**Student Dropout Prediction: Data Exploration Report**

**1. Introduction**

In this analysis, I explored factors that influence student performance and how they relate to dropout rates. My goal here is clear: to use the findings to suggest interventions that could help reduce dropout rates and improve overall student retention. The dataset I’ve worked with includes a broad range of information about students—covering demographics, academic performance, and socioeconomic backgrounds—which allows me to develop a well-rounded view of their experiences.

**Key Insights:**

**1. Demographics**:

- Majority of students are single, indicating that personal life events like marriage may impact academic performance.

**2. Academic Performance**:

- High admission grades don’t guarantee success; early support is crucial.

- Students enrolled in more courses may face burnout, highlighting the need for course load management and counseling.

**3. Socioeconomic Factors**:

- Students from privileged backgrounds perform better, emphasizing the need for support for less privileged students.

- Financial aid (scholarships) positively impacts student success.

**4. Behavioral Patterns**:

- Evening classes are crucial for students with work or other responsibilities, indicating the need for flexible scheduling.

- Displaced students and those with special needs are at higher risk of dropping out, requiring targeted support.

**5. Economic Conditions:**

- Economic instability (unemployment, inflation) increases dropout risk, indicating the need for financial aid and institutional support during tough times.

**6. Visual Insights:**

- A positive correlation exists between admission grades and academic performance, but early intervention is needed for students who struggle initially.

**7. Interventions:**

- Early academic support, expanded financial aid, and economic monitoring can help reduce dropout rates.

**Key Insights**

**Chi-square tests conducted to assess the relationship between categorical variables and dropout rates.**

**Academic Performance:**

- Admission Grade Category: There is a strong relationship between admission grades and dropout rates (Chi-square = 73.14, p < 0.001). Lower admission grades tend to correlate with a higher likelihood of dropping out.

- Curricular Units 1st Sem (Grade) Category: The first-semester grades show an extremely strong association with dropout rates (Chi-square = 1093.68, p < 0.001). Students with poor first-semester performance are significantly more likely to drop out.

- Curricular Units 2nd Sem (Grade) Category: A similar strong association is observed with second-semester grades (Chi-square = 1533.74, p < 0.001). Consistently poor academic performance across both semesters significantly increases the risk of dropping out.

**Economic Factors:**

- Unemployment Rate Category: A moderate relationship is found between unemployment rates and dropout rates (Chi-square = 6.23, p = 0.044). Higher unemployment rates may increase dropout likelihood, possibly due to financial challenges.

- Inflation Rate Category: Inflation rates also show a significant relationship with dropout rates (Chi-square = 7.51, p = 0.023). Higher inflation may increase dropout risks.

- GDP Category: There is a strong association between GDP and dropout rates (Chi-square = 17.09, p < 0.001). Economic downturns are likely to increase the risk of students dropping out.

**Age Category:**

- Age Category: A strong relationship exists between age and dropout rates (Chi-square = 444.17, p < 0.001). Certain age groups may face more external pressures (e.g., family, work) contributing to their higher dropout likelihood.

**Conclusion:**

- Primary Drivers: Academic performance, especially poor grades in both the first and second semesters, is the strongest predictor of student dropout.

- Secondary Drivers: Economic conditions, including unemployment, inflation, and GDP, alongside age, also play significant roles in increasing dropout risks.

**Hypotheses Formulation**

Based on the exploratory analysis of the dataset, the following hypotheses have been formulated to investigate the factors influencing student dropout rates:

1. **Hypothesis 1**: Students with lower admission grades are more likely to drop out compared to those with higher admission grades.
2. **Hypothesis 2**: There is a negative correlation between first-semester grades and dropout rates; specifically, students with lower first-semester grades will have higher dropout rates.
3. **Hypothesis 3**: Financial support, such as scholarships, positively impacts student retention, leading to lower dropout rates among scholarship holders compared to those without financial aid.
4. **Hypothesis 4**: Students from disadvantaged socioeconomic backgrounds (as indicated by parental occupation) are more likely to drop out than those from more privileged backgrounds.
5. **Hypothesis 5**: Higher levels of unemployment and inflation in the broader economic environment are associated with increased dropout rates among students.
6. **Hypothesis 6**: Students enrolled in evening classes are more likely to drop out than those attending daytime classes, potentially due to conflicting responsibilities.
7. **Hypothesis 7**: Age influences dropout rates, with younger students being less likely to drop out than older students who may face additional external pressures.