Business Report

ML 2 Coded Project

PGPDSBA

Chithira Raj

Table of Contents

List of Tables

List of Figures

1. 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.

2. Objective

In FY 2016, the OFLC processed 775,979 employer applications for 1,699,957 positions for temporary and permanent labour 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.

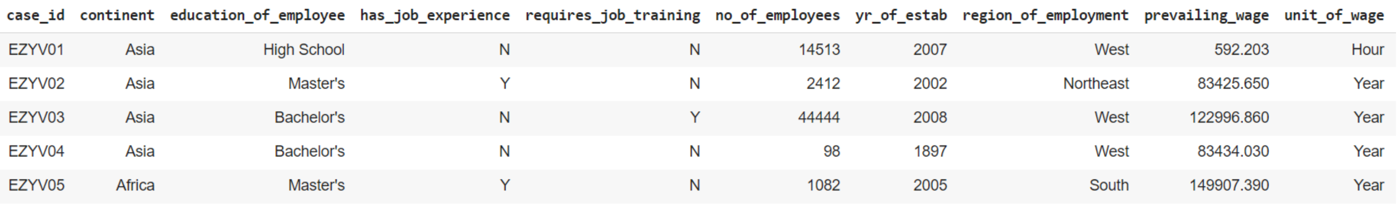
3. Data Dictionary

S.No.	Variables	Description
1	case_id	ID of each visa application
2	continent	Information of continent the employee
3	education_of_employee	Information of education of the employee
4	has_job_experience	Does the employee has any job experience? Y= Yes; N = No
5	requires_job_training	Does the employee require any job training? Y = Yes; N = No
6	no_of_employees	Number of employees in the employer's company
7	yr_of_estab	Year in which the employer's company was established
8	region_of_employment	Information of foreign worker's intended region of employment in the US.
9	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.
10	unit_of_wage	Unit of prevailing wage. Values include Hourly, Weekly, Monthly, and Yearly.
11	full_time_position	Is the position of work full-time? Y = Full-Time Position; N = Part-Time Position
12	case_status	Flag indicating if the Visa was certified or denied

Table 1: Data Dictionary

4. Data Overview

4.1. Import libraries and load the data



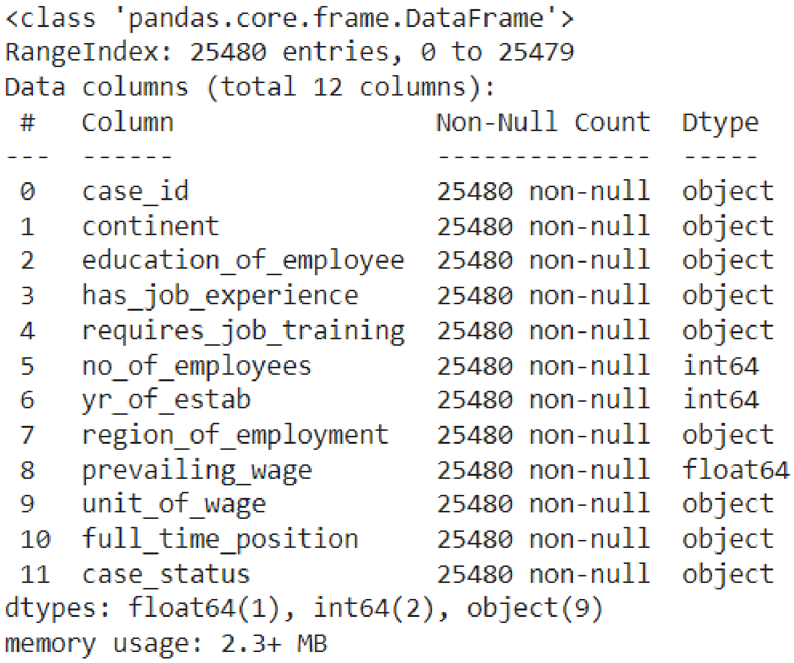
case_id	continent	education_of_employee	has_job_experience	requires_job_training	no_of_employees	yr_of_estab	region_of_employment	prevailing_wage	unit_of_wage
EZYV01	Asia	High School	N	N	14513	2007	West	592.203	Hour
EZYV02	Asia	Master's	Y	N	2412	2002	Northeast	83425.650	Year
EZYV03	Asia	Bachelor's	N	Y	44444	2008	West	122996.860	Year
EZYV04	Asia	Bachelor's	N	N	98	1897	West	83434.030	Year
EZYV05	Africa	Master's	Y	N	1082	2005	South	149907.390	Year

