Database 4 (Emphasis on economic factors, very large number of instances)

Predict students' dropout and academic success

Investigating the Impact of Social and Economic Factors

https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention

INTRODUCTION :

- 4424 rows and 23 feature columns
- It includes demographic, social economic and academic performance data.
- Which factors are linked with student dropout or completion ?
- How different features interact with each other ?

GOAL OF RESEARCH :

1. Predict student retention : identify student risk factors for drop outs to take early interventions
2. Improve academic performance : educational institutions could better understand their student's academic performance and identify areas of improvement from both individual and institutional perspective.
3. Using demographic information : could motivate institutions to develop specific initiatives to help certain groups more easily to access higher education.

IMPLEMENTATION :

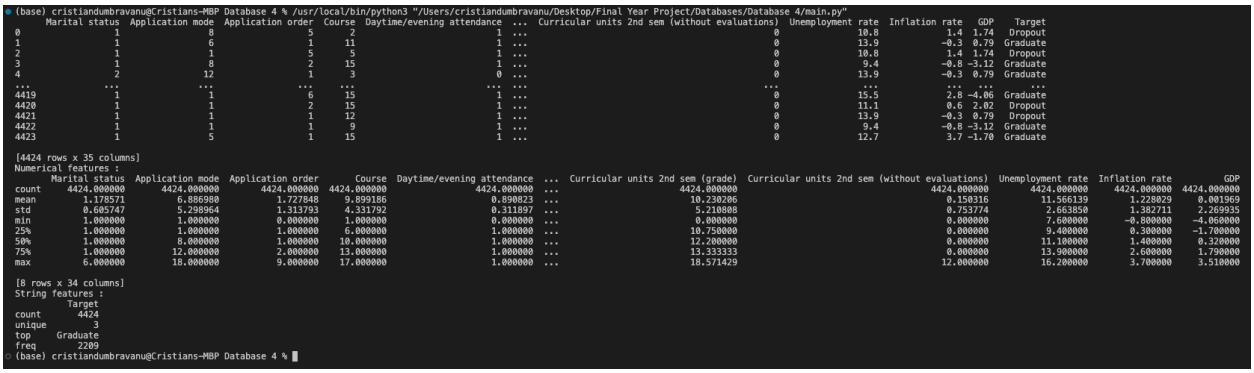
- They have :
  - Unemployment rate, inflation rate, and GDP from their region to help understand how economic factors play a role in academic performance.

```
import numpy as np
import pandas as pd
import sklearn.linear_model, sklearn.datasets
from sklearn.preprocessing import StandardScaler, MinMaxScaler
import matplotlib.pyplot as plt
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import mean_squared_error, r2_score, mean_a
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.impute import SimpleImputer
pd.options.mode.chained_assignment = None

rawData = pd.read_csv('dataset.csv')

print(rawData)

print("Numerical features : ")
print(rawData.select_dtypes(include=np.number).describe())
print("String features : ")
print(rawData.select_dtypes(exclude=np.number).describe())
```



```
(base) cristiandumbravanu@Cristians-MBP Database 4 % /usr/local/bin/python3 "/Users/cristiandumbravanu/Desktop/Final Year Project/Databases/Database 4/main.py"
      Marital status  Application mode  Application order  Course  Daytime/evening attendance  ...  Curricular units 2nd sem (without evaluations)  Unemployment rate  Inflation rate   GDP    Target
0                  1                 8                  5       2                           1  ...                                               0               10.8             1.4  1.74   Dropout
1                  1                 6                  1      11                           1  ...                                               0               13.9            -0.3  0.79  Graduate
2                  1                 1                  5       5                           1  ...                                               0               10.8             1.4  1.74   Dropout
3                  1                 8                  2      15                           1  ...                                               0                9.4            -0.8 -3.12  Graduate
4                  2                12                  1       3                           0  ...                                               0               13.9            -0.3  0.79  Graduate
...              ...               ...                ...     ...                         ...  ...                                             ...                ...             ...   ...       ...
4419               1                 1                  6      15                           1  ...                                               0               15.5             2.8 -4.06  Graduate
4420               1                 1                  2      15                           1  ...                                               0               11.1             0.6  2.02   Dropout
4421               1                 1                  1      12                           1  ...                                               0               13.9            -0.3  0.79   Dropout
4422               1                 1                  1       9                           1  ...                                               0                9.4            -0.8 -3.12  Graduate
4423               1                 5                  1      15                           1  ...                                               0               12.7             3.7 -1.70  Graduate

[4424 rows x 35 columns]
Numerical features :
       Marital status  Application mode  Application order       Course  Daytime/evening attendance  ...  Curricular units 2nd sem (grade)  Curricular units 2nd sem (without evaluations)  Unemployment rate  Inflation rate          GDP
count     4424.000000       4424.000000        4424.000000  4424.000000                 4424.000000  ...                       4424.000000                                     4424.000000     4424.000000  4424.000000  4424.000000
mean         1.178571          6.886980           1.727848     9.899186                    0.890823  ...                         10.230206                                        0.150316       11.566139     1.228029     0.001969
std          0.605747          5.298964           1.313793     4.331792                    0.311897  ...                          5.210808                                        0.753774        2.663850     1.382711     2.269935
min          1.000000          1.000000           0.000000     1.000000                    0.000000  ...                          0.000000                                        0.000000        7.600000    -0.800000    -4.060000
25%          1.000000          1.000000           1.000000     6.000000                    1.000000  ...                         10.750000                                        0.000000        9.400000     0.300000    -1.700000
50%          1.000000          8.000000           1.000000    10.000000                    1.000000  ...                         12.200000                                        0.000000       11.100000     1.400000     0.320000
75%          1.000000         12.000000           2.000000    13.000000                    1.000000  ...                         13.333333                                        0.000000       13.900000     2.600000     1.790000
max          6.000000         18.000000           9.000000    17.000000                    1.000000  ...                         18.571429                                       12.000000       16.200000     3.700000     3.510000

[8 rows x 34 columns]
String features :
          Target
count       4424
unique         3
top     Graduate
freq        2209
(base) cristiandumbravanu@Cristians-MBP Database 4 %
```