ML PROJECT PREDICTING STUDENTS DROPOUT AND ACADEMIC SUCCESS In this project we are going to learn what aspect of the students life effects more on their studies. To prepare the dataset for analysis, we will initially install specific Python packages including numpy, pandas, seaborn, and matplotlib for data processing and analysis. For machine learning model, we will install Extreme Gradient Boosting (XGB) library and use binary logistics.

Importing necessary libraries

```
import warnings
warnings.filterwarnings("ignore")

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
import xgboost as xgb
from sklearn.metrics import accuracy_score
```

Reading the dataset

```
df=pd.read_csv("dataset.csv")
```

```
df
```

```
      Marital status  Application mode  Application order  Course  \
0                  1                 8                  5       2
1                  1                 6                  1      11
2                  1                 1                  5       5
3                  1                 8                  2      15
4                  2                12                  1       3
...              ...               ...                ...     ...
4419               1                 1                  6      15
4420               1                 1                  2      15
4421               1                 1                  1      12
4422               1                 1                  1       9
4423               1                 5                  1      15

      Daytime/evening attendance  Previous qualification  Nacionality
\
0                              1                       1            1

1                              1                       1            1

2                              1                       1            1

3                              1                       1            1
```

```
4                                  0                              1             1

...                              ...                            ...           ...

4419                               1                              1             1

4420                               1                              1            19

4421                               1                              1             1

4422                               1                              1             1

4423                               1                              1             9


      Mother's qualification  Father's qualification  Mother's 
occupation  \
0                         13                      10                  
6   
1                          1                       3                  
4   
2                         22                      27                 
10   
3                         23                      27                  
6   
4                         22                      28                 
10   
...                      ...                     ...                 
...   
4419                       1                       1                  
6   
4420                       1                       1                 
10   
4421                      22                      27                 
10   
4422                      22                      27                  
8   
4423                      23                      27                  
6   

      ...  Curricular units 2nd sem (credited)  \
0     ...                                    0   
1     ...                                    0   
2     ...                                    0   
3     ...                                    0   
4     ...                                    0   
...   ...                                  ...   
4419  ...                                    0   
```

```
4420  ...                                0
4421  ...                                0
4422  ...                                0
4423  ...                                0

      Curricular units 2nd sem (enrolled)  \
0                                       0
1                                       6
2                                       6
3                                       6
4                                       6
...                                   ...
4419                                    6
4420                                    6
4421                                    8
4422                                    5
4423                                    6

      Curricular units 2nd sem (evaluations)  \
0                                          0
1                                          6
2                                          0
3                                         10
4                                          6
...                                      ...
4419                                       8
4420                                       6
4421                                       9
4422                                       6
4423                                       6

      Curricular units 2nd sem (approved)  Curricular units 2nd sem 
(grade)  \
0                                       0
0.000000
1                                       6
13.666667
2                                       0
0.000000
3                                       5
12.400000
4                                       6
13.000000
...                                   ...
...
4419                                    5
12.666667
4420                                    2
11.000000
4421                                    1
```

```
13.500000
4422                                     5
12.000000
4423                                     6
13.000000

      Curricular units 2nd sem (without evaluations)  Unemployment
rate  \
0                                                  0
10.8
1                                                  0
13.9
2                                                  0
10.8
3                                                  0
9.4
4                                                  0
13.9
...                                              ...                ..
.
4419                                               0
15.5
4420                                               0
11.1
4421                                               0
13.9
4422                                               0
9.4
4423                                               0
12.7

      Inflation rate   GDP    Target
0                1.4  1.74   Dropout
1               -0.3  0.79  Graduate
2                1.4  1.74   Dropout
3               -0.8 -3.12  Graduate
4               -0.3  0.79  Graduate
...              ...   ...       ...
4419             2.8 -4.06  Graduate
4420             0.6  2.02   Dropout
4421            -0.3  0.79   Dropout
4422            -0.8 -3.12  Graduate
4423             3.7 -1.70  Graduate

[4424 rows x 35 columns]
```

```
#shows first five rows of dataset
df.head()
```

```
   Marital status  Application mode  Application order  Course  \
0               1                 8                  5       2
1               1                 6                  1      11
2               1                 1                  5       5
3               1                 8                  2      15
4               2                12                  1       3

   Daytime/evening attendance  Previous qualification  Nacionality  \
0                           1                       1            1
1                           1                       1            1
2                           1                       1            1
3                           1                       1            1
4                           0                       1            1

   Mother's qualification  Father's qualification  Mother's occupation
...  \
0                      13                      10                    6
...
1                       1                       3                    4
...
2                      22                      27                   10
...
3                      23                      27                    6
...
4                      22                      28                   10
...

   Curricular units 2nd sem (credited)  Curricular units 2nd sem
(enrolled)  \
0                                    0
0
1                                    0
6
2                                    0
6
3                                    0
6
4                                    0
6

   Curricular units 2nd sem (evaluations)  \
0                                       0
1                                       6
2                                       0
3                                      10
4                                       6

   Curricular units 2nd sem (approved)  Curricular units 2nd sem
(grade)  \
0                                    0
```

```
0.000000
1                                    6
13.666667
2                                    0
0.000000
3                                    5
12.400000
4                                    6
13.000000

   Curricular units 2nd sem (without evaluations)  Unemployment
rate  \
0                                               0              10.8

1                                               0              13.9

2                                               0              10.8

3                                               0               9.4

4                                               0              13.9


   Inflation rate   GDP    Target
0             1.4  1.74   Dropout
1            -0.3  0.79  Graduate
2             1.4  1.74   Dropout
3            -0.8 -3.12  Graduate
4            -0.3  0.79  Graduate

[5 rows x 35 columns]
```

```
#shows last five rows of datset
df.tail()
```

```
      Marital status  Application mode  Application order  Course  \
4419               1                 1                  6      15
4420               1                 1                  2      15
4421               1                 1                  1      12
4422               1                 1                  1       9
4423               1                 5                  1      15

      Daytime/evening attendance  Previous qualification  Nacionality
\
4419                           1                       1            1

4420                           1                       1           19

4421                           1                       1            1
```

```
4422                          1                       1            1

4423                          1                       1            9


      Mother's qualification  Father's qualification  Mother's 
occupation  \
4419                       1                       1                 
6   
4420                       1                       1                
10   
4421                      22                      27                
10   
4422                      22                      27                 
8   
4423                      23                      27                 
6   

      ...  Curricular units 2nd sem (credited)  \
4419  ...                                    0   
4420  ...                                    0   
4421  ...                                    0   
4422  ...                                    0   
4423  ...                                    0   

      Curricular units 2nd sem (enrolled)  \
4419                                    6   
4420                                    6   
4421                                    8   
4422                                    5   
4423                                    6   

      Curricular units 2nd sem (evaluations)  \
4419                                       8   
4420                                       6   
4421                                       9   
4422                                       6   
4423                                       6   

      Curricular units 2nd sem (approved)  Curricular units 2nd sem 
(grade)  \
4419                                    5                 
12.666667   
4420                                    2                 
11.000000   
4421                                    1                 
13.500000   
4422                                    5                 
```

```
12.000000
4423                                    6
13.000000

      Curricular units 2nd sem (without evaluations)  Unemployment
rate  \
4419                                               0
15.5
4420                                               0
11.1
4421                                               0
13.9
4422                                               0
9.4
4423                                               0
12.7

      Inflation rate   GDP    Target
4419             2.8 -4.06  Graduate
4420             0.6  2.02   Dropout
4421            -0.3  0.79   Dropout
4422            -0.8 -3.12  Graduate
4423             3.7 -1.70  Graduate

[5 rows x 35 columns]
```

```
#shows sample of dataset
df.sample()
```

```
     Marital status  Application mode  Application order  Course  \
840               1                 1                  1       9

     Daytime/evening attendance  Previous qualification
Nacionality  \
840                           1                       1           14


     Mother's qualification  Father's qualification  Mother's
occupation  ...  \
840                       3                       3
10  ...

     Curricular units 2nd sem (credited)  Curricular units 2nd sem
(enrolled)  \
840                                    0
5

     Curricular units 2nd sem (evaluations)  \
840                                       5
```

```
     Curricular units 2nd sem (approved)  Curricular units 2nd sem
(grade)  \
840                                    0
0.0

     Curricular units 2nd sem (without evaluations)  Unemployment rate
\
840                                               0                8.9


     Inflation rate   GDP   Target
840             1.4  3.51  Dropout

[1 rows x 35 columns]
```

```
#shows the number of rows and columns
df.shape
```

```
(4424, 35)
```

```
df.size
```

```
154840
```

```
#info shows all the information of the dataset i.e. how many rows and
columns are present in the dataset and the datatypes
#of the them,if there are any missing or null values present we get to
know all of this from df.info()
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 35 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   Nacionality                                     4424 non-null
int64
```

```
 7   Mother's qualification                          4424 non-null
int64
 8   Father's qualification                          4424 non-null
int64
 9   Mother's occupation                             4424 non-null
int64
 10  Father's occupation                             4424 non-null
int64
 11  Displaced                                       4424 non-null
int64
 12  Educational special needs                       4424 non-null
int64
 13  Debtor                                          4424 non-null
int64
 14  Tuition fees up to date                         4424 non-null
int64
 15  Gender                                          4424 non-null
int64
 16  Scholarship holder                              4424 non-null
int64
 17  Age at enrollment                               4424 non-null
int64
 18  International                                   4424 non-null
int64
 19  Curricular units 1st sem (credited)             4424 non-null
int64
 20  Curricular units 1st sem (enrolled)             4424 non-null
int64
 21  Curricular units 1st sem (evaluations)          4424 non-null
int64
 22  Curricular units 1st sem (approved)             4424 non-null
int64
 23  Curricular units 1st sem (grade)                4424 non-null
float64
 24  Curricular units 1st sem (without evaluations)  4424 non-null
int64
 25  Curricular units 2nd sem (credited)             4424 non-null
int64
 26  Curricular units 2nd sem (enrolled)             4424 non-null
int64
 27  Curricular units 2nd sem (evaluations)          4424 non-null
int64
 28  Curricular units 2nd sem (approved)             4424 non-null
int64
 29  Curricular units 2nd sem (grade)                4424 non-null
float64
 30  Curricular units 2nd sem (without evaluations)  4424 non-null
int64
 31  Unemployment rate                               4424 non-null
float64
```

```
 32  Inflation rate                                  4424 non-null
float64
 33  GDP                                             4424 non-null
float64
 34  Target                                          4424 non-null
object
dtypes: float64(5), int64(29), object(1)
memory usage: 1.2+ MB
```

```
#describe shows the statistical structure of the dataset,we get to see
the mean,standard value
df.describe()
```

```
       Marital status  Application mode  Application order
Course  \
count     4424.000000       4424.000000        4424.000000
4424.000000
mean         1.178571          6.886980           1.727848
9.899186
std          0.605747          5.298964           1.313793
4.331792
min          1.000000          1.000000           0.000000
1.000000
25%          1.000000          1.000000           1.000000
6.000000
50%          1.000000          8.000000           1.000000
10.000000
75%          1.000000         12.000000           2.000000
13.000000
max          6.000000         18.000000           9.000000
17.000000

       Daytime/evening attendance  Previous qualification  Nacionality
\
count                 4424.000000             4424.000000  4424.000000

mean                     0.890823                2.531420     1.254521

std                      0.311897                3.963707     1.748447

min                      0.000000                1.000000     1.000000

25%                      1.000000                1.000000     1.000000

50%                      1.000000                1.000000     1.000000

75%                      1.000000                1.000000     1.000000

max                      1.000000               17.000000    21.000000
```

```
       Mother's qualification  Father's qualification  Mother's
occupation  \
count             4424.000000             4424.000000
4424.000000
mean                12.322107               16.455244
7.317812
std                  9.026251               11.044800
3.997828
min                  1.000000                1.000000
1.000000
25%                  2.000000                3.000000
5.000000
50%                 13.000000               14.000000
6.000000
75%                 22.000000               27.000000
10.000000
max                 29.000000               34.000000
32.000000

       ...  Curricular units 1st sem (without evaluations)  \
count  ...                                       4424.000000
mean   ...                                          0.137658
std    ...                                          0.690880
min    ...                                          0.000000
25%    ...                                          0.000000
50%    ...                                          0.000000
75%    ...                                          0.000000
max    ...                                         12.000000

       Curricular units 2nd sem (credited)  \
count                          4424.000000
mean                              0.541817
std                               1.918546
min                               0.000000
25%                               0.000000
50%                               0.000000
75%                               0.000000
max                              19.000000

       Curricular units 2nd sem (enrolled)  \
count                          4424.000000
mean                              6.232143
std                               2.195951
min                               0.000000
25%                               5.000000
50%                               6.000000
75%                               7.000000
max                              23.000000
```

```
       Curricular units 2nd sem (evaluations)  \
count                             4424.000000
mean                                 8.063291
std                                  3.947951
min                                  0.000000
25%                                  6.000000
50%                                  8.000000
75%                                 10.000000
max                                 33.000000

       Curricular units 2nd sem (approved)  Curricular units 2nd sem
(grade)  \
count                          4424.000000
4424.000000
mean                              4.435805
10.230206
std                               3.014764
5.210808
min                               0.000000
0.000000
25%                               2.000000
10.750000
50%                               5.000000
12.200000
75%                               6.000000
13.333333
max                              20.000000
18.571429

       Curricular units 2nd sem (without evaluations)  Unemployment
rate  \
count                                     4424.000000
4424.000000
mean                                         0.150316
11.566139
std                                          0.753774
2.663850
min                                          0.000000
7.600000
25%                                          0.000000
9.400000
50%                                          0.000000
11.100000
75%                                          0.000000
13.900000
max                                         12.000000
16.200000

       Inflation rate          GDP
count     4424.000000  4424.000000
```

```
mean         1.228029     0.001969
std          1.382711     2.269935
min         -0.800000    -4.060000
25%          0.300000    -1.700000
50%          1.400000     0.320000
75%          2.600000     1.790000
max          3.700000     3.510000

[8 rows x 34 columns]
```

```
#shows null values in the dataset
df.isnull().sum()
```

```
Marital status                                    0
Application mode                                  0
Application order                                 0
Course                                            0
Daytime/evening attendance                        0
Previous qualification                            0
Nacionality                                       0
Mother's qualification                            0
Father's qualification                            0
Mother's occupation                               0
Father's occupation                               0
Displaced                                         0
Educational special needs                         0
Debtor                                            0
Tuition fees up to date                           0
Gender                                            0
Scholarship holder                                0
Age at enrollment                                 0
International                                     0
Curricular units 1st sem (credited)               0
Curricular units 1st sem (enrolled)               0
Curricular units 1st sem (evaluations)            0
Curricular units 1st sem (approved)               0
Curricular units 1st sem (grade)                  0
Curricular units 1st sem (without evaluations)    0
Curricular units 2nd sem (credited)               0
Curricular units 2nd sem (enrolled)               0
Curricular units 2nd sem (evaluations)            0
Curricular units 2nd sem (approved)               0
Curricular units 2nd sem (grade)                  0
Curricular units 2nd sem (without evaluations)    0
Unemployment rate                                 0
Inflation rate                                    0
GDP                                               0
Target                                            0
dtype: int64
```

```python
#this will show if there are any duplicate values present in the dataset
df.duplicated()
```

