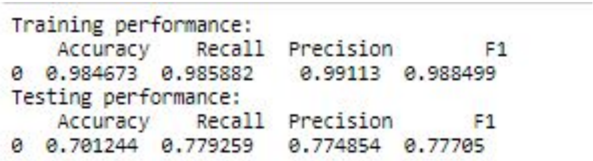


```
Training performance:
    Accuracy  Recall  Precision   F1
0        1.0     1.0        1.0  1.0
Testing performance:
    Accuracy    Recall  Precision        F1
0  0.657367  0.738287   0.746186  0.742215
```

- Decision tree is overfitting the training data with the perfect metrics.

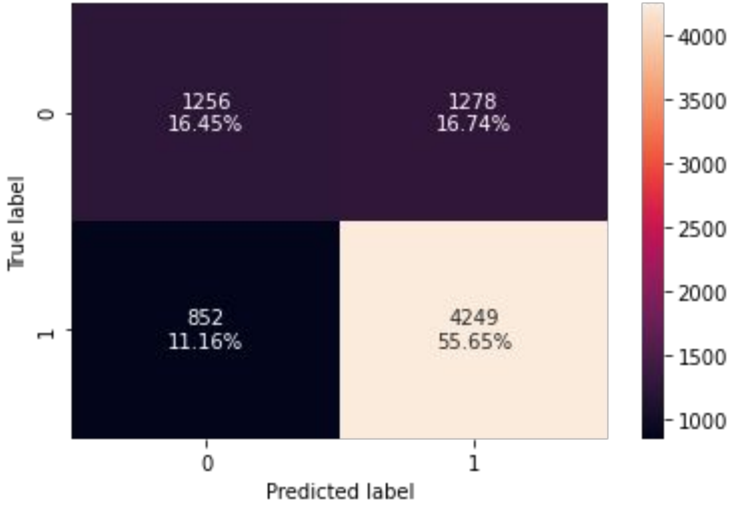
Bagging Classifier:

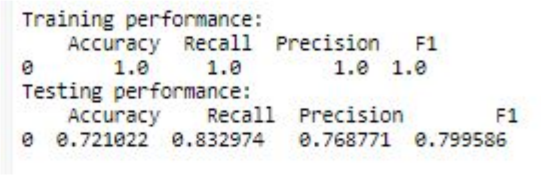



```
Training performance:
   Accuracy    Recall  Precision        F1
0  0.984673  0.985882    0.99113  0.988499
Testing performance:
   Accuracy    Recall  Precision       F1
0  0.701244  0.779259   0.774854  0.77705
```

- Still overfitting the data, but not as much as decision tree.

Random Forest:

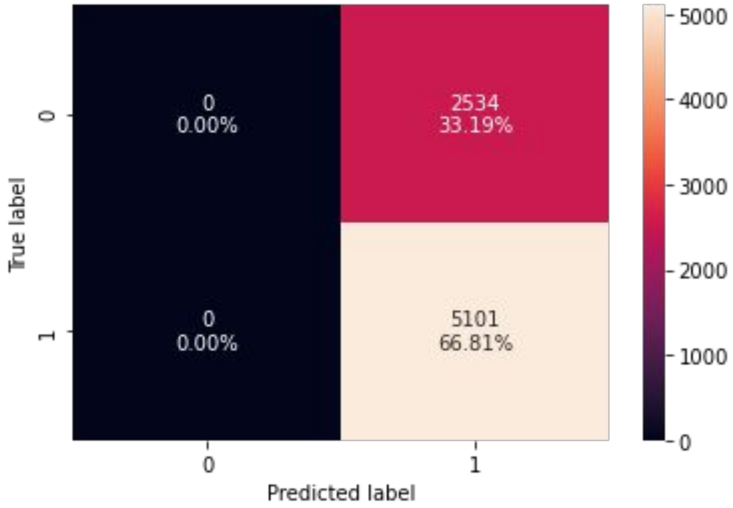


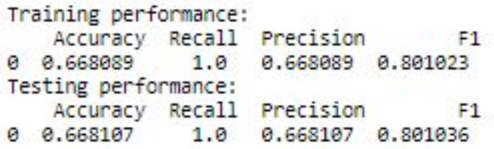


```
Training performance:
    Accuracy  Recall   Precision   F1
0        1.0     1.0         1.0  1.0
Testing performance:
    Accuracy     Recall  Precision         F1
0  0.721022  0.832974   0.768771   0.799586
```

- Still overfitting the data, but that is to be expected from default random forest made of default decision trees which will run data to perfection.

