**Gender Bias in Hiring**

**Name:** Vaishno Lukin Bussa

**1. Problem Framing & Hypothesis**

This project explores whether gender bias exists in hiring decisions based on HR data. The dataset includes features such as Gender, Age, Education, Experience, Performance Score, Department, and the final hiring decision.

**Goals:**

* Identify if gender significantly impacts hiring outcomes.
* Quantify the influence of performance, education, and other factors.

**Hypotheses:**

* H₀ (Null Hypothesis): Gender does not affect hiring.
* H₁ (Alternative Hypothesis): Gender influences hiring outcomes.

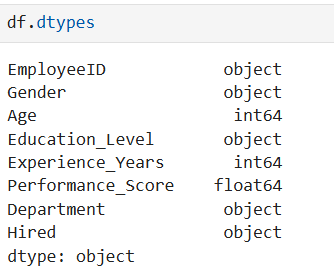
**2. Descriptive Analysis**

Observations from the Dataset:

* Gender distribution is approximately balanced.
* Most hired candidates have performance scores above 3.5.
* No missing values in the dataset.
* Age and experience range is realistic and consistent with job-level expectations.

Key Distributions Analyzed:

* Age
* Experience
* Performance Score

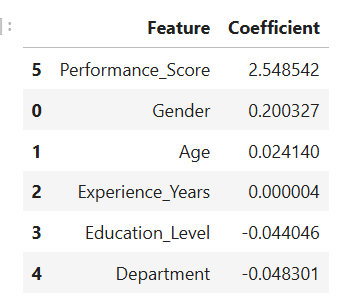


**3. Predictive Analysis**

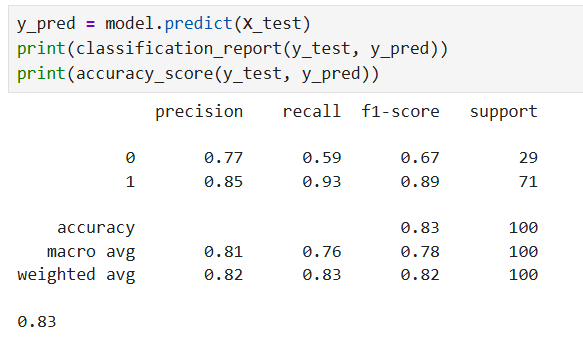
Two models were trained to predict hiring:

* Logistic Regression
* Random Forest Classifier

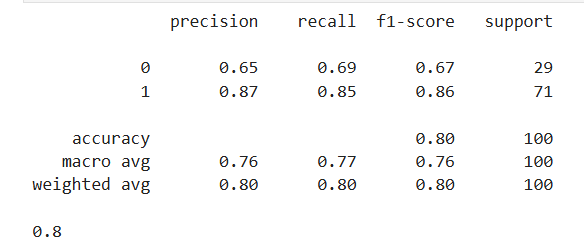
Model Coefficients:



Logistic Regression:



Random Forest Classifier:

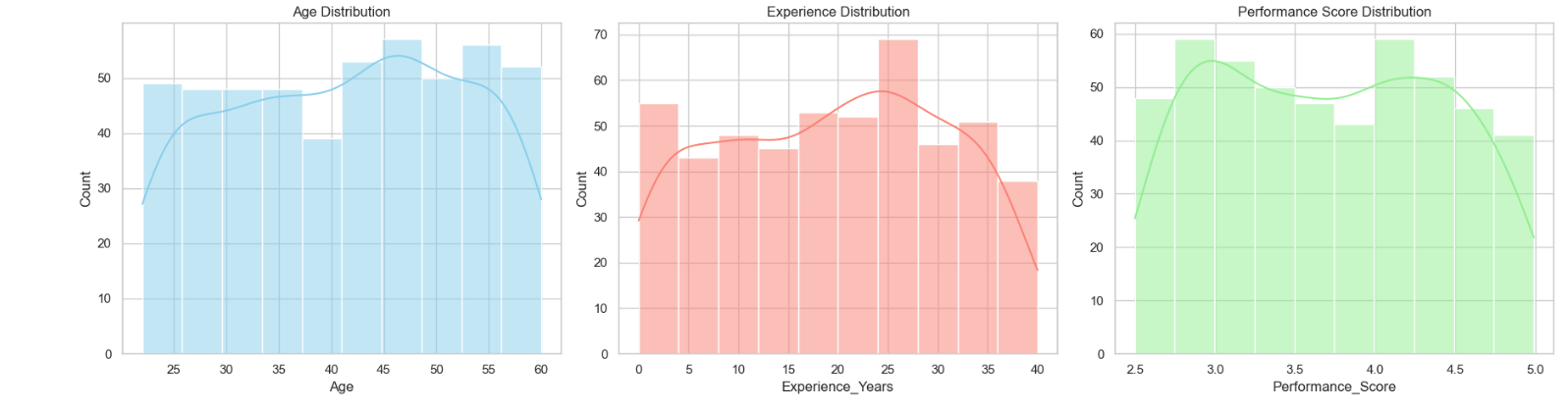


**4. Diagnostic Analysis**

To dive deeper into hiring patterns, we segmented and compared various groups to identify any underlying trends or anomalies in the hiring process.

