Course Final Project : Unsupervised Machine Learning (Clustering)

Student Dropout and Academic Success

Objective of the Analysis

The main objective of this project is to develop a machine learning model capable of clustering into one of three categories. Original dataset has three different target class: Dropout, Enrolled, or Graduate. The goal is to explore various clustering techniques and analyze their effectiveness in identifying patterns within the student data. While the original dataset includes a predefined target variable, it is only used for evaluation rather than training the clustering models.

Dataset Overview

The dataset was acquired from the UCI Machine Learning Repository, created from a higher education institution's student records. The data was gathered from several disjoint databases related to students enrolled in different undergraduate degrees. This dataset aims to help predict student dropout and academic success. It help identify at-risk students, ultimately contributing to efforts aimed at reducing academic dropout and failure in higher education. This dataset consist :

- Number of Instances: 4424 student records
- Number of Features: 36, covering a wide range of attributes such as academic path, demographics, and socio-economic factors.
- Target/Predictor: Three category classification (dropout, enrolled, and graduate) representing student risk at the end of the normal duration of the course. This column will not be used in clustering model. However will used as reference to compared with model result

Data Exploration and Preprocessing

Exploratory Data Analysis (EDA)

- No missing values were found in the dataset.
- Feature correlations were examined, revealing some highly correlated columns.
- All features were numeric, which streamlined the clustering process.

Feature Engineering and Transformation

- Most columns were integers, so they were converted to float data types for consistency.
- Some features exhibited high skewness, which was corrected using logarithmic transformation.
- Standard Scaling was applied to normalize the data, ensuring all features were on a comparable scale.
- Stratified Shuffle Split was implemented to maintain the class distribution when splitting data for evaluation.

Model Training

Three different unsupervised machine learning models used. :

1. KMeans
2. Gaussian Mixture
3. Agglomerative Clustering

The same dataset was used for all three models, and the predicted cluster labels were appended to the dataset. To compare model performance, value counts for each model's predicted clusters and the original target labels were stored in a separate DataFrame for evaluation.

Model Evaluation

All three clustering models produced similar results, indicating that the dataset contains clear structure and patterns. K-Means and GMM showed particularly close results, suggesting that both approaches capture similar underlying distributions in the data. When comparing model predictions to the target column, discrepancies were observed, but a notable pattern emerged where one cluster consistently had a much higher representation compared to the other two.

```
result_df
```

	Target(Reference)	kmeans_cluster	gmm_cluster	agglomerative_cluster
0	2209	3503	3496	3390
1	1421	741	748	719
2	794	180	180	315

Final Model Recommendation

While the initial clustering models provided valuable insights, there are several ways to enhance model performance:

1. Dimensionality Reduction: Applying Principal Component Analysis (PCA)
2. Hyperparameter Tuning: Adjusting the number of clusters, covariance type (for GMM), and linkage criteria (for Agglomerative Clustering) may improve results.
3. Feature Selection: Removing redundant or highly correlated features could improve model interpretability and efficiency.

This project demonstrates the potential of clustering models to uncover patterns in student success and dropout trends. Further refinements and additional data sources could enhance predictive capabilities, supporting educational institutions in proactive decision-making.

Unsupervised Machine Learning: Final Project

Import Libraries

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import OneHotEncoder, LabelEncoder,OrdinalEncoder, StandardScaler, MinMaxScaler
from sklearn.model_selection import StratifiedShuffleSplit, train_test_split, GridSearchCV
from sklearn.cluster import KMeans
from sklearn.cluster import AgglomerativeClustering
from sklearn.mixture import GaussianMixture
```

Load Data and Explore

```
# Load dataset

# from google.colab import drive

# drive.mount('/content/drive',force_remount=True)

students_data = pd.read_csv('/content/drive/My Drive/Machine_Learning/Coursera/IBM_Machine_Learning/Unsupervised_Machine_Learning/Final_P
students_data.head()
```

	Marital Status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Previous qualification (grade)	Nacionality	Mother's qualification	Father qualificati
0	1	17	5	171	1	1	122.0	1	19	
1	1	15	1	9254	1	1	160.0	1	1	
2	1	1	5	9070	1	1	122.0	1	37	
3	1	17	2	9773	1	1	122.0	1	38	
4	2	39	1	8014	0	1	100.0	1	37	

5 rows × 37 columns