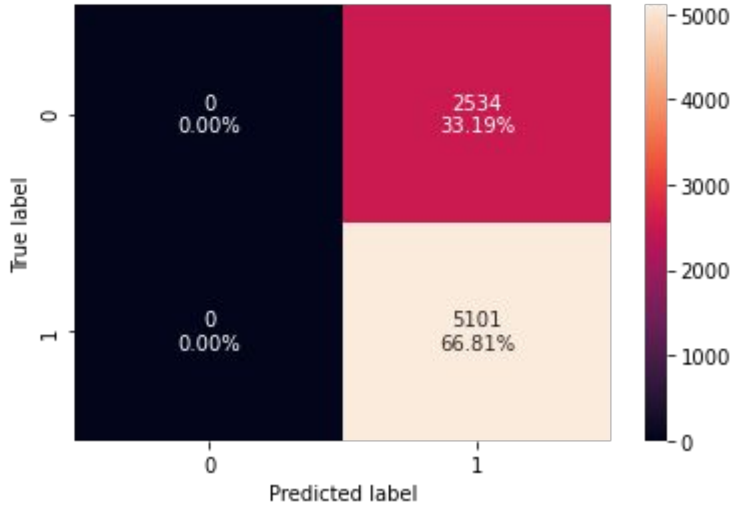
Tuned Decision Tree:

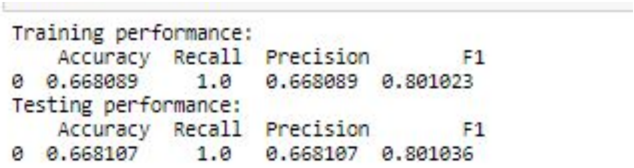



```
Training performance:
   Accuracy  Recall  Precision        F1
0  0.668089     1.0   0.668089  0.801023
Testing performance:
   Accuracy  Recall  Precision        F1
0  0.668107     1.0   0.668107  0.801036
```

- Overfitting has been reduced with similar scores across both training and test data. We will continue to try to increase predictive True Positive.

Tuned Bagging Classifier:

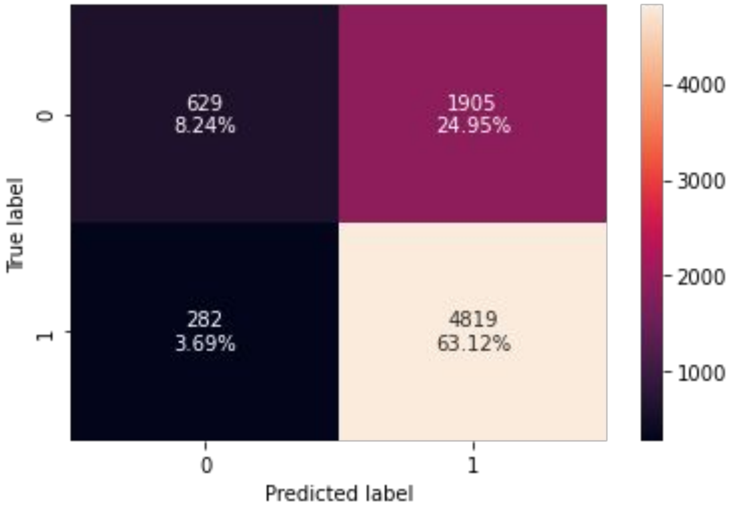


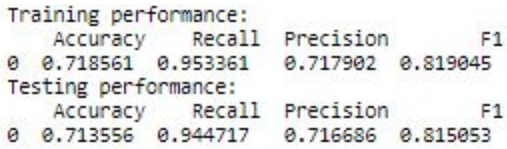


```
Training performance:
    Accuracy  Recall  Precision        F1
0  0.668089     1.0   0.668089  0.801023
Testing performance:
    Accuracy  Recall  Precision        F1
0  0.668107     1.0   0.668107  0.801036
```

- Same results as tuned decision tree.

Tuned Random Forest:

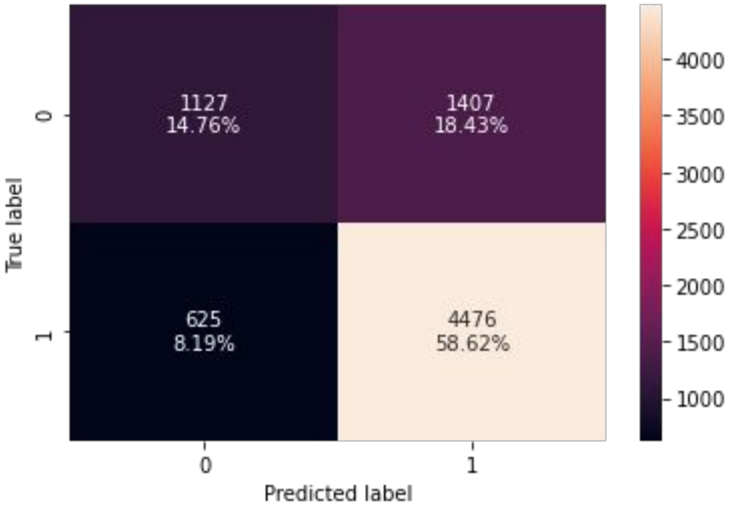


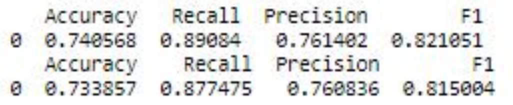


```
Training performance:
    Accuracy    Recall  Precision        F1
0  0.718561  0.953361   0.717902  0.819045
Testing performance:
    Accuracy    Recall  Precision        F1
0  0.713556  0.944717   0.716686  0.815053
```

- Overfitting has been lowered.
- Accuracy has been improved compared to decision tree and bagging classifier.
- F1 and recall scores are acceptable.

