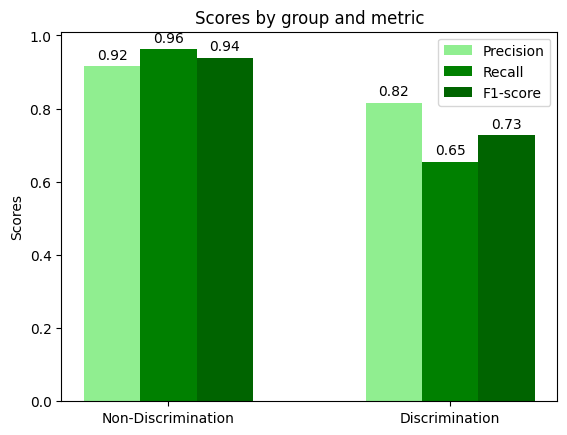
1. **Classification Report and Confusion Matrix**:
   * The model shows an overall accuracy of 90%, with a precision of 92% for predicting non-discrimination ('disc\_hire' = 0) and 82% for predicting discrimination ('disc\_hire' = 1).
   * The recall is 96% for non-discrimination and 65% for discrimination, indicating the model is better at identifying non-discriminatory cases.
   * The F1-score, which is a balance between precision and recall, is 94% for non-discrimination and 73% for discrimination.
2. **Coefficients of the Logistic Regression Model**:
   * The coefficients represent the relationship between each explanatory variable and the likelihood of experiencing hiring discrimination.
   * Positive coefficients increase the probability of experiencing discrimination, while negative coefficients decrease this probability.
   * Significant variables include:
     + **emp\_fin** (employment finance): Positive coefficient, suggesting higher values in this variable increase the likelihood of discrimination.
     + **disc\_wage** (discrimination in wage): Very high positive coefficient, indicating a strong association with experiencing hiring discrimination.
     + **disc\_social** (discrimination in social aspects): Positive coefficient, implying an association with higher chances of discrimination.
     + **marriage**, **income\_quartile**, **birth\_region**, **edu\_cat**: Negative coefficients, suggesting these factors might decrease the likelihood of discrimination.
3. **Confusion Matrix Analysis**:
   * 533 true negatives (correctly predicted non-discriminatory cases).
   * 93 true positives (correctly predicted discriminatory cases).
   * 49 false negatives (discriminatory cases predicted as non-discriminatory).
   * 21 false positives (non-discriminatory cases predicted as discriminatory).

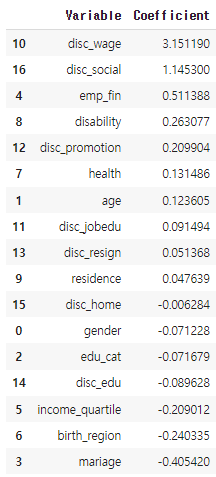
These results indicate that variables related to financial employment status, perceived wage discrimination, and social discrimination are strongly associated with the experience of hiring discrimination. However, other factors like marital status, income, birth region, and education category appear to reduce the likelihood of experiencing hiring discrimination. ​



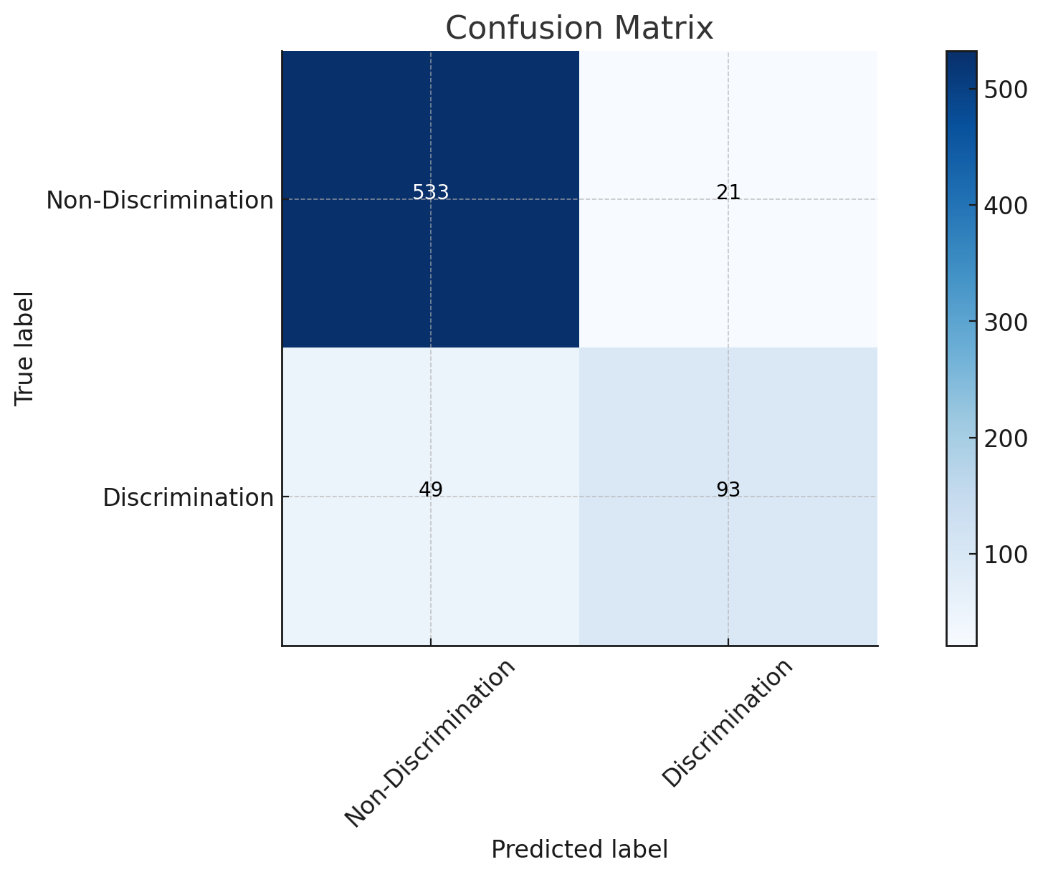
[1. Classification Report and Confusion Matrix]