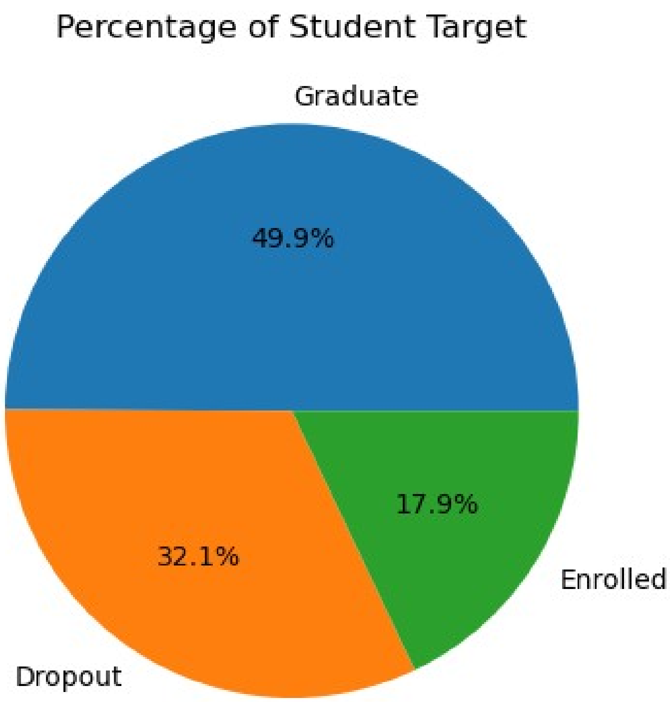
```
0       False
1       False
2       False
3       False
4       False
        ...  
4419    False
4420    False
4421    False
4422    False
4423    False
Length: 4424, dtype: bool
```

There are no null and duplicated values in the data.

Data Visualisation(EDA)

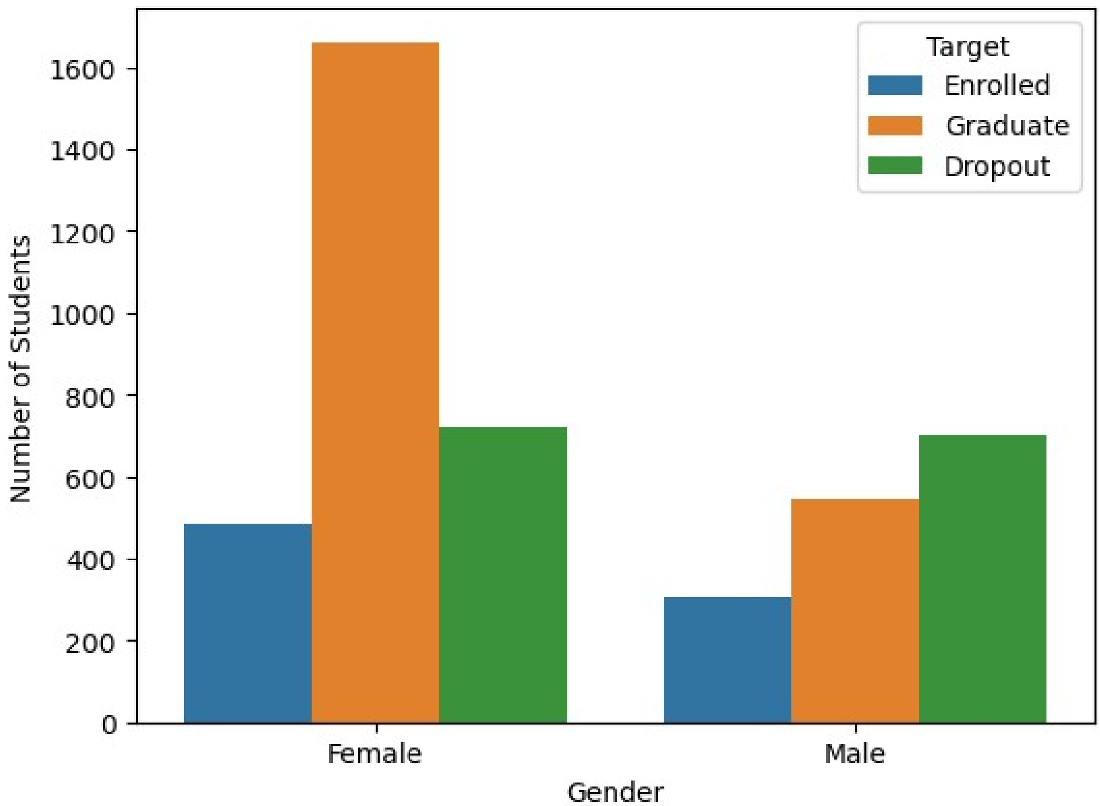
Since most of the variables in the data set are categorical, we will mainly use bar graphs to visualize them. However, for variables that are discrete or continuous, we will use distribution plots to display their distribution. Additionally, we will utilize correlation heatmaps to examine the relationships between variables in the data.

```python
student_target=df['Target'].value_counts()
plt.pie(student_target,labels=student_target.index,autopct='%1.1f%%')
plt.title('Percentage of Student Target')
plt.show()
```



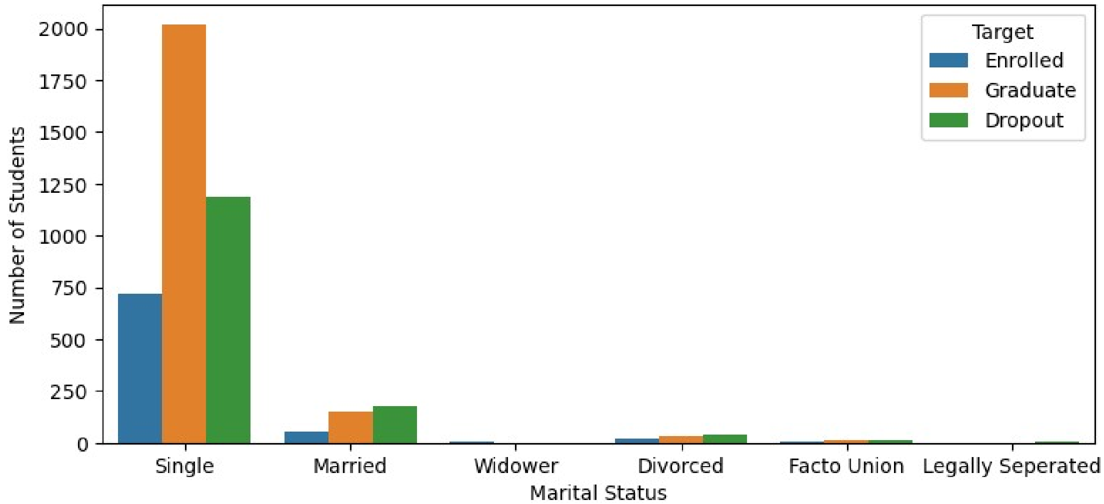
Approximately 50% of students in the data have graduated.

```
sns.countplot(data=df,x='Gender',hue='Target',hue_order=['Enrolled','G
raduate','Dropout'])
plt.xticks(ticks=[0,1],labels=['Female','Male'])
plt.ylabel('Number of Students')
plt.show()
```



