Exploratory Data Analysis for Academic Success Classification Model

**1. Target Distribution Across Gender**

* The visualization shows the count of ‘Target’ classes (‘Graduate’, ‘Dropout’, and ‘Enrolled’) for each ‘Gender’.
* **Observation**:

- Both male and female distributions indicate that the majority belong to the ‘Graduate’ class.

- The ‘Dropout’ and ‘Enrolled’ categories are relatively fewer for both genders, consistent with the overall class imbalance in the dataset.

**2. Target Distribution Across Courses**

* A stacked bar chart illustrates the distribution of ‘Target’ classes for each ‘Course’.
* **Observation**:

- Certain courses have a higher proportion of ‘Graduates’ compared to ‘Dropouts’ or ‘Enrolled’.

- The pattern suggests that course-specific factors may significantly influence academic outcomes.

**3. Grades and Target Relationship**

* Violin plots depict the distribution of ‘1st semester grades’ and ‘2nd semester grades’ for each ‘Target’ class.
* **Observation**:

- ‘Graduate’ students typically have higher grades in both semesters compared to ‘Dropout’ and ‘Enrolled’ students.

- The ‘Dropout’ class exhibits a wider spread of grades, indicating variability in academic performance.

**4. Correlation Among Curricular Unit Features**

* A heatmap shows the correlations between features related to ‘curricular units’.
* **Observation**:

- High positive correlations exist between features like ‘enrolled’, ‘evaluations’, and ‘approved’ for both semesters.

- Grades show moderate correlations with the number of approved’ units, suggesting performance linkage.

**5. Parental Occupation and Target**

* A scatter plot visualizes the effect of ‘Father's occupation’ and ‘Mother's occupation’ on ‘Target’.
* **Observation**:

- Clustering patterns indicate that certain combinations of parental occupations are associated with specific ‘Target’ classes.

- The plot also highlights outliers, potentially identifying unique parental occupational impacts.

**6. Effect of Inflation Rate on Target**

* A count plot shows the distribution of ‘Target’ classes across different `Inflation rate` levels.
* **Observation**:

- ‘Graduates’ dominate across most inflation rate levels.

- Inflation rate seems to have a relatively uniform distribution across ‘Dropout’ and ‘Enrolled’ categories, with no clear trend.

**7. Effect of GDP on Target**

* A heatmap illustrates the proportion of ‘Target’ classes for different GDP percentages.
* **Observation**:

- Certain GDP ranges have a higher proportion of ‘Graduates’ compared to ‘Dropouts’ or ‘Enrolled’.

- The plot indicates that GDP may have a moderate influence on academic outcomes, potentially reflecting socioeconomic impacts.

**Insights from EDA**

* **Gender-Based Distribution**: Both genders follow the general trend of the dataset, with ‘Graduate’ being the dominant class.
* **Course Influence**: Some courses contribute significantly to the ‘Graduate’ class, highlighting the role of academic programs in student outcomes.
* **Grades as Predictors**: The strong association of grades with ‘Target’ classes underscores their importance in predictive modeling.
* **Parental Occupation**: The clustering patterns in parental occupations suggest that socioeconomic factors play a role in academic success.
* **External Factors**: While inflation rate shows limited differentiation across classes, GDP exhibits a stronger link to academic outcomes.

This exploratory analysis provides valuable insights for feature engineering and model development, focusing on academic, demographic, and socioeconomic predictors of success.