Figure 1: Data Overview

4.2. Check the structure of data

Shape of the dataset: 25480 rows and 12 columns

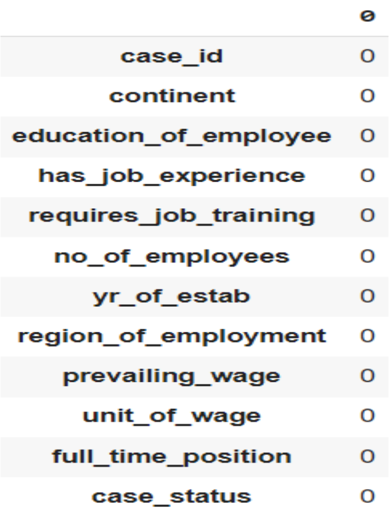
4.3. Check the types of the data



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25480 entries, 0 to 25479
Data columns (total 12 columns):
 #   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)
memory usage: 2.3+ MB
```

Figure 2: Datatypes

4.4. Check for and treat (if needed) missing values



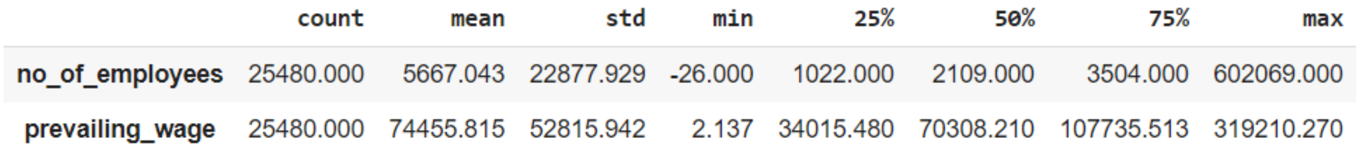
	0
case_id	0
continent	0
education_of_employee	0
has_job_experience	0
requires_job_training	0
no_of_employees	0
yr_of_estab	0
region_of_employment	0
prevailing_wage	0
unit_of_wage	0
full_time_position	0
case_status	0

Figure 3: Missing values check

4.5. Data Duplicates

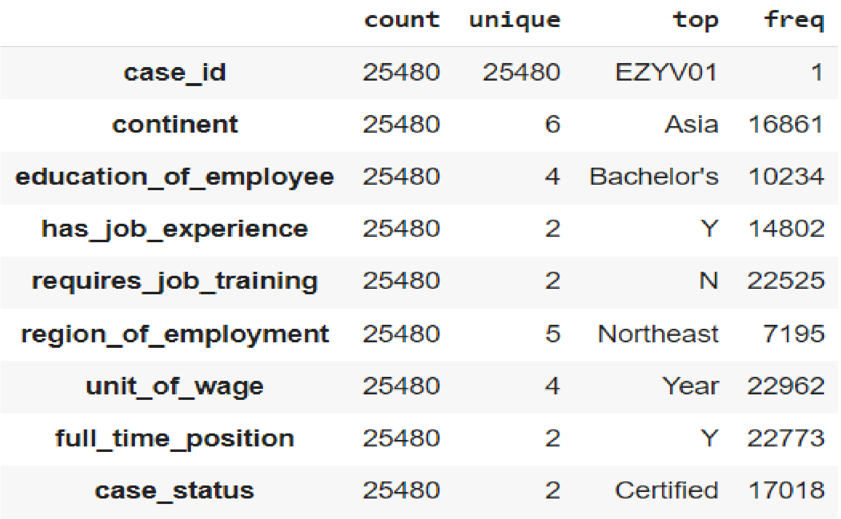
There are no duplicate rows.

4.6. Statistical Summary



	count	mean	std	min	25%	50%	75%	max
no_of_employees	25480.000	5667.043	22877.929	-26.000	1022.000	2109.000	3504.000	602069.000
prevailing_wage	25480.000	74455.815	52815.942	2.137	34015.480	70308.210	107735.513	319210.270

Figure 4: Statistical Summary - Numeric



	count	unique	top	freq
case_id	25480	25480	EZYV01	1
continent	25480	6	Asia	16861
education_of_employee	25480	4	Bachelor's	10234
has_job_experience	25480	2	Y	14802
requires_job_training	25480	2	N	22525
region_of_employment	25480	5	Northeast	7195
unit_of_wage	25480	4	Year	22962
full_time_position	25480	2	Y	22773
case_status	25480	2	Certified	17018

Figure 5: Statistical Summary - Object

4.7. Insights

- The dataset contains 25,480 cases with a wide range of employee numbers and prevailing wages, indicating significant variability.
- Data quality issues were identified, including negative employee counts.
- Most cases are from Asia (66.2%) and involve employees with a Bachelor's degree (40.2%).
- A significant majority of positions are full-time (89.4%), and prevailing wages are reported in various units (e.g., yearly, hourly), with annual wages being the most common (90.1%).
- Certified cases make up 66.8% of the dataset, with categorical variables converted to binary numeric values for further analysis.
- The code snippet effectively converts categorical variables into binary numeric variables, which is often necessary for machine learning models or statistical analysis.

5. Exploratory Data Analysis

5.1. Univariate Analysis

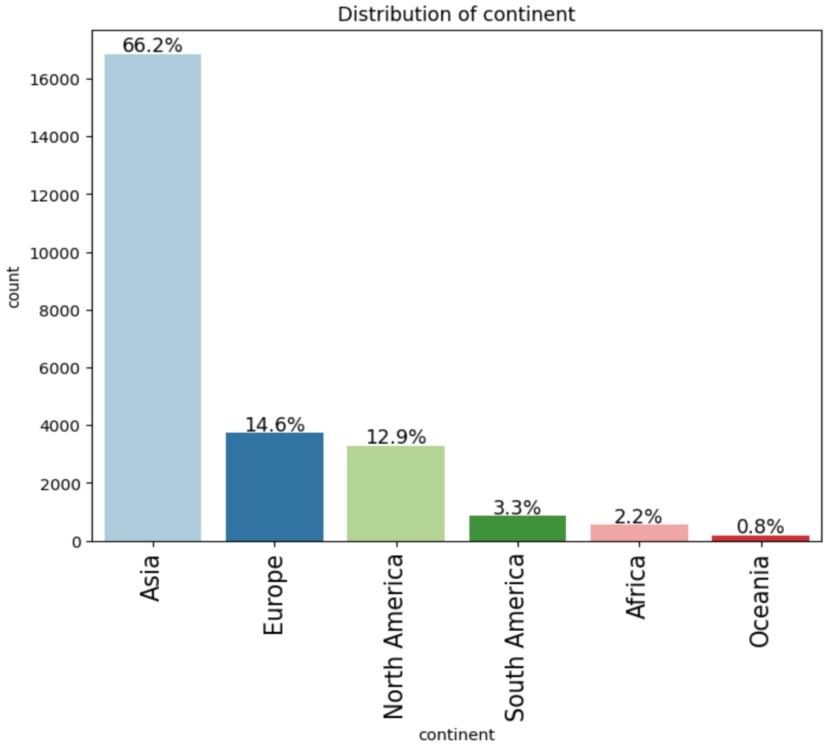