1. **Precision:**
   * **Definition:** Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It answers the question: "Of all the instances the model labeled as positive, how many are actually positive?"
   * **Formula:** Precision = True Positives / (True Positives + False Positives)
   * **Interpretation:**
     + In the context of hiring discrimination, if 'Discrimination' is considered the positive class, then precision tells you the proportion of cases that were correctly identified as 'Discrimination' out of all the cases that the model predicted as 'Discrimination'.
     + A high precision means that when the model predicts 'Discrimination', it is usually correct. This is important in scenarios where the cost of a false positive (incorrectly labeling a case as 'Discrimination') is high.
     + In the context of hiring discrimination, if 'Discrimination' is considered the positive class, then precision tells you the proportion of cases that were correctly identified as 'Discrimination' out of all the cases that the model predicted as 'Discrimination'.
     + A high precision means that when the model predicts 'Discrimination', it is usually correct. This is important in scenarios where the cost of a false positive (incorrectly labeling a case as 'Discrimination') is high.
2. **Recall (Sensitivity)**:
   * **Definition**: Recall is the ratio of correctly predicted positive observations to all the actual positive observations. In other words, it answers the question: "Out of all the actual positives, how many were identified correctly by the model?"
   * **Formula**: Recall = True Positives / (True Positives + False Negatives)
   * **Interpretation**:
     + A high recall indicates a low rate of false negatives (i.e., few actual positives are missed by the model).
     + It's crucial in scenarios where missing a positive is significantly worse than wrongly labeling a negative as a positive. For example, in medical diagnoses, a high recall is essential because missing a disease can be more harmful than a false alarm.
3. **F1-Score**:
   * **Definition**: The F1-score is the harmonic mean of precision and recall. It is a single metric that balances both the precision (the ratio of correctly predicted positive observations to the total predicted positives) and the recall.
   * **Formula**: F1-Score = 2 \* (Precision \* Recall) / (Precision + Recall)
   * **Interpretation**:
     + The F1-score ranges from 0 to 1, where 1 is the best possible score (perfect precision and recall) and 0 is the worst.
     + A high F1-score indicates a balance between precision and recall, meaning the model effectively identifies positive cases and minimizes false positives.
     + This metric is particularly useful when the costs of false positives and false negatives are roughly equal, or when the distribution of positive and negative classes is imbalanced.

In summary, recall measures how well a model identifies all relevant instances, while the F1-score provides a balance between precision and recall, considering both false positives and false negatives. Both are crucial for understanding the effectiveness of a classification model beyond mere accuracy.



[2. Coefficients of the Logistic Regression Model]



[3. Confusion Matrix Analysis]

The confusion matrix for the logistic regression model has been successfully visualized. This matrix provides a clear representation of the model's performance in classifying cases of hiring discrimination:

* **True Negatives (Top-Left Quadrant)**: 533 instances where the model correctly predicted non-discrimination.
* **False Positives (Top-Right Quadrant)**: 21 instances where the model incorrectly predicted discrimination.
* **False Negatives (Bottom-Left Quadrant)**: 49 instances where the model incorrectly predicted non-discrimination.
* **True Positives (Bottom-Right Quadrant)**: 93 instances where the model correctly predicted discrimination.