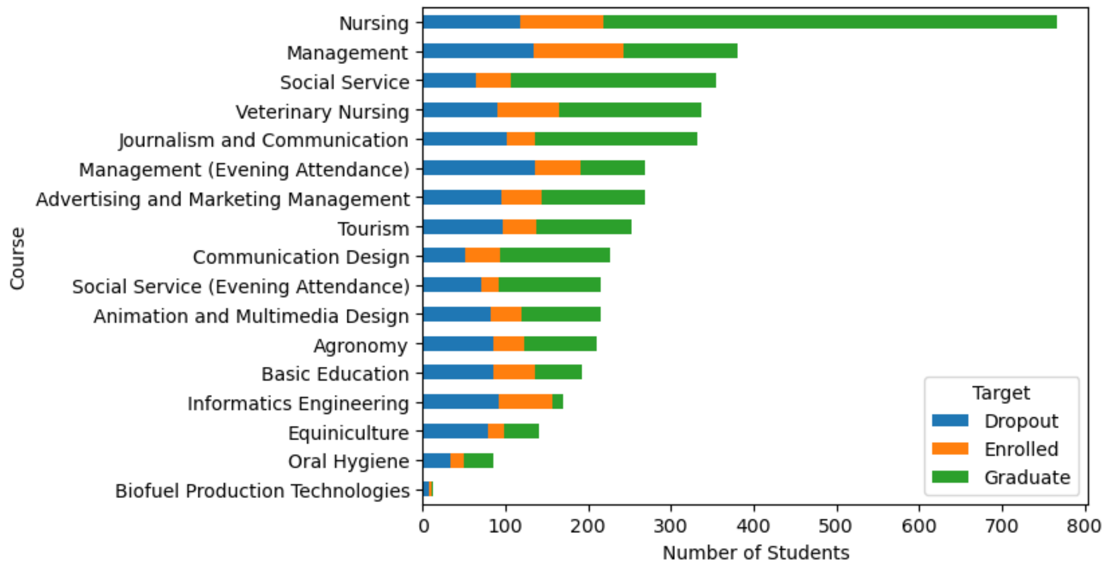
According to the data, a higher number of graduates are female. However, females also have the highest number of dropouts, although the difference compared to males is small.

```
plt.figure(figsize=(9,4))
sns.countplot(data=df,x='Marital status',hue='Target',hue_order=['Enrolled','Graduate','Dropout'])
plt.xticks(ticks=[0,1,2,3,4,5],
labels=['Single','Married','Widower','Divorced','Facto Union','Legally Seperated'])
plt.xlabel('Marital Status')
plt.ylabel('Number of Students')
plt.show()
```



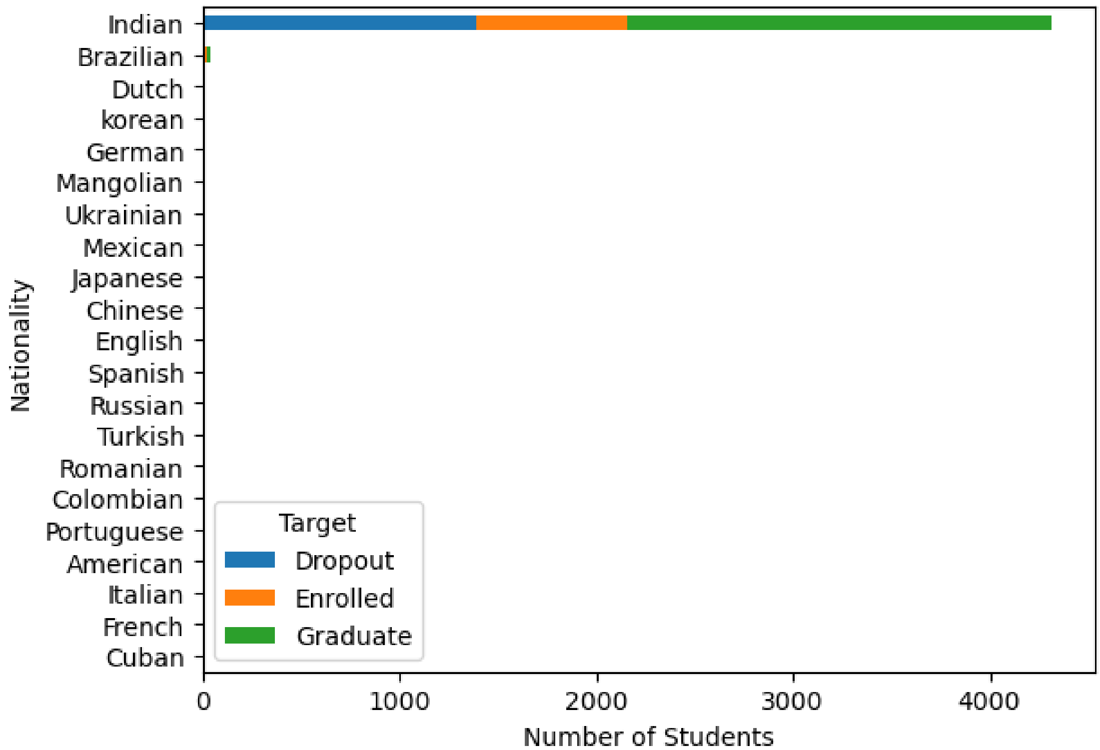
Regarding marital status, the majority of both graduates and dropouts are single.

```
student_course=df.groupby(['Course','Target']).size().reset_index().pivot(columns='Target',index='Course',values=0)

# Rename the index of the DataFrame
student_course=student_course.rename(index={1:'Biofuel Production Technologies',2:'Animation and Multimedia Design',3:'Social Service (Evening Attendance)',4:'Agronomy',5:'Communication Design',6:'Veterinary Nursing',7:'Informatics Engineering',8:'Equiniculture',9:'Management',10:'Social Service',11:'Tourism',12:'Nursing',13:'Oral Hygiene',14:'Advertising and Marketing Management',15:'Journalism and Communication',16:'Basic Education',17:'Management (Evening Attendance)'})
student_course_total=student_course.sum(axis=1)
student_course_sorted=student_course_total.sort_values(ascending=True)
student_course.loc[student_course_sorted.index].plot(kind='barh',stacked=True)
plt.xlabel('Number of Students')
plt.show()
```



Nursing course produced the highest number of graduates while management course has the highest number of dropouts.

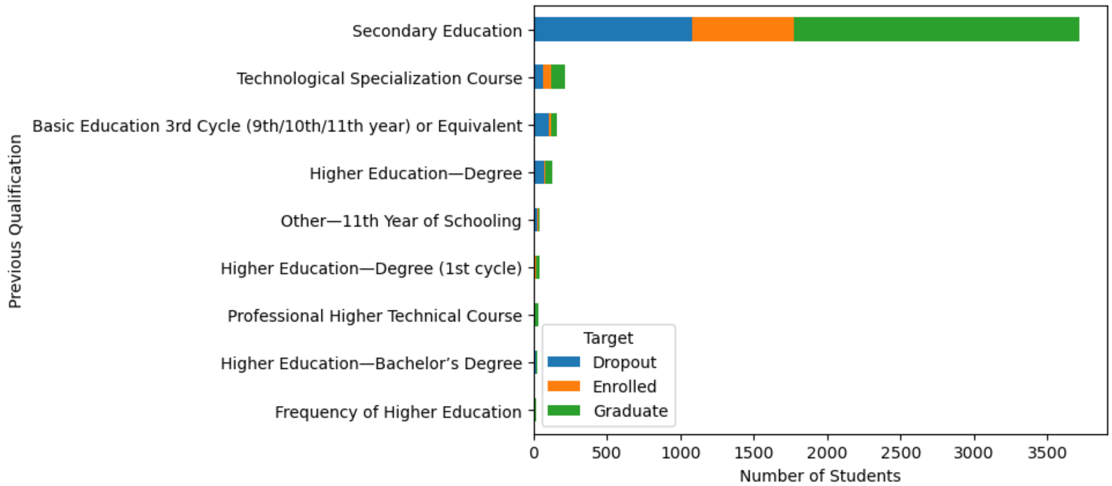
```
nationality=df.groupby(['Nacionality',
'Target']).size().reset_index().pivot(columns='Target',index='Nacional
ity',values=0)

# Rename the index of the DataFrame
nationality=nationality.rename(index={1:'Indian',2:'Chinese',3:'korean
',4:'Japanese',5:'French',6:'Italian',7:'American',8:'English',9:'Germ
an',10:'Mangolian',11:'Spanish',12:'Dutch',13:'Portuguese',14:'Brazili
an',15:'Russian',16:'Mexican',17:'Turkish',18:'Ukrainian',19:'Romanian
',20:'Colombian',21:'Cuban'})
nationality_total=nationality.sum(axis=1)
nationality_sorted=nationality_total.sort_values(ascending=True)
nationality.loc[nationality_sorted.index].plot(kind='barh',
stacked=True)
plt.xlabel('Number of Students')
plt.ylabel('Nationality')
plt.show()
```



The plot shows that the majority of the students in the dataset are Indian, which accounts for the highest frequency among all the nationalities.