Adaboost:

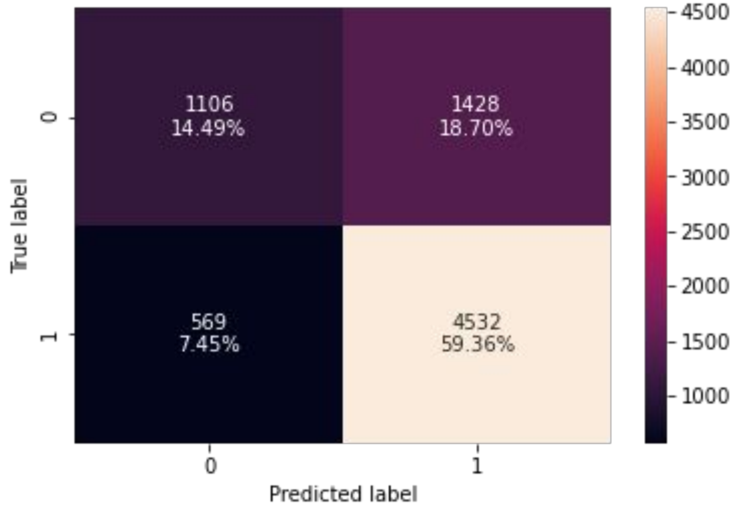


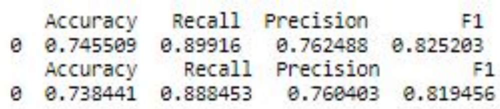


```
   Accuracy   Recall  Precision        F1
0  0.740568  0.89084   0.761402  0.821051
   Accuracy    Recall  Precision        F1
0  0.733857  0.877475   0.760836  0.815004
```

- Accuracy is improving slightly
- Good general model in predicting True positive and True negative.

Adaboost Tuning Model:

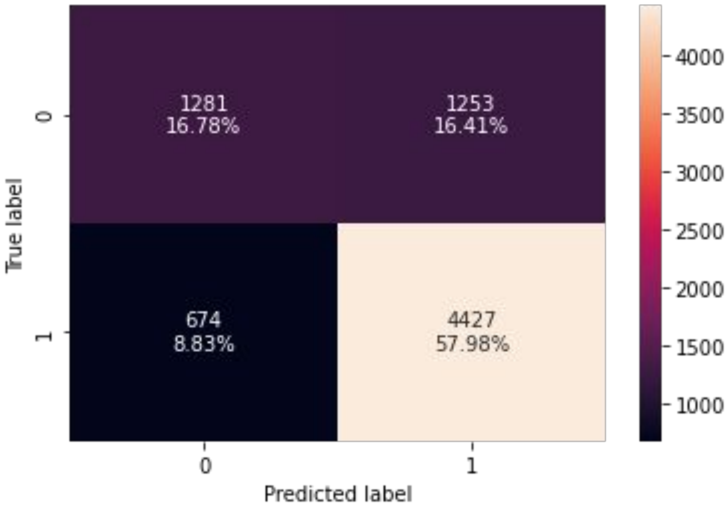


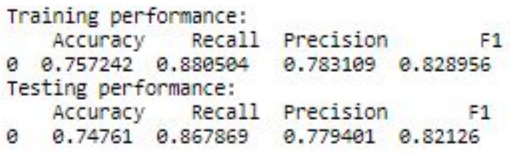


```
   Accuracy    Recall  Precision        F1
0  0.745509  0.89916   0.762488  0.825203
   Accuracy     Recall  Precision        F1
0  0.738441  0.888453   0.760403  0.819456
```

- Similar to regular Adaboost model with good predicting across both test and training data.
- Adaboost tuning model has increased True positive prediction.

Gradient Boosting:

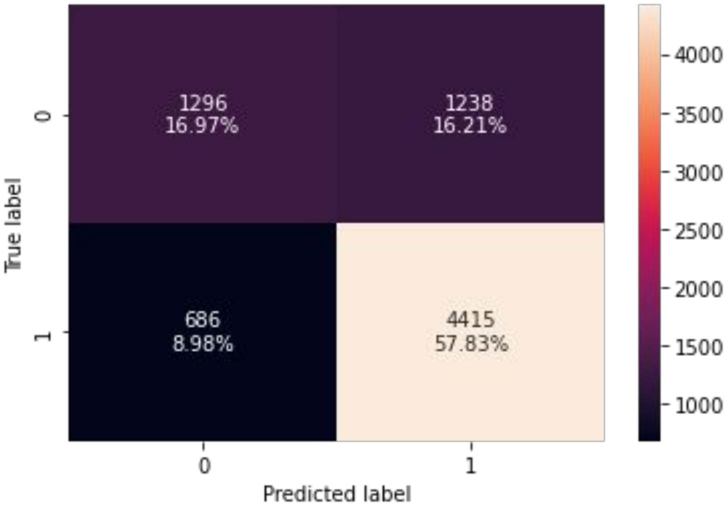


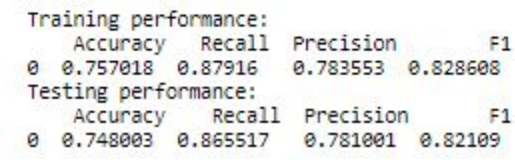


```
Training performance:
    Accuracy    Recall  Precision        F1
0  0.757242  0.880504   0.783109  0.828956
Testing performance:
    Accuracy    Recall  Precision        F1
0   0.74761  0.867869   0.779401   0.82126
```

- Similar to regular Adaboost model with good predicting across both test and training data.

Gradient Boost Tuned:

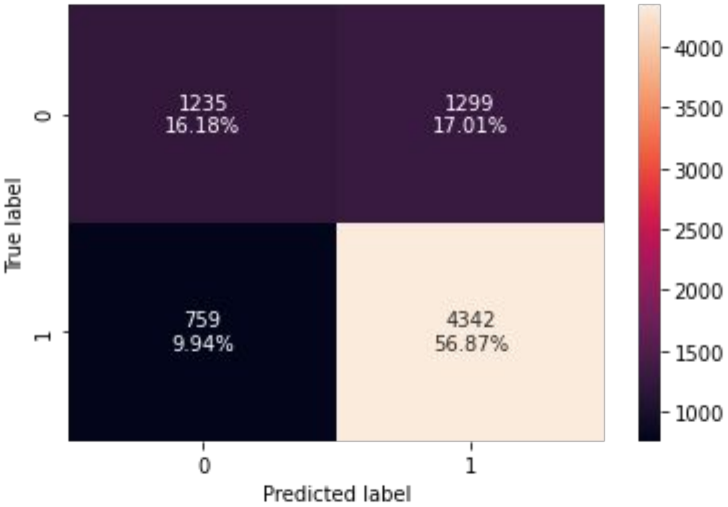


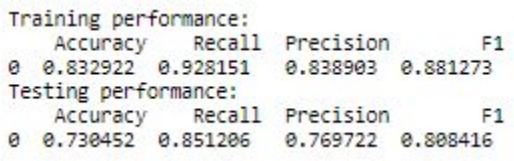


```
Training performance:
    Accuracy   Recall  Precision        F1
0  0.757018  0.87916   0.783553  0.828608
Testing performance:
    Accuracy    Recall  Precision       F1
0  0.748003  0.865517   0.781001  0.82109
```

- Tuned model similar to regular gradient boost model.

XGB Boost Model:

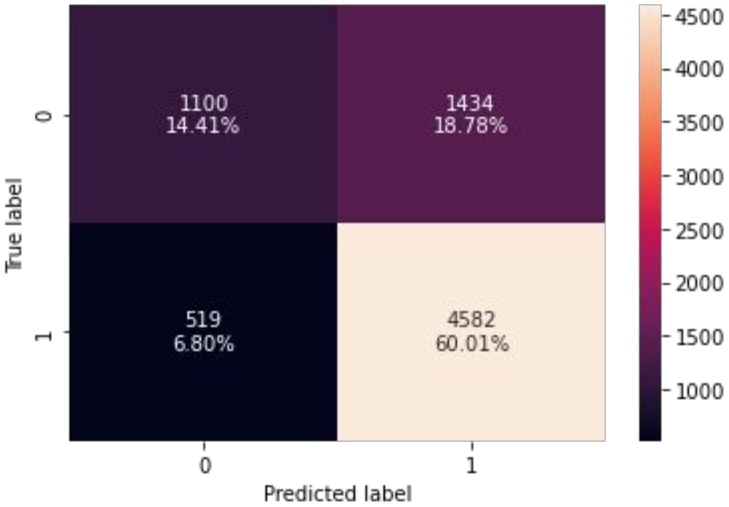


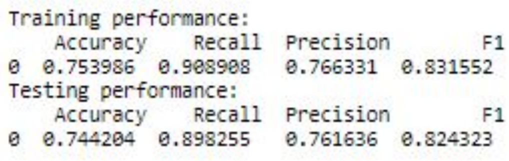


```
Training performance:
   Accuracy    Recall  Precision        F1
0  0.832922  0.928151   0.838903  0.881273
Testing performance:
   Accuracy    Recall  Precision        F1
0  0.730452  0.851206   0.769722  0.808416
```

- XGB boost is slightly overfitting the training data.

