Project Outline

0. Imports and random state

In [1]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, f1_score, precision_score, recall_score, matthews_corrcoef
from sklearn.model_selection import GridSearchCV, train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import RobustScaler, StandardScaler
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.calibration import LabelEncoder


my_random_state = 69
```

1. Team

- Itmam Alam
- Akos Papp

2. Data

1. Link to dataset(s) https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success
2. Download the necessary files, describe the attributes in the notebook including classes/labels

Features

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Marital Status	Feature	Integer	Marital Status	1 – single 2 – married 3 – widower 4 – divorced 5 – facto union 6 – legally separated		no
Application mode	Feature	Integer		1 - 1st phase - general contingent 2 - Ordinance No. 612/93 5 - 1st phase - special contingent (Azores Island) 7 - Holders of other higher courses 10 - Ordinance No. 854-B/99 15 - International student (bachelor) 16 - 1st phase - special contingent (Madeira Island) 17 - 2nd phase - general contingent 18 - 3rd phase - general contingent 26 - Ordinance No. 533-A/99, item b2) (Different Plan) 27 - Ordinance No. 533-A/99, item b3 (Other Institution) 39 - Over 23 years old 42 - Transfer 43 - Change of course 44 - Technological specialization diploma holders 51 - Change of institution/course 53 - Short cycle diploma holders 57 - Change of institution/course (International)		no
Application order	Feature	Integer		Application order (between 0 - first choice; and 9 last choice)		no
Course	Feature	Integer		33 - Biofuel Production Technologies 171 - Animation and Multimedia Design 8014 - Social Service (evening attendance) 9003 - Agronomy 9070 - Communication Design 9085 - Veterinary Nursing 9119 - Informatics Engineering 9130 - Equinculture 9147 - Management 9238 - Social Service 9254 - Tourism 9500 - Nursing 9556 - Oral Hygiene 9670 - Advertising and Marketing Management 9773 - Journalism and Communication 9853 - Basic Education 9991 - Management (evening attendance)		no
Daytime/evening attendance	Feature	Integer		1 – daytime 0 - evening		no
Previous qualification	Feature	Integer	Education Level	1 - Secondary education 2 - Higher education - bachelor's degree 3 - Higher education - degree 4 - Higher education - master's 5 - Higher education - doctorate 6 - Frequency of higher education 9 - 12th year of schooling - not completed 10 - 11th year of schooling - not completed 12 - Other - 11th year of schooling 14 - 10th year of schooling 15 - 10th year of schooling - not completed 19 - Basic education 3rd cycle (9th/10th/11th year) or equiv. 38 - Basic education 2nd cycle (6th/7th/8th year) or equiv. 39 - Technological specialization course 40 - Higher education - degree (1st cycle) 42 - Professional higher technical course 43 - Higher education - master (2nd cycle)		no
Previous qualification (grade)	Feature	Continuous		Grade of previous qualification (between 0 and 200)		no
Nacionality	Feature	Integer	Nationality	1 - Portuguese; 2 - German; 6 - Spanish; 11 - Italian; 13 - Dutch; 14 - English; 17 - Lithuanian; 21 - Angolan; 22 - Cape Verdean; 24 - Guinean; 25 - Mozambican; 26 - Santomean; 32 - Turkish; 41 - Brazilian; 62 - Romanian; 100 - Moldova (Republic of); 101 - Mexican; 103 - Ukrainian; 105 - Russian; 108 - Cuban; 109 - Colombian		no
Mother's qualification	Feature	Integer	Education Level	1 - Secondary Education - 12th Year of Schooling or Eq. 2 - Higher Education - Bachelor's Degree 3 - Higher Education - Degree 4 - Higher Education - Master's 5 - Higher Education - Doctorate 6 - Frequency of Higher Education 9 - 12th Year of Schooling - Not Completed 10 - 11th Year of Schooling - Not Completed 11 - 7th Year (Old) 12 - Other - 11th Year of Schooling 14 - 10th Year of Schooling 18 - General commerce course 19 - Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv. 22 - Technical-professional course 26 - 7th year of schooling 27 - 2nd cycle of the general high school course 29 - 9th Year of Schooling - Not Completed 30 - 8th year of schooling 34 - Unknown 35 - Can't read or write 36 - Can read without having a 4th year of schooling 37 - Basic education 1st cycle (4th/5th year) or equiv. 38 - Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv. 39 - Technological specialization course 40 - Higher education - degree (1st cycle) 41 - Specialized higher studies course 42 - Professional higher technical course 43 - Higher Education - Master (2nd cycle) 44 - Higher Education - Doctorate (3rd cycle)		no
Father's qualification	Feature	Integer	Education Level	1 - Secondary Education - 12th Year of Schooling or Eq. 2 - Higher Education - Bachelor's Degree 3 - Higher Education - Degree 4 - Higher Education - Master's 5 - Higher Education - Doctorate 6 - Frequency of Higher Education 9 - 12th Year of		no

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
				Schooling - Not Completed 10 - 11th Year of Schooling - Not Completed 11 - 7th Year (Old) 12 - Other - 11th Year of Schooling 13 - 2nd year complementary high school course 14 - 10th Year of Schooling 18 - General commerce course 19 - Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv. 20 - Complementary High School Course 22 - Technical-professional course 25 - Complementary High School Course - not concluded 26 - 7th year of schooling 27 - 2nd cycle of the general high school course 29 - 9th Year of Schooling - Not Completed 30 - 8th year of schooling 31 - General Course of Administration and Commerce 33 - Supplementary Accounting and Administration 34 - Unknown 35 - Can't read or write 36 - Can read without having a 4th year of schooling 37 - Basic education 1st cycle (4th/5th year) or equiv. 38 - Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv. 39 - Technological specialization course 40 - Higher education - degree (1st cycle) 41 - Specialized higher studies course 42 - Professional higher technical course 43 - Higher Education - Master (2nd cycle) 44 - Higher Education - Doctorate (3rd cycle)		
Mother's occupation	Feature	Integer	Occupation	0 - Student 1 - Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers 2 - Specialists in Intellectual and Scientific Activities 3 - Intermediate Level Technicians and Professions 4 - Administrative staff 5 - Personal Services, Security and Safety Workers and Sellers 6 - Farmers and Skilled Workers in Agriculture, Fisheries and Forestry 7 - Skilled Workers in Industry, Construction and Craftsmen 8 - Installation and Machine Operators and Assembly Workers 9 - Unskilled Workers 10 - Armed Forces Professions 90 - Other Situation 99 - (blank) 122 - Health professionals 123 - teachers 125 - Specialists in information and communication technologies (ICT) 131 - Intermediate level science and engineering technicians and professions 132 - Technicians and professionals, of intermediate level of health 134 - Intermediate level technicians from legal, social, sports, cultural and similar services 141 - Office workers, secretaries in general and data processing operators 143 - Data, accounting, statistical, financial services and registry-related operators 144 - Other administrative support staff 151 - personal service workers 152 - sellers 153 - Personal care workers and the like 171 - Skilled construction workers and the like, except electricians 173 - Skilled workers in printing, precision instrument manufacturing, jewelers, artisans and the like 175 - Workers in food processing, woodworking, clothing and other industries and crafts 191 - cleaning workers 192 - Unskilled workers in agriculture, animal production, fisheries and forestry 193 - Unskilled workers in extractive industry, construction, manufacturing and transport 194 - Meal preparation assistants		no
Father's occupation	Feature	Integer	Occupation	0 - Student 1 - Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers 2 - Specialists in Intellectual and Scientific Activities 3 - Intermediate Level Technicians and Professions 4 - Administrative staff 5 - Personal Services, Security and Safety Workers and Sellers 6 - Farmers and Skilled Workers in Agriculture, Fisheries and Forestry 7 - Skilled Workers in Industry, Construction and Craftsmen 8 - Installation and Machine Operators and Assembly Workers 9 - Unskilled Workers 10 - Armed Forces Professions 90 - Other Situation 99 - (blank) 101 - Armed Forces Officers 102 - Armed Forces Sergeants 103 - Other Armed Forces personnel 112 - Directors of administrative and commercial services 114 - Hotel, catering, trade and other services directors 121 - Specialists in the physical sciences, mathematics, engineering and related techniques 122 - Health professionals 123 - teachers 124 - Specialists in finance, accounting, administrative organization, public and commercial relations 131 - Intermediate level science and engineering technicians and professions 132 - Technicians and professionals, of intermediate level of health 134 - Intermediate level technicians from legal, social, sports, cultural and similar services 135 - Information and communication technology technicians 141 - Office workers, secretaries in general and data processing operators 143 - Data, accounting, statistical, financial services and registry-related operators 144 - Other administrative support staff 151 - personal service workers 152 - sellers 153 - Personal care workers and the like 154 - Protection and security services personnel 161 - Market-oriented farmers and skilled agricultural and animal production workers 163 - Farmers, livestock keepers, fishermen, hunters and gatherers, subsistence 171 - Skilled construction workers and the like, except electricians 172 - Skilled workers in metallurgy, metalworking and similar 174 - Skilled workers in electricity and electronics 175 - Workers in food processing, woodworking, clothing and other industries and crafts 181 - Fixed plant and machine operators 182 - assembly workers 183 - Vehicle drivers and mobile equipment operators 192 - Unskilled workers in agriculture, animal production, fisheries and forestry 193 - Unskilled workers in extractive industry, construction, manufacturing and transport 194 - Meal preparation assistants 195 - Street vendors (except food) and street service providers		no
Admission grade	Feature	Continuous		Admission grade (between 0 and 200)		no
Displaced	Feature	Integer		1 – yes 0 – no		no
Educational special needs	Feature	Integer		1 – yes 0 – no		no
Debtor	Feature	Integer		1 – yes 0 – no		no
Tuition fees up to date	Feature	Integer		1 – yes 0 – no		no
Gender	Feature	Integer	Gender	1 – male 0 – female		no
Scholarship holder	Feature	Integer		1 – yes 0 – no		no
Age at enrollment	Feature	Integer	Age	Age of studend at enrollment		no
International	Feature	Integer		1 – yes 0 – no		no
Curricular units 1st sem (credited)	Feature	Integer		Number of curricular units credited in the 1st semester		no
Curricular units 1st sem (enrolled)	Feature	Integer		Number of curricular units enrolled in the 1st semester		no
Curricular units 1st sem (evaluations)	Feature	Integer		Number of evaluations to curricular units in the 1st semester		no
Curricular units 1st sem (approved)	Feature	Integer		Number of curricular units approved in the 1st semester		no
Curricular units 1st sem (grade)	Feature	Integer		Grade average in the 1st semester (between 0 and 20)		no
Curricular units 1st sem (without evaluations)	Feature	Integer		Number of curricular units without evalutions in the 1st semester		no

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Curricular units 2nd sem (credited)	Feature	Integer		Number of curricular units credited in the 2nd semester		no
Curricular units 2nd sem (enrolled)	Feature	Integer		Number of curricular units enrolled in the 2nd semester		no
Curricular units 2nd sem (evaluations)	Feature	Integer		Number of evaluations to curricular units in the 2nd semester		no
Curricular units 2nd sem (approved)	Feature	Integer		Number of curricular units approved in the 2nd semester		no
Curricular units 2nd sem (grade)	Feature	Integer		Grade average in the 2nd semester (between 0 and 20)		no
Curricular units 2nd sem (without evaluations)	Feature	Integer		Number of curricular units without evalutions in the 1st semester		no
Unemployment rate	Feature	Continuous		Unemployment rate (%)		no
Inflation rate	Feature	Continuous		Inflation rate (%)		no
GDP	Feature	Continuous		GDP		no
Target	Target	Categorical		Target. The problem is formulated as a three category classification task (dropout, enrolled, and graduate) at the end of the normal duration of the course		no

In [ ]:
```
data = pd.read_csv('data.csv', sep=';')