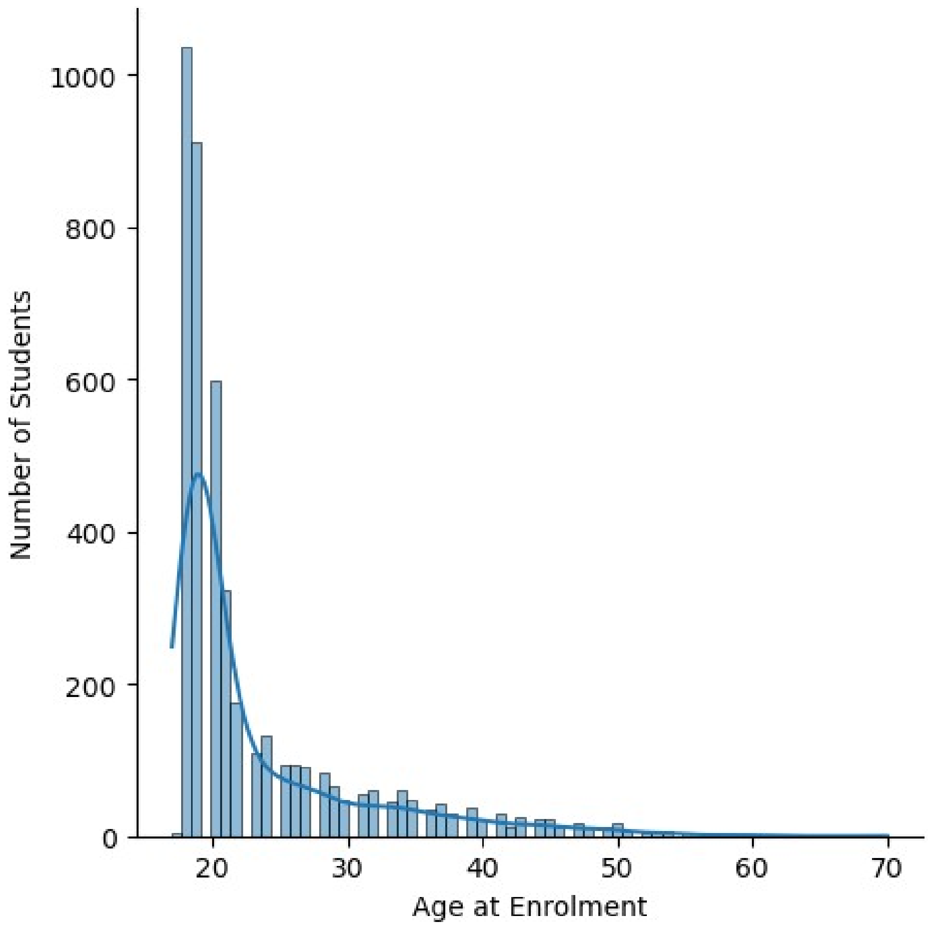
```
student_prequal=df.groupby(['Previous qualification','Target']).size().reset_index().pivot(columns='Target', index='Previous qualification',values=0)

# Rename the index of the DataFrame
student_prequal=student_prequal.rename(index={1:'Secondary Education',2:'Higher Education—Bachelor’s Degree',3:'Higher Education—Degree',4:'Higher Education—Master’s Degree',5:'Higher Education—Doctorate',6:'Frequency of Higher Education',7:'12th Year of Schooling—Not Completed',8:'11th Year of Schooling—Not Completed',9:'Other—11th Year of Schooling',10:'10th Year of Schooling',
                                              11:'10th Year of Schooling—Not Completed',12:'Basic Education 3rd Cycle (9th/10th/11th year) or Equivalent',13:'Basic Education 2nd Cycle (6th/7th/8th year) or Equivalent',14:'Technological Specialization Course',15:'Higher Education—Degree (1st cycle)',16:'Professional Higher Technical Course',17:'Higher Education—Master’s Degree (2nd Cycle)'})
student_prequal_total=student_prequal.sum(axis=1)
student_prequal_sorted=student_prequal_total.sort_values(ascending=True)
student_prequal_top=student_prequal_sorted[8:]
student_prequal.loc[student_prequal_top.index].plot(kind='barh', stacked=True)
plt.xlabel('Count')
```

```
plt.xlabel('Number of Students')
plt.ylabel('Previous Qualification')
plt.show()
```



Most of the students in the data finished secondary education.

```
sns.displot(data=df,x='Age at enrollment',kde=True)
df['Age at enrollment'].describe()
plt.xlabel('Age at Enrolment')
plt.ylabel('Number of Students')
plt.show()
```

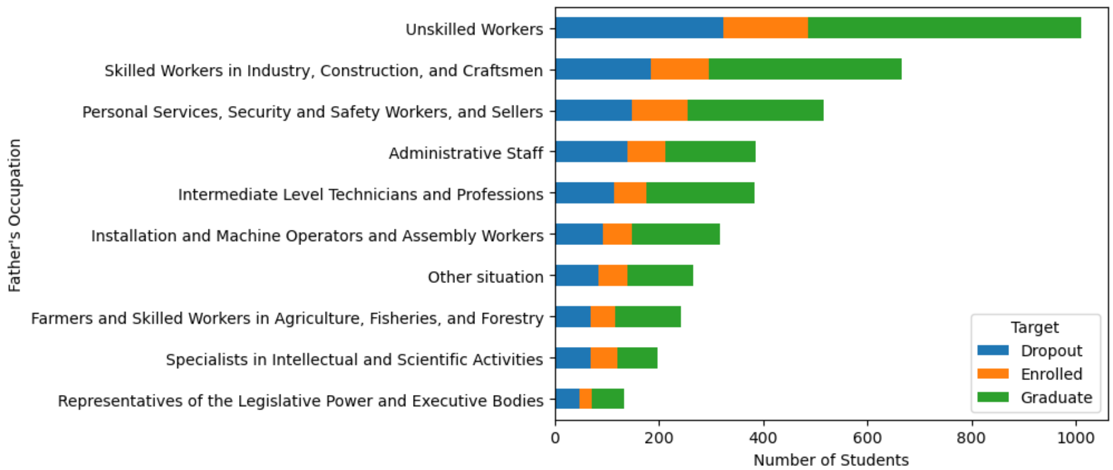


The distribution of age at enrolment is positively skewed, indicating that the majority of students enrolled at a relatively young age. The mean age at enrolment is approximately 23 years old, with the most frequent age range falling between 19 to 25 years old.

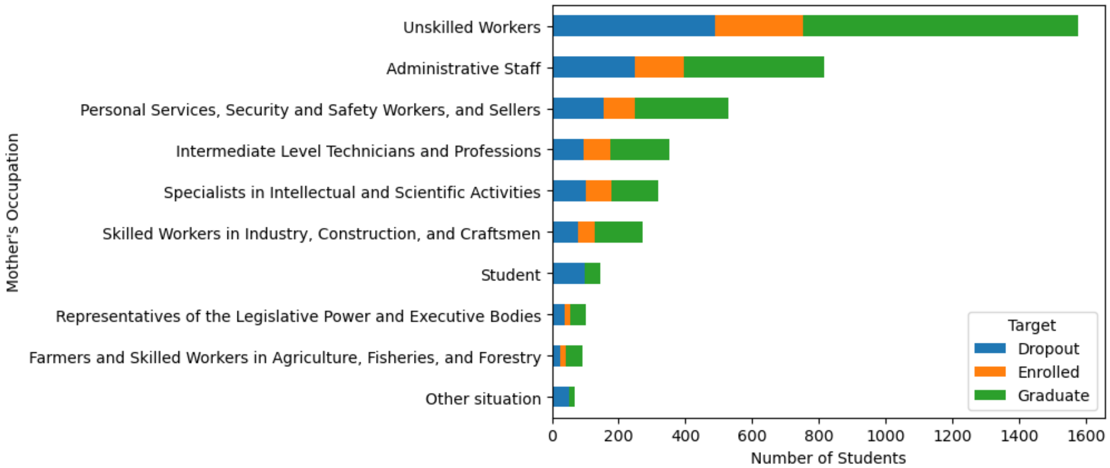
The below graphs are to see the father's and mother's occupation

```
student_foccupation=df.groupby(["Father's occupation",
'Target']).size().reset_index().pivot(columns='Target',index="Father's
occupation",values=0)
student_foccupation=student_foccupation.rename(index={1:'Student',2:'R
epresentatives of the Legislative Power and Executive
Bodies',3:'Specialists in Intellectual and Scientific
Activities',4:'Intermediate Level Technicians and
Professions',5:'Administrative Staff',6:'Personal Services, Security
and Safety Workers, and Sellers',
                                                      7:'Farmers and
Skilled Workers in Agriculture, Fisheries, and Forestry',8:'Skilled
Workers in Industry, Construction, and Craftsmen',9:'Installation and
Machine Operators and Assembly Workers',10:'Unskilled
Workers',11:'Other situation'})
student_foccupation_total=student_foccupation.sum(axis=1)
student_foccupation_sorted=student_foccupation_total.sort_values(ascen
```

```
ding=True)
student_foccupation_top10=student_foccupation_sorted[36:]
student_foccupation.loc[student_foccupation_top10.index].plot(kind='ba
rh', stacked=True)
plt.xlabel('Number of Students')
plt.ylabel("Father's Occupation")
plt.show()
```

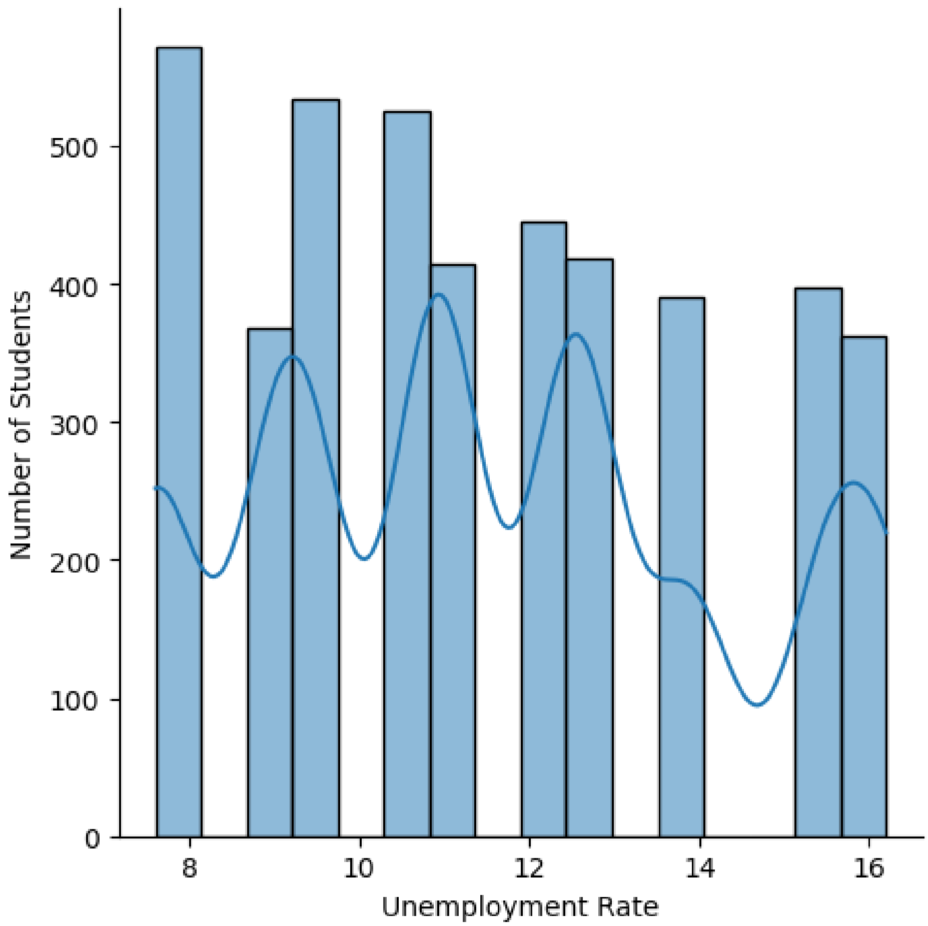



```
student_moccupation=df.groupby(["Mother's occupation",
'Target']).size().reset_index().pivot(columns='Target',
index="Mother's occupation", values=0)
student_moccupation=student_moccupation.rename(index={1:'Student',2:'R
epresentatives of the Legislative Power and Executive
Bodies',3:'Specialists in Intellectual and Scientific
Activities',4:'Intermediate Level Technicians and
Professions',5:'Administrative Staff',6:'Personal Services, Security
and Safety Workers, and Sellers',7:'Farmers and Skilled Workers in
Agriculture, Fisheries, and Forestry',8:'Skilled Workers in Industry,
Construction, and Craftsmen',9:'Installation and Machine Operators and
Assembly Workers',10:'Unskilled Workers',11:'Armed Forces
Professions',12:'Other situation'})
student_moccupation_total = student_moccupation.sum(axis=1)
student_moccupation_sorted =
student_moccupation_total.sort_values(ascending=True)
student_moccupation_top10 = student_moccupation_sorted[22:]
student_moccupation.loc[student_moccupation_top10.index].plot(kind='ba
rh', stacked=True)
plt.xlabel('Number of Students')
plt.ylabel("Mother's Occupation")
plt.show()
```



Highest number of students who graduated and dropped out have parents who are unskilled workers.

```
sns.displot(data=df,x="Unemployment rate",kde=True)
df['Unemployment rate'].describe()
plt.xlabel('Unemployment Rate')
plt.ylabel('Number of Students')
plt.show()
```



The majority of the data points in the unemployment rate distribution fall within the range of 9 to 13.