Figure 6: continent

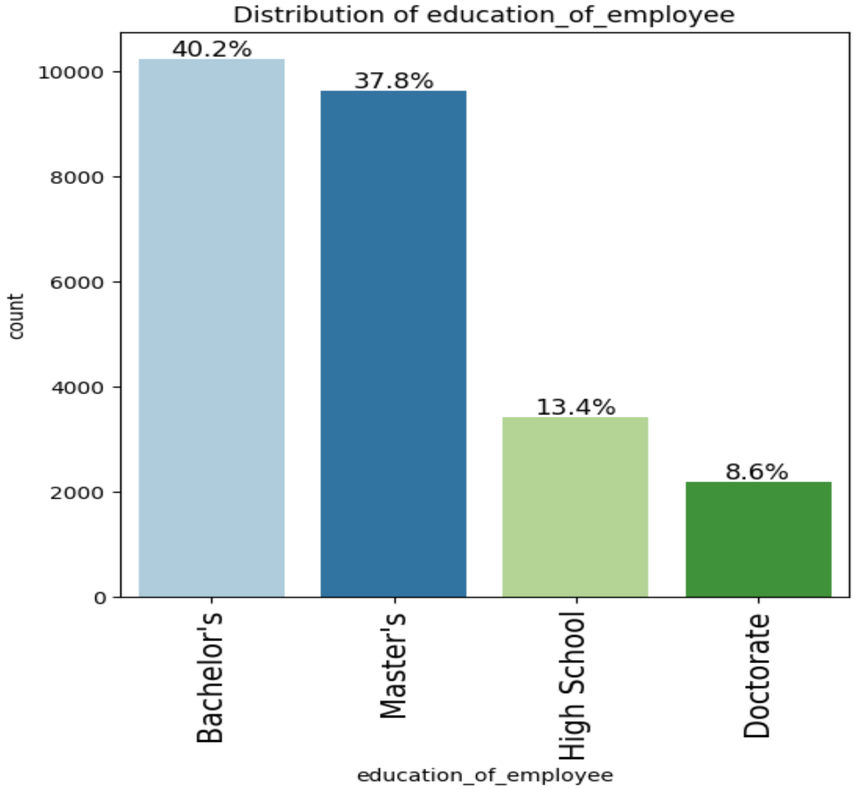


Figure 7: Education of Employee

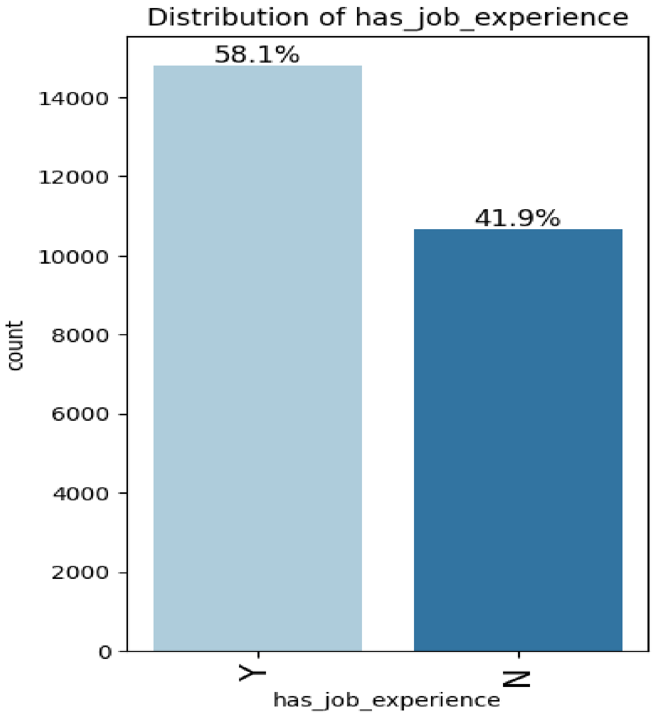


Figure 8: Job Experience

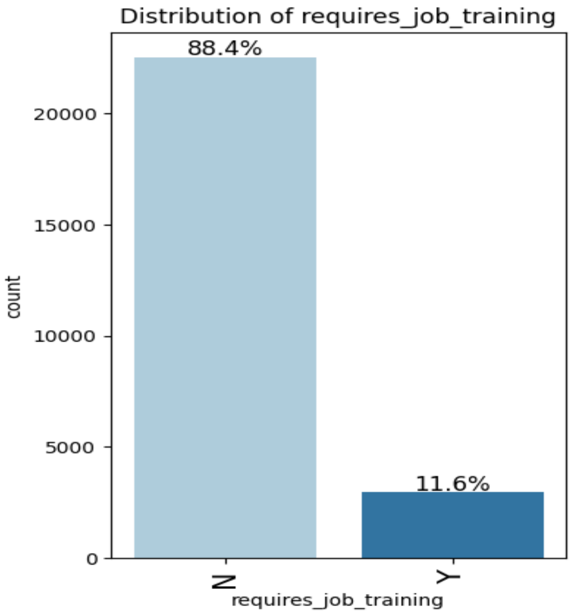


Figure 9: Requires Job Training

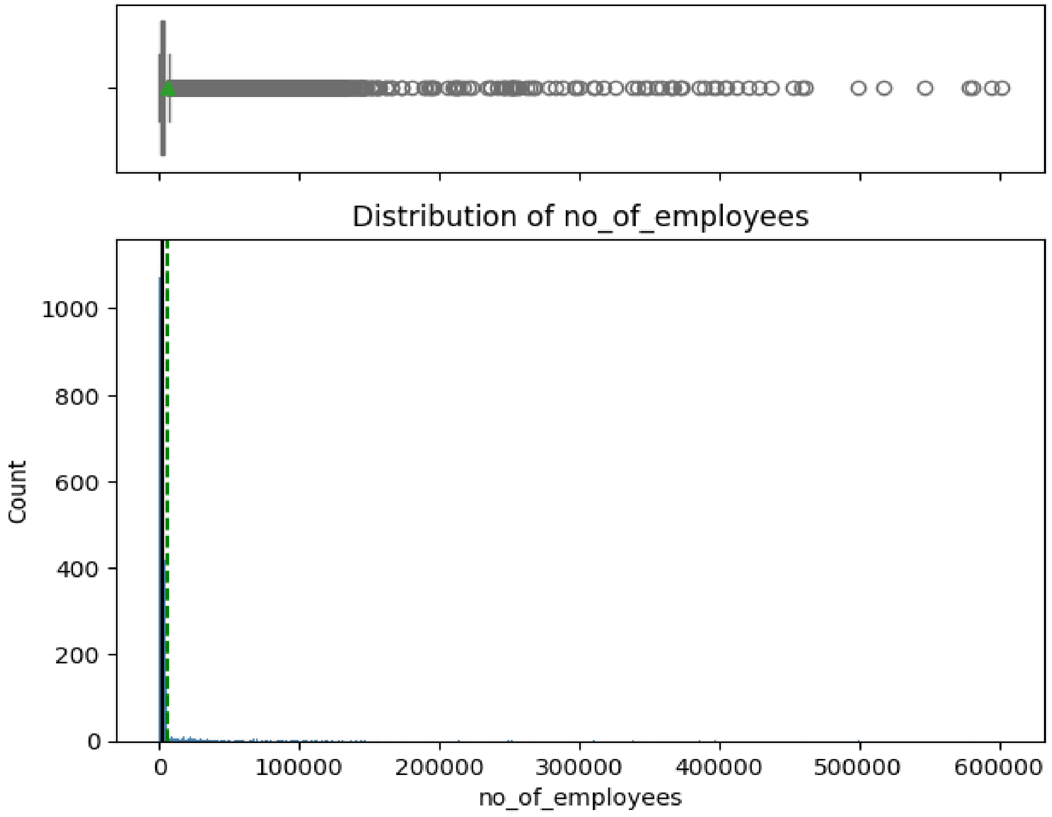


Figure 10: Number of Employees

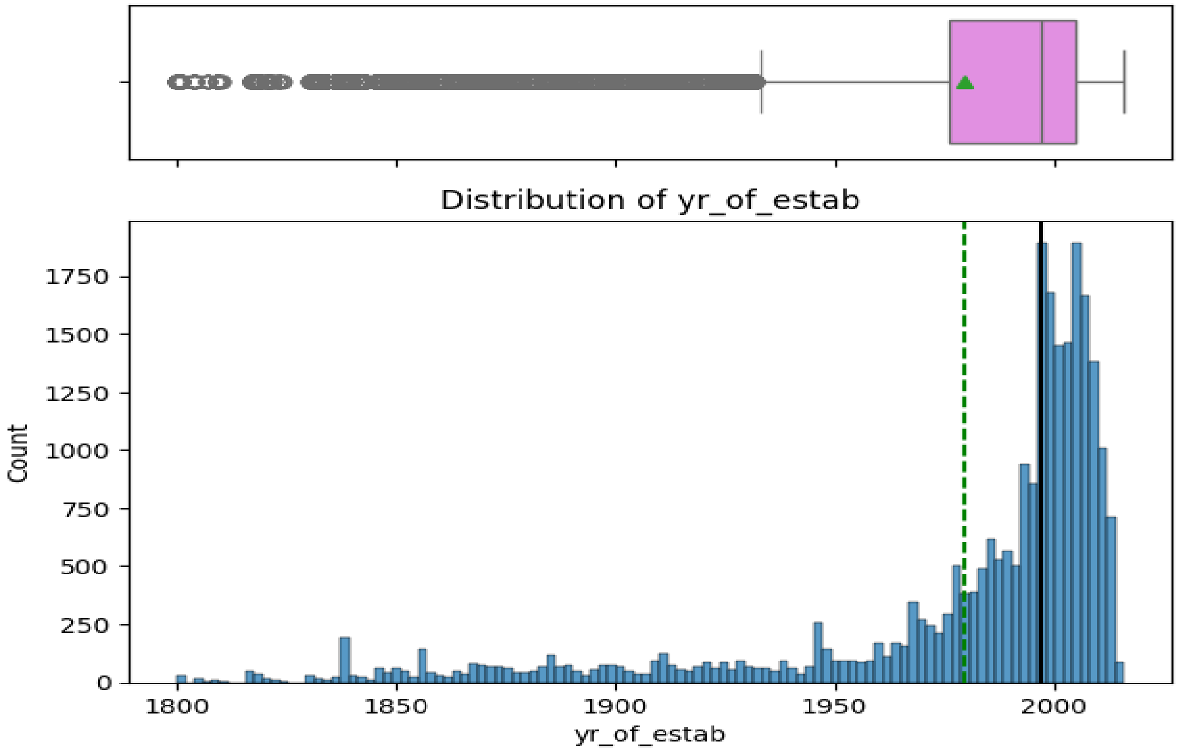


Figure 11: year of establishment

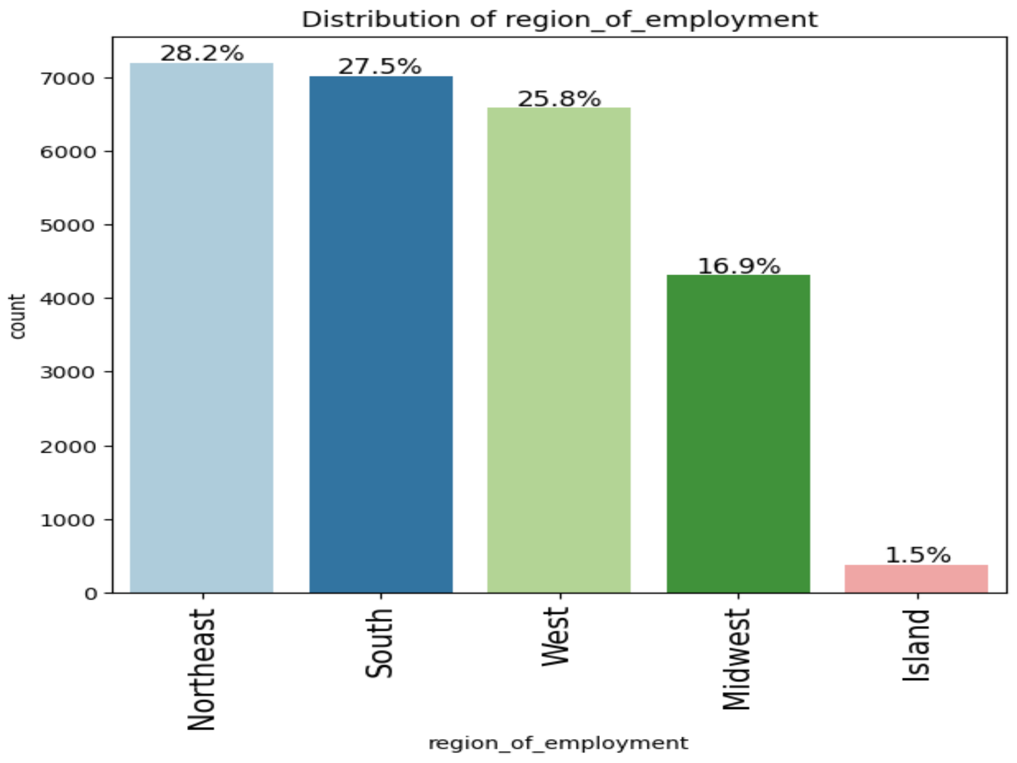


Figure 12: Region

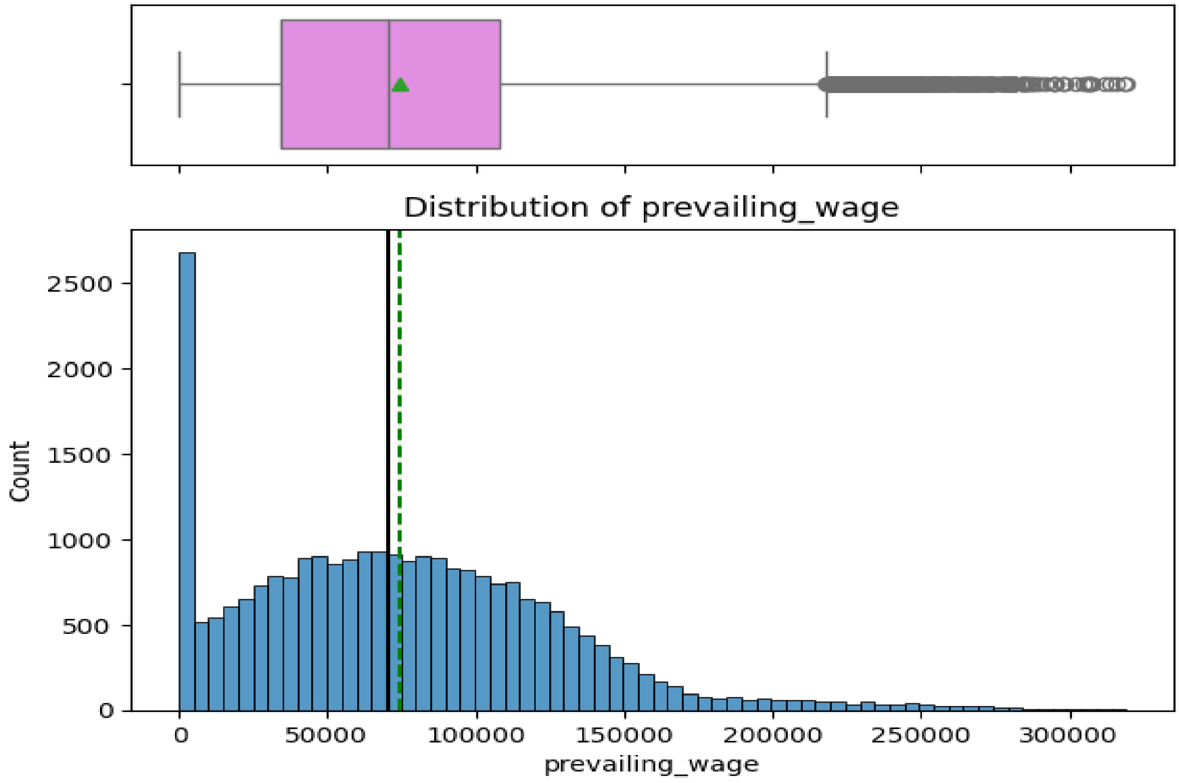


Figure 13: Wage

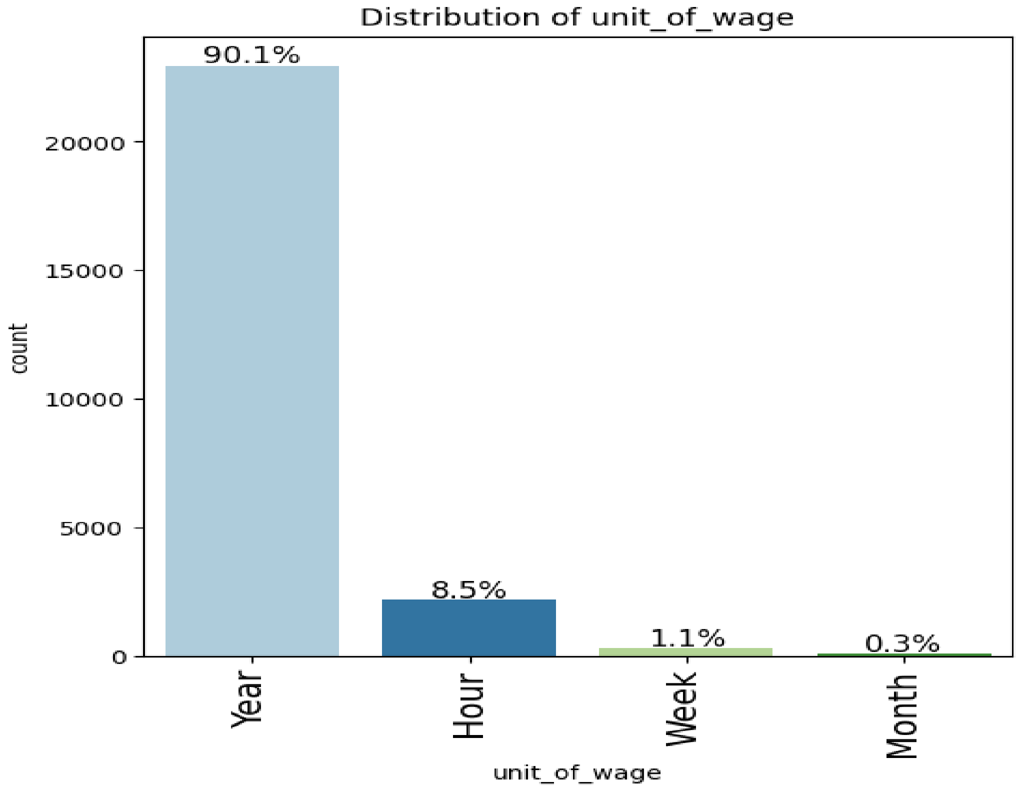


Figure 14: Wage Unit

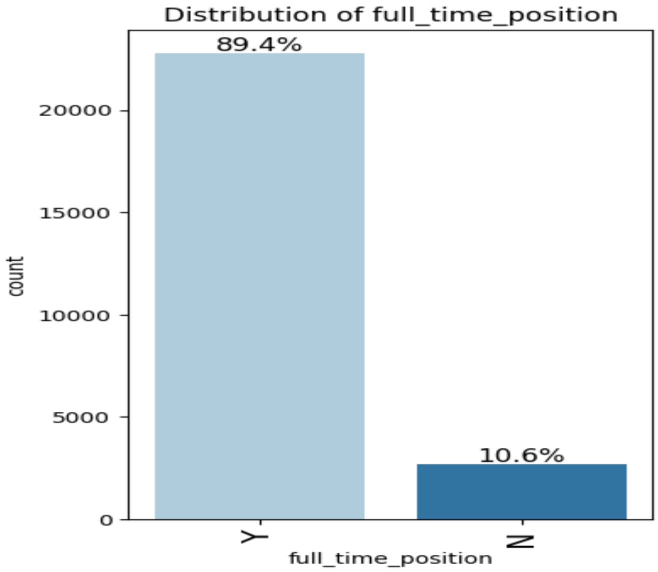


Figure 15: Full Time Position

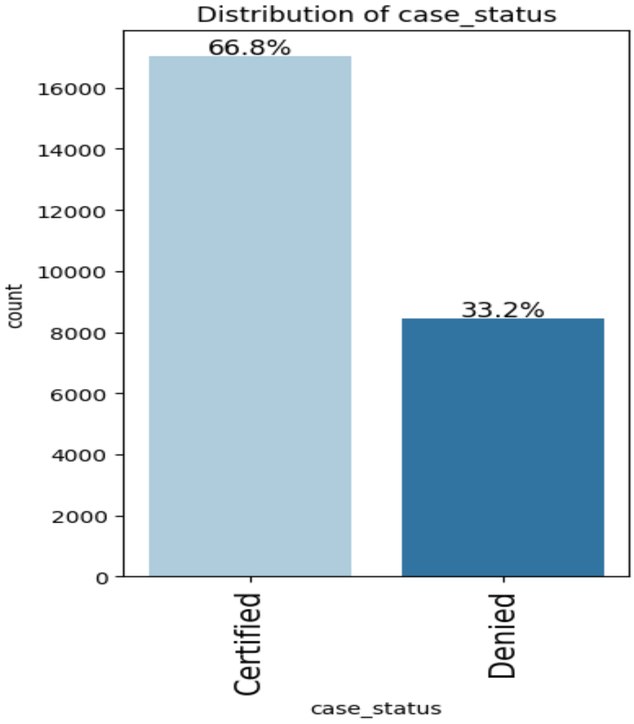


Figure 16: Case Status

Insights

- Number of employees distribution suggests that most companies have a relatively small number of employees, but there are a few large companies that significantly impact the overall distribution
- The histogram shows a significant peak in the number of companies established between 1950 and 2000. This suggests that a large portion of the dataset represents relatively newer companies.
- The majority of cases (66.2%) originate from Asia, and the most common educational background among employees is a Bachelor's degree (40.2%).
- Most positions are full-time (89.4%), and reported wages vary by unit (e.g., annual, hourly), with annual wages being the predominant measure (90.1%).
- Approximately two-thirds (66.8%) of the cases are certified, and categorical variables have been converted to binary numeric values for subsequent analysis

5.2. Bivariate Analysis

Correlation Check

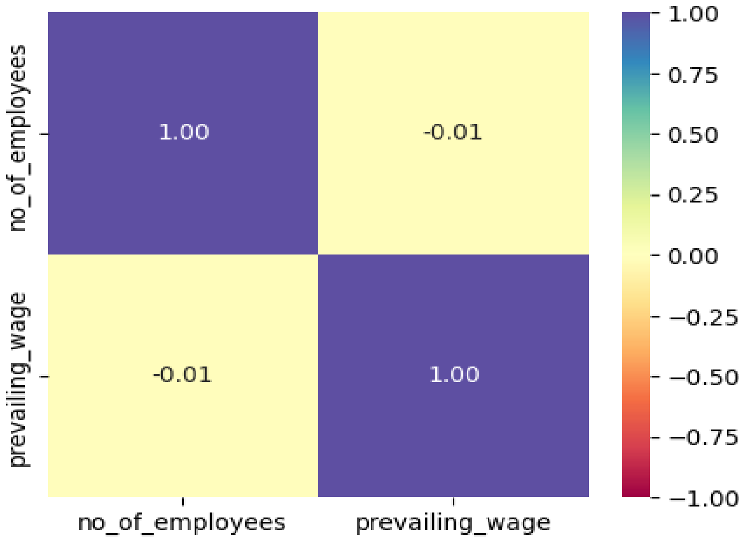


Figure 17: Heatmap

Insights

- Positive correlation between job experience and case status indicates that having job experience is strongly positively correlated with a certified case status.
- Positions requiring job training are more likely to be full-time positions.
- the number of employees and prevailing wage do not seem to be strong predictors of the other variables.

Wage vs Job Experience

Hourly

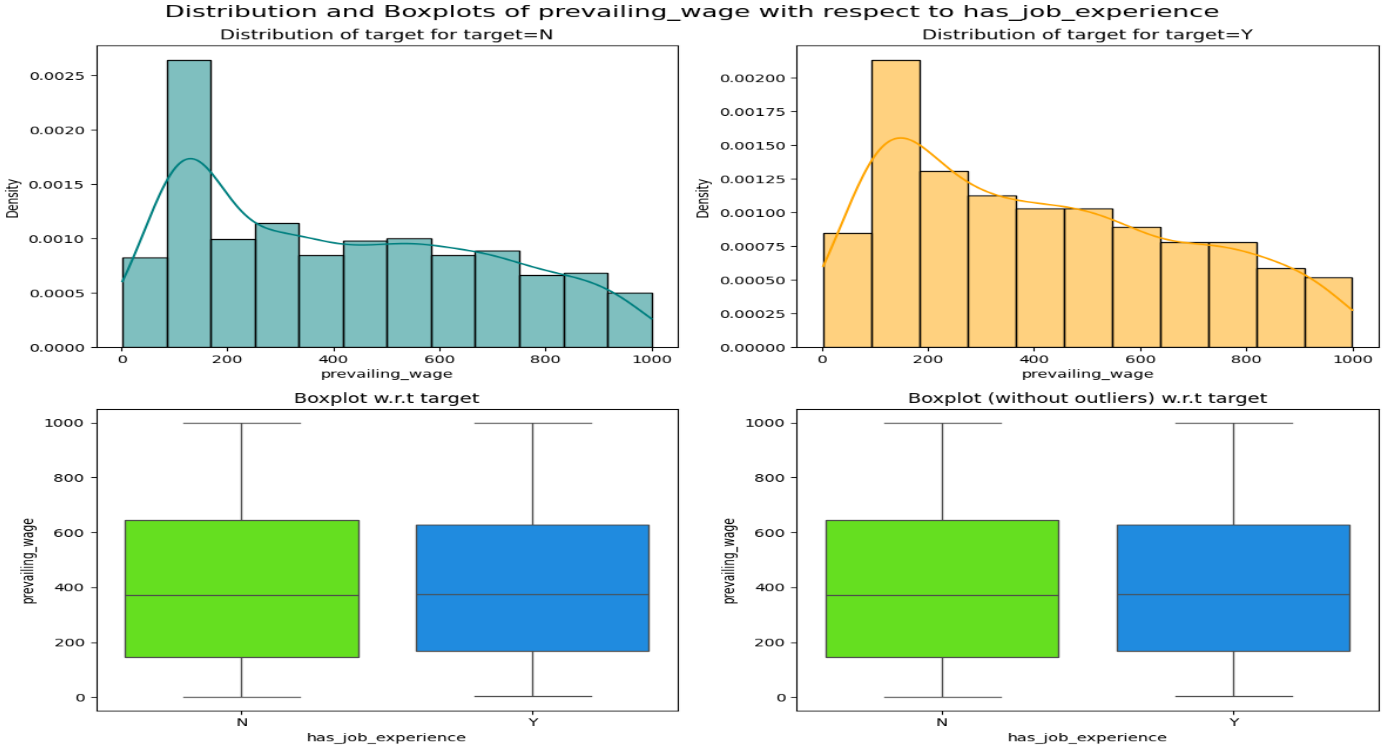


Figure 18: Hourly Wage vs. Job Experience

Weekly

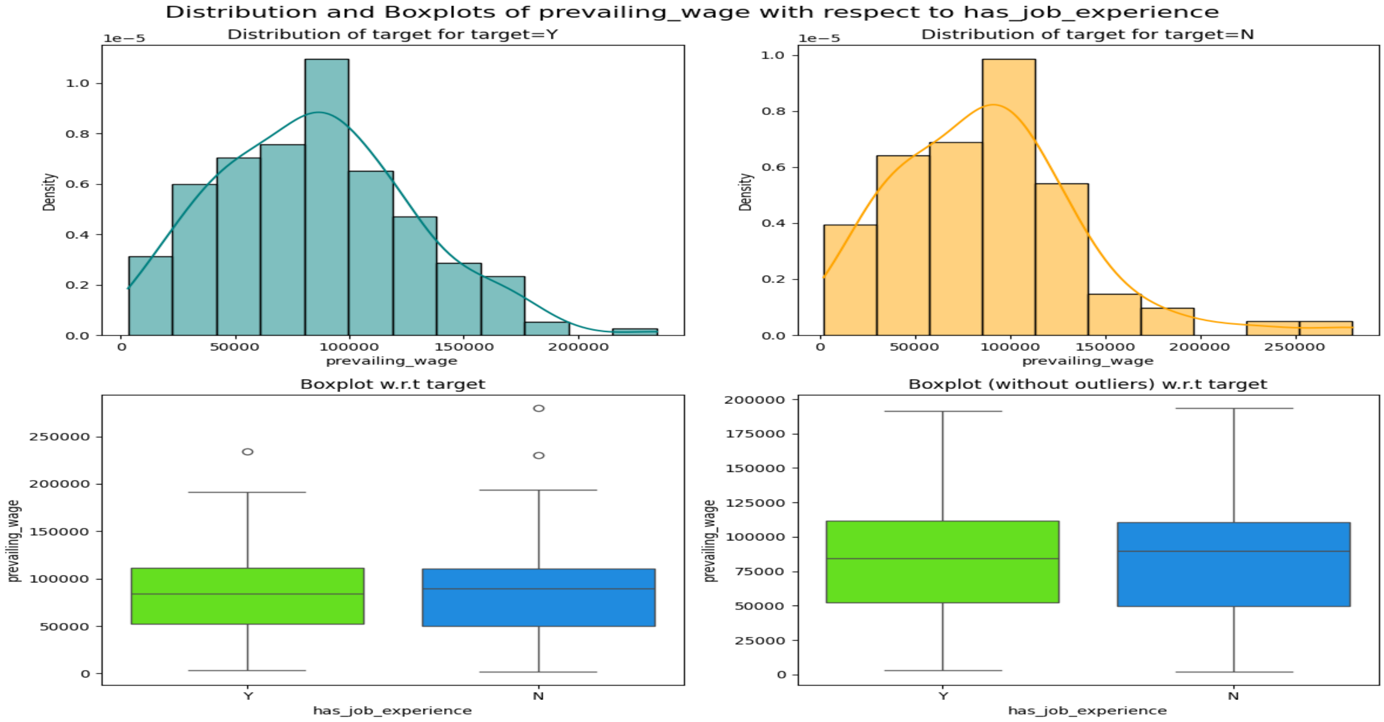


Figure 19: Weekly Wage vs Job Experience

Monthly

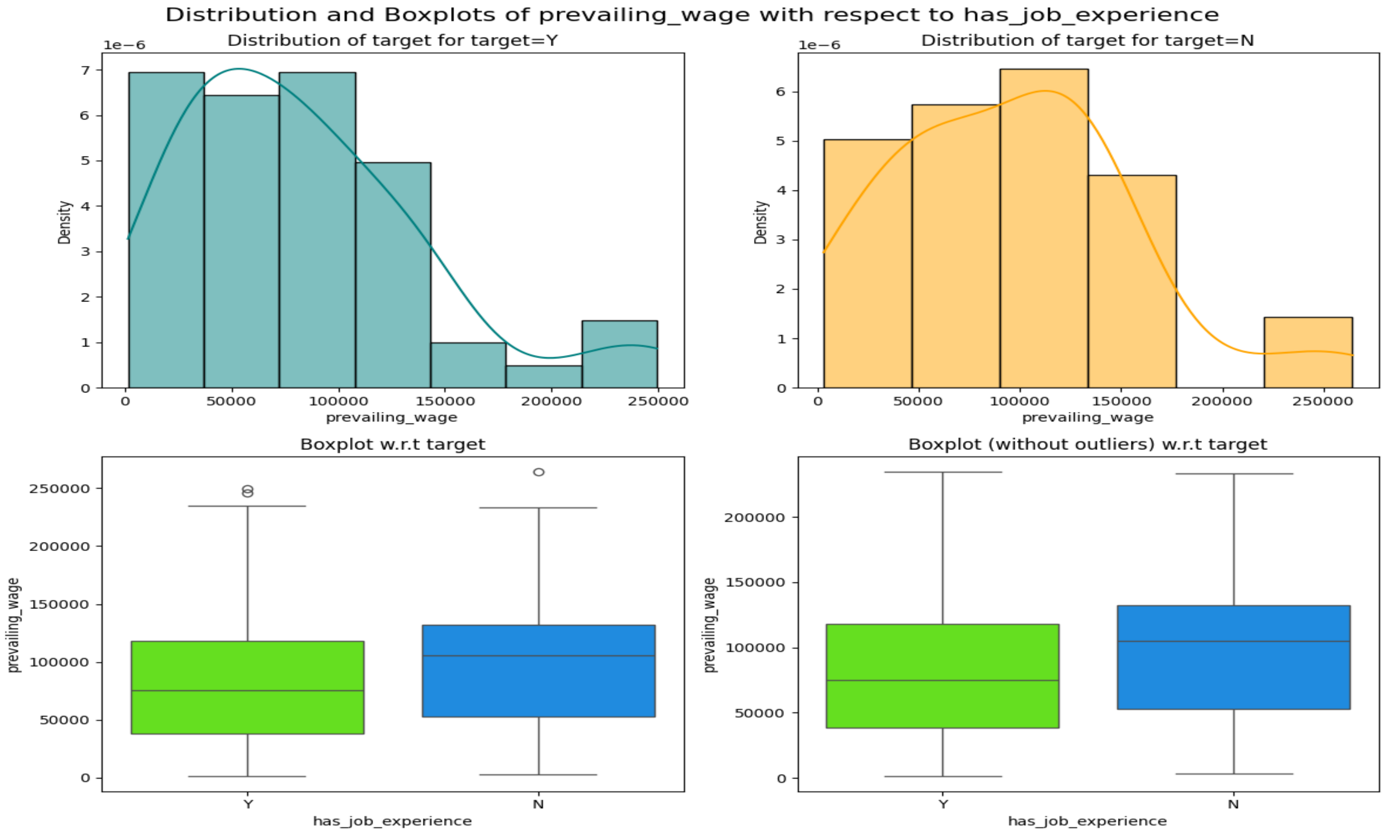


Figure 20: Monthly Wage vs Job Experience

Yearly

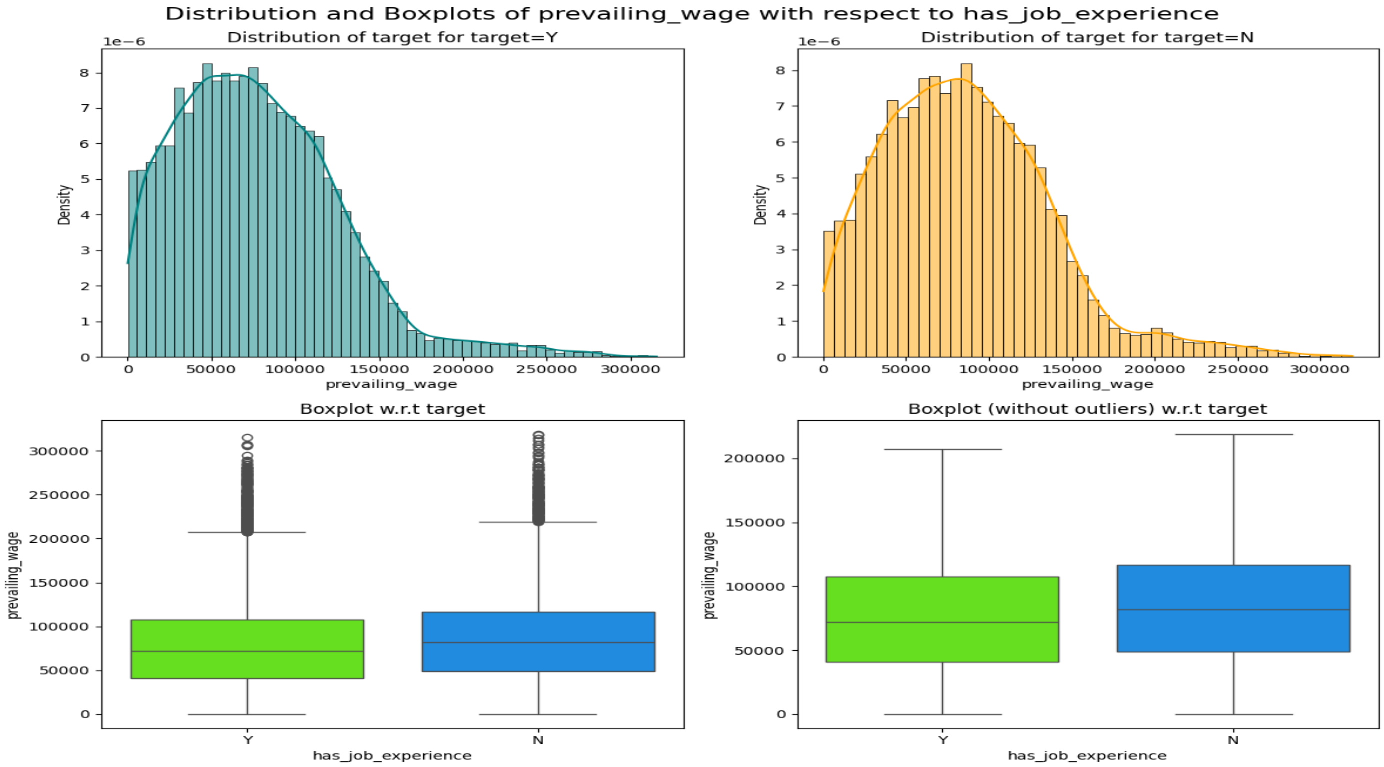


