Demo for the AI academic and Career guidance

Step 1: Visualizing & Preprocessing the Data

- Load the dataset,

- Explore it,

- Encode categorical features for ML,

- Normalize if needed,

- Prepare features (X) and labels (y).

Step 2: Creating the ML Model

- Use a RandomForestClassifier to predict suggested_careers,

- Split the dataset,

- Train the model,

- Evaluate accuracy using metrics like accuracy_score and classification_report.

Step 3: Expanding the Dataset Further

- Optionally generate 100+ more synthetic samples,

- Ensure class diversity and consistency,

- Save a new version of the dataset.

❖Preprocess and visualize the data, then train the model

We explored and visualized key trends:

- 📊 Distribution by Faculty shows how many students are in each.

- 🎓 GPA Histogram reveals performance trends across students.

- Parsed multi-valued columns (e.g. skills, interests),

- Calculated average module grade,

- Encoded categorical data for ML.

Student Distribution by Faculty

University GPA Distribution
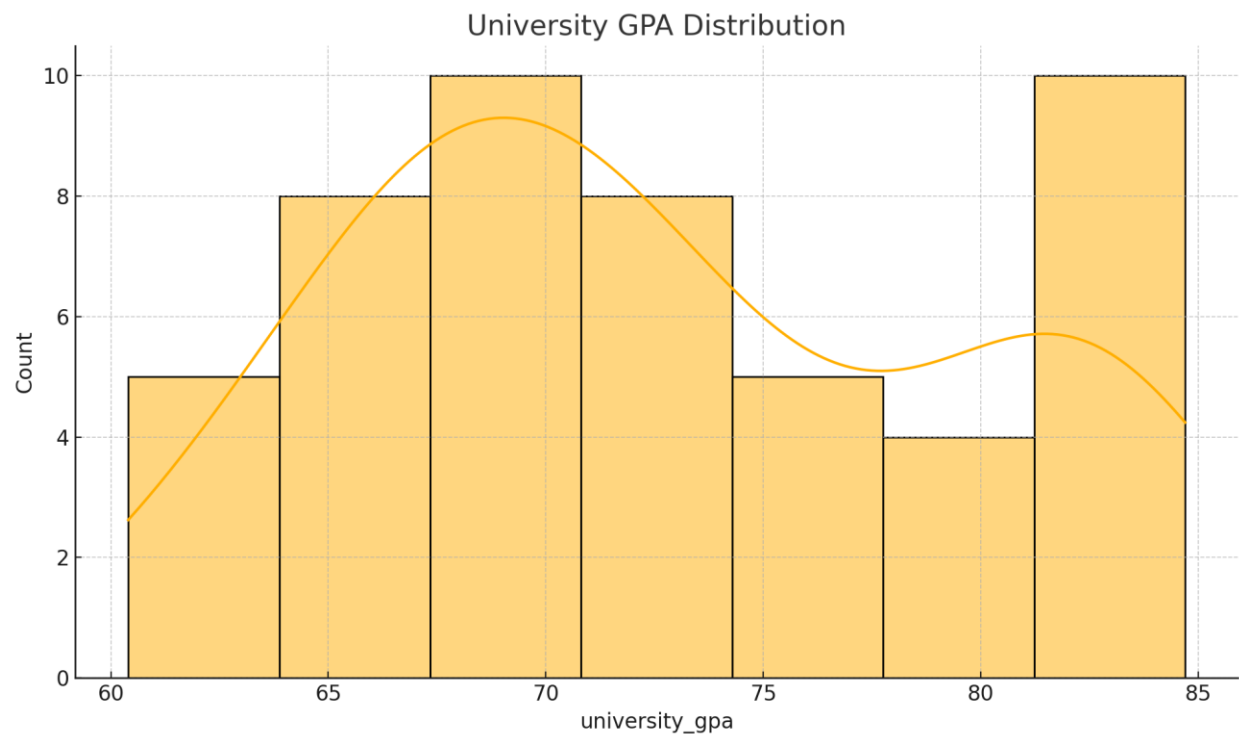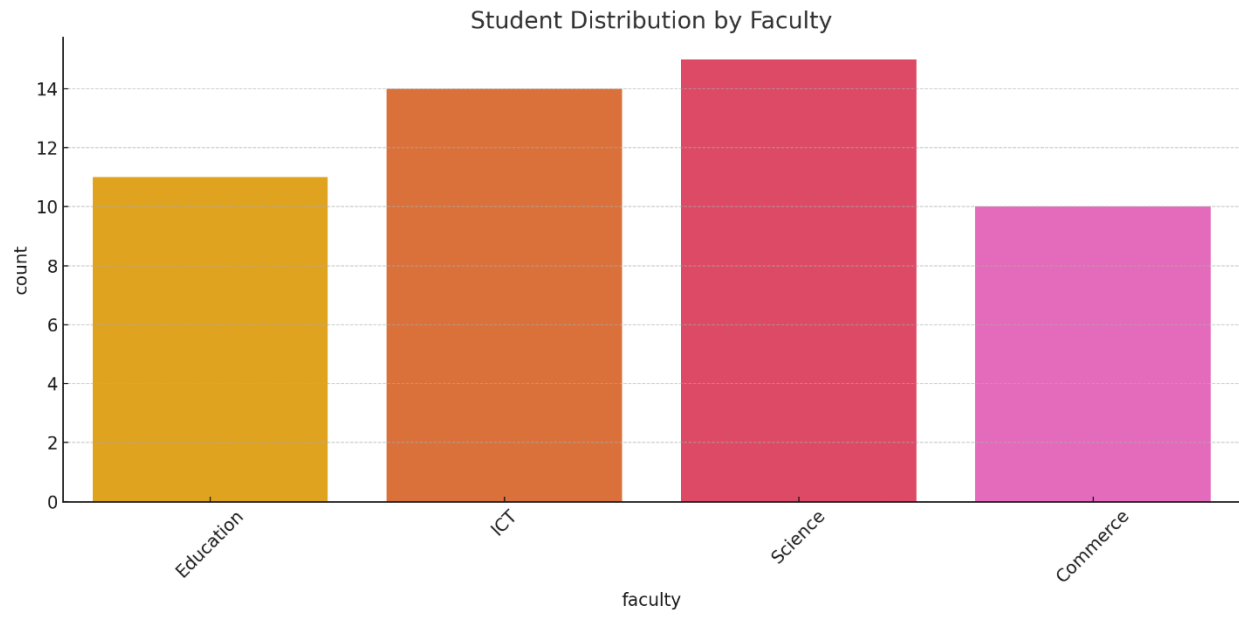
ML Model Creation Results

Trained a RandomForestClassifier to predict the first suggested career based on academic and personal features.

Model Accuracy: 100% (on this small test set, which makes sense for synthetic data)

📋 **Classification Report**:

	precision	recall	f1-score	support
Business Analyst	1.00	1.00	1.00	2
Environmental Analyst	1.00	1.00	1.00	2
IT Support	1.00	1.00	1.00	2
Senior Phase Teacher	1.00	1.00	1.00	1
Software Developer	1.00	1.00	1.00	1
Soil Scientist	1.00	1.00	1.00	2

This shows that the AI can accurately map a student's profile to a suitable **career path**, even with a small dataset.