# some of the columns names contain '\t' characters, which we need to remove
data.columns = data.columns.str.replace('\t', '', regex=False)

X = data[data.columns[:-1]]
Y = data['Target']
```

3. Initial standard analysis with `sample`, `head`, `info`, `describe` (and `unique` values where appropriate!)

In [3]:
```
print("info")
print(data.info())
print("head")
print(data.head())
print("sample")
print(data.sample())
print("describe")
print(data.describe())

for col in data.columns:
    print(f"unique values in {col}")
    print(data[col].unique())

# Display the first few rows of X and Y
print("\nFirst 5 rows of X:")
print(X[:5])

print("\nFirst 5 rows of Y:")
print(Y[:5])

# Display information about the features (X)
print("\nShape of X:")
print(X.shape)

# Display unique values for the target (Y)
print("\nUnique values in Y:")
print(np.unique(Y, return_counts=True))
```

```
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
None
head
   Marital status  Application mode  Application order  Course  \
0               1                17                  5     171   
1               1                15                  1    9254   
2               1                 1                  5    9070   
3               1                17                  2    9773   
4               2                39                  1    8014   

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

   Previous qualification (grade)  Nacionality  Mother's qualification  \
0                           122.0            1                      19   
1                           160.0            1                       1   
2                           122.0            1                      37   
3                           122.0            1                      38   
4                           100.0            1                      37   

   Father's qualification  ...  Curricular units 2nd sem (credited)  \
0                      12  ...                                    0   
1                       3  ...                                    0   
2                      37  ...                                    0   
3                      37  ...                                    0   
4                      38  ...                                    0   

   Curricular units 2nd sem (enrolled)  \
0                                    0   
1                                    6   
2                                    6   
3                                    6   
4                                    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 37 columns]
sample
      Marital status  Application mode  Application order  Course  \
3151               1                 1                  1     171

      Daytime/evening attendance  Previous qualification  \
3151                           1                       1

      Previous qualification (grade)  Nacionality  Mother's qualification  \
3151                           159.0            1                       3

      Father's qualification  ...  Curricular units 2nd sem (credited)  \
3151                       1  ...                                    0

      Curricular units 2nd sem (enrolled)  \
3151                                    0

      Curricular units 2nd sem (evaluations)  \
3151                                       0

      Curricular units 2nd sem (approved)  Curricular units 2nd sem (grade)  \
3151                                    0                               0.0

      Curricular units 2nd sem (without evaluations)  Unemployment rate  \
3151                                               0               12.4

      Inflation rate   GDP   Target
3151             0.5  1.79  Dropout

[1 rows x 37 columns]
describe
       Marital status  Application mode  Application order       Course  \
count     4424.000000       4424.000000        4424.000000  4424.000000
mean         1.178571         18.669078           1.727848  8856.642631
std          0.605747         17.484682           1.313793  2063.566416
min          1.000000          1.000000           0.000000    33.000000
25%          1.000000          1.000000           1.000000  9085.000000
50%          1.000000         17.000000           1.000000  9238.000000
75%          1.000000         39.000000           2.000000  9556.000000
max          6.000000         57.000000           9.000000  9991.000000

       Daytime/evening attendance  Previous qualification  \
count                 4424.000000             4424.000000
mean                     0.890823                4.577758
std                      0.311897               10.216592
min                      0.000000                1.000000
25%                      1.000000                1.000000
50%                      1.000000                1.000000
75%                      1.000000                1.000000
max                      1.000000               43.000000

       Previous qualification (grade)  Nacionality  Mother's qualification  \
count                     4424.000000  4424.000000             4424.000000
mean                       132.613314     1.873192               19.561935
std                         13.188332     6.914514               15.603186
min                         95.000000     1.000000                1.000000
25%                        125.000000     1.000000                2.000000
50%                        133.100000     1.000000               19.000000
75%                        140.000000     1.000000               37.000000
max                        190.000000   109.000000               44.000000

       Father's qualification  ...  \
count             4424.000000  ...
mean                22.275316  ...
std                 15.343108  ...
min                  1.000000  ...
25%                  3.000000  ...
50%                 19.000000  ...
75%                 37.000000  ...
max                 44.000000  ...