Figure 21: Yearly Wage vs Job Experience

Wage vs Job Training

Hourly

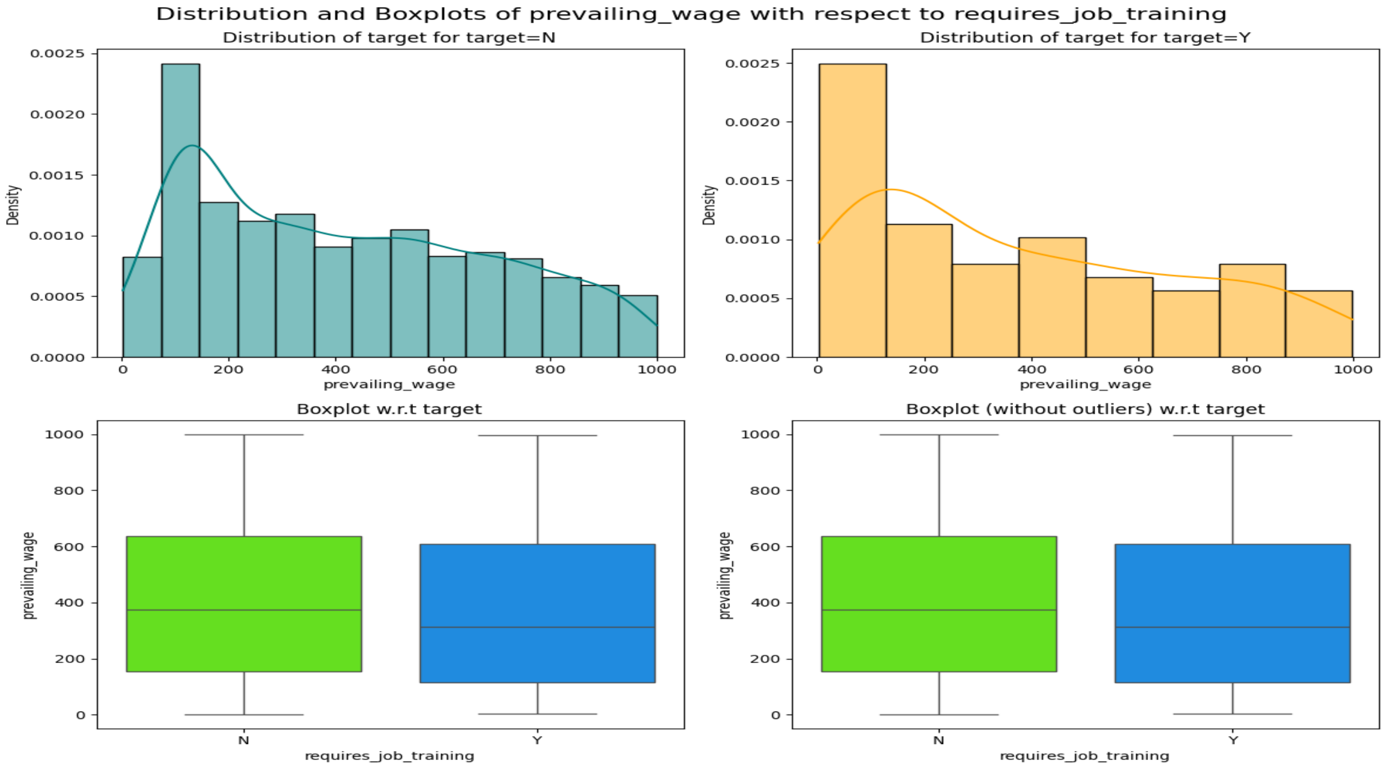


Figure 22: Hourly Wage vs. Job Training

Weekly

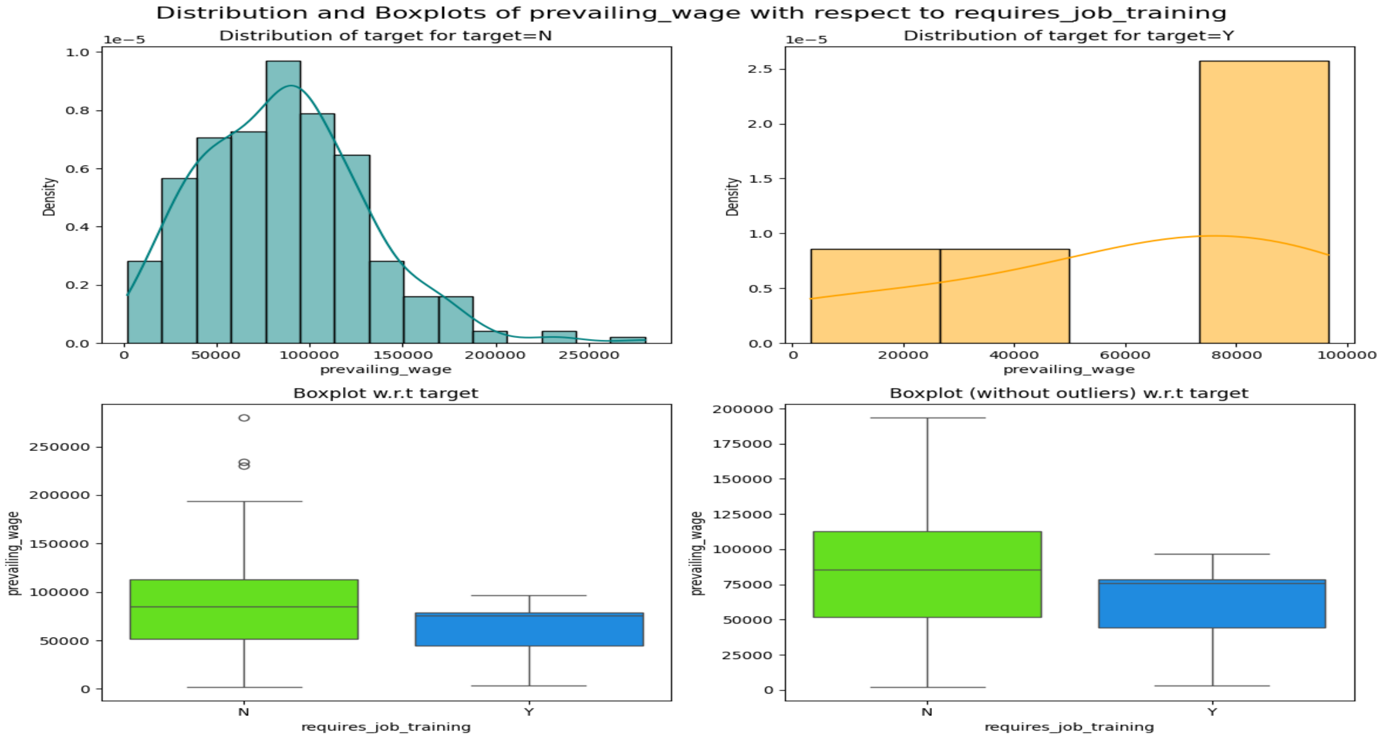


Figure 23: Weekly Wage vs Job Training

Monthly

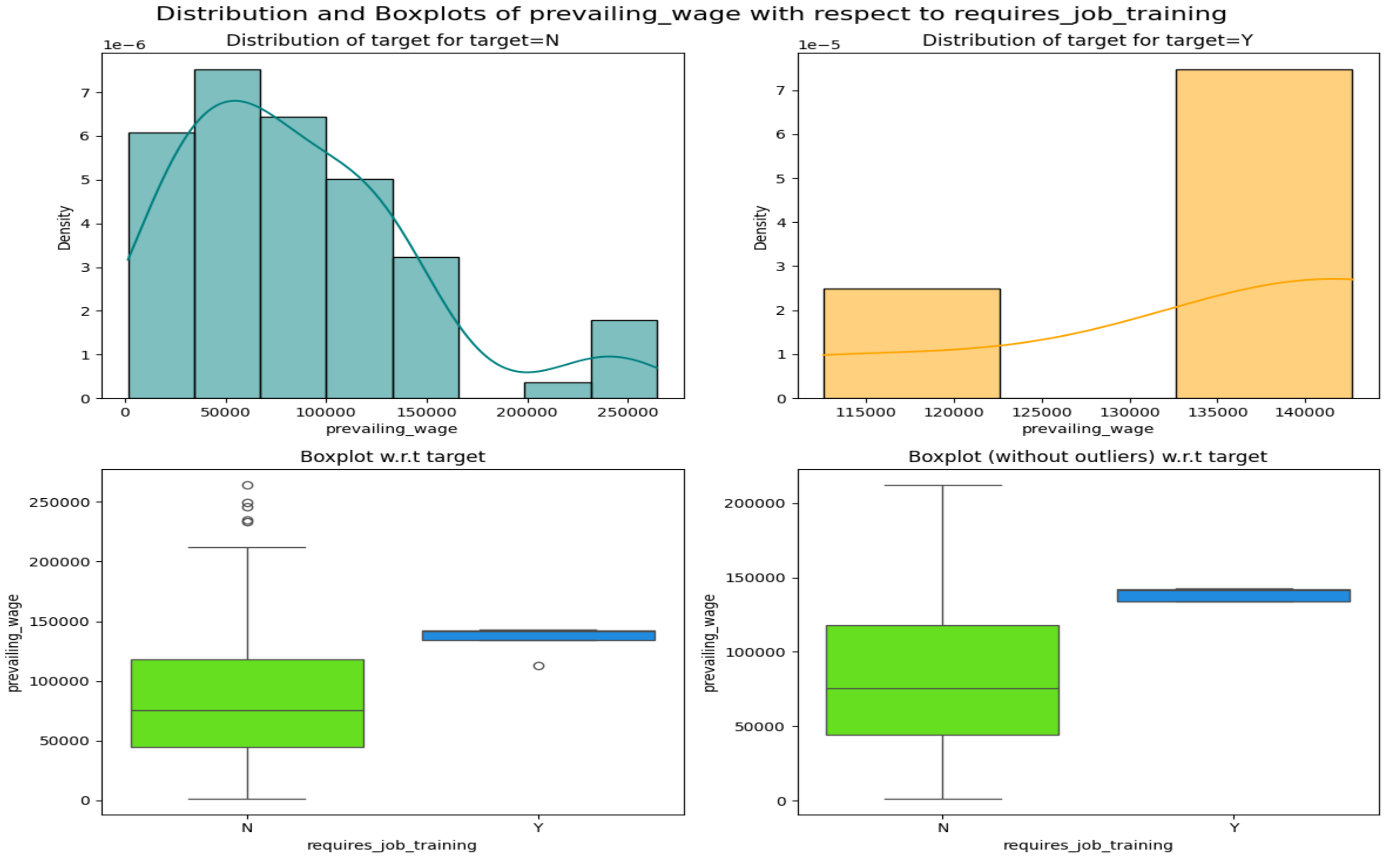


Figure 24: Monthly Wage vs Job Training

Yearly

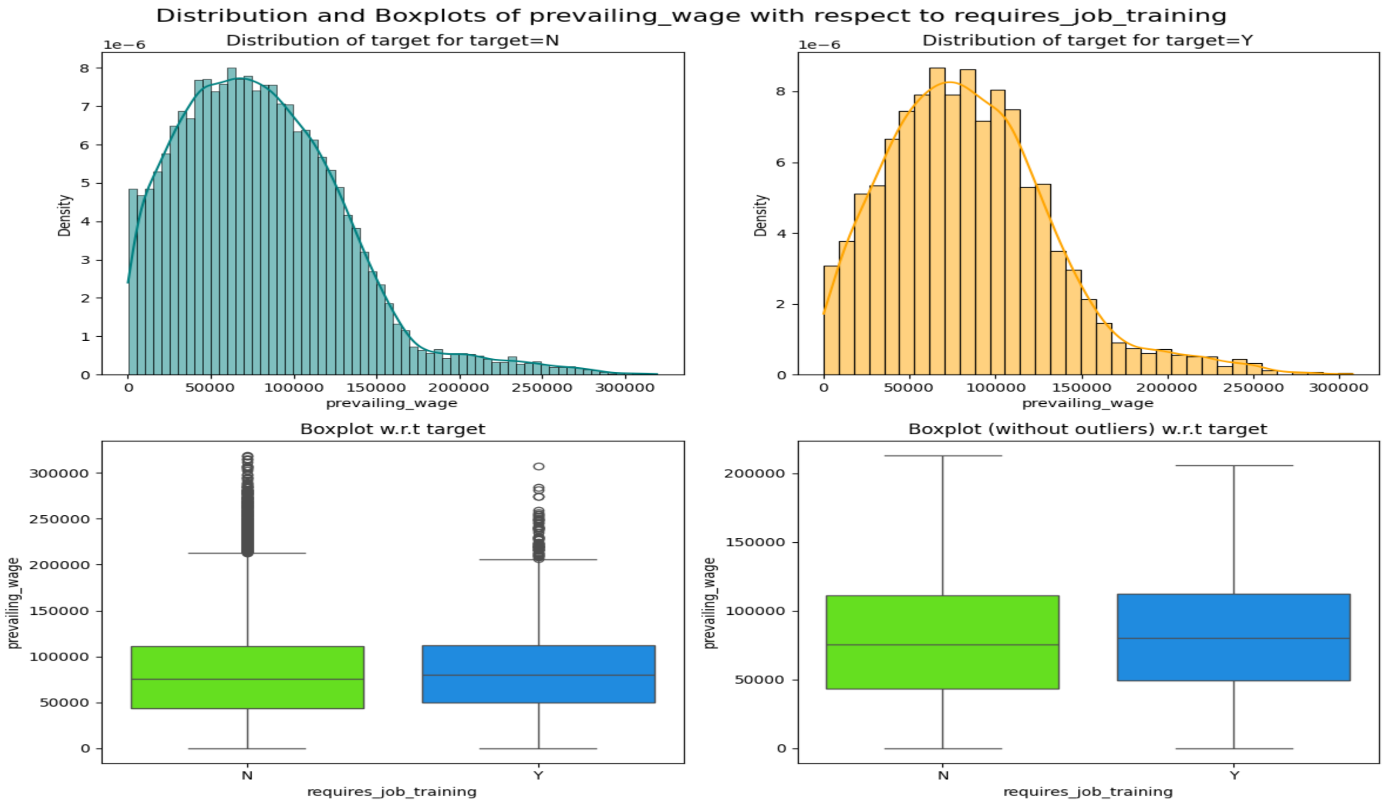


Figure 25: Yearly Wage vs Job Training

Wage vs Full Time Position

Hourly

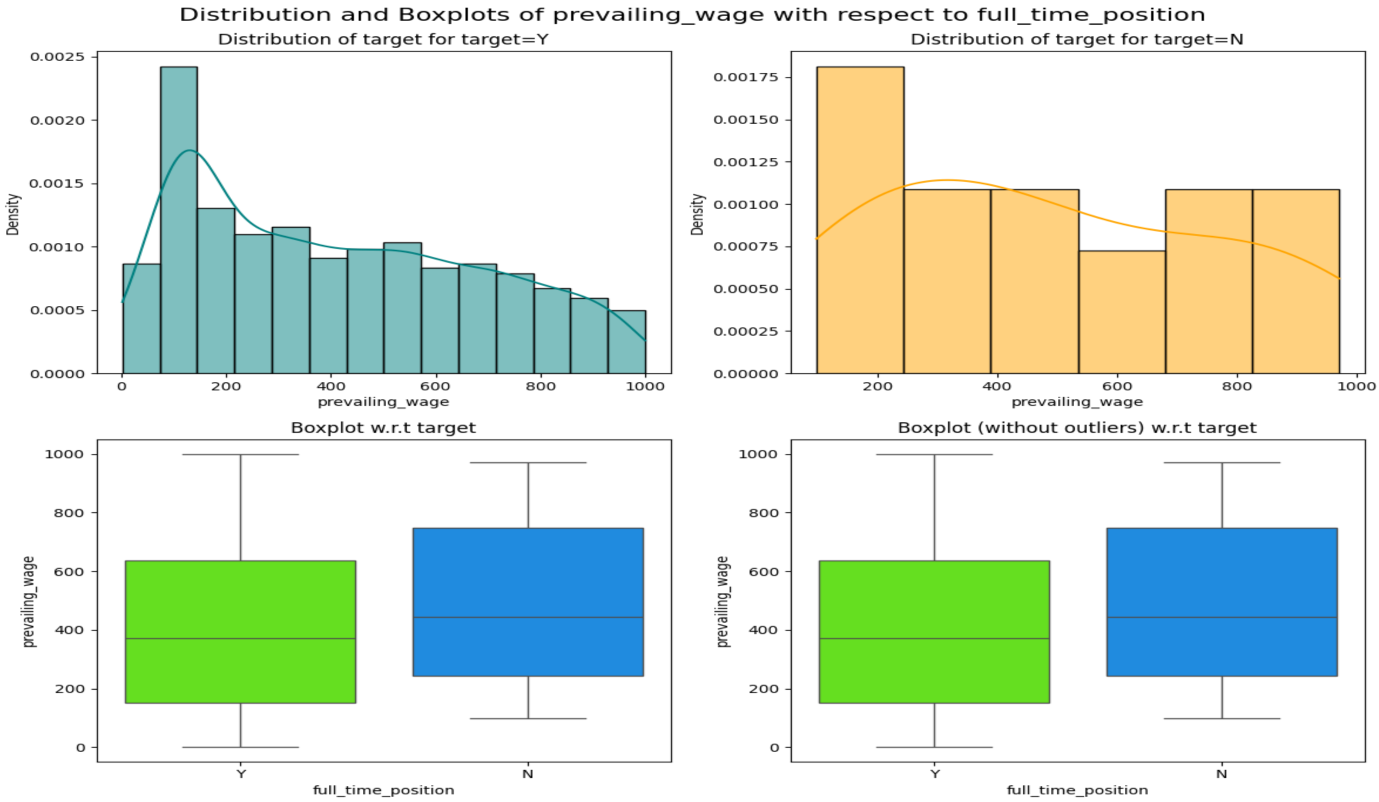


Figure 26: Hourly Wage vs. Full Time Position

Weekly

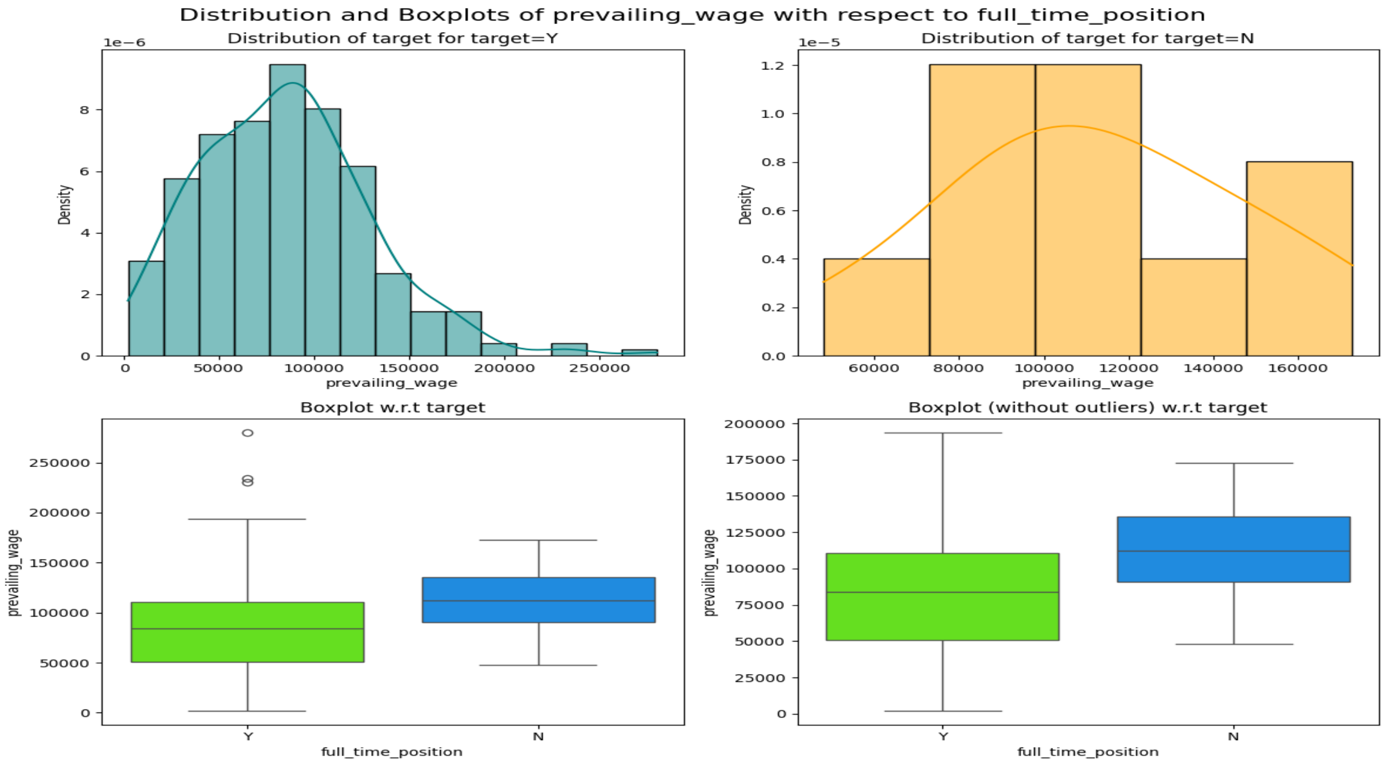


Figure 27: Weekly Wage vs Full Time Position

Monthly

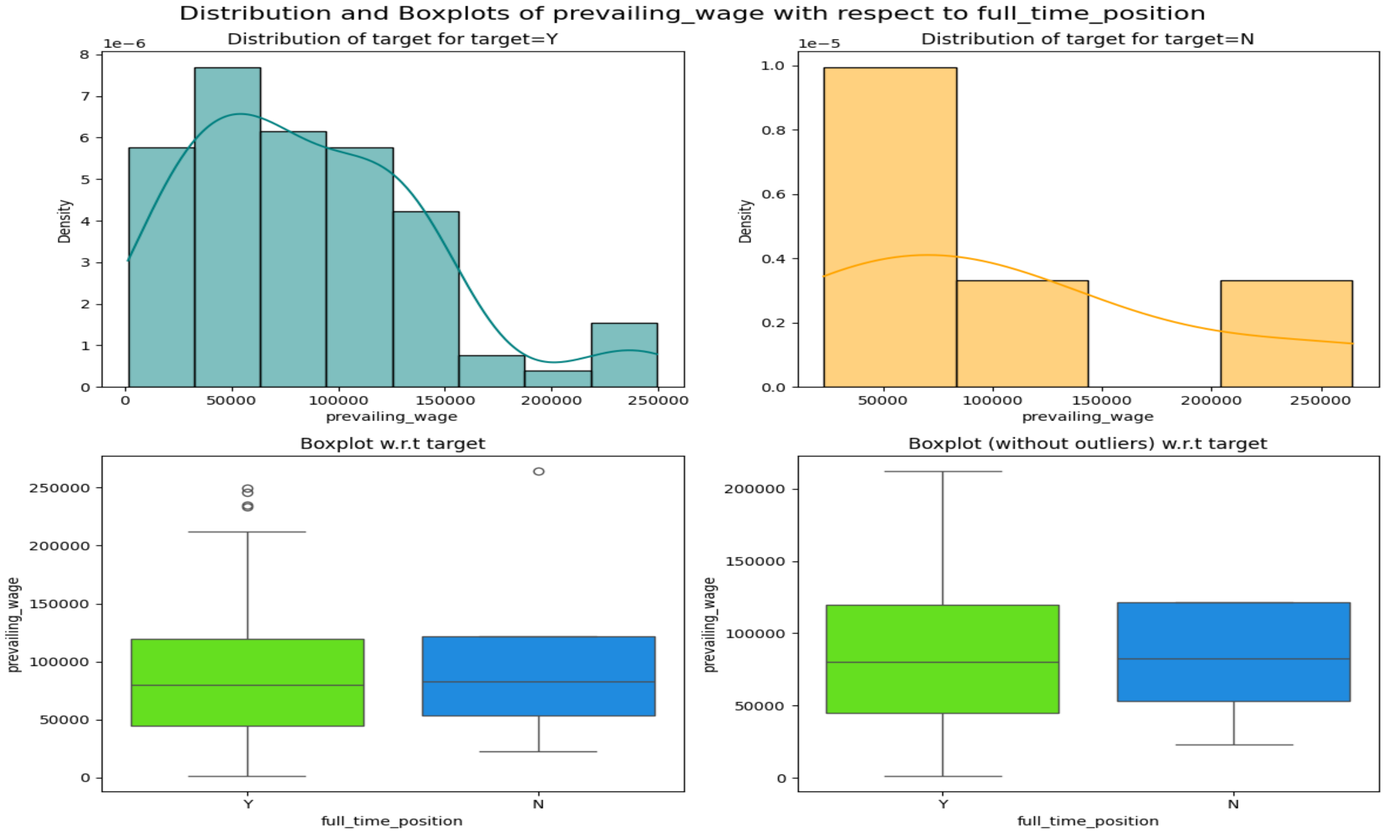


Figure 28: Monthly Wage vs Full Time Position

Yearly

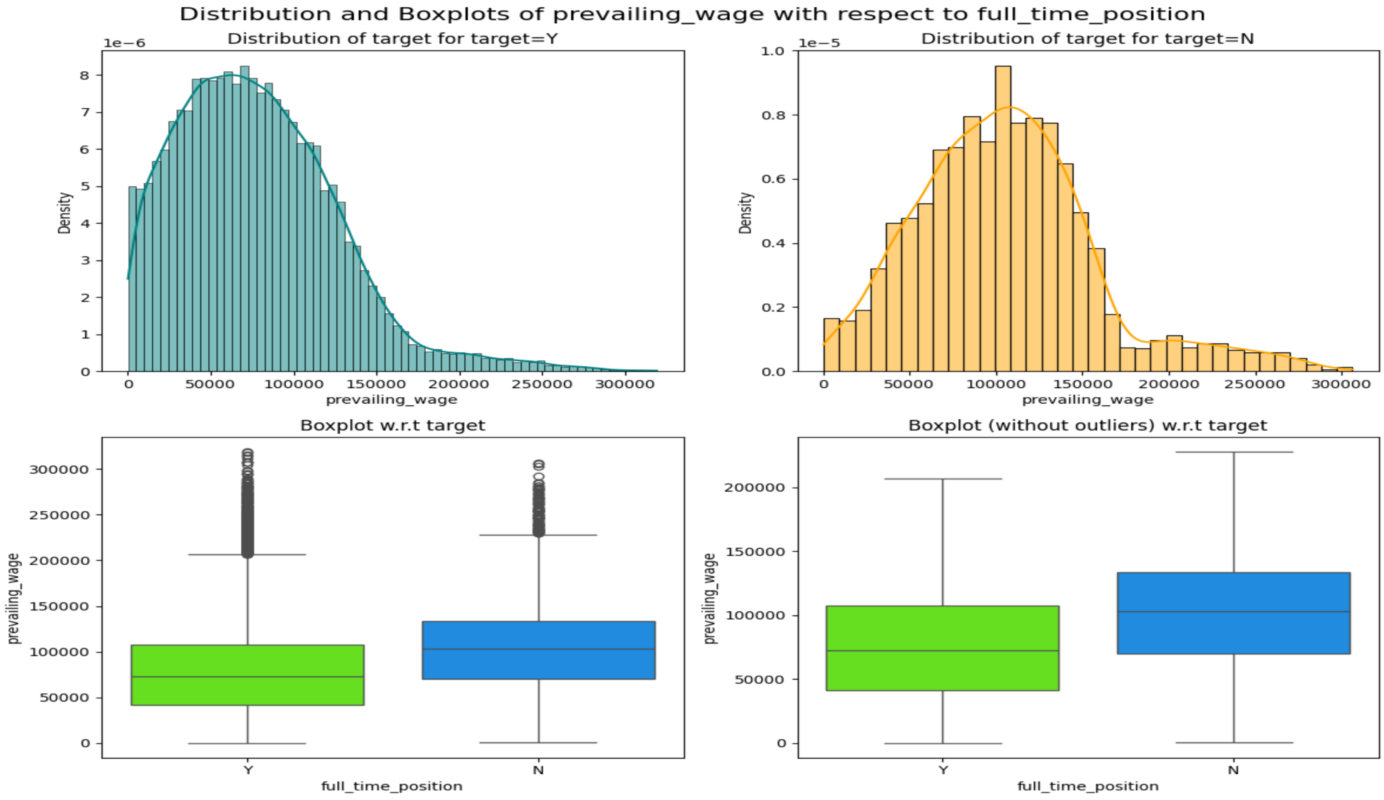


Figure 29: Yearly Wage vs Full Time Position

Wage vs Case Status

Hourly

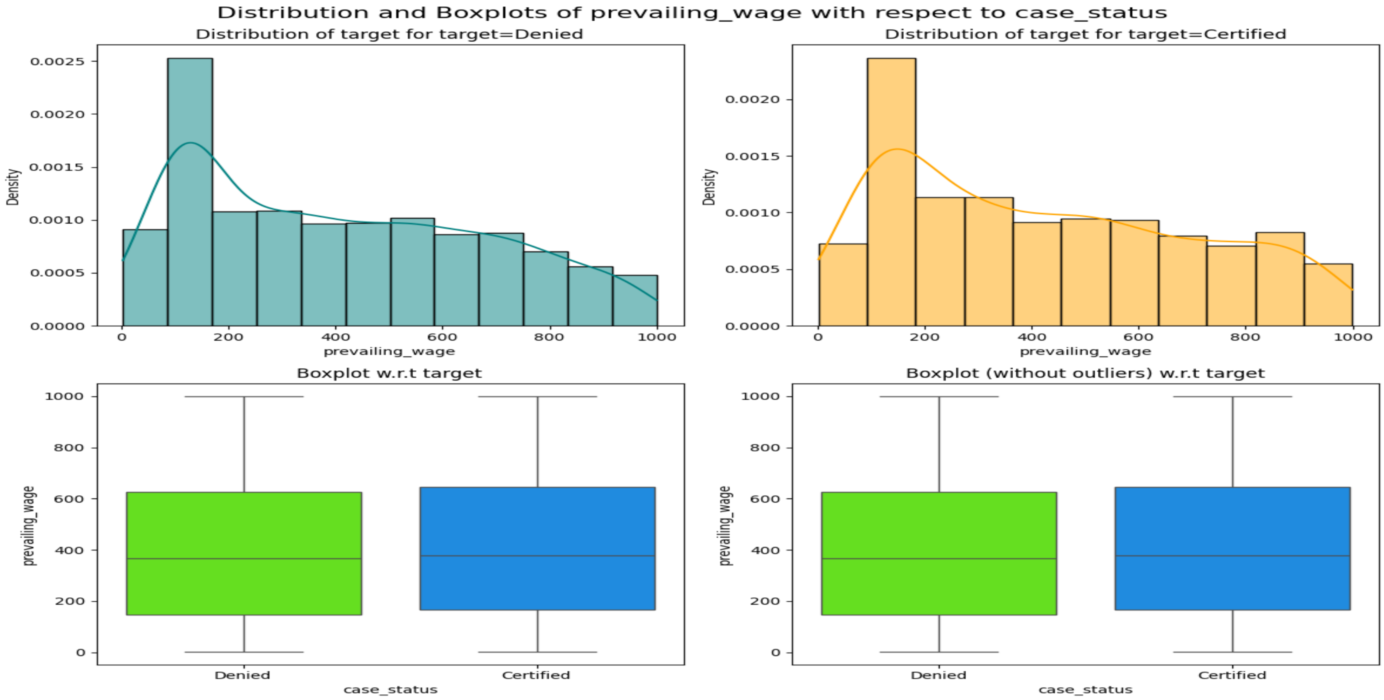


Figure 30: Hourly Wage vs. Case Status

Weekly

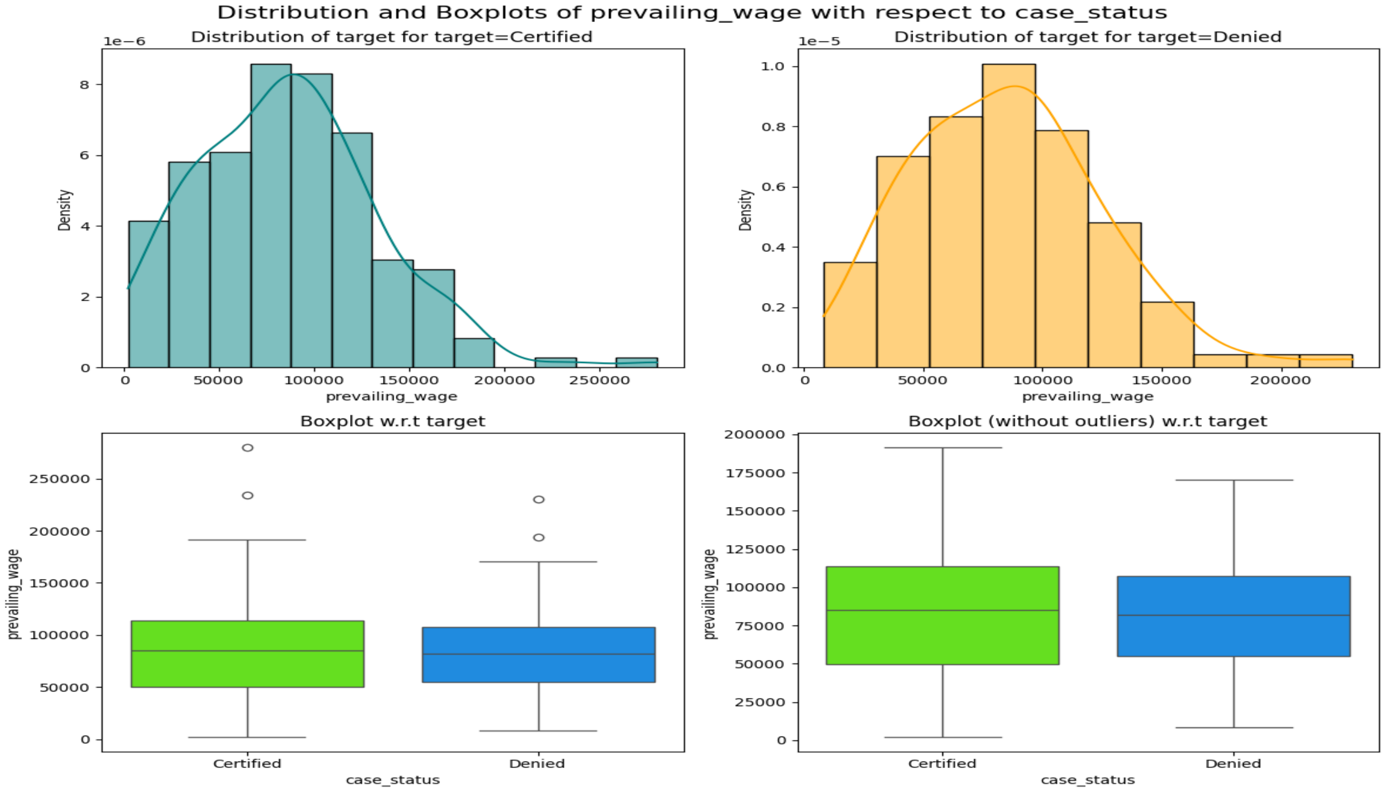


Figure 31: Weekly Wage vs Case Status

Monthly

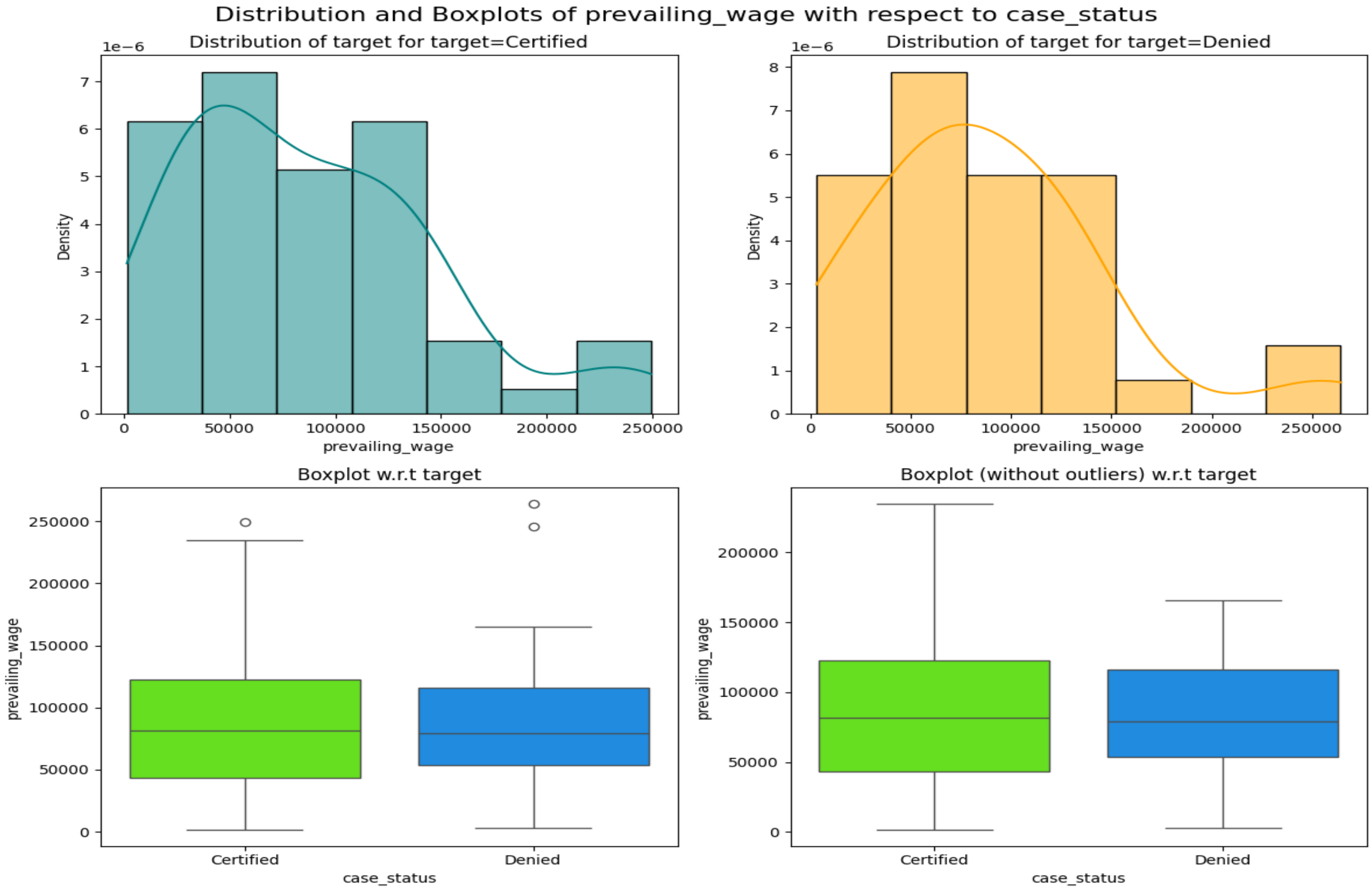


Figure 32: Monthly Wage vs Case Status

Yearly

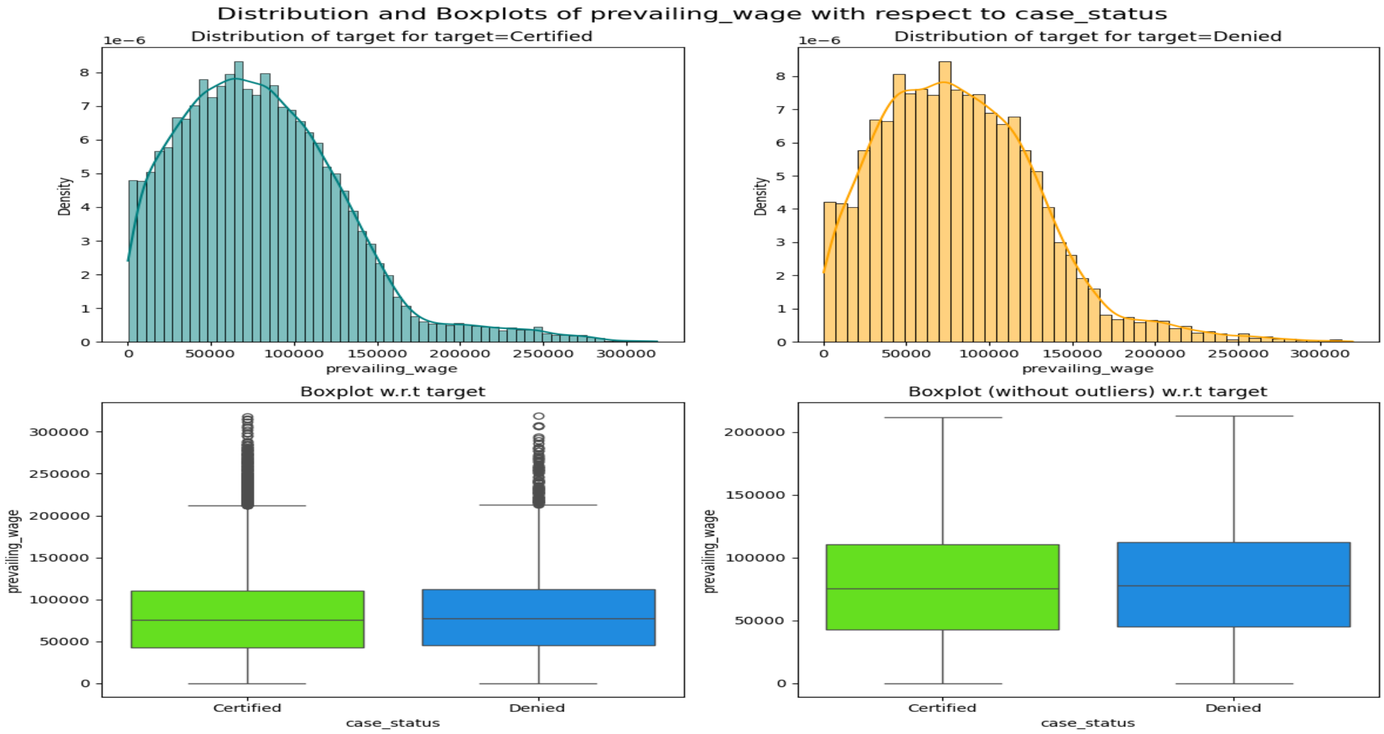


Figure 33: Yearly Wage vs Case Status

Case Status vs Job Experience

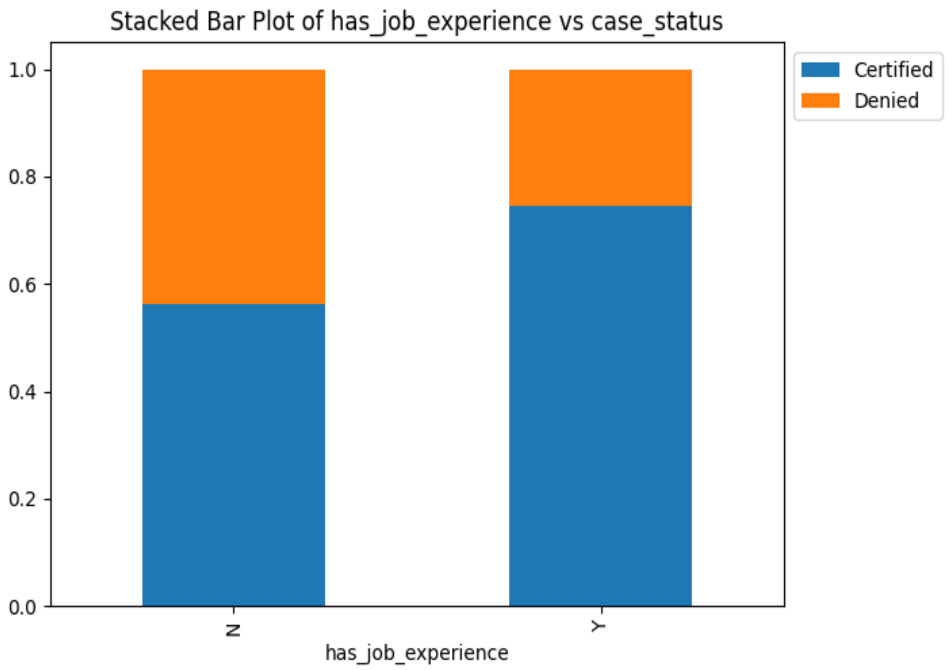


Figure 34: Job Experience vs. Case Status

Case Status vs Job Training

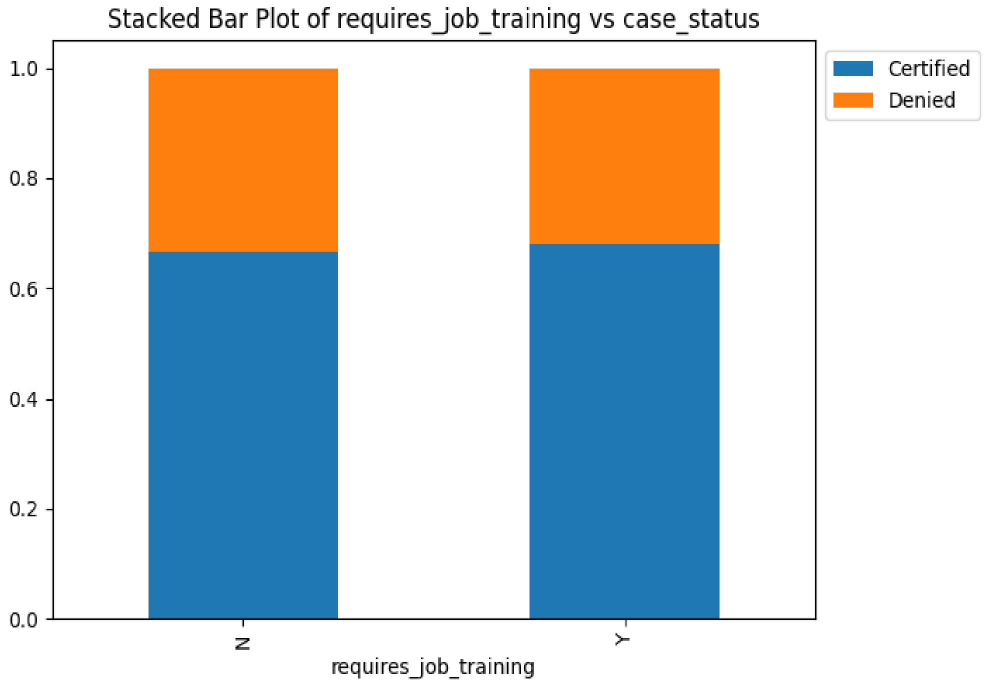


Figure 35: Job Training vs Case Status

Case Status vs Full Time Position

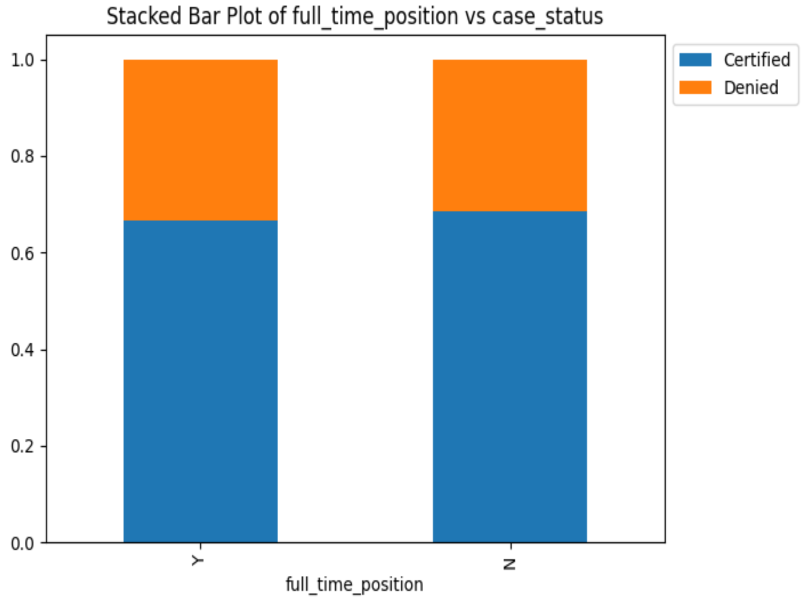


Figure 36: Full Time Position vs Case Status

Case Status vs Region

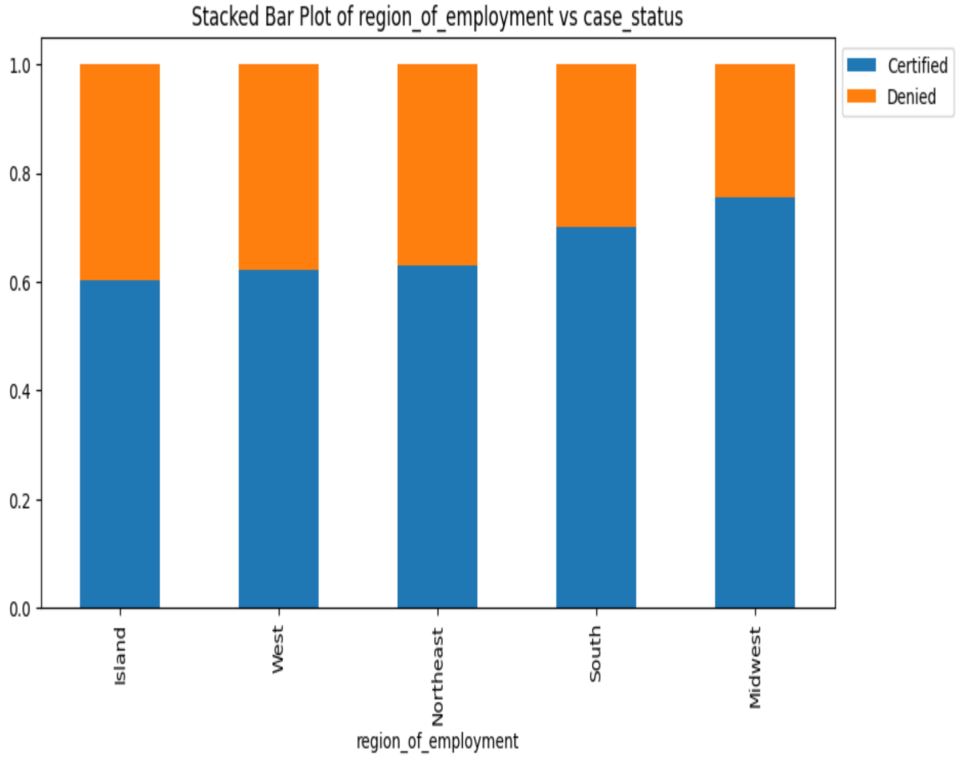


Figure 37: Region vs Case Status

Case Status vs Education

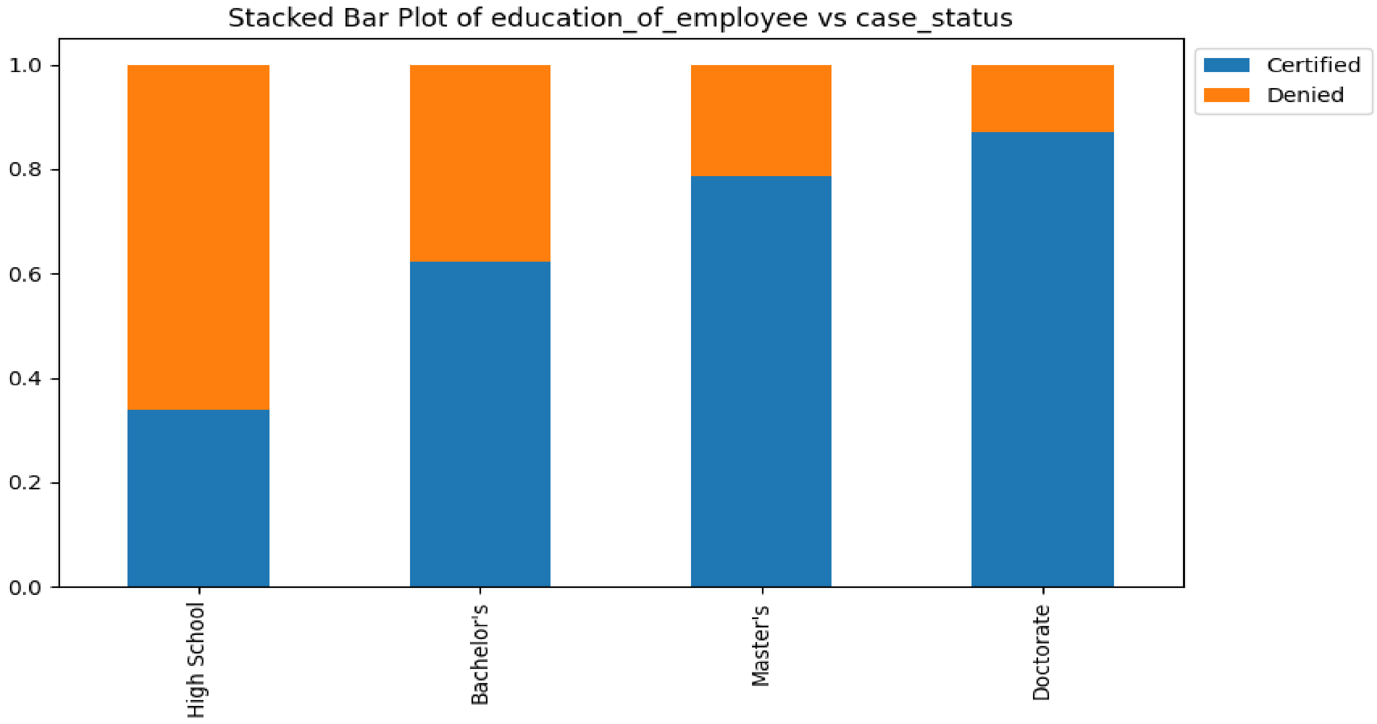


Figure 38: Education vs Case Status

Case Status vs Continent

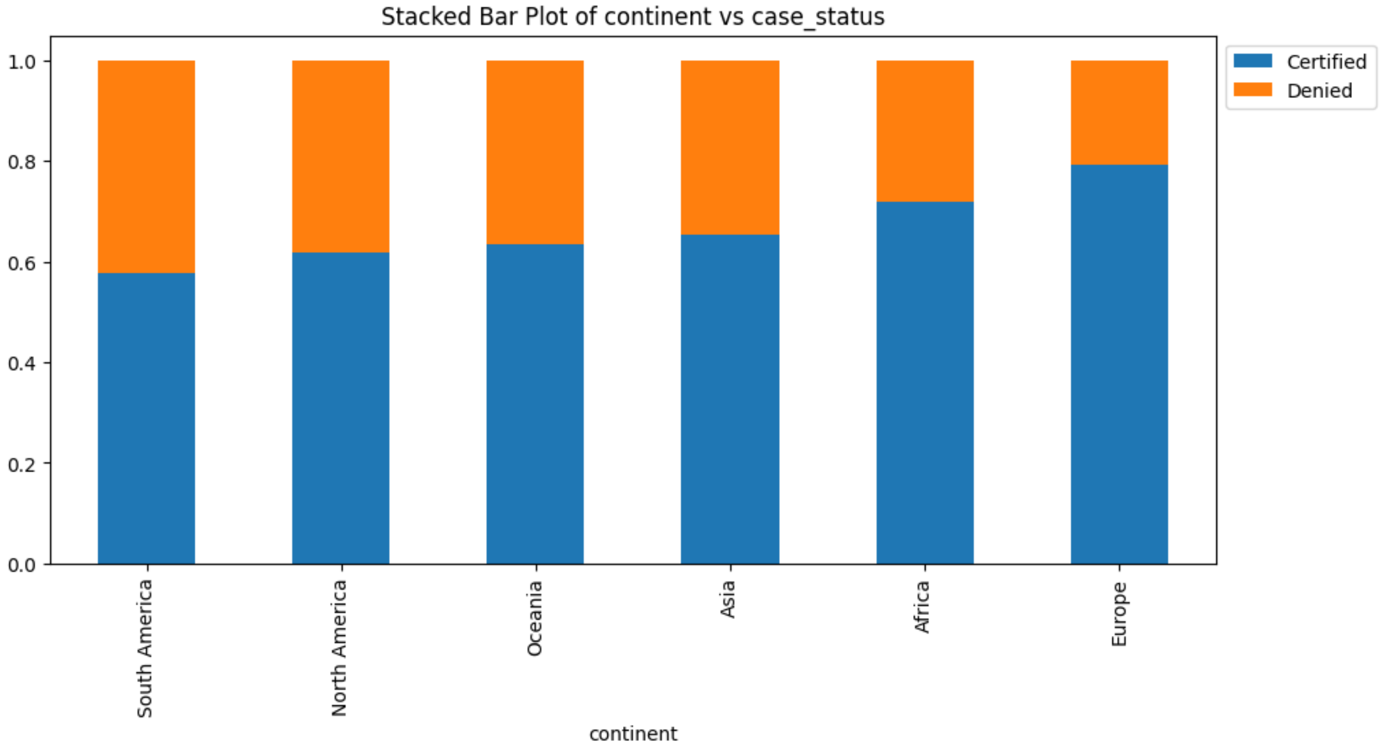


Figure 39: Continent vs Case Status

Insights

- Wage Differences: There is minimal variation in wages between individuals with job experience and those without.
- Impact of Job Training: Weekly wages exhibit a slight variation for individuals requiring job training.
- Full-Time Employment: Individuals in full-time positions earn marginally more compared to those in non-full-time roles.
- Certification Rates: The likelihood of visa case certification is higher for individuals with job experience.