```
# Display basic dataset information
students_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 37 columns):
 #   Column                                    Non-Null Count  Dtype
---  ------                                    --------------  -----
 0   Marital Status                            4424 non-null   int64
 1   Application mode                          4424 non-null   int64
 2   Application order                         4424 non-null   int64
 3   Course                                    4424 non-null   int64
 4   Daytime/evening attendance                4424 non-null   int64
 5   Previous qualification                    4424 non-null   int64
 6   Previous qualification (grade)            4424 non-null   float64
 7   Nacionality                               4424 non-null   int64
 8   Mother's qualification                    4424 non-null   int64
 9   Father's qualification                    4424 non-null   int64
 10  Mother's occupation                       4424 non-null   int64
 11  Father's occupation                       4424 non-null   int64
 12  Admission grade                           4424 non-null   float64
 13  Displaced                                 4424 non-null   int64
 14  Educational special needs                 4424 non-null   int64
 15  Debtor                                    4424 non-null   int64
 16  Tuition fees up to date                   4424 non-null   int64
 17  Gender                                    4424 non-null   int64
 18  Scholarship holder                        4424 non-null   int64
 19  Age at enrollment                         4424 non-null   int64
 20  International                             4424 non-null   int64
 21  Curricular units 1st sem (credited)       4424 non-null   int64
 22  Curricular units 1st sem (enrolled)       4424 non-null   int64
 23  Curricular units 1st sem (evaluations)    4424 non-null   int64
 24  Curricular units 1st sem (approved)       4424 non-null   int64
 25  Curricular units 1st sem (grade)          4424 non-null   float64
```

```
 26  Curricular units 1st sem (without evaluations)  4424 non-null   int64
 27  Curricular units 2nd sem (credited)             4424 non-null   int64
 28  Curricular units 2nd sem (enrolled)             4424 non-null   int64
 29  Curricular units 2nd sem (evaluations)          4424 non-null   int64
 30  Curricular units 2nd sem (approved)             4424 non-null   int64
 31  Curricular units 2nd sem (grade)                4424 non-null   float64
 32  Curricular units 2nd sem (without evaluations)  4424 non-null   int64
 33  Unemployment rate                               4424 non-null   float64
 34  Inflation rate                                  4424 non-null   float64
 35  GDP                                             4424 non-null   float64
 36  Target                                          4424 non-null   object
dtypes: float64(7), int64(29), object(1)
memory usage: 1.2+ MB
```