XGB Boost Tuning:

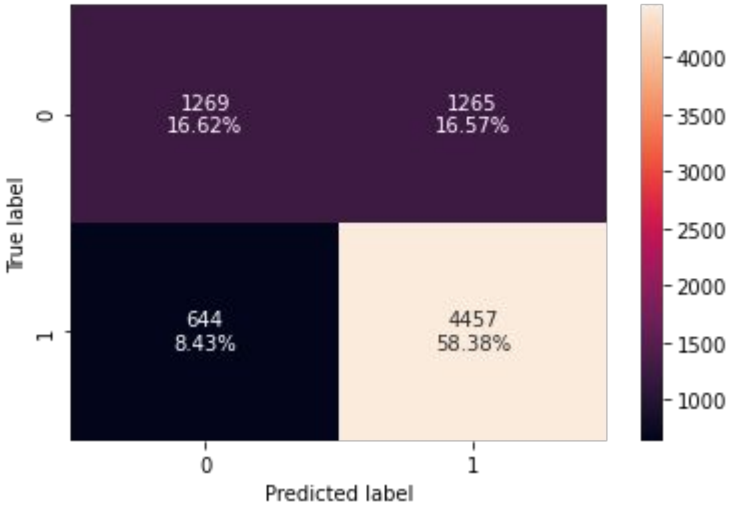


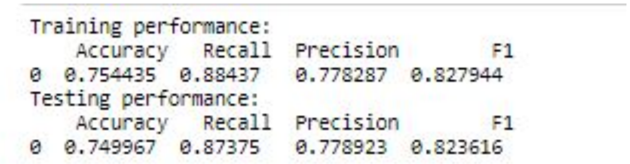


```
Training performance:
   Accuracy    Recall  Precision        F1
0  0.753986  0.908908   0.766331  0.831552
Testing performance:
   Accuracy    Recall  Precision        F1
0  0.744204  0.898255   0.761636  0.824323
```

- Tuned model reduced overfitting and giving good metrics.

Stacking Classifier Model:

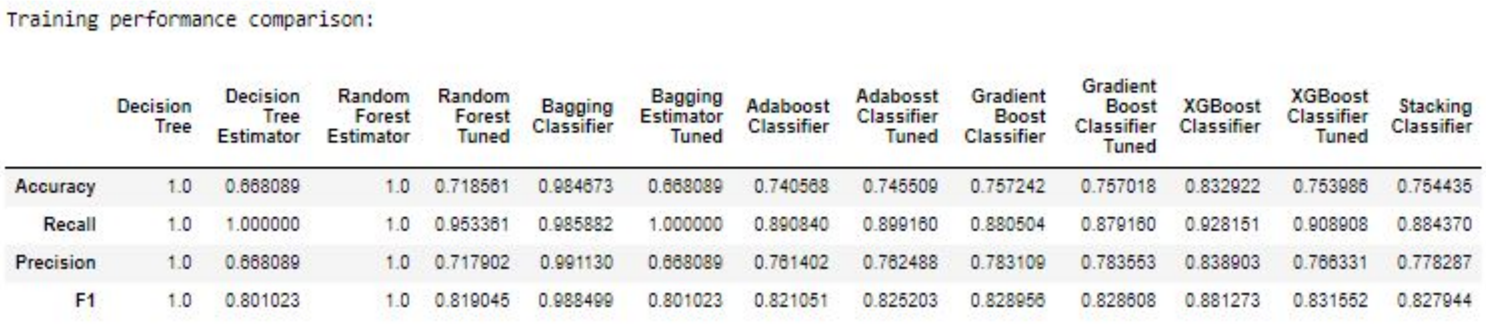




```
Training performance:
    Accuracy   Recall  Precision        F1
0  0.754435  0.88437   0.778287  0.827944
Testing performance:
    Accuracy   Recall  Precision        F1
0  0.749967  0.87375   0.778923  0.823616
```

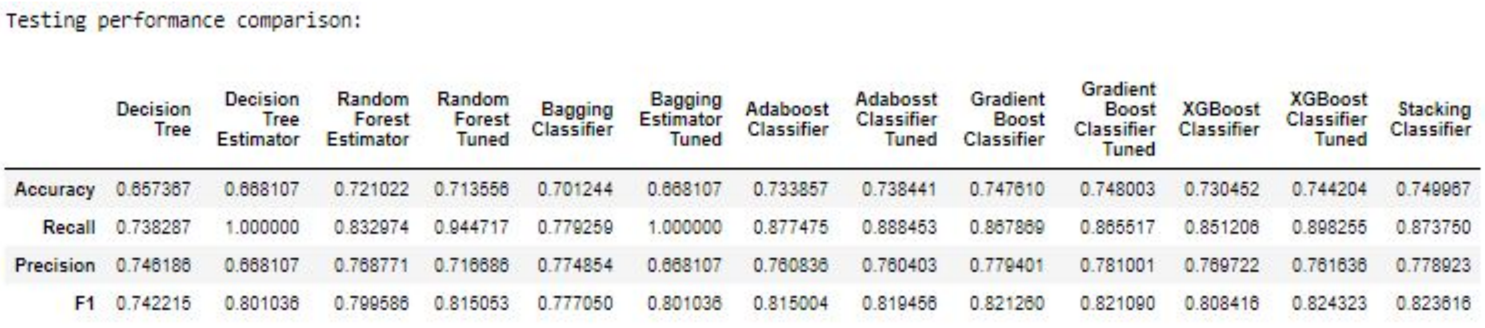
- Similar performance to other tuned models.