**Distribution analysis:**

The following histograms help us identify population trends and potential anomalies in the data

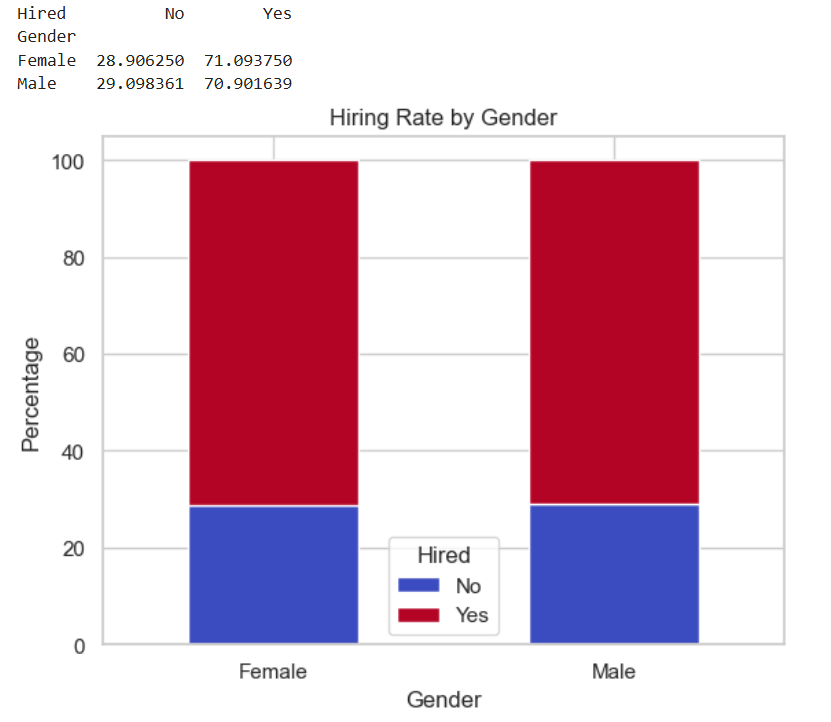


Age: Candidates are evenly spread across 25–60 years, with a slight peak around 45–50, indicating a balanced and mature applicant pool.

Experience: Most candidates have 10–30 years of experience, peaking near 25, suggesting a focus on mid-to-senior level roles.

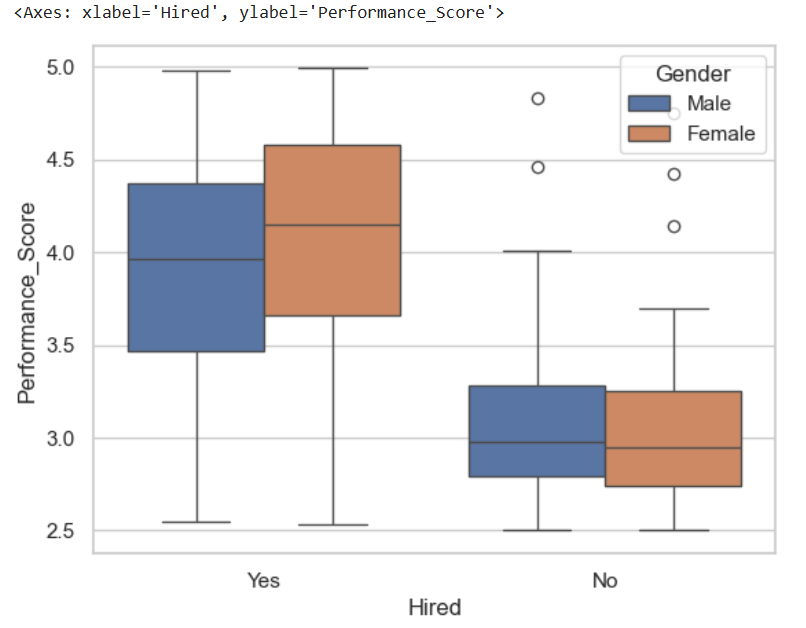
Performance Score: Slight bimodal distribution with peaks around 3.0 and 4.0+, hinting at distinct performance groups within the candidate pool.

**Hiring Rate by Gender:**



**Observation: Female candidates were hired at a slightly higher rate than male candidates a difference of only ~0.19%, which is likely not statistically significant.**

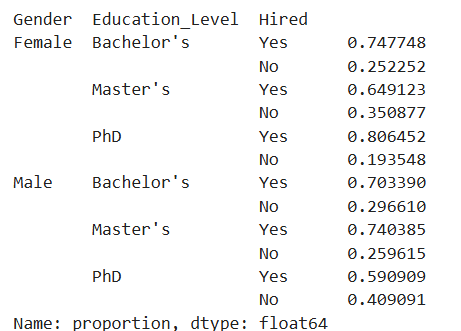
**Performance vs Hiring**

****

* **Hired candidates had consistently higher performance scores.**
* **Both genders followed a similar performance trend, indicating performance is a strong hiring driver, not gender.**

**Experience and Education**

* **Across all education levels, hiring rates between genders were consistent.**
* **Higher education (e.g., PhD) didn't guarantee higher hiring odds again showing performance score outweighed education level**

****

**Anomaly Detection:**

**Using logistic regression, we noticed:**

* **Performance\_Score had the strongest positive coefficient (+2.55).**
* **Gender had a small positive coefficient (+0.20), meaning being male slightly increased the hiring odds after controlling for other features a minor inconsistency with the raw percentage findings.**

**5.Prescriptive Insights**

Based on the findings:

**Recommendations:**

* Implement blind hiring where gender is hidden during initial screening.
* Standardize hiring process to ensure fairness across departments.
* Focus on performance-based metrics over other factors.

**Key point:**

Raw hiring percentages alone can be misleading. Predictive models and grouped analysis reveal performance dominates hiring decisions, while gender plays a subtle role that may or may not be significant depending on the context.

**6.Storytelling and delivery**

Although it appears that a slightly higher percentage of female candidates were hired (71.09%) compared to males (70.90%), the difference is very small less than 1%. This kind of minor gap can happen just by chance and doesn’t strongly suggest bias.

To understand whether gender truly affects hiring, we used a statistical method called logistic regression. This helps us check if gender still makes a difference after considering other important factors like performance, experience, and education. That way, we don't jump to conclusions based only on raw percentages