- Job Training and Full-Time Positions: The requirement for job training and full-time employment status do not significantly influence the likelihood of visa case certification or denial.
- Regional Variations: Visa case denials are less common in the Midwest and island regions; however, these areas report fewer cases overall.
- Educational Impact: Higher educational attainment correlates with increased chances of visa case certification. Specifically, doctorate holders experience fewer denials, whereas high school graduates face higher denial rates.
- European Trends: Europe exhibits lower rejection rates for visa cases. Wage Influence: Wage levels have a limited effect on visa case status.

6. Data Preprocessing

6.1. Missing Value treatment

There are no missing values.

6.2. Duplicate value check

There are no duplicate rows.

6.3. Outlier Detection

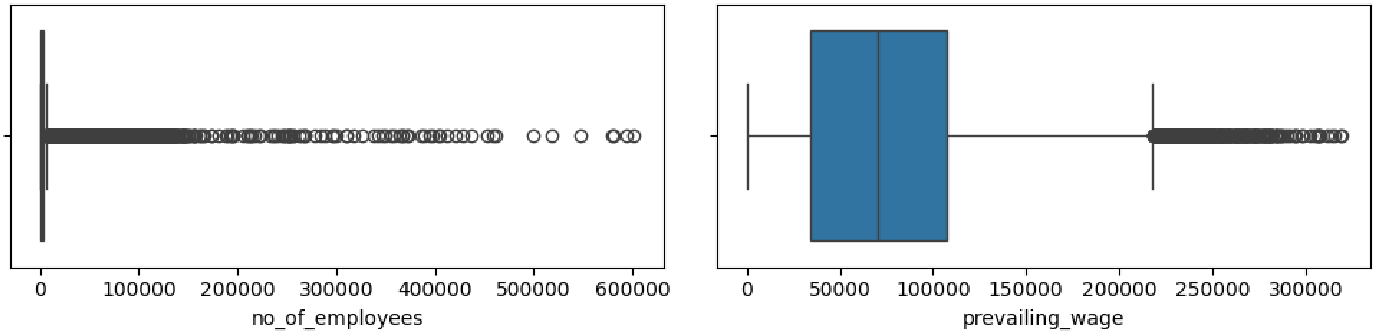


Figure 40: Outliers

There are outliers in the few columns. We have a few options for handling these outliers:

- Use the IQR (Interquartile Range) to determine the lower and upper bounds of the column and either replace or remove the outliers.
- The number of employees and individual wages are influenced by the scale of the company. Without domain knowledge, it is challenging to accurately identify and remove outliers related to these factors.

6.4. Feature Engineering

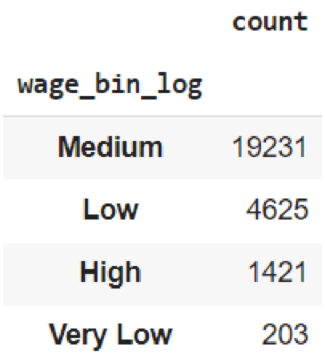
6.4.1. Convert Wages to a Common Unit

To compare wages across different units (hourly, monthly, yearly), we might need to normalize them to a common unit, such as yearly wages.

- Hourly: Wage is multiplied by 40 (hours per week) and then by 52 (weeks per year).
- Weekly: Wage is multiplied by 52 (weeks per year).
- Monthly: Wage is multiplied by 12 (months per year).
- Yearly: No conversion needed.

6.4.2. Logarithmic Binning

If there is significant skewness in the data, logarithmic binning can be effective for creating ranges that accommodate large variations in wage.



	count
wage_bin_log	
Medium	19231
Low	4625
High	1421
Very Low	203

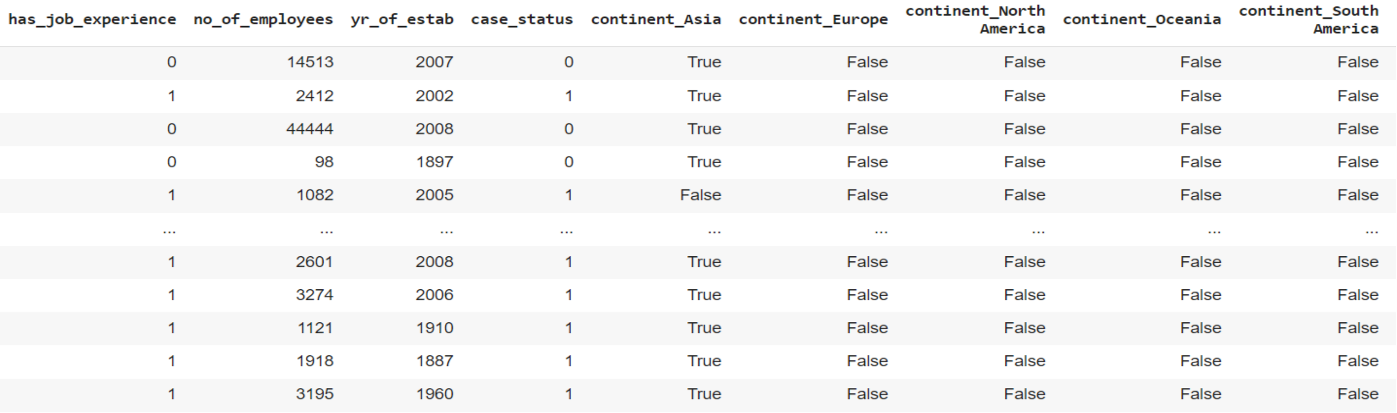
Figure 41: Wage Bins

- Removing features from the dataset that have constant values and those that do not positively impact the prediction model.
- Features removed: requires_job_training, prevailing_wage, unit_of_wage, full_time_position, wage_in_yearly

6.5. Data Preparation for Modeling