Model Comparison:



Training performance comparison:

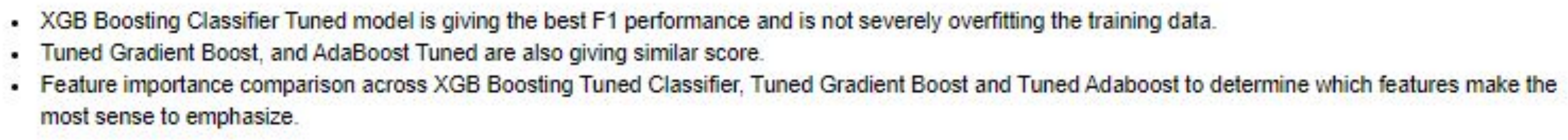
	Decision Tree	Decision Tree Estimator	Random Forest Estimator	Random Forest Tuned	Bagging Classifier	Bagging Estimator Tuned	Adaboost Classifier	Adabosst Classifier Tuned	Gradient Boost Classifier	Gradient Boost Classifier Tuned	XGBoost Classifier	XGBoost Classifier Tuned	Stacking Classifier
Accuracy	1.0	0.668089	1.0	0.718561	0.984673	0.668089	0.740568	0.745509	0.757242	0.757018	0.832922	0.753986	0.754435
Recall	1.0	1.000000	1.0	0.953361	0.985882	1.000000	0.890840	0.899160	0.880504	0.879160	0.928151	0.908908	0.884370
Precision	1.0	0.668089	1.0	0.717902	0.991130	0.668089	0.761402	0.762488	0.783109	0.783553	0.838903	0.766331	0.778287
F1	1.0	0.801023	1.0	0.819045	0.988499	0.801023	0.821051	0.825203	0.828956	0.828608	0.881273	0.831552	0.827944



Testing performance comparison:

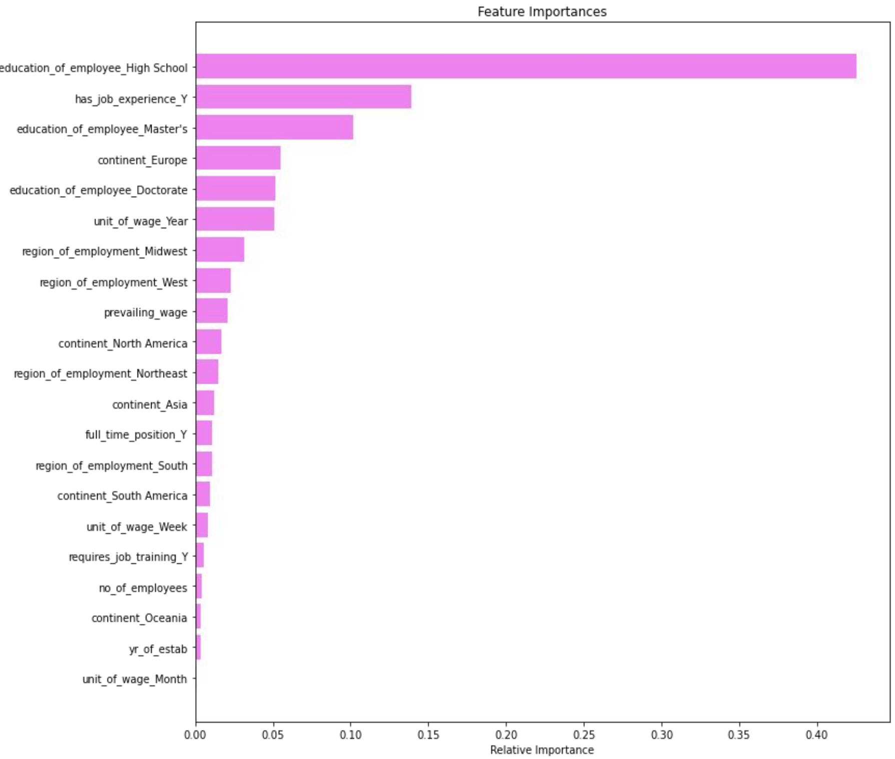
	Decision Tree	Decision Tree Estimator	Random Forest Estimator	Random Forest Tuned	Bagging Classifier	Bagging Estimator Tuned	Adaboost Classifier	Adabosst Classifier Tuned	Gradient Boost Classifier	Gradient Boost Classifier Tuned	XGBoost Classifier	XGBoost Classifier Tuned	Stacking Classifier
Accuracy	0.657367	0.668107	0.721022	0.713556	0.701244	0.668107	0.733857	0.738441	0.747610	0.748003	0.730452	0.744204	0.749967
Recall	0.738287	1.000000	0.832974	0.944717	0.779259	1.000000	0.877475	0.888453	0.867869	0.865517	0.851206	0.898255	0.873750
Precision	0.746186	0.668107	0.768771	0.716686	0.774854	0.668107	0.760836	0.760403	0.779401	0.781001	0.769722	0.761636	0.778923
F1	0.742215	0.801036	0.799586	0.815053	0.777050	0.801036	0.815004	0.819456	0.821260	0.821090	0.808416	0.824323	0.823616