```
# Generate descriptive statistics for the dataset and transpose the result for better readability
students_data.describe().T  # Transpose the summary statistics to display features as rows
```

Marital Status	4424.0	1.178571	0.605747	1.00	1.00	1.000000	1.000000	6.000000
Application mode	4424.0	18.669078	17.484682	1.00	1.00	17.000000	39.000000	57.000000
Application order	4424.0	1.727848	1.313793	0.00	1.00	1.000000	2.000000	9.000000
Course	4424.0	8856.642631	2063.566416	33.00	9085.00	9238.000000	9556.000000	9991.000000
Daytime/evening attendance	4424.0	0.890823	0.311897	0.00	1.00	1.000000	1.000000	1.000000
Previous qualification	4424.0	4.577758	10.216592	1.00	1.00	1.000000	1.000000	43.000000
Previous qualification (grade)	4424.0	132.613314	13.188332	95.00	125.00	133.100000	140.000000	190.000000
Nacionality	4424.0	1.873192	6.914514	1.00	1.00	1.000000	1.000000	109.000000
Mother's qualification	4424.0	19.561935	15.603186	1.00	2.00	19.000000	37.000000	44.000000
Father's qualification	4424.0	22.275316	15.343108	1.00	3.00	19.000000	37.000000	44.000000
Mother's occupation	4424.0	10.960895	26.418253	0.00	4.00	5.000000	9.000000	194.000000
Father's occupation	4424.0	11.032324	25.263040	0.00	4.00	7.000000	9.000000	195.000000
Admission grade	4424.0	126.978119	14.482001	95.00	117.90	126.100000	134.800000	190.000000
Displaced	4424.0	0.548373	0.497711	0.00	0.00	1.000000	1.000000	1.000000
Educational special needs	4424.0	0.011528	0.106760	0.00	0.00	0.000000	0.000000	1.000000
Debtor	4424.0	0.113698	0.317480	0.00	0.00	0.000000	0.000000	1.000000
Tuition fees up to date	4424.0	0.880651	0.324235	0.00	1.00	1.000000	1.000000	1.000000
Gender	4424.0	0.351718	0.477560	0.00	0.00	0.000000	1.000000	1.000000
Scholarship holder	4424.0	0.248418	0.432144	0.00	0.00	0.000000	0.000000	1.000000
Age at enrollment	4424.0	23.265145	7.587816	17.00	19.00	20.000000	25.000000	70.000000
International	4424.0	0.024864	0.155729	0.00	0.00	0.000000	0.000000	1.000000
Curricular units 1st sem (credited)	4424.0	0.709991	2.360507	0.00	0.00	0.000000	0.000000	20.000000
Curricular units 1st sem (enrolled)	4424.0	6.270570	2.480178	0.00	5.00	6.000000	7.000000	26.000000
Curricular units 1st sem (evaluations)	4424.0	8.299051	4.179106	0.00	6.00	8.000000	10.000000	45.000000
Curricular units 1st sem (approved)	4424.0	4.706600	3.094238	0.00	3.00	5.000000	6.000000	26.000000
Curricular units 1st sem (grade)	4424.0	10.640822	4.843663	0.00	11.00	12.285714	13.400000	18.875000
Curricular units 1st sem (without evaluations)	4424.0	0.137658	0.690880	0.00	0.00	0.000000	0.000000	12.000000
Curricular units 2nd sem (credited)	4424.0	0.541817	1.918546	0.00	0.00	0.000000	0.000000	19.000000
Curricular units 2nd sem (enrolled)	4424.0	6.232143	2.195951	0.00	5.00	6.000000	7.000000	23.000000
Curricular units 2nd sem (evaluations)	4424.0	8.063291	3.947951	0.00	6.00	8.000000	10.000000	33.000000
Curricular units 2nd sem (approved)	4424.0	4.435805	3.014764	0.00	2.00	5.000000	6.000000	20.000000
Curricular units 2nd sem (grade)	4424.0	10.230206	5.210808	0.00	10.75	12.200000	13.333333	18.571429
Curricular units 2nd sem (without evaluations)	4424.0	0.150316	0.753774	0.00	0.00	0.000000	0.000000	12.000000
Unemployment rate	4424.0	11.566139	2.663850	7.60	9.40	11.100000	13.900000	16.200000
Inflation rate	4424.0	1.228029	1.382711	-0.80	0.30	1.400000	2.600000	3.700000
GDP	4424.0	0.001969	2.269935	-4.06	-1.70	0.320000	1.790000	3.510000

Exploratary Data Analysis and Feature Engineering

```
Start coding or generate with AI.
```

```
# Check Target Column. We will drop this column later. We will try to see whether we can get similar propotion with 3 cluster.
students_data.Target.value_counts(normalize=True)
```

Target	proportion
Graduate	0.499322
Dropout	0.321203
Enrolled	0.179476

dtype: float64