1. Our goal is to facilitate the process of visa approvals and predict the case status based on the factors affecting the application.
2. Before building the model, we'll need to encode the categorical features.
3. We will split the data into training, validation and testing sets to evaluate the model built on the training data.

6.5.1. Encoding Categorical Features



has_job_experience	no_of_employees	yr_of_estab	case_status	continent_Asia	continent_Europe	continent_North America	continent_Oceania	continent_South America
0	14513	2007	0	True	False	False	False	False
1	2412	2002	1	True	False	False	False	False
0	44444	2008	0	True	False	False	False	False
0	98	1897	0	True	False	False	False	False
1	1082	2005	1	False	False	False	False	False
...	...	...	...	...	...	...	...	...
1	2601	2008	1	True	False	False	False	False
1	3274	2006	1	True	False	False	False	False
1	1121	1910	1	True	False	False	False	False
1	1918	1887	1	True	False	False	False	False
1	3195	1960	1	True	False	False	False	False

Figure 42: Encoding

6.5.2. Train – Test Split

- Number of rows in train data = 15288
- Number of rows in validation data = 5096
- Number of rows in test data = 5096

7. Model building – Original Data

7.1. Model Evaluation Criterion

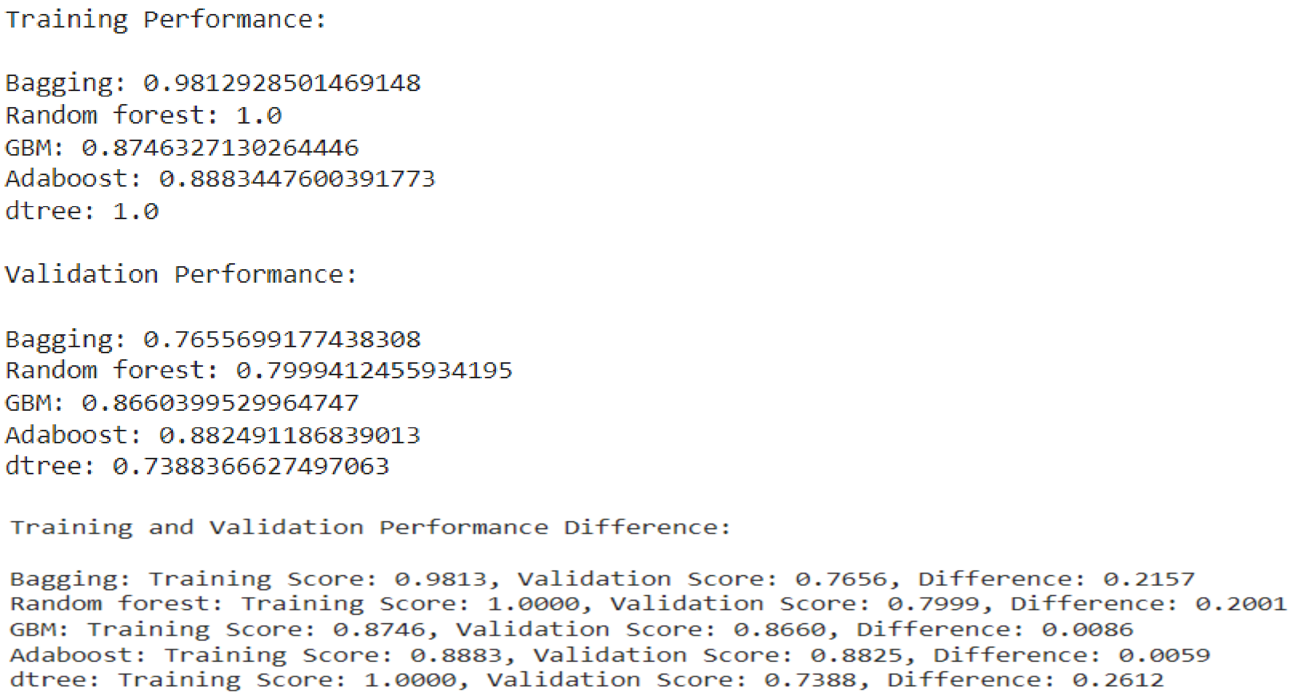
The model can make incorrect predictions in the following ways:

Predicting that a visa application will be certified when it is actually denied. Predicting that a visa application will be denied when it is actually certified. In this context, the more critical scenario is:

Predicting that a visa application will be denied when it is actually certified. This situation represents a missed opportunity to approve a valid visa application, potentially impacting valuable applicants. To mitigate this risk, it is essential to reduce False Negatives:

We should aim to maximize Recall. By increasing Recall, the model will better identify applications that should be certified (true positives), thus minimizing the number of valid applications incorrectly classified as denied (false negatives). This approach helps in accurately approving deserving visa applications based on factors like continent, education level, job experience, number of employees, establishment year, region of employment, prevailing wage, and unit of wage.

7.2. Bagging Classifier, Random Forest Classifier, Gradient Boosting Classifier, AdaBoost Classifier, Decision Tree Classifier



