

▼ TC7 Lab 09

Name: Hitesh Choudhary



PRN No. 20200802146

```
import random
import pandas as pd
data = {
    "resume_id": range(1, 31),
    "gender": [random.choice(["Male", "Female"]) for _ in range(30)],
    "race": [random.choice(["White", "Black", "Asian"]) for _ in range(30)],
    "education": [random.choice(["Bachelor's", "Master's", "Ph.D."]) for _ in range(30)],
    "experience": [random.randint(1, 10) for _ in range(30)],
    "skills": [random.choice(["Java", "Python", "C++"]) for _ in range(30)],
    "hired": [random.choice(["Yes", "No"]) for _ in range(30)]
}

df = pd.DataFrame(data)

df.to_csv("hiring_recommendation_dataset.csv", index=False)
display(df)
```



	resume_id	gender	race	education	experience	skills	hired	
0	1	Male	White	Bachelor's	9	C++	Yes	
1	2	Male	Black	Master's	6	Python	No	
2	3	Male	Asian	Master's	8	Python	Yes	
3	4	Female	Black	Master's	10	Python	Yes	
4	5	Male	White	Bachelor's	7	Python	Yes	
5	6	Female	Black	Ph.D.	3	Python	Yes	
6	7	Female	Asian	Ph.D.	9	C++	No	
7	8	Male	White	Master's	10	Java	No	
8	9	Male	White	Bachelor's	5	C++	No	
9	10	Female	White	Master's	5	Python	No	
10	11	Male	Asian	Ph.D.	2	Java	No	

```
import pandas as pd
import random

df = pd.read_csv("hiring_recommendation_dataset.csv")

male_hired = df[(df["gender"] == "Male") & (df["hired"] == "Yes")].shape[0]
female_hired = df[(df["gender"] == "Female") & (df["hired"] == "Yes")].shape[0]

white_hired = df[(df["race"] == "White") & (df["hired"] == "Yes")].shape[0]
black_hired = df[(df["race"] == "Black") & (df["hired"] == "Yes")].shape[0]
asian_hired = df[(df["race"] == "Asian") & (df["hired"] == "Yes")].shape[0]

total_male = df[df["gender"] == "Male"].shape[0]
total_female = df[df["gender"] == "Female"].shape[0]
total_white = df[df["race"] == "White"].shape[0]
total_black = df[df["race"] == "Black"].shape[0]
total_asian = df[df["race"] == "Asian"].shape[0]

# Calculate disparate impact for gender and race.
disparate_impact_gender = female_hired / total_female / (male_hired / total_male)
disparate_impact_race_white = white_hired / total_white / (black_hired / total_black)
disparate_impact_race_black = black_hired / total_black / (white_hired / total_white)
disparate_impact_race_asian = asian_hired / total_asian / (white_hired / total_white)

print(f"Disparate Impact (Gender): {disparate_impact_gender}")
print(f"Disparate Impact (Race - White): {disparate_impact_race_white}")
print(f"Disparate Impact (Race - Black): {disparate_impact_race_black}")
print(f"Disparate Impact (Race - Asian): {disparate_impact_race_asian}")

weight_gender = total_male / (2 * total_female)
weight_race_white = total_black / (2 * total_black)
weight_race_black = total_white / (2 * total_white)
weight_race_asian = total_asian / (2 * total_asian)

# Apply reweighting to the dataset
df["weight"] = 1.0 # Initialize weights
```

```
df["weight"] = 1.0 # initialize weights
df.loc[df["gender"] == "Male", "weight"] = weight_gender
df.loc[df["race"] == "White", "weight"] = weight_race_white
df.loc[df["race"] == "Black", "weight"] = weight_race_black
df.loc[df["race"] == "Asian", "weight"] = weight_race_asian
```

Disparate Impact (Gender): 1.6813186813186813
Disparate Impact (Race - White): 0.5892857142857142
Disparate Impact (Race - Black): 1.696969696969697
Disparate Impact (Race - Asian): 0.9333333333333335