       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 36 columns]
unique values in Marital status
[1 2 4 3 5 6]
unique values in Application mode
[17 15  1 39 18 53 44 51 43  7 42 16  5  2 10 57 26 27]
unique values in Application order
[5 1 2 4 3 6 9 0]
unique values in Course
[ 171 9254 9070 9773 8014 9991 9500 9238 9670 9853 9085 9130 9556 9147
 9003   33 9119]
unique values in Daytime/evening attendance
[1 0]
unique values in Previous qualification
[ 1 19 42 39 10  3 40  2  4 12 43 15  6  9 38  5 14]
unique values in Previous qualification (grade)
[122.  160.  100.  133.1 142.  119.  137.  138.  139.  136.  133.  110.
 149.  127.  135.  140.  125.  126.  151.  115.  150.  143.  130.  120.
 103.  154.  132.  167.  129.  141.  116.  148.  118.  106.  121.  114.
 124.  123.  113.  111.  131.  158.  146.  117.  153.  178.   99.  134.
 128.  170.  155.  145.  152.  112.  107.  156.  188.   96.  161.  166.
 147.  144.  102.  101.  180.  172.  105.  108.  165.  190.  162.  164.
 163.  159.  117.4 175.  133.8 176.  168.  139.3  97.  157.  140.8 184.4
 148.9 109.  174.  182.  138.6  95.  154.4 163.3 145.7 123.9 124.4 169.
 177.  138.7 119.1 118.9 126.6]
unique values in Nacionality
[  1  62   6  41  26 103  13  25  21 101  11  22  32 100  24 109   2 108
 105  14  17]
unique values in Mother's qualification
[19  1 37 38  3  4 42  2 34 12 40  9  5 39 11 41 30 14 35 36  6 10 29 43
 18 22 27 26 44]
unique values in Father's qualification
[12  3 37 38  1 19  5  4 34  2 39 11  9 36 26 40 14 20 35 41 22 13 29 43
 18 42 10  6 30 25 44 33 27 31]
unique values in Mother's occupation
[  5   3   9   7   4   1 125   0   6   2  90   8 141 175  99 191 151 194
 192 132 152 134  10 143 123 173 193 122 144 131 171 153]
unique values in Father's occupation
[  9   3   7  10   5   8   4   1   2 124   6   0  90 175 121  99 144 195
 192 161 193 151 182 132 131 194 163 135 143 171 103 172 152 183 122 102
 181 134 123 112 153 174 141 114 101 154]
unique values in Admission grade
[127.3 142.5 124.8 119.6 141.5 114.8 128.4 113.1 129.3 123.  130.6 119.3
 130.2 111.8 137.1 120.7 137.4 136.3 124.6 120.3 121.8 125.5 114.9 123.9
 157.  116.4 131.  122.1 118.8 150.  130.  138.8 134.5 131.4 102.5 128.8
 122.9 113.9 120.  121.1 120.4 100.6 121.4 109.7 134.1 127.6 132.4 133.4
 126.1 113.5 121.3 159.3 129.1 155.3 139.8 115.2 131.9 126.  120.9 128.2
```