```
Training Performance:

Bagging: 0.9812928501469148
Random forest: 1.0
GBM: 0.8746327130264446
Adaboost: 0.8883447600391773
dtree: 1.0

Validation Performance:

Bagging: 0.7655699177438308
Random forest: 0.7999412455934195
GBM: 0.8660399529964747
Adaboost: 0.882491186839013
dtree: 0.7388366627497063

Training and Validation Performance Difference:

Bagging: Training Score: 0.9813, Validation Score: 0.7656, Difference: 0.2157
Random forest: Training Score: 1.0000, Validation Score: 0.7999, Difference: 0.2001
GBM: Training Score: 0.8746, Validation Score: 0.8660, Difference: 0.0086
Adaboost: Training Score: 0.8883, Validation Score: 0.8825, Difference: 0.0059
dtree: Training Score: 1.0000, Validation Score: 0.7388, Difference: 0.2612
```

Figure 43: Training and Validation Performance

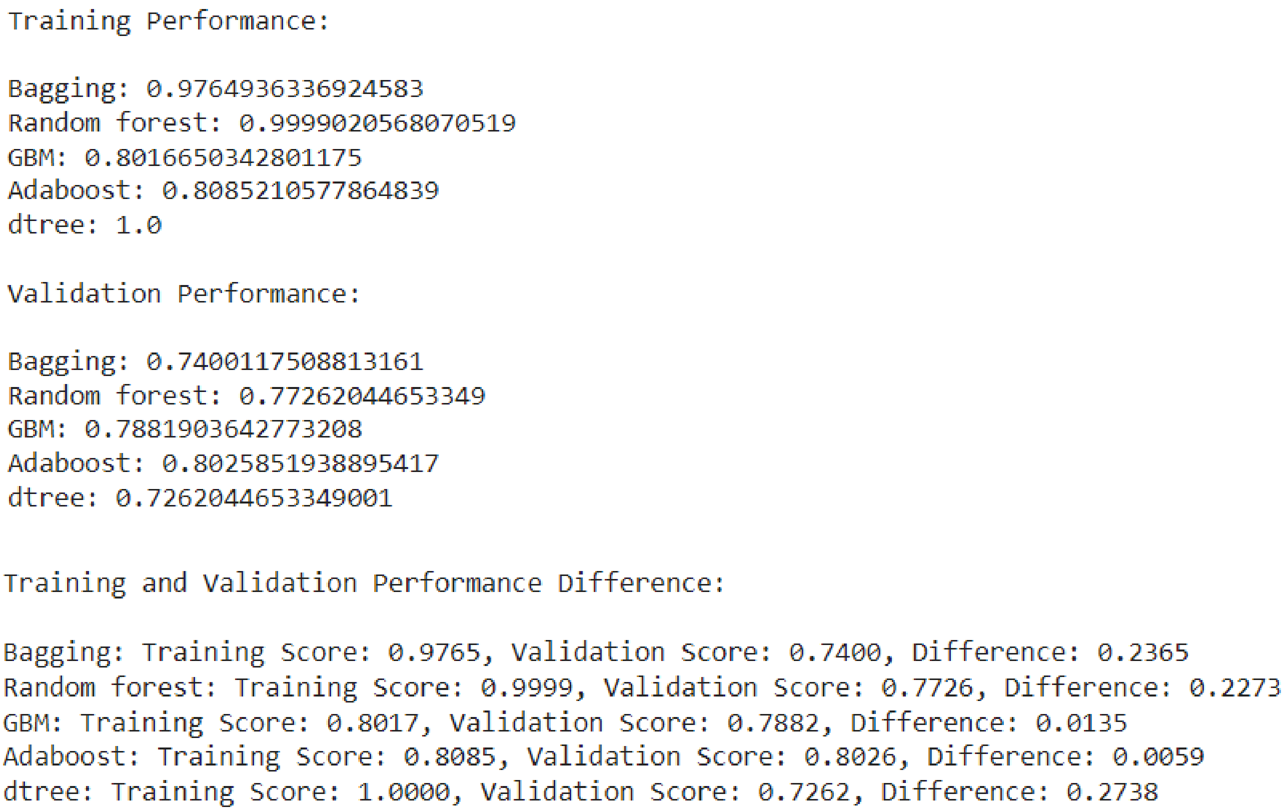
GBM has the best performance followed by AdaBoost model as per the validation performance.

8. Model Building- Oversampled Data

Before Oversampling, counts of label 'Yes': 10210
Before Oversampling, counts of label 'No': 5078

After Oversampling, counts of label 'Yes': 10210
After Oversampling, counts of label 'No': 10210

After Oversampling, the shape of train_X: (20420, 18)
After Oversampling, the shape of train_y: (20420,)



```
Training Performance:

Bagging: 0.9764936336924583
Random forest: 0.9999020568070519
GBM: 0.8016650342801175
Adaboost: 0.8085210577864839
dtree: 1.0

Validation Performance:

Bagging: 0.7400117508813161
Random forest: 0.77262044653349
GBM: 0.7881903642773208
Adaboost: 0.8025851938895417
dtree: 0.7262044653349001


Training and Validation Performance Difference:

Bagging: Training Score: 0.9765, Validation Score: 0.7400, Difference: 0.2365
Random forest: Training Score: 0.9999, Validation Score: 0.7726, Difference: 0.2273
GBM: Training Score: 0.8017, Validation Score: 0.7882, Difference: 0.0135
Adaboost: Training Score: 0.8085, Validation Score: 0.8026, Difference: 0.0059
dtree: Training Score: 1.0000, Validation Score: 0.7262, Difference: 0.2738
```

Figure 44: Training and Validation Performance – Oversampled

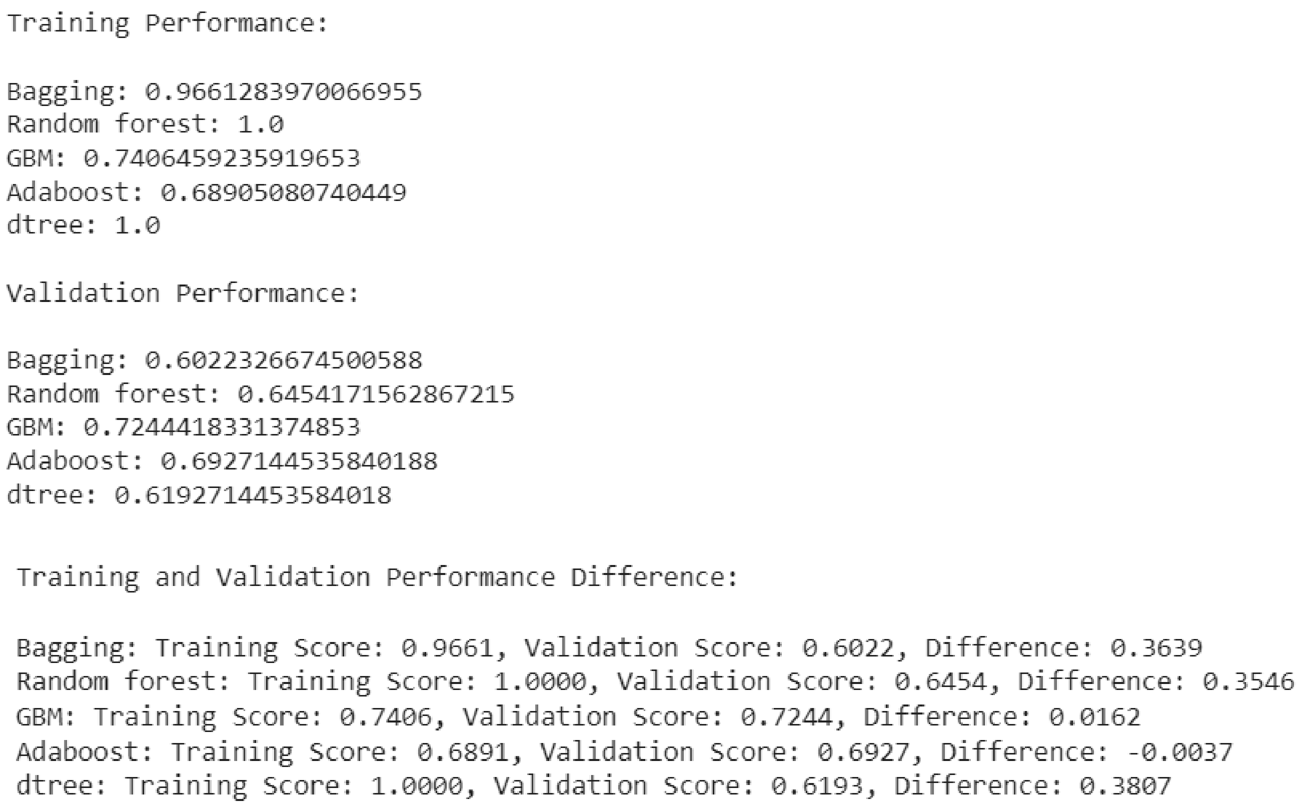
GBM and AdaBoost remain the best-performing models.

9. Model Building – Under sampled Data

Before Under Sampling, counts of label 'Yes': 10210
Before Under Sampling, counts of label 'No': 5078

After Under Sampling, counts of label 'Yes': 5078
After Under Sampling, counts of label 'No': 5078

After Under Sampling, the shape of train_X: (10156, 18)
After Under Sampling, the shape of train_y: (10156,)