Model Comparison:

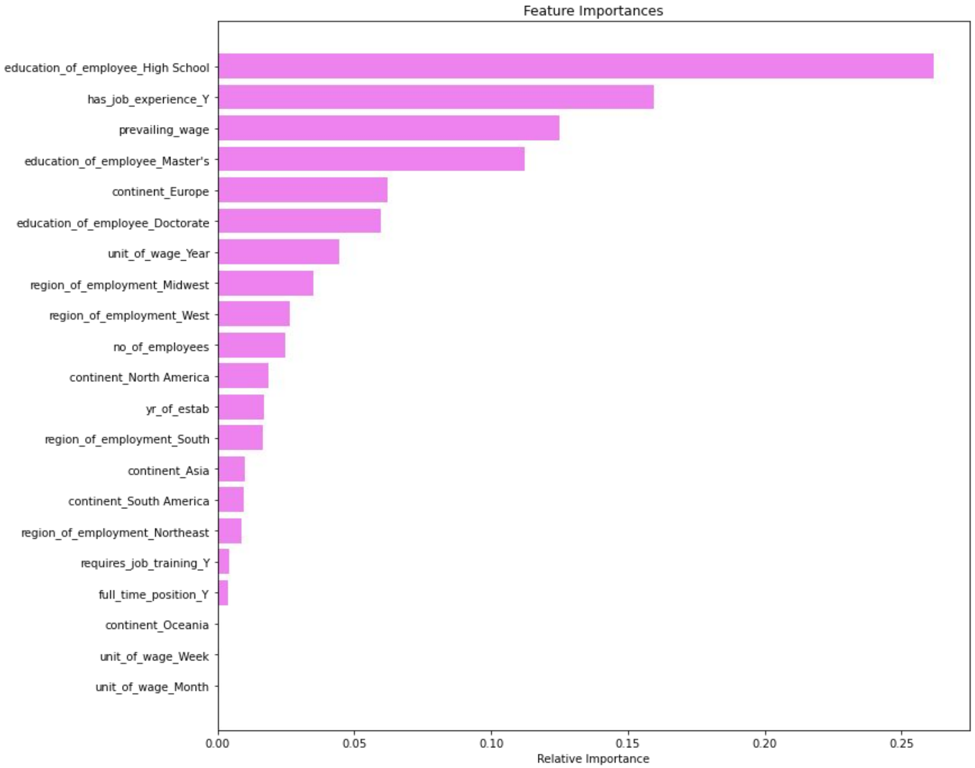


- XGB Boosting Classifier Tuned model is giving the best F1 performance and is not severely overfitting the training data.
- Tuned Gradient Boost, and AdaBoost Tuned are also giving similar score.
- Feature importance comparison across XGB Boosting Tuned Classifier, Tuned Gradient Boost and Tuned Adaboost to determine which features make the most sense to emphasize.