Correlation and Correlation Heatmap

```
corr=df.corr()
corr
```

```
                                            Marital status  \
Marital status                                    1.000000
Application mode                                  0.224855
Application order                                -0.125854
Course                                            0.018925
Daytime/evening attendance                       -0.274939
Previous qualification                            0.120925
Nacionality                                      -0.020722
Mother's qualification                            0.185522
Father's qualification                            0.128326
Mother's occupation                               0.069734
Father's occupation                               0.024351
Displaced                                        -0.234886
Educational special needs                        -0.028343
```

```
Debtor                                              0.034304
Tuition fees up to date                            -0.087158
Gender                                             -0.014738
Scholarship holder                                 -0.053765
Age at enrollment                                   0.522717
International                                      -0.027905
Curricular units 1st sem (credited)                 0.061209
Curricular units 1st sem (enrolled)                 0.052107
Curricular units 1st sem (evaluations)              0.058030
Curricular units 1st sem (approved)                -0.031027
Curricular units 1st sem (grade)                   -0.059811
Curricular units 1st sem (without evaluations)      0.034711
Curricular units 2nd sem (credited)                 0.062831
Curricular units 2nd sem (enrolled)                 0.039026
Curricular units 2nd sem (evaluations)              0.022784
Curricular units 2nd sem (approved)                -0.043739
Curricular units 2nd sem (grade)                   -0.071506
Curricular units 2nd sem (without evaluations)      0.020426
Unemployment rate                                  -0.020338
Inflation rate                                      0.008761
GDP                                                -0.027003

                                                Application mode  \
Marital status                                          0.224855
Application mode                                        1.000000
Application order                                      -0.246497
Course                                                 -0.085116
Daytime/evening attendance                             -0.268616
Previous qualification                                  0.433028
Nacionality                                            -0.001360
Mother's qualification                                  0.092867
Father's qualification                                  0.072798
Mother's occupation                                     0.033489
Father's occupation                                     0.001253
Displaced                                              -0.263079
Educational special needs                              -0.030868
Debtor                                                  0.114348
Tuition fees up to date                                -0.127339
Gender                                                  0.147226
Scholarship holder                                     -0.152818
Age at enrollment                                       0.450700
International                                           0.005050
Curricular units 1st sem (credited)                     0.238269
Curricular units 1st sem (enrolled)                     0.159547
Curricular units 1st sem (evaluations)                  0.219154
Curricular units 1st sem (approved)                    -0.023713
Curricular units 1st sem (grade)                       -0.106213
Curricular units 1st sem (without evaluations)          0.040255
Curricular units 2nd sem (credited)                     0.228973
Curricular units 2nd sem (enrolled)                     0.127461
```

```
Curricular units 2nd sem (evaluations)                    0.164992
Curricular units 2nd sem (approved)                      -0.065203
Curricular units 2nd sem (grade)                         -0.104424
Curricular units 2nd sem (without evaluations)            0.042009
Unemployment rate                                         0.091567
Inflation rate                                           -0.019613
GDP                                                      -0.014563

                                                Application order
Course  \
Marital status                                            -0.125854
0.018925
Application mode                                          -0.246497 -
0.085116
Application order                                          1.000000
0.118928
Course                                                     0.118928
1.000000
Daytime/evening attendance                                 0.158657 -
0.070232
Previous qualification                                    -0.199029 -
0.158382
Nacionality                                               -0.029385 -
0.004761
Mother's qualification                                    -0.061719
0.058909
Father's qualification                                    -0.049936
0.045659
Mother's occupation                                       -0.046591
0.029672
Father's occupation                                       -0.029754
0.016489
Displaced                                                  0.332362
0.006142
Educational special needs                                  0.025597 -
0.001886
Debtor                                                    -0.072151 -
0.053149
Tuition fees up to date                                    0.055891
0.029099
Gender                                                    -0.089559 -
0.111383
Scholarship holder                                         0.073709
0.051668
Age at enrollment                                         -0.271154 -
0.036929
International                                             -0.028801 -
0.004662
Curricular units 1st sem (credited)                       -0.133354 -
0.140546
```

```
Curricular units 1st sem (enrolled)                         -0.016808
0.112285
Curricular units 1st sem (evaluations)                      -0.092156
0.025970
Curricular units 1st sem (approved)                          0.035580
0.077038
Curricular units 1st sem (grade)                             0.058308
0.179482
Curricular units 1st sem (without evaluations)              -0.031699 -
0.060483
Curricular units 2nd sem (credited)                         -0.125815 -
0.120390
Curricular units 2nd sem (enrolled)                          0.028878
0.185879
Curricular units 2nd sem (evaluations)                      -0.055089
0.049236
Curricular units 2nd sem (approved)                          0.071793
0.120000
Curricular units 2nd sem (grade)                             0.055517
0.178997
Curricular units 2nd sem (without evaluations)              -0.015757 -
0.013984
Unemployment rate                                           -0.098419 -
0.050116
Inflation rate                                              -0.011133
0.028775
GDP                                                          0.030201 -
0.012518

                                                 Daytime/evening
attendance  \
Marital status                                                       -
0.274939
Application mode                                                     -
0.268616
Application order
0.158657
Course                                                               -
0.070232
Daytime/evening attendance
1.000000
Previous qualification                                               -
0.103022
Nacionality
0.024433
Mother's qualification                                               -
0.195346
Father's qualification                                               -
0.137769
Mother's occupation                                                  -
```

0.037986
Father's occupation
0.000845
Displaced
0.251767
Educational special needs
0.031017
Debtor
0.006658
Tuition fees up to date
0.038799
Gender -
0.012326
Scholarship holder
0.093912
Age at enrollment -
0.462280
International
0.027973
Curricular units 1st sem (credited) -
0.127466
Curricular units 1st sem (enrolled) -
0.043056
Curricular units 1st sem (evaluations) -
0.045889
Curricular units 1st sem (approved)
0.016935
Curricular units 1st sem (grade)
0.063974
Curricular units 1st sem (without evaluations)
0.045630
Curricular units 2nd sem (credited) -
0.111953
Curricular units 2nd sem (enrolled)
0.000371
Curricular units 2nd sem (evaluations)
0.014610
Curricular units 2nd sem (approved)
0.034022
Curricular units 2nd sem (grade)
0.050493
Curricular units 2nd sem (without evaluations) -
0.004229
Unemployment rate
0.061974
Inflation rate -
0.024043
GDP
0.022929

\	Previous qualification
Marital status	0.120925
Application mode	0.433028
Application order	-0.199029
Course	-0.158382
Daytime/evening attendance	-0.103022
Previous qualification	1.000000
Nacionality	-0.038997
Mother's qualification	0.018868
Father's qualification	0.013152
Mother's occupation	0.006190
Father's occupation	0.005381
Displaced	-0.149356
Educational special needs	-0.015015
Debtor	0.117447
Tuition fees up to date	-0.095246
Gender	0.089952
Scholarship holder	-0.085668
Age at enrollment	0.249821
International	-0.033498
Curricular units 1st sem (credited)	0.159940
Curricular units 1st sem (enrolled)	0.080860
Curricular units 1st sem (evaluations)	0.129364
Curricular units 1st sem (approved)	-0.005295

```
Curricular units 1st sem (grade)                              -0.034252

Curricular units 1st sem (without evaluations)                 0.018276

Curricular units 2nd sem (credited)                            0.138463

Curricular units 2nd sem (enrolled)                            0.056450

Curricular units 2nd sem (evaluations)                         0.101501

Curricular units 2nd sem (approved)                           -0.037265

Curricular units 2nd sem (grade)                              -0.038765

Curricular units 2nd sem (without evaluations)                 0.024186

Unemployment rate                                              0.096914

Inflation rate                                                -0.056388

GDP                                                            0.053968


                                                     Nacionality  \
Marital status                                         -0.020722
Application mode                                       -0.001360
Application order                                      -0.029385
Course                                                 -0.004761
Daytime/evening attendance                              0.024433
Previous qualification                                 -0.038997
Nacionality                                             1.000000
Mother's qualification                                 -0.043847
Father's qualification                                 -0.088892
Mother's occupation                                     0.044123
Father's occupation                                     0.024538
Displaced                                              -0.010774
Educational special needs                              -0.002399
Debtor                                                  0.070860
Tuition fees up to date                                -0.041721
Gender                                                 -0.025462
Scholarship holder                                     -0.018468
Age at enrollment                                      -0.008241
International                                           0.911724
Curricular units 1st sem (credited)                     0.006604
Curricular units 1st sem (enrolled)                    -0.008011
Curricular units 1st sem (evaluations)                  0.005640
Curricular units 1st sem (approved)                     0.000935
Curricular units 1st sem (grade)                        0.002578
Curricular units 1st sem (without evaluations)          0.026203
```

```
Curricular units 2nd sem (credited)              -0.000747
Curricular units 2nd sem (enrolled)              -0.020103
Curricular units 2nd sem (evaluations)           -0.018023
Curricular units 2nd sem (approved)              -0.014142
Curricular units 2nd sem (grade)                 -0.005409
Curricular units 2nd sem (without evaluations)   -0.012052
Unemployment rate                                -0.006013
Inflation rate                                   -0.012331
GDP                                               0.044563

                                               Mother's qualification
\
Marital status                                               0.185522

Application mode                                             0.092867

Application order                                           -0.061719

Course                                                       0.058909

Daytime/evening attendance                                  -0.195346

Previous qualification                                       0.018868

Nacionality                                                 -0.043847

Mother's qualification                                       1.000000

Father's qualification                                       0.524529

Mother's occupation                                          0.295178

Father's occupation                                          0.115989

Displaced                                                   -0.075864

Educational special needs                                   -0.019808

Debtor                                                       0.018776

Tuition fees up to date                                     -0.022861

Gender                                                      -0.062374

Scholarship holder                                           0.048225

Age at enrollment                                            0.279921
```

International	-0.038672
Curricular units 1st sem (credited)	0.041610
Curricular units 1st sem (enrolled)	0.050582
Curricular units 1st sem (evaluations)	0.041967
Curricular units 1st sem (approved)	-0.010555
Curricular units 1st sem (grade)	-0.034105
Curricular units 1st sem (without evaluations)	0.003293
Curricular units 2nd sem (credited)	0.036986
Curricular units 2nd sem (enrolled)	0.033070
Curricular units 2nd sem (evaluations)	0.018874
Curricular units 2nd sem (approved)	-0.013161
Curricular units 2nd sem (grade)	-0.028472
Curricular units 2nd sem (without evaluations)	0.020364
Unemployment rate	-0.106107
Inflation rate	0.056653
GDP	-0.079664

