PROJECT: Student Performance Analysis

Step-by-Step Process Breakdown

Data Collection and Cleaning

My first step involved **acquiring raw student data** from a structured dataset found in the **UCI Machine Learning Repository**. Once I had the data, I dedicated significant effort to **cleaning it within Excel**. This meant:

- **Removing any null (empty) values** to make sure my data was complete and accurate.
- **Standardizing various data formats** to maintain consistency across the dataset.
- **Converting text information (categorical values) into numbers** or a more analysis-friendly format.

Feature Engineering

To extract deeper insights and prepare the data for analysis, I performed several **feature engineering** steps:

- I **converted all student scores to a standardized 4-point scale**, allowing for consistent comparisons.
- I **segregated students into distinct groups** based on important attributes such as gender, caste, and identified learning styles, enabling targeted analysis.

Attribute Name	Meaning	Values	Explanation
ge	Gender	M = Male, F = Female	Student gender
cst	Caste	G = General, ST = Scheduled Tribe, SC = Scheduled Caste, OBC = Other Backward Class, MOBC = Minorities/C	Caste category
tnp	10th Grade Performance	Best, Vg = Very Good, Good, Pass, Fail	Student's 10th grade result category
twp	12th Grade Performance	Same as above	Student's 12th grade result category
iap	Internal Assessment Performance	Same as above	Internal test performance
esp	End-Semester Performance	Same as above	Final academic performance
arr	Regular Attendance	Y = Yes, N = No	Whether student maintained regular attendance
ms	Marital Status	Married, Unmarried	Student's marital status
ls	Learning Style	T = Theoretical, V = Visual	Preferred learning method
as	Academic Support Type	Free, Paid	Whether student received free or paid tuition
fmi	Family Monthly Income	Vh = Very High, High, Am = Above Medium, Medium, Low	Income bracket of family
fs	Family Size	Large, Average, Small	Number of family members
fq	Father's Qualification	Il = Illiterate, Um = Upto Middle School, 10 = 10th, 12 = 12th, Degree, Pg = Postgraduate	Father's education level
mq	Mother's Qualification	Same as above	Mother's education level
fo	Father's Occupation	Service, Business, Retired, Farmer, Others	Occupation type
mo	Mother's Occupation	Service, Business, Retired, Housewife, Others	Occupation type
nf	Number of Family Members	Large, Average, Small	Similar to fs, could be redundant or a separate self-repor
sh	Student Health	Good, Average, Poor	Health condition of the student
ss	Schooling System	Govt, Private	Type of school attended (in school education)
me	Medium of Education	Eng = English, Asm = Assamese, Hin = Hindi, Ben = Bengali	Language medium in school
tt	Travel Time	Large, Average, Small	Time taken to reach college
atd	Attendance	Good, Average, Poor	Actual attendance in college classes

Category	Value
Best	4
Very Good	3
Good	2
Pass	1

Data Analysis

With the data prepared, I proceeded with the **data analysis** phase:

- I **leveraged Excel's pivot table functionality** to evaluate various key metrics. Specifically, I analyzed:
    - The **average end-semester scores**, disaggregated by gender, caste, and different study types.
    - The **relationship between various learning styles and overall academic performance**.
- I also employed a range of **Excel formulas, including AVERAGEIF and COUNTIF**, combined with strategic sorting and filtering, to **identify significant trends and uncover valuable insights** within the student data.

Key Performance Indicators (KPIs) Summary

To provide a concise overview of the project's findings, I created a dedicated **KPIs sheet**:

- On this sheet, I **summarized key average performance metrics**, drawing from both the original raw data and the newly derived information.
- I **designed this section as a clean, easy-to-reference table**, ensuring that important data points were immediately accessible.

KPIs	
Category	Value
Average 10th Score out of 4	2.24
Average 12th Score out of 4	2.28
Average Internal Score out of 4	2.55
Average End Sem Score out of 4	2.24
% Female Candidates	45.04
Average Female Candidate End Sem Score out of 4	2.32
% Male Candidate	54.96
Average Male Candidate End Sem Score out of 4	2.17

Designing the Interactive Dashboard

I wrapped up the project with a dynamic **dashboard sheet** that visually shows all my findings:

- I **compiled a series of summary visuals**, primarily using **bar and column charts** along with **comparison tables**, to present the findings effectively.
- To ensure clarity and ease of understanding, I **grouped charts logically by their analysis types** (e.g., charts related to gender, caste, or learning style were placed together).
- Ensured **clean visual separation and clearly labeled sections**, creating a streamlined and highly comprehensible user experience.
- Added **slicers** to ensure the dashboard is **interactive** and helps the user to retrieve and segregate the desired information.