

## **Project Title: Evaluating the Impact of Interventions on Student Success**

The project aims to assess the effectiveness of interventions in improving the success of students and reduce disparities in student success, considering biased outcomes for Black and Hispanic students.

### **1. Data Generation:**

- Generate a synthetic educational dataset with demographics, test scores, parental information, and biased outcomes. You can use the attached code or generate your own dataset.
- Include variables related to intervention status and key outcome measures.

### **2. Data Exploration:**

- Analyze the dataset to understand its structure and biases.
- Visualize the data to identify patterns and disparities among different student groups.

### **3. Hypothesis Formulation:**

- Formulate hypotheses related to the impact of interventions on student success.
- Hypotheses should consider disparities among ethnic groups.

### **4. Simulation Setup:**

- Define the simulation framework, including parameters for the interventions and their impact. You can use any approach including Agent Based Modeling, Machine Learning, Scenario Analysis, Reinforcement Learning,
- Create a model for simulating the outcomes of different interventions on students.

### **5. Intervention Scenarios:**

- Simulate multiple intervention scenarios, such as tutoring, mentorship programs, financial aid, etc.
- Apply interventions to appropriate student groups, considering demographics.

### **6. Outcome Assessment:**

- Assess the outcomes of each intervention scenario on student success.
- Measure success in terms of graduation status, college enrollment, and employment.

### **7. Statistical Analysis:**

- Conduct statistical analyses to evaluate the effectiveness of each intervention.
- Compare the outcomes of different interventions and their impact on biased groups.

### **8. Ethical Considerations:**

- Examine the ethical implications of bias in the dataset.
- Consider how the simulation results should be interpreted and reported.

### **9. Reporting and Visualization:**

- Present the findings, including the impact of interventions on student success.
- Visualize the results to make them accessible to stakeholders.

### **10. Policy Recommendations:**

- Based on the findings, provide policy recommendations to reduce disparities and improve student success.

### **11. Sensitivity Analysis:**

- Perform sensitivity analyses to assess the robustness of the results to changes in assumptions and biases.

```

import pandas as pd
import numpy as np
import random

# Set the number of students
num_students = 5000

# Create empty lists to store the data
student_id = []
age = []
gender = []
ethnicity = []
test_scores_math = []
test_scores_english = []
parental_education = []
parental_income = []
intervention_group = []
attendance_rate = []
behavioral_problems = []
extracurricular_activities = []
graduation_status = []
college_enrollment = []
employment_status = []

# Introduce bias for African American and Hispanic students
for _ in range(num_students):
    student_id.append(len(student_id) + 1)
    age.append(np.random.randint(6, 18))
    gender.append(random.choice(['Male', 'Female']))
    ethnicity_choice = np.random.choice(['Caucasian', 'African American',
    'Hispanic', 'Asian', 'Other'])
    ethnicity.append(ethnicity_choice)
    test_scores_math.append(np.random.normal(70, 10))
    test_scores_english.append(np.random.normal(75, 8))
    parental_education.append(random.choice(['High School', 'Bachelor',
    'Master', 'PhD']))
    parental_income.append(np.random.normal(50000, 10000))
    intervention_group.append(random.choice([0, 1, 1, 1]))
    attendance_rate.append(np.random.uniform(0.8, 1.0))
    behavioral_problems.append(np.random.randint(0, 11))
    extracurricular_activities.append(np.random.randint(0, 6))

# Introduce bias based on ethnicity
if ethnicity_choice in ['African American', 'Hispanic']:

```

```

        # African American and Hispanic students have higher chances of
        not graduating, not enrolling in college, and being unemployed
        graduation_status_choice = random.choices(['Did not graduate',
        'Graduated'], weights=[0.7, 0.3])[0]
        college_enrollment_choice = random.choices(['Not enrolled',
        'Enrolled'], weights=[0.7, 0.3])[0]
        employment_status_choice = random.choices(['Unemployed',
        'Employed'], weights=[0.7, 0.3])[0]
    else:
        # For other ethnicities, use default values
        graduation_status_choice = 'Graduated'
        college_enrollment_choice = 'Enrolled'
        employment_status_choice = 'Employed'

    graduation_status.append(graduation_status_choice)
    college_enrollment.append(college_enrollment_choice)
    employment_status.append(employment_status_choice)

# Create a DataFrame
data = {
    'Student_ID': student_id,
    'Age': age,
    'Gender': gender,
    'Ethnicity': ethnicity,
    'Test_Scores_Math': test_scores_math,
    'Test_Scores_English': test_scores_english,
    'Parental_Education': parental_education,
    'Parental_Income': parental_income,
    'Intervention_Group': intervention_group,
    'Attendance_Rate': attendance_rate,
    'Behavioral_Problems': behavioral_problems,
    'Extracurricular_Activities': extracurricular_activities,
    'Graduation_Status': graduation_status,
    'College_Enrollment': college_enrollment,
    'Employment_Status': employment_status
}

df = pd.DataFrame(data)

# Display the first few rows of the dataset
print(df.head())

# Save the dataset to a CSV file
df.to_csv('biased_outcomes_simulated_education_data.csv', index=False)

```