\	Father's qualification
Marital status	0.128326
Application mode	0.072798
Application order	-0.049936
Course	0.045659
Daytime/evening attendance	-0.137769
Previous qualification	0.013152
Nacionality	-0.088892

Mother's qualification	0.524529
Father's qualification	1.000000
Mother's occupation	0.207067
Father's occupation	0.184001
Displaced	-0.055007
Educational special needs	0.000917
Debtor	-0.006125
Tuition fees up to date	-0.018033
Gender	-0.073614
Scholarship holder	0.107134
Age at enrollment	0.190410
International	-0.086503
Curricular units 1st sem (credited)	0.039259
Curricular units 1st sem (enrolled)	0.036564
Curricular units 1st sem (evaluations)	0.037225
Curricular units 1st sem (approved)	0.006787
Curricular units 1st sem (grade)	-0.006245
Curricular units 1st sem (without evaluations)	-0.017785
Curricular units 2nd sem (credited)	0.041695
Curricular units 2nd sem (enrolled)	0.023635
Curricular units 2nd sem (evaluations)	0.009471
Curricular units 2nd sem (approved)	0.006052
Curricular units 2nd sem (grade)	-0.006508
Curricular units 2nd sem (without evaluations)	-0.008493

```
Unemployment rate                                              -0.075417

Inflation rate                                                  0.056661

GDP                                                            -0.070200


                                                Mother's
occupation  ...  \
Marital status
0.069734  ...
Application mode
0.033489  ...
Application order                                          -
0.046591  ...
Course
0.029672  ...
Daytime/evening attendance                                 -
0.037986  ...
Previous qualification
0.006190  ...
Nacionality
0.044123  ...
Mother's qualification
0.295178  ...
Father's qualification
0.207067  ...
Mother's occupation
1.000000  ...
Father's occupation
0.724007  ...
Displaced                                                  -
0.038521  ...
Educational special needs                                  -
0.010175  ...
Debtor
0.108151  ...
Tuition fees up to date                                    -
0.004395  ...
Gender                                                     -
0.022324  ...
Scholarship holder
0.092487  ...
Age at enrollment
0.092257  ...
International
0.041414  ...
Curricular units 1st sem (credited)
0.002390  ...
```

```
Curricular units 1st sem (enrolled)
0.014607  ...
Curricular units 1st sem (evaluations)
0.019332  ...
Curricular units 1st sem (approved)
0.015198  ...
Curricular units 1st sem (grade)
0.016955  ...
Curricular units 1st sem (without evaluations)            -
0.012569  ...
Curricular units 2nd sem (credited)                       -
0.002057  ...
Curricular units 2nd sem (enrolled)
0.009287  ...
Curricular units 2nd sem (evaluations)
0.011546  ...
Curricular units 2nd sem (approved)
0.022309  ...
Curricular units 2nd sem (grade)
0.035230  ...
Curricular units 2nd sem (without evaluations)            -
0.004903  ...
Unemployment rate                                         -
0.011772  ...
Inflation rate
0.015014  ...
GDP
0.091880  ...

                                                Curricular units 1st
sem (without evaluations)  \
Marital status
0.034711
Application mode
0.040255
Application order
-0.031699
Course
-0.060483
Daytime/evening attendance
0.045630
Previous qualification
0.018276
Nacionality
0.026203
Mother's qualification
0.003293
Father's qualification
-0.017785
Mother's occupation
```

-0.012569
Father's occupation
-0.035299
Displaced
-0.021671
Educational special needs
-0.012324
Debtor
0.001812
Tuition fees up to date
-0.049775
Gender
-0.006302
Scholarship holder
-0.057770
Age at enrollment
0.057470
International
0.031222
Curricular units 1st sem (credited)
0.116262
Curricular units 1st sem (enrolled)
0.129337
Curricular units 1st sem (evaluations)
0.241800
Curricular units 1st sem (approved)
-0.013360
Curricular units 1st sem (grade)
-0.071660
Curricular units 1st sem (without evaluations)
1.000000
Curricular units 2nd sem (credited)
0.117359
Curricular units 2nd sem (enrolled)
0.109924
Curricular units 2nd sem (evaluations)
0.144683
Curricular units 2nd sem (approved)
-0.013070
Curricular units 2nd sem (grade)
-0.061482
Curricular units 2nd sem (without evaluations)
0.583261
Unemployment rate
-0.045144
Inflation rate
-0.052534
GDP
-0.144673

```
                                                Curricular units 2nd
sem (credited)  \
Marital status
0.062831
Application mode
0.228973
Application order
-0.125815
Course
-0.120390
Daytime/evening attendance
-0.111953
Previous qualification
0.138463
Nacionality
-0.000747
Mother's qualification
0.036986
Father's qualification
0.041695
Mother's occupation
-0.002057
Father's occupation
-0.014596
Displaced
-0.091738
Educational special needs
-0.021671
Debtor
0.025414
Tuition fees up to date
0.014204
Gender
0.018737
Scholarship holder
-0.076480
Age at enrollment
0.207561
International
0.002573
Curricular units 1st sem (credited)
0.944811
Curricular units 1st sem (enrolled)
0.753747
Curricular units 1st sem (evaluations)
0.522187
Curricular units 1st sem (approved)
0.607661
Curricular units 1st sem (grade)
0.113937
```

```
Curricular units 1st sem (without evaluations)
0.117359
Curricular units 2nd sem (credited)
1.000000
Curricular units 2nd sem (enrolled)
0.676258
Curricular units 2nd sem (evaluations)
0.430978
Curricular units 2nd sem (approved)
0.519081
Curricular units 2nd sem (grade)
0.129770
Curricular units 2nd sem (without evaluations)
0.070148
Unemployment rate
0.010580
Inflation rate
0.014490
GDP
-0.024491

                                                Curricular units 2nd
sem (enrolled)  \
Marital status
0.039026
Application mode
0.127461
Application order
0.028878
Course
0.185879
Daytime/evening attendance
0.000371
Previous qualification
0.056450
Nacionality
-0.020103
Mother's qualification
0.033070
Father's qualification
0.023635
Mother's occupation
0.009287
Father's occupation
0.005548
Displaced
-0.041823
Educational special needs
-0.028777
Debtor
```

```
-0.029436
Tuition fees up to date
0.085918
Gender
-0.124227
Scholarship holder
0.026416
Age at enrollment
0.085914
International
-0.013577
Curricular units 1st sem (credited)
0.644826
Curricular units 1st sem (enrolled)
0.942627
Curricular units 1st sem (evaluations)
0.611842
Curricular units 1st sem (approved)
0.733772
Curricular units 1st sem (grade)
0.406167
Curricular units 1st sem (without evaluations)
0.109924
Curricular units 2nd sem (credited)
0.676258
Curricular units 2nd sem (enrolled)
1.000000
Curricular units 2nd sem (evaluations)
0.604821
Curricular units 2nd sem (approved)
0.703258
Curricular units 2nd sem (grade)
0.395135
Curricular units 2nd sem (without evaluations)
0.067697
Unemployment rate
0.064436
Inflation rate
0.016844
GDP
-0.007592

                                                Curricular units 2nd
sem (evaluations)  \
Marital status
0.022784
Application mode
0.164992
Application order
-0.055089
```

```
Course
0.049236
Daytime/evening attendance
0.014610
Previous qualification
0.101501
Nacionality
-0.018023
Mother's qualification
0.018874
Father's qualification
0.009471
Mother's occupation
0.011546
Father's occupation
0.000833
Displaced
-0.038839
Educational special needs
-0.010851
Debtor
0.024201
Tuition fees up to date
0.063482
Gender
-0.041789
Scholarship holder
-0.021410
Age at enrollment
0.056286
International
-0.004399
Curricular units 1st sem (credited)
0.427845
Curricular units 1st sem (enrolled)
0.599567
Curricular units 1st sem (evaluations)
0.778863
Curricular units 1st sem (approved)
0.539934
Curricular units 1st sem (grade)
0.487236
Curricular units 1st sem (without evaluations)
0.144683
Curricular units 2nd sem (credited)
0.430978
Curricular units 2nd sem (enrolled)
0.604821
Curricular units 2nd sem (evaluations)
1.000000
```

```
Curricular units 2nd sem (approved)
0.463535
Curricular units 2nd sem (grade)
0.453394
Curricular units 2nd sem (without evaluations)
0.144877
Unemployment rate
0.045808
Inflation rate
-0.012643
GDP
-0.004854

                                                Curricular units 2nd
sem (approved)  \
Marital status
-0.043739
Application mode
-0.065203
Application order
0.071793
Course
0.120000
Daytime/evening attendance
0.034022
Previous qualification
-0.037265
Nacionality
-0.014142
Mother's qualification
-0.013161
Father's qualification
0.006052
Mother's occupation
0.022309
Father's occupation
0.023651
Displaced
0.063698
Educational special needs
-0.016315
Debtor
-0.146977
Tuition fees up to date
0.291921
Gender
-0.224266
Scholarship holder
0.202704
Age at enrollment
```

```
-0.112052
International
-0.010565
Curricular units 1st sem (credited)
0.490478
Curricular units 1st sem (enrolled)
0.673341
Curricular units 1st sem (evaluations)
0.442265
Curricular units 1st sem (approved)
0.904002
Curricular units 1st sem (grade)
0.673335
Curricular units 1st sem (without evaluations)
-0.013070
Curricular units 2nd sem (credited)
0.519081
Curricular units 2nd sem (enrolled)
0.703258
Curricular units 2nd sem (evaluations)
0.463535
Curricular units 2nd sem (approved)
1.000000
Curricular units 2nd sem (grade)
0.760804
Curricular units 2nd sem (without evaluations)
-0.061567
Unemployment rate
0.048805
Inflation rate
-0.024566
GDP
0.022427

                                                Curricular units 2nd
sem (grade)  \
Marital status
-0.071506
Application mode
-0.104424
Application order
0.055517
Course
0.178997
Daytime/evening attendance
0.050493
Previous qualification
-0.038765
Nacionality
-0.005409
```

```
Mother's qualification
-0.028472
Father's qualification
-0.006508
Mother's occupation
0.035230
Father's occupation
0.036711
Displaced
0.069087
Educational special needs
-0.012761
Debtor
-0.139424
Tuition fees up to date
0.296480
Gender
-0.199133
Scholarship holder
0.181227
Age at enrollment
-0.173419
International
0.001460
Curricular units 1st sem (credited)
0.132971
Curricular units 1st sem (enrolled)
0.361959
Curricular units 1st sem (evaluations)
0.355036
Curricular units 1st sem (approved)
0.685560
Curricular units 1st sem (grade)
0.837170
Curricular units 1st sem (without evaluations)
-0.061482
Curricular units 2nd sem (credited)
0.129770
Curricular units 2nd sem (enrolled)
0.395135
Curricular units 2nd sem (evaluations)
0.453394
Curricular units 2nd sem (approved)
0.760804
Curricular units 2nd sem (grade)
1.000000
Curricular units 2nd sem (without evaluations)
-0.079216
Unemployment rate
0.001462
```