```
feature_columns = students_data.select_dtypes(exclude=['object']).columns.tolist()
students_data[feature_columns].info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 36 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Marital Status                                  4424 non-null   int64
 1   Application mode                                4424 non-null   int64
 2   Application order                               4424 non-null   int64
 3   Course                                          4424 non-null   int64
 4   Daytime/evening attendance                      4424 non-null   int64
 5   Previous qualification                          4424 non-null   int64
 6   Previous qualification (grade)                  4424 non-null   float64
 7   Nacionality                                     4424 non-null   int64
 8   Mother's qualification                          4424 non-null   int64
 9   Father's qualification                          4424 non-null   int64
 10  Mother's occupation                             4424 non-null   int64
 11  Father's occupation                             4424 non-null   int64
 12  Admission grade                                 4424 non-null   float64
 13  Displaced                                       4424 non-null   int64
 14  Educational special needs                       4424 non-null   int64
 15  Debtor                                          4424 non-null   int64
 16  Tuition fees up to date                         4424 non-null   int64
 17  Gender                                          4424 non-null   int64
 18  Scholarship holder                              4424 non-null   int64
 19  Age at enrollment                               4424 non-null   int64
 20  International                                   4424 non-null   int64
 21  Curricular units 1st sem (credited)             4424 non-null   int64
 22  Curricular units 1st sem (enrolled)             4424 non-null   int64
 23  Curricular units 1st sem (evaluations)          4424 non-null   int64
 24  Curricular units 1st sem (approved)             4424 non-null   int64
 25  Curricular units 1st sem (grade)                4424 non-null   float64
 26  Curricular units 1st sem (without evaluations)  4424 non-null   int64
 27  Curricular units 2nd sem (credited)             4424 non-null   int64
 28  Curricular units 2nd sem (enrolled)             4424 non-null   int64
 29  Curricular units 2nd sem (evaluations)          4424 non-null   int64
 30  Curricular units 2nd sem (approved)             4424 non-null   int64
 31  Curricular units 2nd sem (grade)                4424 non-null   float64
 32  Curricular units 2nd sem (without evaluations)  4424 non-null   int64
 33  Unemployment rate                               4424 non-null   float64
 34  Inflation rate                                  4424 non-null   float64
 35  GDP                                             4424 non-null   float64
dtypes: float64(7), int64(29)
memory usage: 1.2 MB
```

```
# The correlation matrix
corr_mat = students_data[feature_columns].corr()

# Strip out the diagonal values for the next step
for x in range(len(students_data[feature_columns].columns)):
    corr_mat.iloc[x,x] = 0.0