```
 120.1 100.  134.  130.8 135.8 111.7 132.9 115.5 106.  117.  110.2 155.7
 180.4 110.  161.  117.6 128.7 112.2 100.8 105.  114.  137.  124.9 134.3
 111.5 160.  117.4 122.2 118.2 106.7 108.2 107.  136.1 115.3 140.4 113.4
 118.6 122.3 127.9 117.1 145.3 122.6 128.  123.7 131.7 133.2 109.3 113.
 157.9 112.1 174.7 110.1  99.7 121.  119.1 124.7 117.2 131.5 121.7 123.4
 132.8 108.7 138.1 126.5 127.4 123.6 122.  125.8 123.3 124.4 170.  121.5
 108.  132.3 148.  113.3 133.  140.  128.3 104.  135.  126.6 129.  162.3
 163.4 122.8 118.  129.8 152.  131.8 105.9 132.1 129.5 126.7 116.5 149.8
 115.1 124.5 136.  126.3 126.9 145.  115.8 147.  122.5 117.5 127.5 133.3
  97.  112.  130.5 141.7 119.7 119.4 155.  122.7 117.9 116.8 125.4 127.2
 103.4 123.2 124.1  99.5 110.8 118.9 121.6 149.2 126.2 127.8 113.7 117.8
 136.7 144.7 142.3 143.  100.1 101.  116.  135.6 118.7 125.7 107.1 127.
 154.4 116.1 118.5 146.7 124.3 137.8 147.8 155.6 130.9 125.  136.8 151.
 103.5 134.4 132.5 114.7 166.9 125.9 178.3 135.1 136.2 116.3 124.2 127.1
 172.  128.5 131.2 112.9 140.9 148.4 129.7 119.9 141.  116.6 140.2 146.2
 156.1 115.  158.7 111.9 114.2  96.  128.9 131.3 139.  120.8 150.5 129.4
 114.4 114.5 130.3 114.6 129.9 114.3 152.4 155.1 153.2 125.2 141.3 104.5
 113.2 123.8 133.8 132.7 106.6 124.  116.9 130.4 132.  143.7 152.8 133.5
 101.8 121.9 148.8 126.8 115.9 132.2 183.5 100.9 110.3 137.2 118.4 144.9
 146.8 138.  119.8  95.  144.2 135.3 143.9 140.7 118.1 106.5 137.5 102.
 133.9 134.7 125.3 106.1 139.9 109.  109.5 131.6 136.4 132.6 158.  144.4
 143.3 112.3 180.  136.5 137.6 138.4 123.5 142.8 155.5 163.5 161.9 166.6
 137.7 152.1 162.9 146.5 190.  144.3 148.3 107.5 138.5 113.6 143.2 118.3
 115.6 112.4 154.  111.3 162.5 119.  131.1 162.  110.6 156.9 159.1 149.
 129.6 133.6 146.9 109.1 129.2 154.1 147.2 151.6 111.6 117.3 115.7 134.8
 112.6 128.1 143.4  96.7 143.1 143.5 151.1 112.5 133.1 151.3 113.8 107.7
 128.6 134.6 135.7 135.5 127.7 140.1 105.5 120.2 145.6 142.  107.8 109.4
 143.6 119.5 175.6 136.9 101.7 112.7 105.8 126.4 164.9 157.5 146.6 142.2
 135.9 106.8 139.7 146.  108.8 134.9 136.6 159.  105.4 139.1 184.  108.3
 147.6 125.1 150.3 120.5 120.6 168.5 108.5 138.2 144.  152.3 169.7 117.7
 159.7 159.9 153.5 123.1 171.2 153.9 111.   98.9 139.4 154.8  95.5 111.1
 147.7 138.3 141.8 125.6 144.1 122.4 116.2 139.2 116.7 155.8  97.2 156.
 163.7 156.3 153.  141.4 168.  101.3  98.7 100.7 106.2 103.   98.1 130.7
 164.3  99.  161.2 149.9 110.5 137.9 165.8 152.5 140.5 104.8 140.8 121.2
 133.7 151.9 145.4 115.4  97.4 105.2 134.2 146.3 161.5 148.7  98.  143.8
 138.7 138.6 155.2 106.3  97.5 139.6 150.6 184.4 150.2 146.1 102.6 153.8
 137.3 149.3 119.2 147.3 147.5 110.4 148.9 150.9  99.3 104.1 176.7 156.8
 144.6 153.6 158.3 154.9 100.5 107.2 141.2 150.4 138.9  95.1 157.7 104.6
 178.  161.1 107.6 145.2 160.6 145.8  95.8 105.6 110.7 106.4 150.1 156.5
  98.6 146.4 165.7 156.4 148.5 104.7 151.5 139.3 158.1 163.3 107.3 145.7
 153.1 161.8 150.8 163.6 145.9 105.1 144.8 151.2 167.3 168.2 166.  101.5
 112.8 104.2 102.2 105.3 149.5 130.1 108.6 135.4 149.7 169.2 144.5 153.7
 157.8 152.9  98.5 160.1 160.5 151.4 151.7 176.  142.4 141.6 100.2 152.7
 173.3 157.4 102.4 162.2 159.5 135.2 154.5 105.7 108.4 109.8 165.2 103.6
  99.6 167.1 139.5 154.3 142.7 140.3 103.7 162.4  96.1 109.6 149.4 101.6
 148.6 107.4 168.6 155.9 179.6 141.1 163.  154.7 109.9 154.6 148.2 142.6
 147.9 160.4 111.4 163.1 102.8 162.1 103.8 156.2]
unique values in Displaced
[1 0]
unique values in Educational special needs
[0 1]
unique values in Debtor
[0 1]
unique values in Tuition fees up to date
[1 0]
unique values in Gender
[1 0]
unique values in Scholarship holder
[0 1]
unique values in Age at enrollment
[20 19 45 50 18 22 21 34 37 43 55 39 29 24 27 23 26 33 35 25 44 36 47 28
 38 30 31 32 40 42 48 49 46 41 70 60 53 51 52 54 61 58 59 17 57 62]
unique values in International
[0 1]
unique values in Curricular units 1st sem (credited)
[ 0  2  3  6  7 13  4  1  5 19 11  8 10  9 15 12 14 18 17 16 20]
unique values in Curricular units 1st sem (enrolled)
[ 0  6  5  7  8  1 12 10 18  9 21  3 17 16 11 14 13  2  4 15 19 23 26]
unique values in Curricular units 1st sem (evaluations)
[ 0  6  8  9 10  5  7 14 12 15 13 11  1 17 18 19 21  4 16  3 24  2 22 45
 20 26 29 36 32 23 27 31 28 25 33]
unique values in Curricular units 1st sem (approved)
[ 0  6  5  7  4  1  3  2  8 18 10  9 21 11 13 12 16 14 17 19 15 20 26]
unique values in Curricular units 1st sem (grade)
[ 0.         14.         13.42857143 12.33333333 11.85714286 13.3
 13.875      11.4        13.21428571 10.57142857 13.25       13.2
 12.         13.30625    12.5        11.66666667 11.4375     12.85714286
 13.375      13.29666667 11.6        11.375      12.66666667 12.93333333
 12.83333333 11.33333333 12.4        10.         11.         12.75
 14.8        13.928      13.         11.5        13.51666667 13.66666667
 11.83333333 14.71428571 10.6        13.04444444 14.935      13.13333333
 11.9        12.57142857 11.2        11.975      10.33333333 10.5
 14.84285714 12.2        12.16666667 13.08       10.66666667 11.875
 13.75       14.2        11.96       12.875      14.63157895 13.26666667
 13.85714286 13.4        12.51666667 11.61666667 13.14285714 12.6
 10.8        12.8        14.16666667 12.1        13.46714286 13.2375
 14.83333333 13.188      17.125      13.605      13.86333333 13.33333333
 12.25       13.16666667 12.85555556 14.91875    14.25714286 11.25
 12.45714286 11.68333333 11.75       11.95       14.35       12.61904762
 13.05       12.24285714 12.93375    11.81666667 11.16666667 12.32428571
 11.14285714 12.91428571 13.83333333 13.71428571 12.64285714 12.91666667
 15.625      11.71428571 12.125      13.85       13.74166667 15.33333333
 11.57142857 13.5        12.47142857 12.90909091 12.9        13.22
 13.15142857 15.         14.4        16.         13.57142857 14.82857143
 13.67142857 12.67       14.5        13.68571429 11.55555556 13.79714286
 14.00375    13.8        10.83333333 13.385      15.66666667 14.33333333
 11.42857143 14.72571429 12.86       15.11666667 15.2        16.2
 14.54285714 13.69428571 11.28571429 11.75714286 13.05857143 14.3
 11.8        13.39125    10.69230769 12.58571429 13.98       16.15384615
```

14.03333333	13.41428571	13.54545455	15.13571429	12.28571429	12.74
12.55555556	11.27272727	14.66666667	14.85714286	11.72727273	14.73142857
13.61538462	15.8	14.65428571	12.40142857	11.5875	15.23076923
12.90142857	14.14285714	13.25375	13.78571429	13.91666667	12.08285714
13.81428571	11.3	12.3	12.69428571	14.4125	12.92875
14.40285714	13.38142857	14.13125	13.6	13.84615385	16.4
13.81818182	10.2	11.44444444	13.31	13.48714286	13.96666667
11.98	11.61428571	14.31428571	12.76666667	14.70625	11.81428571
12.77125	11.72142857	14.75	10.75	12.61	13.625
11.69	14.52857143	15.95714286	12.625	13.49666667	11.77777778
12.42857143	12.5075	12.09090909	14.6	13.82857143	12.5375
12.1625	12.45	11.45454545	11.88888889	16.88571429	13.05555556
12.35571429	12.35714286	12.84714286	12.34	13.73125	10.88888889
13.025	12.04285714	12.8375	12.62857143	14.26875	12.22857143
15.375	12.61428571	10.92307692	15.83333333	13.3625	15.4
12.81818182	13.39473684	14.25	13.44444444	11.8375	12.68421053
17.33333333	13.11111111	14.11111111	14.72727273	13.10571429	11.84615385
14.42857143	12.14285714	13.47142857	13.105	15.5	12.22222222
14.125	15.16666667	15.14285714	13.41142857	14.668	13.345
14.53714286	13.08714286	14.14571429	14.45	15.71428571	13.46285714
12.41428571	14.01666667	10.4	13.91428571	13.76111111	13.84285714
14.68285714	18.	11.46153846	14.24444444	12.6	14.01428571
12.8875	12.95	13.09428571	13.7	11.70571429	15.28571429
13.46666667	14.875	14.72857143	14.46	16.5	15.22857143
12.77777778	13.128	11.605	13.18181818	12.98125	11.68
13.175	13.31428571	12.79142857	11.625	11.99444444	14.1
12.39666667	15.01428571	14.05	14.85	14.7875	12.49571429
14.56666667	12.68571429	14.445	13.41666667	13.1	13.53333333
12.71428571	13.30857143	14.07125	14.90142857	12.65	12.13571429
12.62	13.76666667	13.11625	12.7	15.75	14.35555556
17.00555556	12.11428571	13.05875	13.28571429	14.07142857	13.29714286
13.92307692	13.37	13.48571429	12.94	13.25857143	14.65285714
13.72727273	15.07	12.60625	13.59833333	12.04142857	13.76923077
15.44444444	15.57142857	13.63	16.9	12.76428571	14.60428571
13.57625	13.76375	10.71428571	14.35714286	13.92571429	13.44285714
13.125	11.67777778	16.14285714	11.78571429	13.42375	12.46428571
12.375	13.02428571	14.28571429	14.38461538	11.98857143	11.54545455
16.46153846	14.37	14.69625	12.85555556	13.04	15.65714286
13.33142857	13.45454545	15.02857143	11.42105263	13.35	13.88125
13.15	12.34285714	13.45714286	12.90857143	12.81666667	11.61111111
13.97666667	12.45454545	10.98333333	13.38461538	16.11111111	13.58571429
13.34	13.38888889	13.58333333	12.82857143	11.32857143	13.03333333
12.6875	10.46666667	12.83571429	13.24875	12.54545455	11.98333333
10.25	12.51428571	13.30769231	13.34285714	14.27272727	12.1375
13.52857143	13.6375	14.36	14.51428571	14.7	16.57142857
15.145	13.76125	15.30769231	14.75714286	11.91666667	15.63375
13.38125	12.16	13.3425	14.175	14.15384615	12.58823529
13.05375	12.03571429	11.53846154	13.11714286	12.63636364	15.125
13.72571429	15.57	14.11428571	13.07571429	13.51285714	12.288
12.76571429	14.88888889	11.35714286	13.88428571	13.8875	13.4275
13.075	13.49285714	14.58333333	15.66	13.55714286	12.18333333
13.435	13.11666667	13.87714286	12.94428571	14.26666667	12.48375
13.45	12.48571429	14.86857143	13.03	14.57142857	13.3675
12.80833333	13.91857143	13.68333333	14.69285714	14.035	11.7725
12.46	14.80571429	12.725	13.135	12.85	14.14125
14.6625	12.99333333	10.21428571	13.54285714	11.82857143	12.96666667
14.12571429	12.78333333	13.44	12.21428571	12.84615385	13.37142857
13.38571429	13.32857143	14.3775	14.15	14.07222222	12.63333333
14.23333333	15.7	16.33333333	12.5625	13.78125	13.26
15.15555556	12.82875	10.875	12.75833333	13.50714286	13.61428571
11.97142857	12.38571429	14.28333333	13.27272727	15.115	14.56571429
13.79666667	14.96285714	13.64285714	12.15384615	13.40375	16.16666667
11.29444444	13.43333333	12.74285714	14.93571429	12.52857143	12.36363636
15.09090909	13.05882353	13.39166667	14.08571429	10.16666667	12.12142857
14.22222222	12.49285714	15.88571429	14.77777778	12.72727273	18.875
15.08333333	12.44428571	12.51333333	11.15555556	11.575	13.19857143
13.87571429	13.15384615	14.44142857	13.2325	13.07375	13.46428571
14.4025	15.46875	14.98333333	15.88888889	13.98333333	14.84285714
12.55375	12.66571429	12.57894737	12.002	13.86714286	12.675
11.63571429	12.52608696	16.42857143	13.336	14.11571429	12.415
14.31285714	15.42857143	10.85714286	11.73333333	13.04666667	13.11
14.41428571	13.045	12.858	14.2525	14.06	15.07571429
13.77857143	13.07285714	16.71428571	11.65833333	14.625	12.56
13.19	12.77142857	11.71666667	17.	12.21052632	12.46153846
11.90909091	13.46	10.36842105	12.69230769	14.30769231	14.45454545
12.44444444	13.6125	12.41428571	14.06666667	13.55555556	13.27625
14.375	12.97166667	10.92857143	12.75714286	13.67	14.13714286
13.55	14.63636364	12.78571429	13.80142857	11.82	13.42222222
14.57285714	12.98285714	13.755	14.71111111	14.61538462	13.67714286
13.95833333	13.3875	12.64833333	13.39	10.9	15.85714286
12.62857143	14.23285714	12.32	12.35	13.09090909	15.94285714
14.575	13.47	13.79857143	10.64285714	14.65	12.26666667
10.375	14.44444444	12.50625	16.85714286	15.11428571	12.04
14.32857143	11.71666667	12.58	14.53571429	11.6875	12.84142857
17.11111111	12.36	15.25	13.81375	11.9625	14.98571429
13.78	13.805	14.42428571	12.07142857	14.11	9.8
14.31666667	12.72857143	13.68142857	11.94	13.89428571	13.46875
13.41625	15.3	14.61666667	12.19230769	13.9	11.61538462
13.99571429	12.88625	14.50125	13.56875	14.15375	12.45428571
13.27166667	13.54833333	13.2575	11.41666667	11.15	11.67857143
14.48571429	12.65714286	13.07142857	13.126	15.69230769	15.05263158
12.93875	12.18	12.71857143	11.64285714	12.5125	13.48333333
12.565	13.06	10.28571429	12.27142857	13.22222222	13.15428571
15.10142857	12.41666667	14.93333333	14.455	12.32857143	13.27142857
12.58333333	13.73166667	12.49166667	10.81818182	12.87857143	14.53846154
12.73333333	13.93714286	13.6875	13.398	11.76923077	13.08166667
14.59444444	13.17647059	12.59125	11.89473684	12.36571429	16.83333333
11.125	14.54	15.67571429	13.18333333	11.9725	13.95714286
12.91	13.58	12.88888889	14.34125	14.20857143	14.30714286