```
Training Performance:

Bagging: 0.9661283970066955
Random forest: 1.0
GBM: 0.7406459235919653
Adaboost: 0.68905080740449
dtree: 1.0

Validation Performance:

Bagging: 0.6022326674500588
Random forest: 0.6454171562867215
GBM: 0.7244418331374853
Adaboost: 0.6927144535840188
dtree: 0.6192714453584018
```

```
Training and Validation Performance Difference:

Bagging: Training Score: 0.9661, Validation Score: 0.6022, Difference: 0.3639
Random forest: Training Score: 1.0000, Validation Score: 0.6454, Difference: 0.3546
GBM: Training Score: 0.7406, Validation Score: 0.7244, Difference: 0.0162
Adaboost: Training Score: 0.6891, Validation Score: 0.6927, Difference: -0.0037
dtree: Training Score: 1.0000, Validation Score: 0.6193, Difference: 0.3807
```

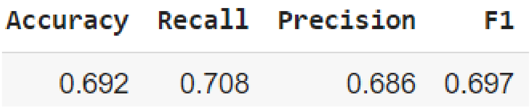
Figure 45: Training and Validation Performance – Under sampled

AdaBoost and GBM remain the best models with small differences between training and validation scores, indicating good generalization. AdaBoost has the smallest performance gap, but GBM has slightly better overall recall, making it the better model in this case.

It was observed that the GBM and AdaBoost models trained on the undersampled and oversampled dataset, demonstrated strong performance on both the training and validation sets. Since models can sometimes overfit after undersampling or oversampling, it’s important to tune them for more generalized performance. We will proceed with tuning these three models using the same datasets (undersampled or oversampled) on which they were originally trained.

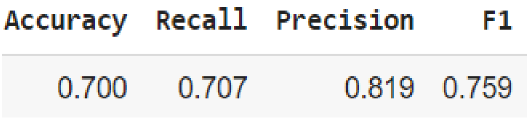
10. Hyperparameter Tuning

10.1. Tuning AdaBoost Classifier model with Under sampled data



Accuracy	Recall	Precision	F1
0.692	0.708	0.686	0.697

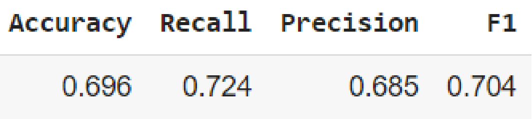
Figure 46: Train Performance: Tuned Adaboost Under sampled



Accuracy	Recall	Precision	F1
0.700	0.707	0.819	0.759

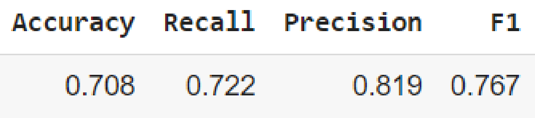
Figure 47: Validation Performance: Tuned Adaboost Under sampled

10.2. Tuning Gradient Boosting model with Under sampled Data



Accuracy	Recall	Precision	F1
0.696	0.724	0.685	0.704

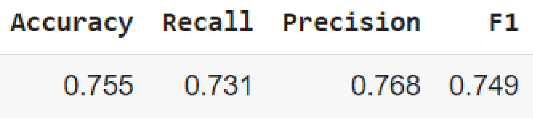
Figure 48:Train Performance: Tuned Gradient Boosting Under sampled



Accuracy	Recall	Precision	F1
0.708	0.722	0.819	0.767

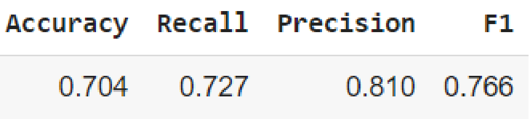
Figure 49:Validation Performance: Tuned Gradient Boosting Under sampled

10.3. Tuning AdaBoost Classifier model with Oversampled data



Accuracy	Recall	Precision	F1
0.755	0.731	0.768	0.749

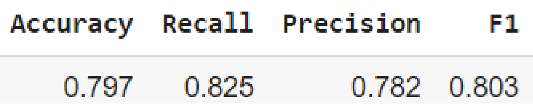
Figure 50: Train Performance: Tuned AdaBoost Classifier Oversampled



Accuracy	Recall	Precision	F1
0.704	0.727	0.810	0.766

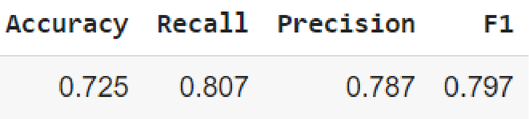
Figure 51: Validation Performance: Tuned AdaBoost Classifier Oversampled

10.4. Tuning Gradient Boosting model with Oversampled data



Accuracy	Recall	Precision	F1
0.797	0.825	0.782	0.803

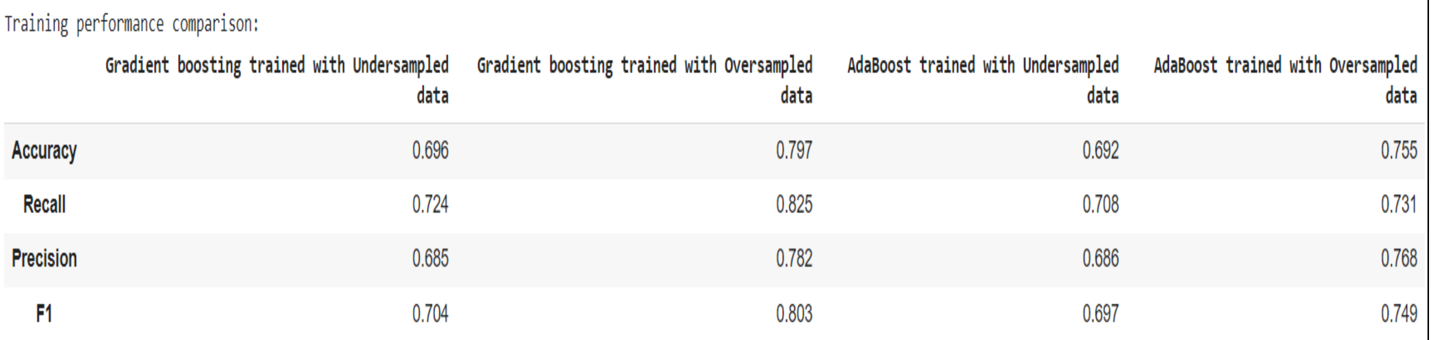
Figure 52: Train Performance: Tuned Gradient Boosting Oversampled



Accuracy	Recall	Precision	F1
0.725	0.807	0.787	0.797

Figure 53: Validation Performance: Tuned Gradient Boosting Oversampled

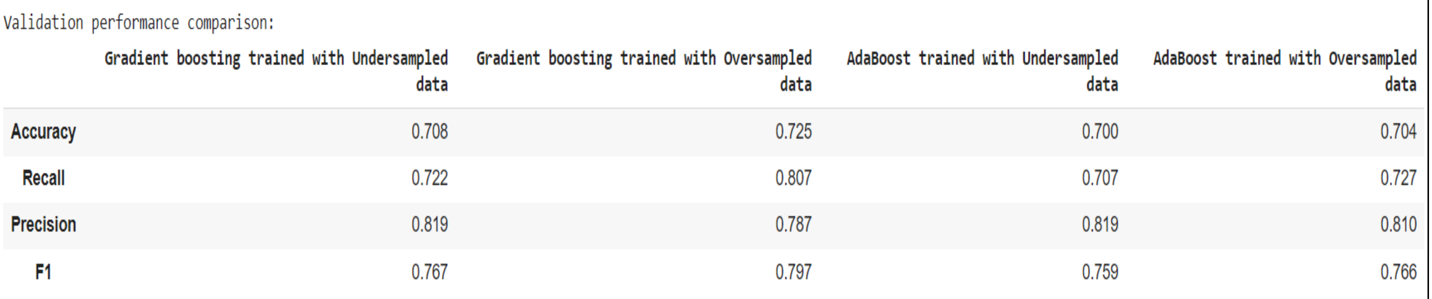
11. Model Comparison and Final Model Selection



Training performance comparison:

	Gradient boosting trained with Undersampled data	Gradient boosting trained with Oversampled data	AdaBoost trained with Undersampled data	AdaBoost trained with Oversampled data
Accuracy	0.696	0.797	0.692	0.755
Recall	0.724	0.825	0.708	0.731
Precision	0.685	0.782	0.686	0.768
F1	0.704	0.803	0.697	0.749

Figure 54: Training Performance



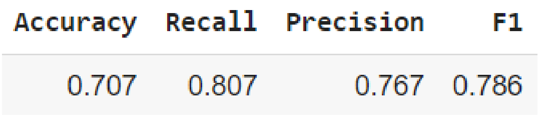
Validation performance comparison:

	Gradient boosting trained with Undersampled data	Gradient boosting trained with Oversampled data	AdaBoost trained with Undersampled data	AdaBoost trained with Oversampled data
Accuracy	0.708	0.725	0.700	0.704
Recall	0.722	0.807	0.707	0.727
Precision	0.819	0.787	0.819	0.810
F1	0.767	0.797	0.759	0.766

Figure 55: Validation Performance

Gradient Boosting trained with Oversampled data is the best overall model, excelling in accuracy, recall, and F1 score, making it the most balanced and effective for this scenario.

Let's check the performance on test set,



Accuracy	Recall	Precision	F1
0.707	0.807	0.767	0.786

Figure 56: Performance on Test Set

While the recall is solid (indicating good sensitivity to actual positives), the low precision suggests that the model struggles with false positives. The overall accuracy and F1 score are reasonable but not exceptional. To improve the model's performance on the test set, focusing on boosting precision (e.g., through tuning thresholds or further refinement of oversampling techniques) might help achieve a better balance.

11.1. Feature Importance

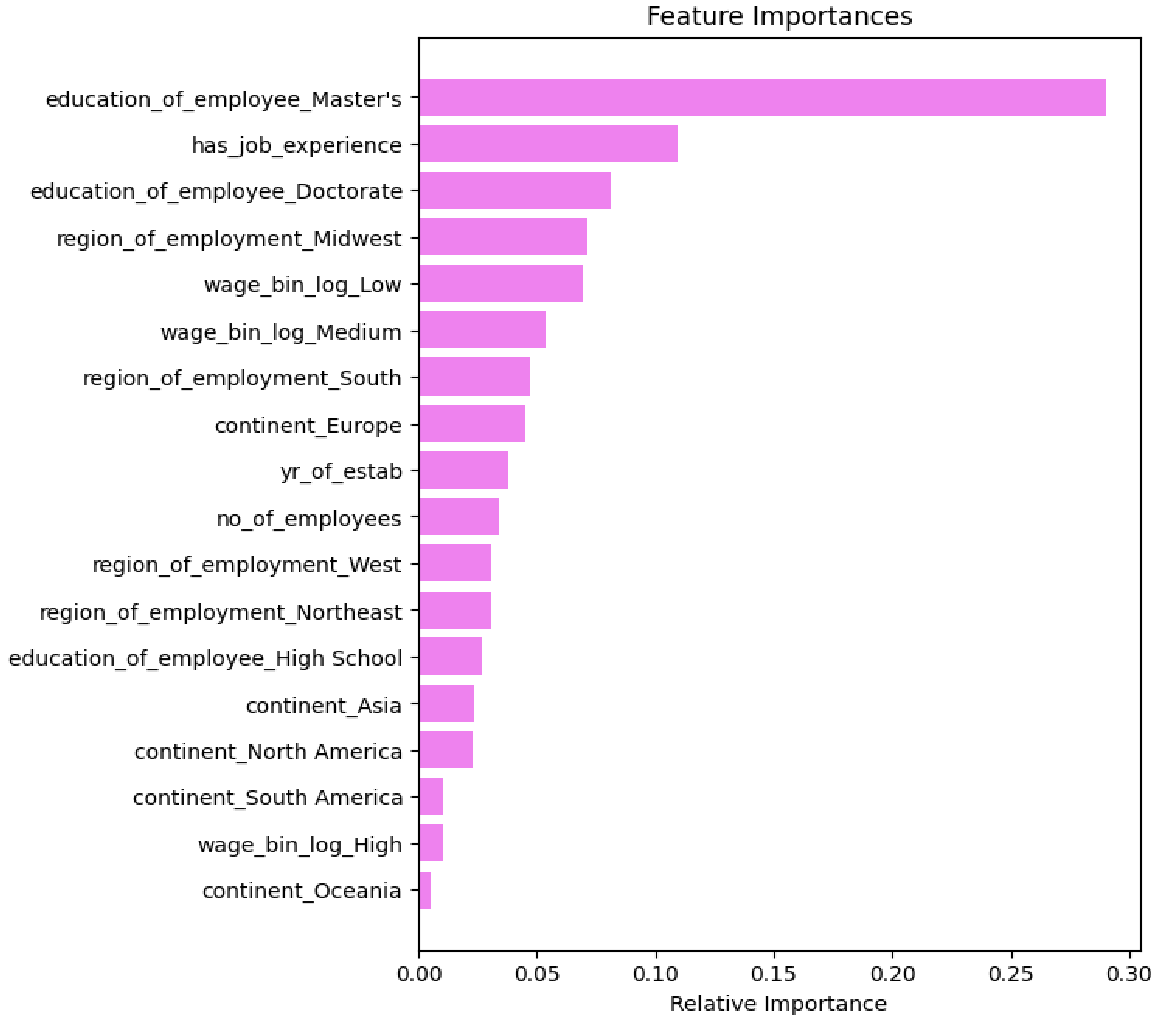


Figure 57: Feature Importance

12. Business Insights and Recommendations

1. Education of Employee - Master's:

   The most important factor driving visa certification decisions.

   - Insight: Applicants with a Master's degree is significantly more likely to have their visa applications approved.
   - Recommendation: Employers should prioritize candidates with Master's degrees for visa applications to increase their chances of approval.

2. Has Job Experience:

   Job experience is the second most important factor.

   - Insight: Candidates with job experience have a high likelihood of visa certification.
   - Recommendation: Employers should focus on hiring candidates with relevant job experience, as this greatly improves the chances of success in visa applications.

3. Education of Employee - Doctorate:

   Doctorate degrees also play a crucial role in visa certification.

- Insight: Applicants with a PhD are very likely to receive visa approval.
- Recommendation: When possible, prioritize PhD holders for applications as their educational background strengthens their case for visa certification.

4. Region of Employment - Midwest:

   The Midwest region significantly impacts visa certification success.

   - Insight: Visa applicants for positions in the Midwest are more likely to have their applications approved.
   - Recommendation: Companies based in the Midwest or planning to hire in this region may benefit from a higher visa approval rate.

5. Wage Bins (Low and Medium):

   Wage categories, especially low and medium wage brackets, are important for visa outcomes.

   - Insight: Offering wages in these categories may influence the likelihood of approval.
   - Recommendation: Ensure that wages offered to visa applicants are competitive and within the accepted industry norms, as this plays a role in decision-making.

6. Region of Employment - South:

   The South is another important region influencing visa outcomes.

   - Insight: Visa applications in the South show notable success rates.
   - Recommendation: Tailor visa applications to regional nuances, highlighting the strategic importance of the location for the company.

Other Factors:

1. Continents (Europe, Asia, North America):

   Employment locations across these continents have a moderate impact on visa certification outcomes.

   - Insight: Regional factors tied to the applicant's continent also influence the visa decision.
   - Recommendation: Employers operating in these continents should factor in regional considerations when applying for visas.

2. Company's Year of Establishment and Number of Employees:

   Established companies and larger organizations show a slight edge in visa certifications.

   - Insight: Company stability and size may positively influence visa decisions.
   - Recommendation: Younger companies may need to provide stronger business justifications to improve their chances.

Recommendations for EasyVisa:

1. Focus on Education: Develop targeted strategies to prioritize applicants with advanced educational qualifications, particularly Master's and Doctorate degrees.

2. Enhance Job Experience Criteria: Emphasize job experience as a key differentiator in visa applications, working with employers to highlight candidates’ relevant work history.

3. Regional and Wage-Based Strategy: Tailor visa applications based on employment region, especially the Midwest and South, and ensure wages align with industry standards.