corr_mat
```

	Marital Status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Previous qualification (grade)	Nacionality	Mothe qualifica
Marital Status	0.000000	0.264006	-0.125854	0.046365	-0.274939	0.062529	-0.022406	-0.008843	0.19
Application mode	0.264006	0.000000	-0.286357	0.065385	-0.304092	0.422411	-0.039020	-0.000661	0.11
Application order	-0.125854	-0.286357	0.000000	0.059507	0.158657	-0.184315	-0.064484	-0.022416	-0.06
Course	0.046365	0.065385	0.059507	0.000000	-0.043151	0.006654	-0.081013	-0.033923	0.05
Daytime/evening attendance	-0.274939	-0.304092	0.158657	-0.043151	0.000000	-0.071871	0.052597	0.018530	-0.20
Previous qualification	0.062529	0.422411	-0.184315	0.006654	-0.071871	0.000000	0.104072	-0.029214	-0.01
Previous qualification (grade)	-0.022406	-0.039020	-0.064484	-0.081013	0.052597	0.104072	0.000000	0.054088	-0.06
Nacionality	-0.008843	-0.000661	-0.022416	-0.033923	0.018530	-0.029214	0.054088	0.000000	-0.04
Mother's qualification	0.193163	0.118974	-0.064956	0.054543	-0.204767	-0.013190	-0.060670	-0.049946	0.00
Father's qualification	0.130353	0.083276	-0.050288	0.050724	-0.139894	-0.006614	-0.035234	-0.085282	0.53
Mother's occupation	0.034994	0.051600	-0.039039	0.031265	-0.019067	0.014822	-0.011415	0.043187	0.07
Father's occupation	0.031609	0.036707	-0.030014	0.028881	-0.015477	0.016263	-0.019284	0.020626	0.05
Admission grade	-0.004771	-0.013271	-0.096930	-0.125058	0.007970	0.184183	0.580444	0.028176	-0.05
Displaced	-0.234886	-0.301197	0.332362	-0.087399	0.251767	-0.115188	-0.011001	-0.007448	-0.08
Educational special needs	-0.028343	-0.030779	0.025597	-0.018795	0.031017	-0.010461	-0.001458	-0.005982	-0.02
Debtor	0.034304	0.122818	-0.072151	-0.032673	0.006658	0.095104	-0.037507	0.051989	0.01
Tuition fees up to date	-0.087158	-0.136262	0.055891	0.018381	0.038799	-0.068453	0.060578	-0.026115	-0.02
Gender	-0.014738	0.160130	-0.089559	-0.099571	-0.012326	0.078684	-0.047874	-0.023258	-0.05
Scholarship holder	-0.053765	-0.163117	0.073709	0.017138	0.093912	-0.070513	0.055965	-0.010490	0.04
Age at enrollment	0.522717	0.516243	-0.271154	0.042994	-0.462280	0.156234	-0.111377	-0.003647	0.29
International	-0.027905	0.003438	-0.028801	-0.026737	0.027973	-0.026226	0.048529	0.790935	-0.03
Curricular units 1st sem (credited)	0.061209	0.247426	-0.133354	-0.096335	-0.127466	0.166025	-0.008872	-0.000370	0.04
Curricular units 1st sem (enrolled)	0.052107	0.164299	-0.016808	0.328461	-0.043056	0.078702	-0.029169	-0.013292	0.05
Curricular units 1st sem (evaluations)	0.058030	0.225938	-0.092156	0.272845	-0.045889	0.130597	-0.070702	-0.007763	0.04
Curricular units 1st sem (approved)	-0.031027	-0.029055	0.035580	0.180500	0.016935	0.022020	0.048410	-0.002268	-0.01
Curricular units 1st sem (grade)	-0.059811	-0.117741	0.058308	0.389349	0.063974	-0.000497	0.059438	0.000869	-0.03
Curricular units 1st sem (without evaluations)	0.034711	0.045828	-0.031699	0.034514	0.045630	0.002887	-0.003926	0.009145	0.00
Curricular units 2nd sem (credited)	0.062831	0.238445	-0.125815	-0.089817	-0.111953	0.143031	-0.018489	-0.007278	0.04
Curricular units 2nd sem (enrolled)	0.039026	0.130046	0.028878	0.401539	0.000371	0.056179	-0.031649	-0.020113	0.03
Curricular units 2nd sem (evaluations)	0.022784	0.167872	-0.055089	0.278797	0.014610	0.114850	-0.061355	-0.025721	0.02

(evaluations)									
Curricular units 2nd sem (approved)	-0.043739	-0.071526	0.071793	0.198032	0.034022	-0.008632	0.050263	-0.017880	-0.01
Curricular units 2nd sem (grade)	-0.071506	-0.115424	0.055517	0.348728	0.050493	0.000942	0.053239	-0.008497	-0.03
Curricular units 2nd sem (without evaluations)	0.020426	0.047983	-0.015757	0.030816	-0.004229	0.005102	-0.019015	-0.014041	0.02
Unemployment rate	-0.020338	0.089080	-0.098419	0.007153	0.061974	0.111958	0.045222	-0.000651	-0.11
Inflation rate	0.008761	-0.016375	-0.011133	0.017710	-0.024043	-0.063736	0.018710	-0.008922	0.05
GDP	-0.027003	-0.022743	0.030201	-0.020265	0.022929	0.064069	-0.052620	0.034478	-0.08

36 rows × 36 columns

```
# Pairwise maximal correlations
corr_mat.abs().max().sort_values()
```

	0
Educational special needs	0.046131
Inflation rate	0.112295
Scholarship holder	0.202704
Gender	0.224266
Application order	0.332362
Unemployment rate	0.335178
GDP	0.335178
Displaced	0.362032
Course	0.401539
Debtor	0.408454
Tuition fees up to date	0.408454
Previous qualification	0.422411
Daytime/evening attendance	0.462280
Application mode	0.516243
Age at enrollment	0.522717
Marital Status	0.522717
Father's qualification	0.535140
Mother's qualification	0.535140
Admission grade	0.580444
Previous qualification (grade)	0.580444
Curricular units 2nd sem (without evaluations)	0.583261
Curricular units 1st sem (without evaluations)	0.583261
Curricular units 2nd sem (evaluations)	0.778863
Curricular units 1st sem (evaluations)	0.778863
Nacionality	0.790935
International	0.790935
Curricular units 1st sem (grade)	0.837170
Curricular units 2nd sem (grade)	0.837170
Curricular units 2nd sem (approved)	0.904002
Curricular units 1st sem (approved)	0.904002
Father's occupation	0.910472
Mother's occupation	0.910472
Curricular units 1st sem (enrolled)	0.942627
Curricular units 2nd sem (enrolled)	0.942627
Curricular units 2nd sem (credited)	0.944811
Curricular units 1st sem (credited)	0.944811

dtype: float64

Check skew values in anticipation of transformations.