```
 14.21428571 13.97142857 15.07777778 13.38333333 13.01125    11.85
 12.32777778 12.07692308 11.23076923 10.98571429 16.13142857 14.83857143
 13.9475     13.06666667 13.74625    13.81714286 15.53846154 14.05714286
 13.29       13.86166667 15.43333333 15.11       15.79857143 13.77571429
 14.71666667 14.368      12.08428571 10.74444444 13.4173913  13.24285714
 13.55714286 14.09333333 13.51428571 12.74428571 13.3125     14.95
 14.99285714 13.01666667 13.32714286 11.832      12.41285714 12.54571429
 12.51428571 13.62857143 13.73684211 10.76923077 12.47619048 13.29166667
 14.24       14.10571429 11.96666667 13.3475     13.17571429 13.74285714
 12.21714286 11.07142857 11.325      11.92307692 14.88875    12.4125
 14.9125    ]
unique values in Curricular units 1st sem (without evaluations)
[ 0  1  2  4  3  6 12 10  7  5  8]
unique values in Curricular units 2nd sem (credited)
[ 0  1  2  5  7  4 10  3 13  9  6 11 12  8 14 15 16 18 19]
unique values in Curricular units 2nd sem (enrolled)
[ 0  6  5  8  7 11 12  9 13 19  3 10  4 17  2  1 14 15 16 23 18 21]
unique values in Curricular units 2nd sem (evaluations)
[ 0  6 10 17  8  5  7 14  9 12 11 13 19  3 15 16  4 18  2 21  1 26 27 22
 20 24 28 23 25 33]
unique values in Curricular units 2nd sem (approved)
[ 0  6  5  8  2  7  4  1  3 10 13 11 19  9 12 17 14 20 16 18]
unique values in Curricular units 2nd sem (grade)
[ 0.         13.66666667 12.4        13.         11.5        14.345
 14.14285714 13.5        14.2        13.21428571 11.         12.
 14.545      12.25       11.425      12.85714286 12.28571429 14.11428571
 13.28571429 12.33333333 13.71666667 10.57142857 13.4        14.375
 13.42857143 10.         11.6        15.8        14.34285714 13.8
 14.33333333 13.2        13.77142857 12.83333333 14.16666667 11.83333333
 13.595      12.9        14.935      12.22142857 12.42857143 10.75
 10.8        13.75       10.85714286 13.33333333 15.05555556 11.66666667
 12.2        13.3        11.33333333 13.08       11.2        13.25
 14.47777778 14.23076923 12.27272727 12.3        15.         12.8
 12.57142857 12.51666667 10.66666667 11.4        14.         11.25
 13.77777778 15.01111111 13.74285714 14.6        11.71428571 12.5
 14.83333333 12.89285714 17.5875    12.66666667 12.33125    14.17857143
 10.83333333 11.8        11.85714286 12.6        12.76111111 15.355
 14.25714286 10.5        13.5625    11.68333333 13.16666667 11.95
 10.4        14.58333333 12.78947368 13.05       12.24285714 12.93375
 14.42857143 12.32428571 12.91428571 12.64285714 12.91666667 15.85714286
 13.3125     14.5        12.19285714 14.9125    11.57142857 14.25
 10.33333333 11.4625    13.55555556 13.6        12.36363636 13.22
 13.15142857 10.6        12.16666667 14.82857143 13.67142857 13.55
 15.42857143 15.3875    13.775      13.81285714 13.83333333 13.385
 14.72571429 13.1375    12.86       13.71428571 16.16888889 10.25
 15.14285714 16.4       14.88181818 14.35555556 13.08333333 13.05857143
 14.05555556 14.66666667 14.57142857 13.39125    13.1        12.30769231
 16.42857143 13.44285714 13.98       15.75       14.03333333 13.41428571
 13.57142857 15.2875    10.97857143 11.11111111 11.16666667 12.14
 12.75       14.73142857 14.8        14.65428571 12.40142857 14.07142857
 15.6        15.4       12.375      13.09444444 13.25375    14.1
 13.91666667 12.08285714 13.81428571 11.3        16.         12.69428571
 16.05       14.0625    14.40285714 13.38142857 13.98375    14.09090909
 13.45454545 12.625     13.26666667 13.31       13.48714286 13.96666667
 13.225      14.0375    14.31428571 12.12857143 11.75       14.4775
 13.54285714 11.875     12.77125    11.72142857 12.525      10.2
 12.61       11.69      11.28571429 15.01818182 15.18888889 13.49666667
 12.5075     13.625     14.71428571 13.82857143 12.61428571 15.2
 13.825      14.55555556 11.7       11.54545455 11.46153846 15.74444444
 15.8125     12.35571429 11.98       12.35714286 13.85714286 13.78888889
 13.01875    13.73125    13.81111111 12.04285714 15.5        12.71428571
 12.62857143 14.26875    13.2125    11.45454545 11.70833333 14.44444444
 13.54545455 13.55473684 11.7375    11.22222222 12.48125    12.38461538
 17.         14.85714286 12.14285714 15.77777778 13.10571429 14.4
 17.6        16.575     11.18181818 11.42857143 13.94444444 13.105
 11.36363636 13.29444444 15.625     12.7        14.63375    13.345
 14.53714286 13.22222222 12.95      14.28571429 13.62777778 14.14444444
 13.16428571 15.33333333 14.01666667 14.10625    12.82375    13.84285714
 11.58333333 12.23333333 12.6       15.35       12.46153846 12.54545455
 12.8875     12.55       13.7       15.11111111 11.70571429 16.075
 13.64285714 14.75      11.125     16.32727273 17.42857143 15.22857143
 13.19333333 12.4125    12.90909091 12.98125    11.68       14.875
 13.31428571 15.57142857 12.79142857 12.43142857 12.94285714 16.0875
 15.1        15.8625    14.67142857 14.56666667 12.68571429 14.445
 14.625      13.41666667 14.8125    13.61428571 14.07125    14.90142857
 12.44555556 13.32857143 14.05142857 13.11625    13.49375    16.58222222
 12.11428571 13.05875    15.1625    13.29714286 13.48571429 13.76
 13.25857143 12.6375    14.65285714 11.14285714 15.07       12.5125
 11.9        11.93333333 12.96818182 13.14285714 12.04142857 15.30769231
 15.16666667 16.375     13.63       12.22857143 16.37777778 12.76428571
 15.68333333 13.57625    15.83333333 13.76375    14.35714286 13.92571429
 13.74444444 15.875     16.2       11.78571429 13.42375    12.46428571
 13.02428571 18.57142857 12.38      11.98857143 12.88888889 13.875
 13.41111111 14.65      14.62222222 13.9125    13.99444444 11.68571429
 15.65714286 13.33142857 15.02857143 12.58333333 11.07692308 13.35
 13.88125    15.06875    12.34285714 14.5125    12.90857143 12.875
 12.05555556 12.55555556 12.04166667 10.98333333 14.6875    10.84
 12.45714286 14.88888889 13.58333333 12.82857143 11.32857143 13.03333333
 10.92307692 14.23125    13.24875    15.61111111 16.14285714 11.625
 12.28125    10.44444444 12.77777778 12.81818182 11.65       14.81818182
 12.91111111 14.65555556 14.3875    14.36       14.51428571 14.7
 15.0625     15.66666667 15.7425    15.075      13.76125    14.07777778
 12.76666667 14.02222222 15.63375    13.125     12.45       13.38125
 11.34       13.3425    13.03125    13.35714286 12.84615385 13.05375
 16.5        15.57      15.91111111 13.07571429 13.51285714 12.288
 13.54444444 13.72727273 11.1       14.86444444 12.16       13.4275
 12.36111111 12.68888889 13.55714286 14.63636364 11.725      12.18333333
 13.435      12.17647059 13.11666667 15.6625    12.94428571 14.26666667
 12.48375    15.3125    13.45      12.9375    13.21875    13.3675
```