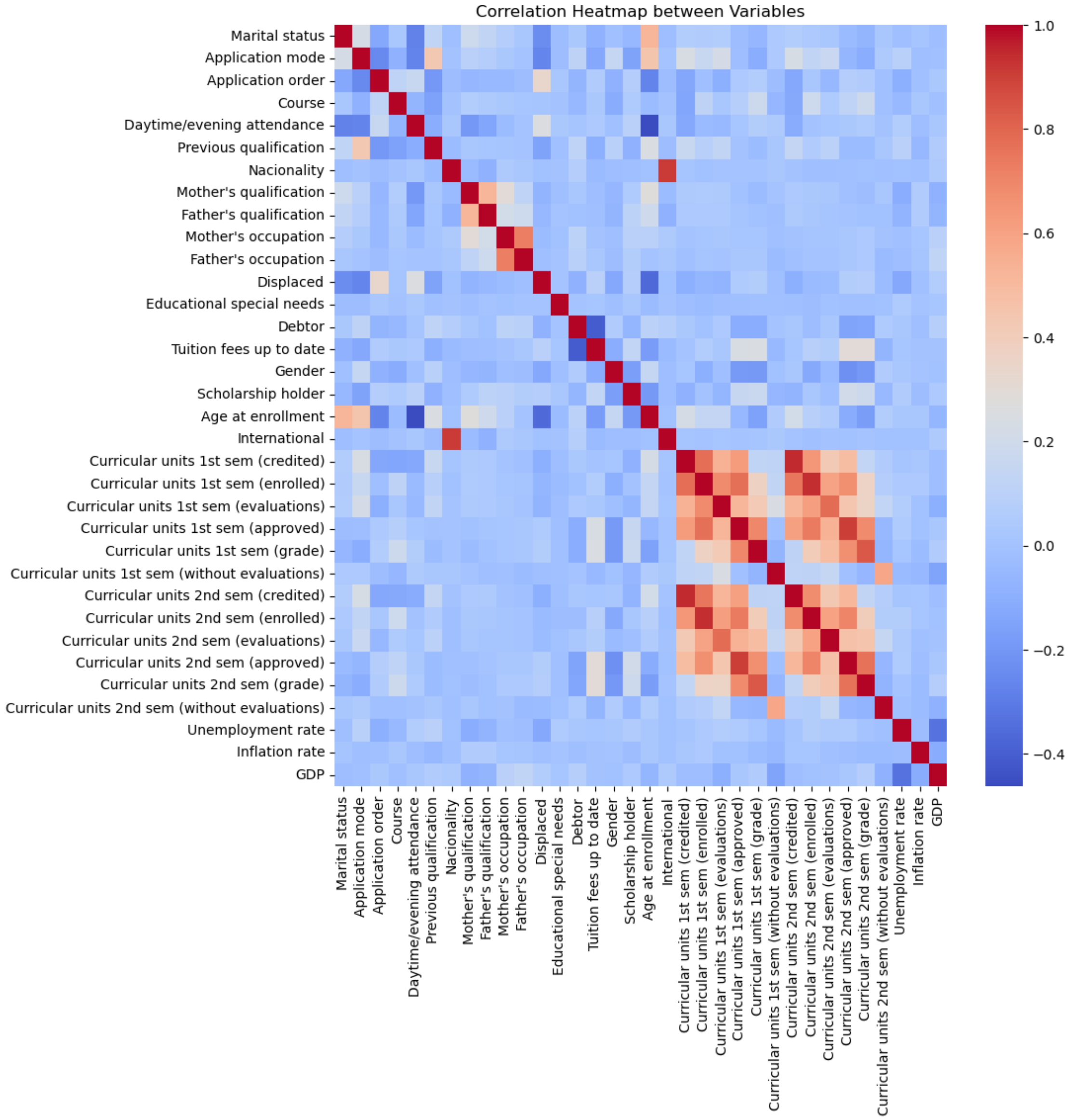
```
Inflation rate
-0.038166
GDP
0.071269

                                                Curricular units 2nd
sem (without evaluations)  \
Marital status
0.020426
Application mode
0.042009
Application order
-0.015757
Course
-0.013984
Daytime/evening attendance
-0.004229
Previous qualification
0.024186
Nacionality
-0.012052
Mother's qualification
0.020364
Father's qualification
-0.008493
Mother's occupation
-0.004903
Father's occupation
-0.044760
Displaced
-0.035959
Educational special needs
-0.007491
Debtor
0.048552
Tuition fees up to date
-0.071817
Gender
0.057223
Scholarship holder
-0.048723
Age at enrollment
0.061654
International
-0.010660
Curricular units 1st sem (credited)
0.055256
Curricular units 1st sem (enrolled)
0.069547
Curricular units 1st sem (evaluations)
```

```
0.134296
Curricular units 1st sem (approved)
-0.053983
Curricular units 1st sem (grade)
-0.066076
Curricular units 1st sem (without evaluations)
0.583261
Curricular units 2nd sem (credited)
0.070148
Curricular units 2nd sem (enrolled)
0.067697
Curricular units 2nd sem (evaluations)
0.144877
Curricular units 2nd sem (approved)
-0.061567
Curricular units 2nd sem (grade)
-0.079216
Curricular units 2nd sem (without evaluations)
1.000000
Unemployment rate
-0.013960
Inflation rate
-0.034391
GDP
-0.080292

                                                Unemployment rate  \
Marital status                                          -0.020338
Application mode                                         0.091567
Application order                                       -0.098419
Course                                                  -0.050116
Daytime/evening attendance                               0.061974
Previous qualification                                   0.096914
Nacionality                                             -0.006013
Mother's qualification                                  -0.106107
Father's qualification                                  -0.075417
Mother's occupation                                     -0.011772
Father's occupation                                     -0.026094
Displaced                                               -0.130327
Educational special needs                                0.046131
Debtor                                                   0.021128
Tuition fees up to date                                  0.013460
Gender                                                   0.022195
Scholarship holder                                       0.055152
Age at enrollment                                        0.025018
International                                           -0.010015
Curricular units 1st sem (credited)                      0.009778
Curricular units 1st sem (enrolled)                      0.038404
Curricular units 1st sem (evaluations)                   0.061545
Curricular units 1st sem (approved)                      0.051286
```

```
Curricular units 1st sem (grade)                          0.014821
Curricular units 1st sem (without evaluations)           -0.045144
Curricular units 2nd sem (credited)                       0.010580
Curricular units 2nd sem (enrolled)                       0.064436
Curricular units 2nd sem (evaluations)                    0.045808
Curricular units 2nd sem (approved)                       0.048805
Curricular units 2nd sem (grade)                          0.001462
Curricular units 2nd sem (without evaluations)           -0.013960
Unemployment rate                                         1.000000
Inflation rate                                           -0.028885
GDP                                                      -0.335178

                                                Inflation rate
GDP
Marital status                                        0.008761 -
0.027003
Application mode                                     -0.019613 -
0.014563
Application order                                    -0.011133
0.030201
Course                                                0.028775 -
0.012518
Daytime/evening attendance                           -0.024043
0.022929
Previous qualification                               -0.056388
0.053968
Nacionality                                          -0.012331
0.044563
Mother's qualification                                0.056653 -
0.079664
Father's qualification                                0.056661 -
0.070200
Mother's occupation                                   0.015014
0.091880
Father's occupation                                   0.008047
0.125574
Displaced                                            -0.012385
0.062875
Educational special needs                             0.004396
0.012016
Debtor                                               -0.021888
0.075050
Tuition fees up to date                              -0.000706 -
0.002768
Gender                                                0.003556 -
0.008108
Scholarship holder                                   -0.031104
0.035650
Age at enrollment                                     0.025377 -
0.064678
```

```
International                                          -0.009642
0.044389
Curricular units 1st sem (credited)                     0.023348 -
0.026513
Curricular units 1st sem (enrolled)                     0.036758 -
0.026262
Curricular units 1st sem (evaluations)                 -0.006604 -
0.099761
Curricular units 1st sem (approved)                    -0.007114
0.018459
Curricular units 1st sem (grade)                       -0.033904
0.054801
Curricular units 1st sem (without evaluations)         -0.052534 -
0.144673
Curricular units 2nd sem (credited)                     0.014490 -
0.024491
Curricular units 2nd sem (enrolled)                     0.016844 -
0.007592
Curricular units 2nd sem (evaluations)                 -0.012643 -
0.004854
Curricular units 2nd sem (approved)                    -0.024566
0.022427
Curricular units 2nd sem (grade)                       -0.038166
0.071269
Curricular units 2nd sem (without evaluations)         -0.034391 -
0.080292
Unemployment rate                                      -0.028885 -
0.335178
Inflation rate                                          1.000000 -
0.112295
GDP                                                    -0.112295
1.000000

[34 rows x 34 columns]
```

```
plt.figure(figsize=(10,10))
sns.heatmap(corr,cmap='coolwarm')
plt.title('Correlation Heatmap between Variables')
plt.show()
```



Correlation between features are low except for Nationality and International. Hence, we can dropped these features in the regression model for predicting student target.

Preprocessing the Data

To predict students' academic success and dropout, we will use logistic regression to determine the target variable using the feature variables.Since the target data contains students who are still enrolled, we will drop them from the dataset and use the data for student who dropped out and graduated.

```
df=df.drop(df[df['Target']=='Enrolled'].index)
df.head()
```

```
   Marital status  Application mode  Application order  Course  \
0               1                 8                  5       2
1               1                 6                  1      11
```

```
2               1                 1                  5       5
3               1                 8                  2      15
4               2                12                  1       3

   Daytime/evening attendance  Previous qualification  Nacionality  \
0                           1                       1            1
1                           1                       1            1
2                           1                       1            1
3                           1                       1            1
4                           0                       1            1

   Mother's qualification  Father's qualification  Mother's occupation
...  \
0                      13                      10                    6
...
1                       1                       3                    4
...
2                      22                      27                   10
...
3                      23                      27                    6
...
4                      22                      28                   10
...

   Curricular units 2nd sem (credited)  Curricular units 2nd sem
(enrolled)  \
0                                    0
0
1                                    0
6
2                                    0
6
3                                    0
6
4                                    0
6

   Curricular units 2nd sem (evaluations)  \
0                                       0
1                                       6
2                                       0
3                                      10
4                                       6

   Curricular units 2nd sem (approved)  Curricular units 2nd sem
(grade)  \
0                                    0
0.000000
1                                    6
13.666667
```

```
2                                    0
0.000000
3                                    5
12.400000
4                                    6
13.000000

   Curricular units 2nd sem (without evaluations)  Unemployment
rate  \
0                                               0              10.8

1                                               0              13.9

2                                               0              10.8

3                                               0               9.4

4                                               0              13.9


   Inflation rate   GDP    Target
0             1.4  1.74   Dropout
1            -0.3  0.79  Graduate
2             1.4  1.74   Dropout
3            -0.8 -3.12  Graduate
4            -0.3  0.79  Graduate

[5 rows x 35 columns]
```