```
skew_columns = (students_data[feature_columns].skew().sort_values(ascending=False))

skew_columns = skew_columns.loc[skew_columns > 0.75]
skew_columns
```

	0
Nacionality	10.703998
Educational special needs	9.154976
Curricular units 1st sem (without evaluations)	8.207403
Curricular units 2nd sem (without evaluations)	7.267701
International	6.104830
Father's occupation	5.395173
Mother's occupation	5.339227
Curricular units 2nd sem (credited)	4.634820
Marital Status	4.399764
Curricular units 1st sem (credited)	4.169049
Previous qualification	2.871207
Debtor	2.434652
Age at enrollment	2.054988
Application order	1.881050
Curricular units 1st sem (enrolled)	1.619041
Scholarship holder	1.164871
Curricular units 1st sem (evaluations)	0.976637
Curricular units 2nd sem (enrolled)	0.788114
Curricular units 1st sem (approved)	0.766262

dtype: float64

```
# Perform log transform on skewed columns
for col in skew_columns.index.tolist():
    students_data[col] = np.log1p(students_data[col])
```

```
# Perform standard scaler
scaler = StandardScaler()
students_data[feature_columns]= scaler.fit_transform(students_data[feature_columns])

students_data[feature_columns]
```

	Marital Status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Previous qualification (grade)	Nacionality	Mother's qualification	quali
0	-0.323463	-0.095470	2.322256	-4.209520	0.350082	-0.395901	-0.804841	-0.153037	-0.036018	
1	-0.323463	-0.209869	-0.606025	0.192580	0.350082	-0.395901	2.076819	-0.153037	-1.189759	
2	-0.323463	-1.010660	2.322256	0.103404	0.350082	-0.395901	-0.804841	-0.153037	1.117723	
3	-0.323463	-0.095470	0.474716	0.444115	0.350082	-0.395901	-0.804841	-0.153037	1.181819	
4	1.787906	1.162916	-0.606025	-0.408389	-2.856470	-0.395901	-2.473171	-0.153037	1.117723	
...	...	...	...	...	...	...	...	...	...	
4419	-0.323463	-1.010660	2.733134	0.444115	0.350082	-0.395901	-0.577342	-0.153037	-1.189759	
4420	-0.323463	-1.010660	0.474716	0.444115	0.350082	-0.395901	-0.956508	9.046312	-1.189759	
4421	-0.323463	-1.010660	-0.606025	0.311805	0.350082	-0.395901	1.621820	-0.153037	1.117723	
4422	-0.323463	-1.010660	-0.606025	0.140722	0.350082	-0.395901	3.593483	-0.153037	1.117723	
4423	-0.323463	-0.495866	-0.606025	0.444115	0.350082	-0.395901	1.470154	5.505993	1.181819	

4424 rows × 36 columns