Model Comparison: Feature Importance

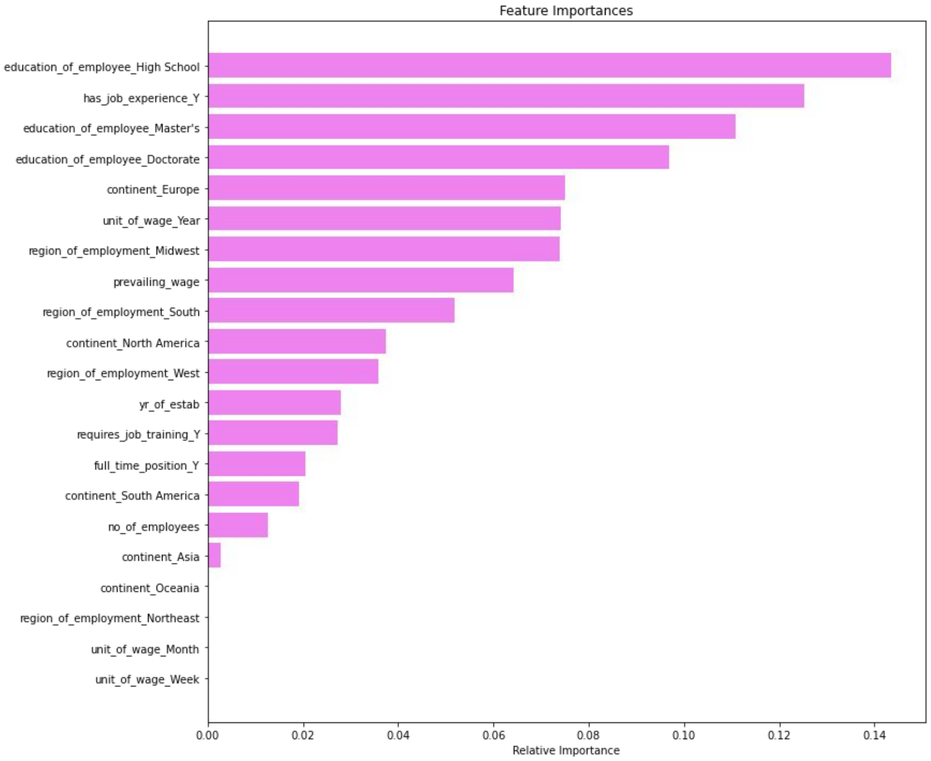


XGB Boost Tuning Model



Gradient Boosting Tuning Model

Model Comparison: Feature Importance

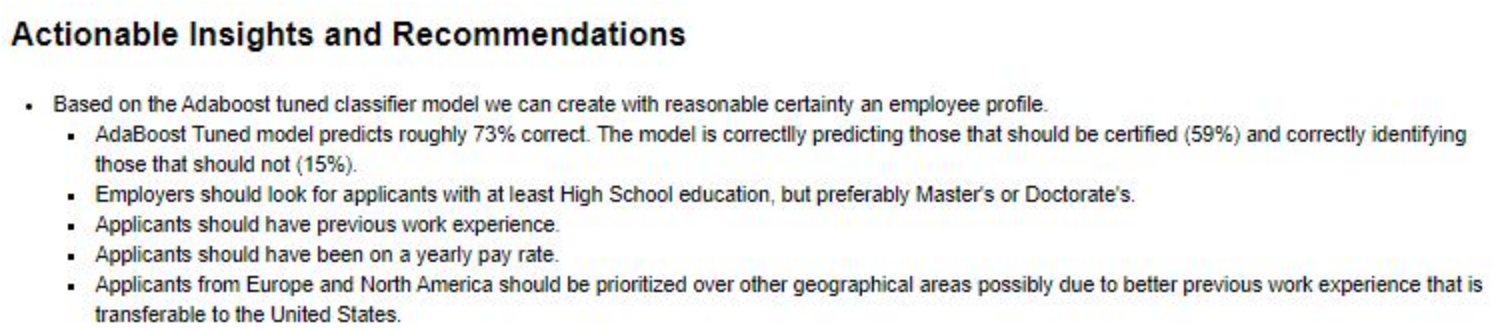


Adaboost Tuning Model

- This model is giving good metric performance as well as showing a more evenly distributed feature importance, instead of relying so heavily on just high school education like the other two tuned models.

- This model is placing importance on employees with previous work experience, multiple levels of education, having a yearly unit of wage, being from Europe or North America, applying to full time positions and prevailing wage.

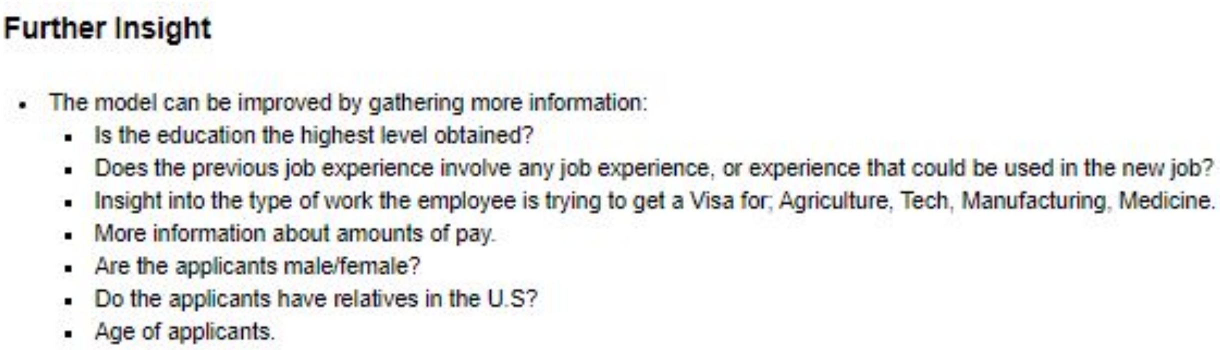
Actionable Insights and Recommendations:



Actionable Insights and Recommendations

- Based on the Adaboost tuned classifier model we can create with reasonable certainty an employee profile.
  - AdaBoost Tuned model predicts roughly 73% correct. The model is correctlly predicting those that should be certified (59%) and correctly identifying those that should not (15%).
  - Employers should look for applicants with at least High School education, but preferably Master's or Doctorate's.
  - Applicants should have previous work experience.
  - Applicants should have been on a yearly pay rate.
  - Applicants from Europe and North America should be prioritized over other geographical areas possibly due to better previous work experience that is transferable to the United States.

Further Insight:



Further Insight

- The model can be improved by gathering more information:
    - Is the education the highest level obtained?
    - Does the previous job experience involve any job experience, or experience that could be used in the new job?
    - Insight into the type of work the employee is trying to get a Visa for; Agriculture, Tech, Manufacturing, Medicine.
    - More information about amounts of pay.
    - Are the applicants male/female?
    - Do the applicants have relatives in the U.S?
    - Age of applicants.