Idea 2

**Research Question**

The structure of the family and the engagement of the parents greatly influence a student's academic path. Assuming that I’m a data analyst at a high school, given the complex interaction between family dynamics and educational outcomes, my research question is: **How do family structure and parental involvement impact academic performance?**

The purpose of this study is to investigate the relationships between important variables and students' grades and general academic success, including family size, parental living arrangements, and parental educational attainment. Educational initiatives and policies can be informed by the insights derived from the findings.

**Data Source**

[The dataset](https://www.kaggle.com/datasets/abdullah0a/student-dropout-analysis-and-prediction-dataset) for this research includes demographic and academic data from 700 students. Key variables include Family\_Size (e.g., GT3 for greater than 3, LE3 for less than or equal to 3), Parental\_Status (A for parents living together, T for living apart), Mother\_Education and Father\_Education (0 to 4 scale), Family\_Support (yes/no), Final\_Grade (G3), along with other relevant indicators of academic performance, such as grades from previous assessments and school absences.

**Data Analysis**

1. Exploratory Data Analysis (EDA)

I will begin by conducting an EDA to understand the distribution of family-related variables (e.g., family size, parental status, and parental education) and their correlation with academic performance (e.g., Final\_Grade). Visualizations such as histograms and scatter plots will illustrate relationships between family characteristics and grades.

Key questions to explore through EDA:

* Do students from smaller or larger families show significant differences in academic performance?
* How does parental education level correlate with students' final grades?

1. Modeling

Multiple Linear Regression Model: I will use a multiple linear regression model to predict academic performance (Final\_Grade) based on family-related factors such as parental education, family size, and parental status. This will allow for a detailed examination of how these independent variables contribute to students' grades.

*Independent Variables:* Family\_Size (GT3/LE3), Parental\_Status (A/T), Mother\_Education, Father\_Education (0-4), Family\_Support (yes/no).

*Dependent Variable*: Final\_Grade (G3).

The model specification will ensure correct assumptions, such as linearity, independence, and normality of residuals, to improve prediction accuracy.

**Application and Implications**

Teachers, legislators, and school administrators must all understand how family structure and parental participation affect students' academic achievement. If there is a high correlation between parental education and improved student outcomes, schools could create programs to more successfully involve parents in their children's education. On the other hand, pupils from less educated homes could profit from extra school-based assistance like tutoring or mentorship programs.

Teachers can create focused interventions to enhance student performance by identifying important family-related characteristics that impact academic success. This is especially important for kids who do not have the best home support systems.

**Ethical and Societal Implications**

To preserve the students' privacy, the data utilized in this analysis has been anonymized. Thoughts about family dynamics and their possible impact on student achievement must take ethics into account. Even while the results can point to differences in parental education or family structure, interventions must be planned in a way that is fair and non-stigmatizing to guarantee that all students get the help they need. Schools must use caution to avoid feeding preconceptions and punishing children for circumstances beyond their control, such as their living situation or family history.

Furthermore, equity should be the first priority in any support measures, guaranteeing that all kids, regardless of family circumstances, have access to the same possibilities. A special focus needs to be placed on preventing the perpetuation of structural disparities in education, especially with regard to students from disadvantaged or minority backgrounds. By giving these students more resources and support, schools should aim to elevate them without putting undue pressure or blame on the family.

---------------------------------------------------------------------------------------------------------------------

Idea 3

**Research Question:**

This study seeks to identify the key factors that influence whether a student would avail of a mentoring program within the UCD School of Medicine. Mentoring programs are crucial for supporting students' growth on the academic, professional, and personal fronts. By identifying the major factors that influence student participation, mentorship programs can be made more successful. Assuming that I’m a data analyst at the career center of the UCD School of Medicine. The research question is: "**What factors most influence a student’s likelihood to participate in the UCD School of Medicine mentoring program?"**

**Data Source:**

[The dataset](https://figshare.com/articles/dataset/Student_Survey_Responses_xlsx/5745729?file=10117953) originates from a survey conducted by the UCD School of Medicine, with responses related to students' demographic information, academic stage, and preferences regarding mentoring. There are more than 100 respondents. The target variable is "I would avail of a mentoring program within the UCD School of Medicine", which is measured on a Likert scale from "Strongly Agree" to "Strongly Disagree."

Key variables for the analysis:

* Demographic Factors: Nationality, Stage of education.
* Mentoring Preferences: Preferences for clinical experience, communication skills, availability of mentors, and shared interests.
* Perceived Support: Academic, professional, and personal support provided by the school.

**Data Analysis:**

1. Exploratory Data Analysis (EDA)

To better understand the distribution of variables, I will perform an EDA on key demographic factors (age, stage of education, gender) and responses to mentoring-related questions. I will use visualizations such as bar charts and heatmaps to explore:

* How do different student groups (by age or stage) perceive the value of mentoring?
* Which mentor characteristics are most valued, and do these preferences vary by demographics?

1. Modelling:

The main model I will use for this analysis will be a decision tree.

*Target variable:* "I would avail of a mentoring program within the UCD School of Medicine" (ordinal variable, or converted to binary).

*Independent variables:*

* Demographic factors (nationality, stage of education).
* Perceived support in academic, professional, and personal development.
* Mentoring preferences (desired qualities in a mentor, desired outcomes of a mentoring program).
* The model will use a decision tree to determine the hierarchy of importance of these variables.

Cross-validation will be used to avoid overfitting, and model accuracy will be assessed through measures like classification accuracy, precision, recall, and the F1 score. Additionally, decision tree pruning techniques will be used to prevent overfitting, ensuring the model generalizes well to unseen data.

**Implications for Stakeholders**

The outcomes of the decision tree model will provide the UCD School of Medicine with useful information. The school can take the following actions by determining the key elements that encourage students to participate in mentorship programs:

Customize your mentoring program: For instance, the school should concentrate on outreach initiatives if the model reveals that students in preclinical or earlier educational levels are more likely to participate in mentorship programs.

Optimize mentor recruitment: The school should concentrate on finding mentors who meet certain requirements, such as clinical experience or strong communication skills, based on the students' preferences for their mentors.

**Ethical, Legal, and Societal Implications**

First off, the UCD School of Medicine needs to make sure that mentoring resources are allocated fairly if the results reveal differences in the involvement in the program across participants based on factors like age, educational attainment, or past experiences. It is imperative that the program neither favor or ignore students in a way that is disproportionately dependent on these demographic characteristics. Ensuring the equity and inclusivity of the mentorship program is one way to address its societal effects. It is crucial to apply the knowledge gained from this analysis to improve the accessibility of mentorship programs for all students, especially those who might feel marginalized or unsupported.

We also need to talk about possible wage differences. The program may inadvertently maintain disparities in access to career counseling, professional networks, and mentorship if it favors students from higher socioeconomic backgrounds (a tendency that may be deduced from demographic proxies like course stage or perceived class rank). As a result, it is morally required to create a mentoring program that provides equal support to kids from all backgrounds.