```
 12.82571429 13.8875     13.68333333 14.9625     14.035      12.33
 12.46       14.80571429 13.7125     13.46153846 13.24444444 12.62142857
 14.14125    14.6625     12.66875    11.38461538 14.22222222 12.9125
 12.96666667 14.36875    12.84285714 11.225      15.125      13.06666667
 12.63636364 13.3875     11.77777778 14.3775     13.04285714 14.15
 12.47142857 14.0475     12.63333333 14.23333333 15.7        13.78125
 14.28625    16.8        12.82875    12.75833333 14.4375     12.72857143
 13.12555556 15.092      14.56571429 14.96285714 17.16666667 14.77272727
 13.06363636 13.40375    14.125      12.74285714 13.12857143 14.93571429
 13.90625    12.41666667 12.4875     14.08571429 11.50625    15.28571429
 14.3        11.90909091 11.44444444 11.7125     15.88571429 11.58
 11.35714286 12.08333333 17.69230769 15.3        11.34714286 12.18181818
 10.42857143 13.375      11.64285714 13.07142857 13.58571429 13.61538462
 12.42       13.0625     12.21428571 13.08571429 13.65625    13.2325
 13.07375    14.78888889 14.4025     16.3875     14.98333333 13.98333333
 11.61538462 11.675      13.75555556 14.38888889 12.55375    12.66571429
 12.11764706 12.002      13.86714286 12.675      11.575      11.63571429
 11.55555556 13.15       12.52608696 16.66666667 13.20714286 14.11571429
 13.11111111 12.415      14.31285714 14.01428571 15.925      11.80833333
 11.88888889 13.30625    13.11       14.41428571 13.045      12.858
 14.2525     11.24285714 15.07571429 13.12222222 12.6875     13.07285714
 12.11666667 12.56       12.77142857 13.18181818 10.72727273 12.2375
 12.10526316 16.6        15.23076923 14.575      12.41428571 11.4375
 14.06666667 13.27625    12.32142857 12.75714286 13.67       11.97142857
 14.13714286 14.1875     14.35       11.82       14.9        12.18888889
 14.95555556 14.4625     14.62777778 14.53333333 14.71875    15.9
 12.20555556 12.62857143 14.23285714 12.22222222 12.575      15.94285714
 10.71428571 13.32625    12.32       14.75625    15.36666667 11.55
 14.86363636 12.07142857 12.78571429 13.8375     15.11428571 11.77142857
 11.78333333 14.32857143 16.33333333 10.868      12.15714286 14.53571429
 12.84142857 13.81375    13.04444444 14.98571429 13.661      14.2725
 13.9375     14.11       14.31666667 10.16666667 13.68142857 12.61666667
 11.56666667 14.3625     13.31375    13.41625    14.61666667 14.06
 12.19230769 13.9        12.06666667 12.88625    13.88888889 14.50125
 14.02625    14.15375    12.45428571 12.94375    14.775      13.2575
 11.9625     12.81428571 13.0375     14.58888889 12.125      12.65714286
 13.126      11.375      13.79857143 14.53846154 12.93875    12.88
 11.8125     12.80555556 13.475      13.26666667 13.48333333 17.71428571
 13.65714286 13.15428571 16.09375    14.93333333 13.91       12.32857143
 12.72727273 14.47142857 12.67142857 12.49166667 11.63636364 11.85625
 14.46153846 13.93714286 14.85       15.62222222 13.5425     11.92857143
 12.38571429 13.472      16.90909091 12.82       12.92307692 12.36571429
 14.54       12.73333333 15.67571429 13.82       13.18333333 12.65125
 13.95714286 12.91       12.93625    14.34125    15.12777778 15.0125
 14.21428571 15.02352941 13.38333333 13.09090909 13.01125    13.65
 11.736      14.51111111 12.6125     15.25       12.09090909 14.2125
 13.51666667 15.52222222 13.9475     13.74625    13.81714286 15.55555556
 11.95625    13.97142857 13.29       15.43333333 13.78       15.11
 15.79857143 15.6        12.48571429 16.28571429 14.71666667 14.44875
 12.08428571 12.15       11.8375     12.5625     13.4173913  13.24285714
 12.05       14.51125    13.51428571 12.74428571 13.53333333 14.75555556
 15.6375     13.92727273 14.7375     14.99285714 13.01666667 13.32714286
 10.375      11.832      12.41285714 12.54571429 12.51428571 12.21052632
 13.29166667 14.24       13.6625     11.96666667 13.3475     14.925
 12.11875    11.325      12.06375    14.88875    11.2625     11.08333333]
unique values in Curricular units 2nd sem (without evaluations)
[ 0  5  2  1  3  6  4 12  7  8]
unique values in Unemployment rate
[10.8 13.9  9.4 16.2 15.5  8.9 12.7 11.1  7.6 12.4]
unique values in Inflation rate
[ 1.4 -0.3 -0.8  0.3  2.8  3.7  0.6  2.6  0.5]
unique values in GDP
[ 1.74  0.79 -3.12 -0.92 -4.06  3.51 -1.7   2.02  0.32  1.79]
unique values in Target
['Dropout' 'Graduate' 'Enrolled']

