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:

- H_0 (Null Hypothesis): Gender does not affect hiring.
- H_1 (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

```
df.dtypes
```

```
EmployeeID           object
Gender               object
Age                   int64
Education_Level      object
Experience_Years      int64
Performance_Score    float64
Department           object
Hired                object
dtype: object
```

3. Predictive Analysis

Two models were trained to predict hiring:

- Logistic Regression
- Random Forest Classifier

Model Coefficients:

	Feature	Coefficient
5	Performance_Score	2.548542
0	Gender	0.200327
1	Age	0.024140
2	Experience_Years	0.000004
3	Education_Level	-0.044046
4	Department	-0.048301

Logistic Regression:

```
              precision    recall  f1-score   support

           0       0.77      0.59      0.67        29
           1       0.85      0.93      0.89        71

    accuracy                           0.83       100
   macro avg       0.81      0.76      0.78       100
weighted avg       0.82      0.83      0.82       100

0.83
```

Random Forest Classifier:

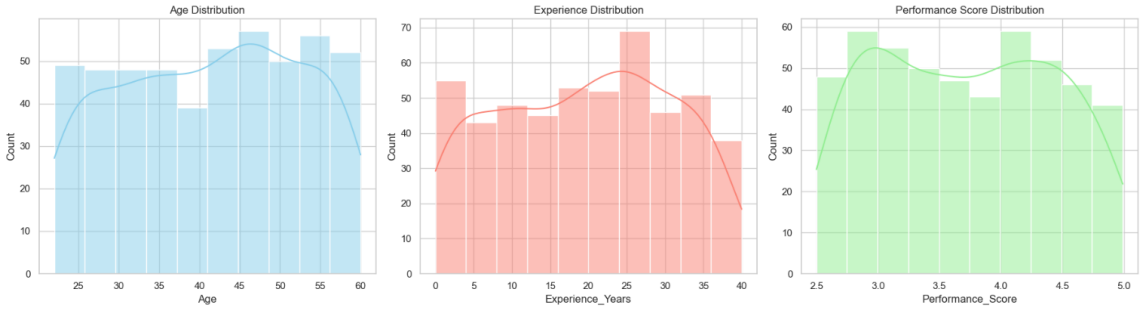
```
              precision    recall  f1-score   support

           0       0.65      0.69      0.67        29
           1       0.87      0.85      0.86        71

    accuracy                           0.80       100
   macro avg       0.76      0.77      0.76       100
weighted avg       0.80      0.80      0.80       100

0.8
```

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.

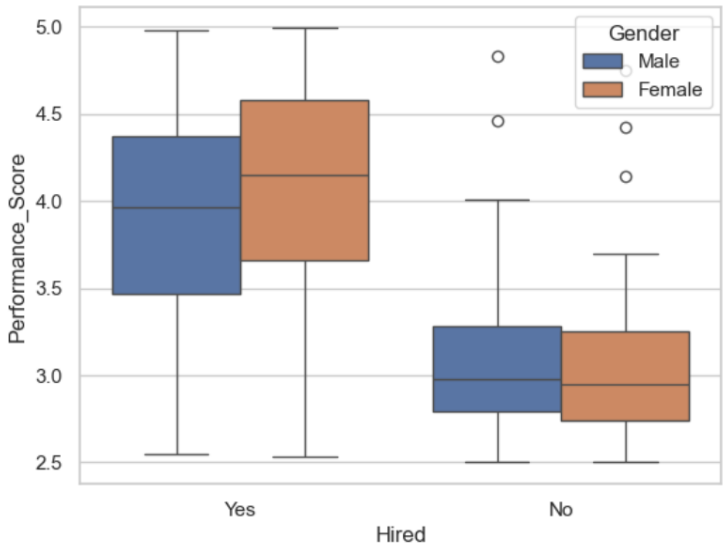
Hiring Rate by Gender:

Hired	No	Yes
Gender		
Female	28.906250	71.093750
Male	29.098361	70.901639

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

```
<Axes: xlabel='Hired', ylabel='Performance_Score'>
```



- 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

```
Gender  Education_Level  Hired
Female  Bachelor's       Yes      0.747748
                         No       0.252252
        Master's         Yes      0.649123
                         No       0.350877
        PhD              Yes      0.806452
                         No       0.193548
Male    Bachelor's       Yes      0.703390
                         No       0.296610
        Master's         Yes      0.740385
                         No       0.259615
        PhD              Yes      0.590909
                         No       0.409091
Name: proportion, dtype: float64
```

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