Converting Target Variable into Numeric Form We will transform the target variable into numeric form using label encoder,a data preprocessing feature from SciKit library.The labels dropout and graduate become 0 and 1, respectively.

```
encoder=LabelEncoder()
df['Target']=encoder.fit_transform(df['Target'])
df.head()
```

```
   Marital status  Application mode  Application order  Course  \
0               1                 8                  5       2
1               1                 6                  1      11
2               1                 1                  5       5
3               1                 8                  2      15
4               2                12                  1       3

   Daytime/evening attendance  Previous qualification  Nacionality  \
0                           1                       1            1
1                           1                       1            1
2                           1                       1            1
```

```
3                           1                        1            1
4                           0                        1            1

   Mother's qualification  Father's qualification  Mother's occupation
...  \
0                      13                      10                    6
...
1                       1                       3                    4
...
2                      22                      27                   10
...
3                      23                      27                    6
...
4                      22                      28                   10
...

   Curricular units 2nd sem (credited)  Curricular units 2nd sem
(enrolled)  \
0                                    0
0
1                                    0
6
2                                    0
6
3                                    0
6
4                                    0
6

   Curricular units 2nd sem (evaluations)  \
0                                       0
1                                       6
2                                       0
3                                      10
4                                       6

   Curricular units 2nd sem (approved)  Curricular units 2nd sem
(grade)  \
0                                    0
0.000000
1                                    6
13.666667
2                                    0
0.000000
3                                    5
12.400000
4                                    6
13.000000

   Curricular units 2nd sem (without evaluations)  Unemployment
```

```
rate  \
0                                                0            10.8

1                                                0            13.9

2                                                0            10.8

3                                                0             9.4

4                                                0            13.9


   Inflation rate   GDP  Target
0             1.4  1.74       0
1            -0.3  0.79       1
2             1.4  1.74       0
3            -0.8 -3.12       1
4            -0.3  0.79       1

[5 rows x 35 columns]
```

Splitting data into x and y We set x and y as the dataframe feature and target variables, respectively. Note that we will drop the Nationality and International columns since they are highly correlated and only one nationality i.e Indian significantly dominates the data. This will prevent bias in the statistical regression.

```
x=df.drop(columns=['Nacionality','International','Target'],axis=1)
y=df['Target']
```

```
x.head()
```

```
   Marital status  Application mode  Application order  Course  \
0               1                 8                  5       2
1               1                 6                  1      11
2               1                 1                  5       5
3               1                 8                  2      15
4               2                12                  1       3

   Daytime/evening attendance  Previous qualification  Mother's
qualification  \
0                           1                       1
13
1                           1                       1
1
2                           1                       1
22
3                           1                       1
23
4                           0                       1
22
```

```
   Father's qualification  Mother's occupation  Father's
occupation  ...  \
0                      10                    6
10  ...
1                       3                    4
4  ...
2                      27                   10
10  ...
3                      27                    6
4  ...
4                      28                   10
10  ...

   Curricular units 1st sem (without evaluations)  \
0                                               0
1                                               0
2                                               0
3                                               0
4                                               0

   Curricular units 2nd sem (credited)  Curricular units 2nd sem
(enrolled)  \
0                                    0
0
1                                    0
6
2                                    0
6
3                                    0
6
4                                    0
6

   Curricular units 2nd sem (evaluations)  \
0                                       0
1                                       6
2                                       0
3                                      10
4                                       6

   Curricular units 2nd sem (approved)  Curricular units 2nd sem
(grade)  \
0                                    0
0.000000
1                                    6
13.666667
2                                    0
0.000000
3                                    5
```

```
12.400000
4                                    6
13.000000

   Curricular units 2nd sem (without evaluations)  Unemployment
rate  \
0                                               0                10.8

1                                               0                13.9

2                                               0                10.8

3                                               0                 9.4

4                                               0                13.9


   Inflation rate   GDP
0             1.4  1.74
1            -0.3  0.79
2             1.4  1.74
3            -0.8 -3.12
4            -0.3  0.79

[5 rows x 32 columns]
```

```
y.head()
```

```
0    0
1    1
2    0
3    1
4    1
Name: Target, dtype: int32
```

Splitting Data into Training and Testing Data To begin with the logistic regression as our machine learning model, we split the data into training and testing data. 80% of the data will be our training model and rest 20% will be the testing model. We choose the third state of the random sampling.

```
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2,random_st
ate=3)
```

```
print(x.shape,xtrain.shape,xtest.shape)
```

```
(3630, 32) (2904, 32) (726, 32)
```

Model Building

Logistic regression will now be implemented using Extreme Gradient Boosting (XGBoost) which is one of the available open source libraries used for regression models. In this case, binary logistic is set for our model with 1000 n_estimators. The n_estimators serves as the number of decision trees or classification considering the data from feature variables.

```
bin_log=xgb.XGBClassifier(objective='binary:logistic',n_estimators=100
0)
bin_log.fit(xtrain,ytrain)
```

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None,
feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None,
importance_type=None,
              interaction_constraints=None, learning_rate=None,
max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan,
monotone_constraints=None,
              n_estimators=1000, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, ...)
```

Data Prediction and Evaluation of the Model We now set the logistic regression model to the training data.

```
target_prediction=bin_log.predict(xtest)
print(target_prediction)
```

```
[1 1 1 1 1 1 1 1 0 1 1 0 0 0 1 1 1 1 0 1 1 0 0 1 1 0 1 1 1 1 1 1 1 1 1
1 0
 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 0 0 1 0 0 0 0 1 1 0 1 0 0 1 0 1 0 0 0 1
1 1
 1 1 1 0 0 1 1 1 0 0 0 1 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 1 1 1 1 0 1 1 0
1 1
 0 0 0 1 0 1 1 1 1 0 0 0 1 1 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 0 1 0
1 0
 1 0 1 1 1 0 1 1 0 1 1 0 0 0 1 0 1 1 1 1 1 1 1 0 1 0 0 1 0 1 1 1 1 0 1
0 1
 1 1 0 1 1 1 0 0 1 1 0 1 0 0 0 0 0 1 0 0 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1
0 1
 1 0 0 0 0 0 0 1 0 0 1 1 1 1 1 1 0 1 0 1 1 1 1 0 1 1 0 1 1 1 1 1 1 0 1
1 0
 1 1 1 0 0 1 0 1 1 1 0 0 1 0 0 1 1 0 0 1 1 1 1 0 0 0 0 1 0 0 1 1 1 1 1
1 0
 1 0 0 0 1 0 1 1 0 1 0 0 0 1 1 1 0 1 0 1 0 1 1 0 1 0 0 0 0 1 0 1 1 0 1
1 1
 1 1 0 1 1 1 1 0 1 0 1 1 1 0 0 1 1 0 0 1 0 0 0 0 1 0 1 1 1 1 1 1 0 1 0
0 1
 1 0 1 1 1 1 1 0 0 0 0 1 0 1 1 1 1 1 0 1 0 1 1 0 0 1 1 1 1 1 0 0 0 1 1
0 1
```

```
 0 1 1 1 1 0 0 1 0 1 1 1 1 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 1 1 1 0
0 0
 0 0 1 0 1 1 1 1 1 1 0 1 1 0 0 1 0 1 1 0 1 0 1 0 1 0 1 0 0 1 1 1 1 0 1
1 1
 1 1 0 1 0 1 0 1 1 1 1 1 1 1 0 0 1 1 0 1 1 1 1 0 1 1 1 0 1 0 0 1 0 1 1
1 1
 1 1 1 1 0 1 1 0 0 1 0 1 0 1 1 1 1 1 0 0 1 1 0 1 0 0 1 0 0 0 0 1 1 1 1
1 0
 0 0 1 0 1 0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 1 0 1 1 1 1
0 0
 0 0 0 1 1 1 0 1 1 1 0 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1
1 0
 1 1 1 1 1 1 0 0 1 1 0 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 0
1 1
 1 1 0 1 0 1 1 1 1 1 1 0 1 1 0 0 1 1 1 1 0 1 0 1 1 0 1 1 0 1 0 1 1 0 1
1 1
 1 1 0 1 0 1 0 0 1 1 0 1 1 1 1 0 0 0 0 0 0 1 0]
```

```
data_accuracy=accuracy_score(ytest,target_prediction)
print("Accuracy:",data_accuracy)
```

```
Accuracy: 0.9008264462809917
```

Conclusion From the analysis we can conclude,the Extreme Gradient Boosting (XGB) gives the accuracy of 90%.Therefore by using this model we can classify whether the student will graduate or dropout

Crosschecking the value predicted by the model with the value present in data

```
df.iloc[192]
```

```
Marital status                                          1.00
Application mode                                        1.00
Application order                                       2.00
Course                                                 14.00
Daytime/evening attendance                              1.00
Previous qualification                                  1.00
Nacionality                                             1.00
Mother's qualification                                  1.00
Father's qualification                                  3.00
Mother's occupation                                     5.00
Father's occupation                                     4.00
Displaced                                               0.00
Educational special needs                               0.00
Debtor                                                  0.00
Tuition fees up to date                                 1.00
Gender                                                  0.00
Scholarship holder                                      0.00
Age at enrollment                                      19.00
International                                           0.00
Curricular units 1st sem (credited)                     0.00
```

```
Curricular units 1st sem (enrolled)                5.00
Curricular units 1st sem (evaluations)             5.00
Curricular units 1st sem (approved)                5.00
Curricular units 1st sem (grade)                  13.00
Curricular units 1st sem (without evaluations)     0.00
Curricular units 2nd sem (credited)                0.00
Curricular units 2nd sem (enrolled)                5.00
Curricular units 2nd sem (evaluations)             5.00
Curricular units 2nd sem (approved)                5.00
Curricular units 2nd sem (grade)                  13.20
Curricular units 2nd sem (without evaluations)     0.00
Unemployment rate                                  9.40
Inflation rate                                    -0.80
GDP                                               -3.12
Target                                             1.00
Name: 232, dtype: float64
```

```
input_data=(1,1,2,14,1,1,1,3,5,4,0,0,0,1,0,0,19,0,5,5,5,13,0,0,5,5,5,1
3.2,0,9.4,-0.8,-3.12)
input_data_as_numpy_array=np.asarray(input_data)
input_data_reshaped=input_data_as_numpy_array.reshape(1,-1)
prediction=bin_log.predict(input_data_reshaped)
print('Prediction:',prediction)
#print("The initial value is ",prediction[0])
```

```
Prediction: [1]
```

As seen above the model has predicted the same value present in the dataset.

Forecasting with unknown data

```
input_data=(1,1,1,9,1,1,22,1,10,11,0,0,0,1,0,0,21,0,5,12,3,12,0,0,5,12
,1,15,3,16,0,-0)
input_data_as_numpy_array=np.asarray(input_data)
input_data_reshaped=input_data_as_numpy_array.reshape(1,-1)
prediction=bin_log.predict(input_data_reshaped)
print('Prediction:',prediction)
```

```
Prediction: [0]
```