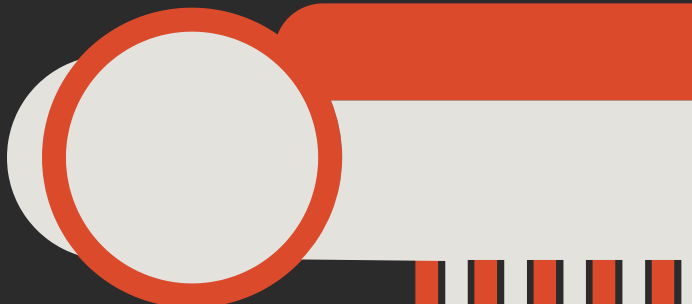




Mini Project

SDG 4 (Quality Education)

**Topic : Predict Student's Drops Out
and Academic Outcomes**





Introduction



What is Student Dropout?

- Many students leave school before completing their education
- Affects both students and society (job opportunities, economy, etc.).

Why is this important?

- High dropout rates lead to unemployment and poverty.
 - Schools need to identify at-risk students early
- 


Objectives

The main objective of this project is to develop a machine learning model that can predict the likelihood of a student dropping out based on socioeconomic factors. Specifically, this project aims to:

- Identify key socioeconomic factors that influence student dropout rates.
- Build a predictive model to estimate dropout risk for individual students.
- Provide insights to educators and policymakers to develop early intervention strategies.



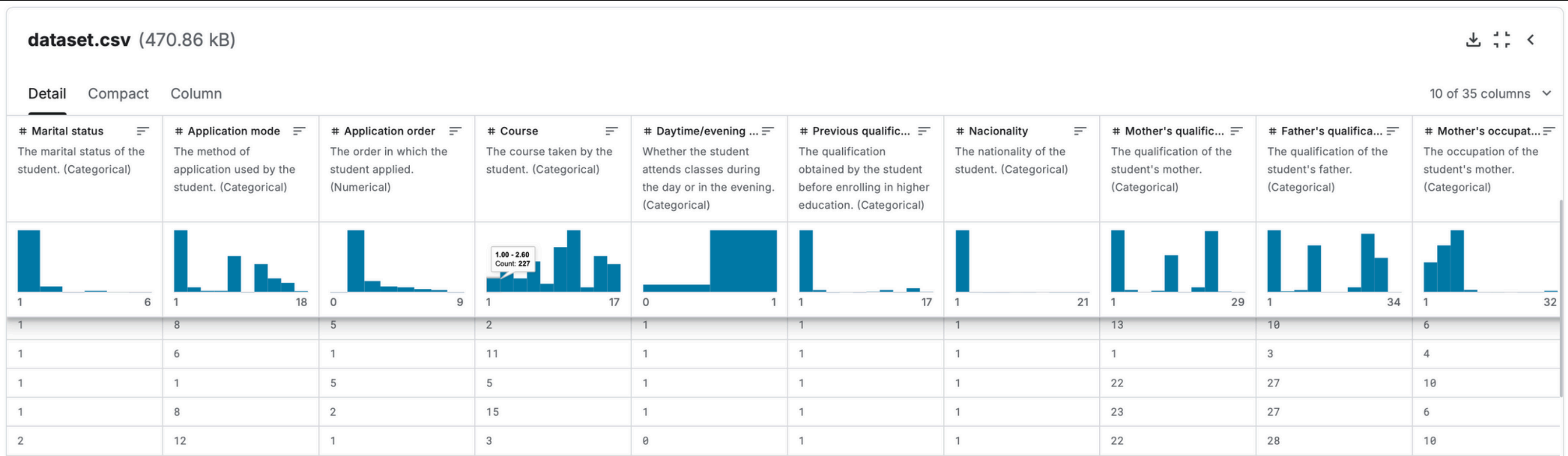
Importance of Dropout Prediction

- **Early Intervention:** Identifying students at risk allows schools to provide targeted support, such as tutoring, counseling, or financial aid.
 - **Improving Student Success:** Helps educators understand the factors leading to dropout, such as academic struggles, financial issues, or lack of engagement.
 - **Institutional Performance & Reputation:** High dropout rates can negatively impact a school's reputation and funding. Reducing dropout rates improves graduation rates and overall institutional success.
 - **Economic & Social Impact:** Students who drop out tend to earn lower wages and face higher unemployment rates. By reducing dropout rates, society benefits from a more skilled and educated workforce, reducing economic disparities.
 - **Data-Driven Decision Making:** Schools can use machine learning models and predictive analytics to analyze patterns and trends in dropout rates. Insights from data help in policy-making, resource allocation, and curriculum development.
- 

Why It Is Feasible

- **Feasibility:** This topic is highly feasible, thanks to the increasing availability of data and advancements in machine learning.
- **Existing Research:** It is not completely new. There is a growing body of research exploring the use of machine learning to predict student performance and dropout rates.
 - * Studies often use various machine learning techniques, including logistic regression, decision trees, random forests, and neural networks, to analyze student data.
 - * Research has also explored the impact of socioeconomic factors like family income, parental education, housing stability, and access to resources on student outcomes.
- **Data Availability:** Many educational institutions and government agencies collect data on student demographics, academic performance, and socioeconomic backgrounds. This data can be used to train and validate machine learning models.
- **Technological Advancement:** The constant advancement of Machine learning algorithms makes it more and more accurate to predict outcomes using large datasets.

About Dataset



This dataset contains data from a higher education institution on various variables related to undergraduate students, including demographics, social-economic factors, and academic performance, to investigate the impact of these factors on student dropout and academic success

METHODOLOGY

- **Data Preprocessing:**
 - Clean the data by handling missing values, removing duplicates, and normalizing features.
 - Modify some column names to make them more consistent and easier to work with
 - Change the data types of columns that should be categorical from 'int' to 'category'
- **EDA:**
 - Examine the relationship between each features with the target variable [e.g., Chi-square Independence Test, Correlation Heat map].
- **Model Selection:**
 - We plan to use machine learning algorithms such as [e.g., Logistic Regression, Random Forests, XGBoost].
 - These models will help us identify patterns and predict dropout likelihood and academic success.

Final Impact

By using a data-driven, ML-based predictive approach, this research can help policymakers allocate resources effectively, prevent student dropouts, and ensure that students receive the necessary support before they disengage from education. This aligns with **SDG 4 (Quality Education)** by improving student retention rates and promoting equitable learning opportunities.