First 5 rows of X:
   Marital status  Application mode  Application order  Course  \
0               1                17                  5     171
1               1                15                  1    9254
2               1                 1                  5    9070
3               1                17                  2    9773
4               2                39                  1    8014

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

   Previous qualification (grade)  Nacionality  Mother's qualification  \
0                           122.0            1                      19
1                           160.0            1                       1
2                           122.0            1                      37
3                           122.0            1                      38
4                           100.0            1                      37

   Father's qualification  ...  \
0                      12  ...
1                       3  ...
2                      37  ...
3                      37  ...
4                      38  ...

   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 36 columns]

First 5 rows of Y:
0     Dropout
1    Graduate
2     Dropout
3    Graduate
4    Graduate
Name: Target, dtype: object

Shape of X:
(4424, 36)

Unique values in Y:
(array(['Dropout', 'Enrolled', 'Graduate'], dtype=object), array([1421,  794, 2209]))
```

3. Data Visualization

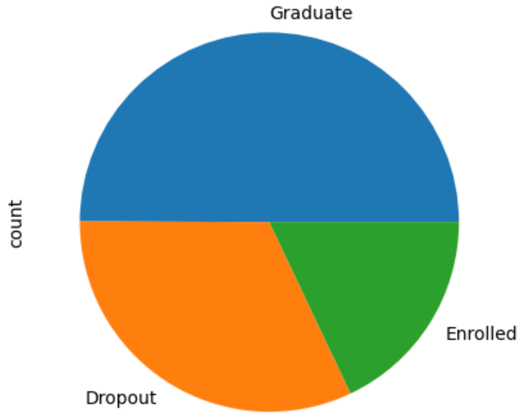
1. 4 different types of plots (correlation, histograms, etc.) including discussion
2. Data imbalance discussion with graphic

Target Imbalance

The data shows a class imbalance with more graduates compared to dropouts and enrolled students. This imbalance should be considered when building predictive models, as it may bias the model towards predicting the majority class (graduates).

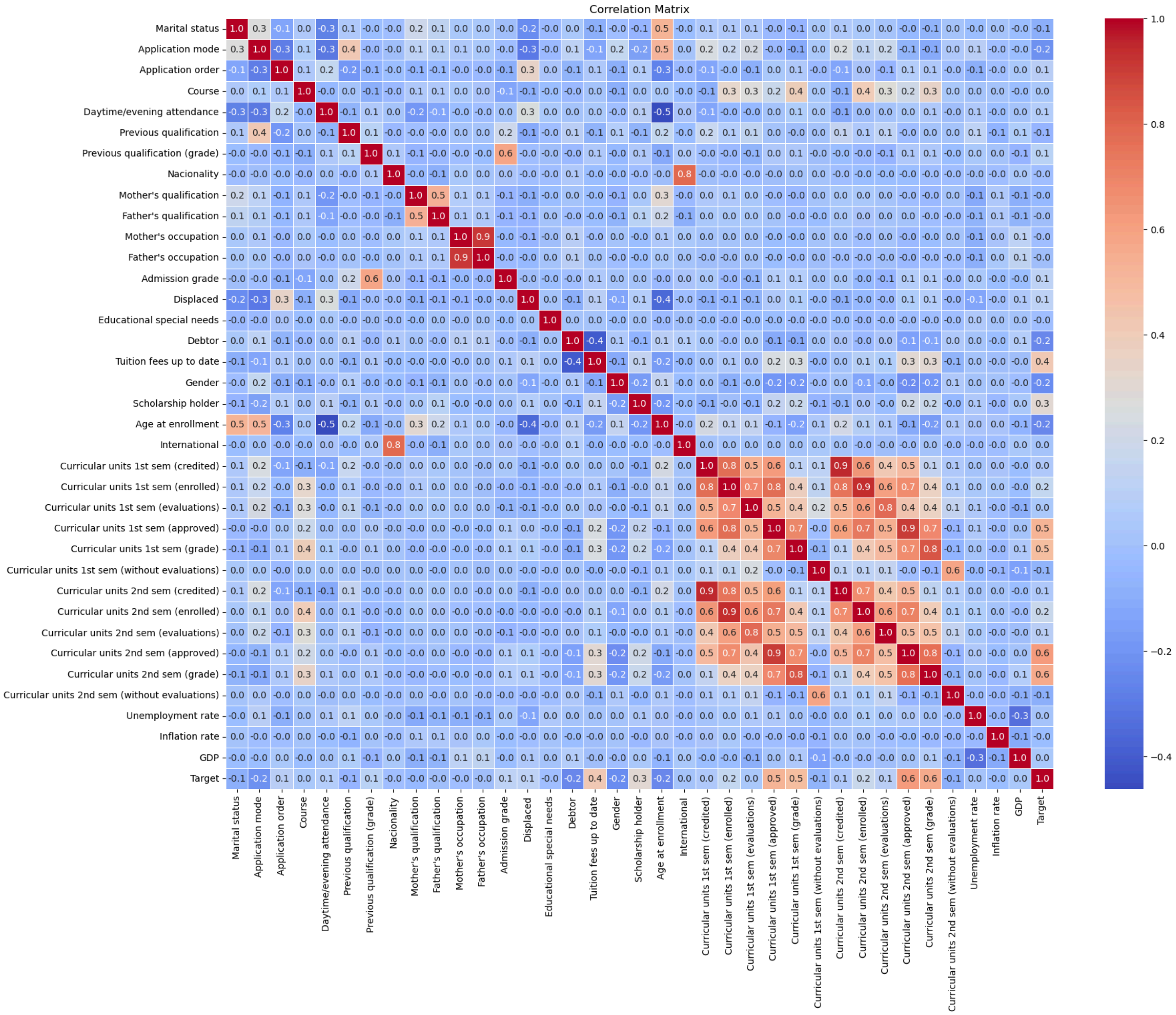
In [4]:
```python
data['Target'].value_counts().plot(kind='pie')
```

Out[4]: `<Axes: ylabel='count'>`



Correlation Matrix

In [5]:
```python
# encoding the target since it is categorical and for correlation we need numerical values
label_encoder = LabelEncoder()
data_with_encoded_target = data.copy()
data_with_encoded_target['Target'] = label_encoder.fit_transform(data['Target'])