```
students_data[feature_columns].describe().T
```

	count	mean	std	min	25%	50%	75%	max
Marital Status	4424.0	-3.244341e-16	1.000113	-0.323463	-0.323463	-0.323463	-0.323463	6.200020
Application mode	4424.0	-1.477621e-16	1.000113	-1.010660	-1.010660	-0.095470	1.162916	2.192505
Application order	4424.0	-1.991575e-16	1.000113	-2.453564	-0.606025	-0.606025	0.474716	3.683829
Course	4424.0	2.288706e-16	1.000113	-4.276402	0.110674	0.184826	0.338945	0.549769
Daytime/evening attendance	4424.0	7.066881e-17	1.000113	-2.856470	0.350082	0.350082	0.350082	0.350082
Previous qualification	4424.0	-8.753295e-17	1.000113	-0.395901	-0.395901	-0.395901	-0.395901	3.120153
Previous qualification (grade)	4424.0	-3.589654e-16	1.000113	-2.852337	-0.577342	0.036907	0.560156	4.351815
Nacionality	4424.0	-2.778569e-16	1.000113	-0.153037	-0.153037	-0.153037	-0.153037	9.132138
Mother's qualification	4424.0	-5.781993e-17	1.000113	-1.189759	-1.125662	-0.036018	1.117723	1.566400
Father's qualification	4424.0	0.000000e+00	1.000113	-1.386793	-1.256427	-0.213496	0.959802	1.416085
Mother's occupation	4424.0	6.986575e-17	1.000113	-2.381500	-0.389656	-0.164014	0.468185	4.144378
Father's occupation	4424.0	3.180096e-16	1.000113	-2.537909	-0.474360	0.128257	0.414362	4.229452
Admission grade	4424.0	-9.708930e-16	1.000113	-2.208378	-0.626926	-0.060642	0.540172	4.352230
Displaced	4424.0	-7.388103e-17	1.000113	-1.101914	-1.101914	0.907512	0.907512	0.907512
Educational special needs	4424.0	3.212219e-17	1.000113	-0.107993	-0.107993	-0.107993	-0.107993	9.259865
Debtor	4424.0	-2.650080e-17	1.000113	-0.358167	-0.358167	-0.358167	-0.358167	2.791994
Tuition fees up to date	4424.0	-1.317010e-16	1.000113	-2.716392	0.368135	0.368135	0.368135	0.368135
Gender	4424.0	1.027910e-16	1.000113	-0.736572	-0.736572	-0.736572	1.357640	1.357640
Scholarship holder	4424.0	-2.328858e-17	1.000113	-0.574914	-0.574914	-0.574914	-0.574914	1.739390
Age at enrollment	4424.0	1.856662e-15	1.000113	-1.019819	-0.608706	-0.418328	0.415030	4.334881
International	4424.0	-4.818328e-17	1.000113	-0.159682	-0.159682	-0.159682	-0.159682	6.262442
Curricular units 1st sem (credited)	4424.0	-1.284887e-17	1.000113	-0.358314	-0.358314	-0.358314	-0.358314	4.666063
Curricular units 1st sem (enrolled)	4424.0	5.203794e-16	1.000113	-4.222784	-0.260006	0.080924	0.376251	3.066514
Curricular units 1st sem (evaluations)	4424.0	-7.388103e-17	1.000113	-3.069388	-0.190988	0.180758	0.477591	2.593958
Curricular units 1st sem (approved)	4424.0	-6.424437e-18	1.000113	-2.019551	-0.185334	0.351140	0.555098	2.341197
Curricular units 1st sem (grade)	4424.0	-1.702476e-16	1.000113	-2.197102	0.074163	0.339635	0.569711	1.700182
Curricular units 1st sem (without evaluations)	4424.0	-8.030546e-18	1.000113	-0.245125	-0.245125	-0.245125	-0.245125	9.079579
Curricular units 2nd sem (credited)	4424.0	7.709324e-17	1.000113	-0.339689	-0.339689	-0.339689	-0.339689	5.240482
Curricular units 2nd sem (enrolled)	4424.0	5.219855e-16	1.000113	-4.325593	-0.268547	0.080493	0.382846	2.870413
Curricular units 2nd sem (evaluations)	4424.0	-8.994212e-17	1.000113	-2.042630	-0.522682	-0.016033	0.490616	6.317081
Curricular units 2nd sem (approved)	4424.0	-1.509743e-16	1.000113	-1.471527	-0.808050	0.187165	0.518904	5.163242
Curricular units 2nd sem (grade)	4424.0	3.212219e-17	1.000113	-1.963489	0.099764	0.378064	0.595585	1.600935
Curricular units 2nd sem (without evaluations)	4424.0	-2.890997e-17	1.000113	-0.238476	-0.238476	-0.238476	-0.238476	8.580943
Unemployment rate	4424.0	-5.460771e-17	1.000113	-1.489043	-0.813253	-0.175007	0.876222	1.739731
Inflation rate	4424.0	1.445498e-16	1.000113	-1.466871	-0.671242	0.124386	0.992345	1.787974
GDP	4424.0	2.569775e-17	1.000113	-1.789667	-0.749872	0.140122	0.787790	1.545607

So far what we have done are:

- We have removed Target columns
- Converted all the feature columns to float columns
- Columns with higher skew(>0.75) have been log tranform
- And all the columns have been rescaled to standard scaller

Now we are ready to perform cluster. Intention to do 3 cluster from this dataset.

Model Clustering

1. KMeans

```
km = KMeans(n_clusters=3, random_state=42)
km = km.fit(students_data[feature_columns])
```