correlation_matrix = data_with_encoded_target.corr()
plt.figure(figsize=(20, 15))
sns.heatmap(correlation_matrix, annot=True,
            cmap='coolwarm', fmt=".1f", linewidths=0.5)
plt.xticks(rotation=90)
plt.yticks(rotation=0)
plt.title('Correlation Matrix')
plt.show()
```



In [ ]:
```python
highly_correlated = []
for i in range(len(correlation_matrix.columns)):
    for j in range(i):
        if abs(correlation_matrix.iloc[i, j]) >= 0.7:
            highly_correlated.append(
                (correlation_matrix.index[i], correlation_matrix.columns[j]))

less_correlated = []
for i in range(len(correlation_matrix.columns)):
    for j in range(i):
        if abs(correlation_matrix.iloc[i, j]) <= 0.3:
            less_correlated.append(
                (correlation_matrix.index[i], correlation_matrix.columns[j]))
target_column = 'Target'
correlation_with_target = data_with_encoded_target.corr(
)[target_column].sort_values(ascending=False)
print("Correlation with Target:>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>")
print(correlation_with_target)
```

```
Correlation with Target:>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
Target                                             1.000000
Curricular units 2nd sem (approved)                0.624157
Curricular units 2nd sem (grade)                   0.566827
Curricular units 1st sem (approved)                0.529123
Curricular units 1st sem (grade)                   0.485207
Tuition fees up to date                            0.409827
Scholarship holder                                 0.297595
Curricular units 2nd sem (enrolled)                0.175847
Curricular units 1st sem (enrolled)                0.155974
Admission grade                                    0.120889
Displaced                                          0.113986
Previous qualification (grade)                     0.103764
Curricular units 2nd sem (evaluations)             0.092721
Application order                                  0.089791
Daytime/evening attendance                         0.075107
Curricular units 2nd sem (credited)                0.054004
Curricular units 1st sem (credited)                0.048150
Curricular units 1st sem (evaluations)             0.044362
GDP                                                0.044135
Course                                             0.034219
Unemployment rate                                  0.008627
International                                      0.003934
Father's qualification                            -0.001393
Father's occupation                               -0.001899
Mother's occupation                               -0.005629
Educational special needs                         -0.007353
Nacionality                                       -0.014801
Inflation rate                                    -0.026874
Mother's qualification                            -0.043178
Previous qualification                            -0.056039
Curricular units 1st sem (without evaluations)    -0.068702
Marital status                                    -0.089804
Curricular units 2nd sem (without evaluations)    -0.094028
Application mode                                  -0.221747
Gender                                            -0.229270
Debtor                                            -0.240999
Age at enrollment                                 -0.243438
Name: Target, dtype: float64
```

The correlation analysis reveals several meaningful relationships between input features and the target variable.

Curricular performance indicators show the strongest correlations:

- Curricular units 2nd sem (approved) (0.62)
- Curricular units 2nd sem (grade) (0.57)
- Curricular units 1st sem (approved) (0.53)
- Curricular units 1st sem (grade) (0.49)

> These suggest that **academic success** in the *early* semesters is a strong predictor of student outcomes.

Financial status indicators also play a notable role:

- Tuition fees up to date (0.41)
- Scholarship holder (0.30)

> Students who are financially stable or receive scholarships are more likely to continue their studies.

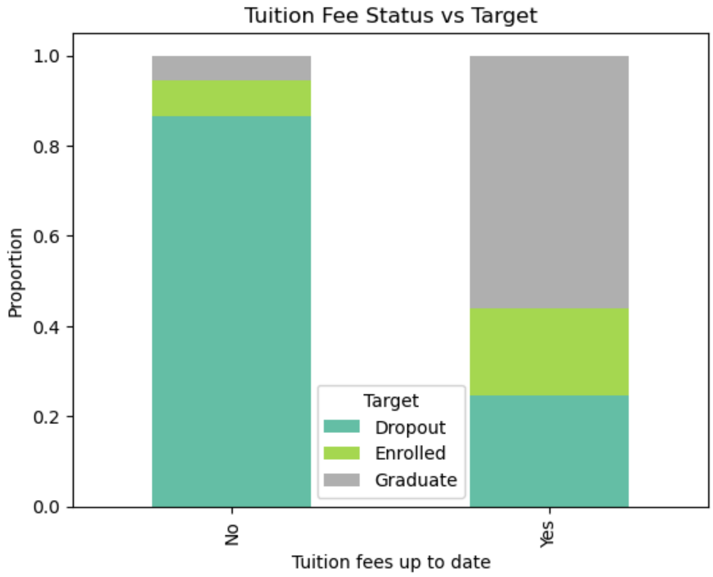
Weak or negligible correlations are seen in variables such as:

- Gender (–0.23)
- Age at enrollment (–0.24)
- Application mode (–0.22)

Marital status, Father’s qualification, and Mother’s qualification all show near-zero correlation.

Tuition Fee correlation

In [7]:
```
tuition_fee_map = {
    1: 'Yes',
    0: 'No',
}

pd.crosstab(data['Tuition fees up to date'].map(tuition_fee_map), data['Target'],
            normalize='index').plot(kind='bar', stacked=True, colormap='Set2')
plt.title('Tuition Fee Status vs Target')
plt.ylabel('Proportion')
plt.xlabel('Tuition fees up to date')
plt.legend(title='Target')
plt.show()
```



As we saw in the correlation matrix, financial factors play a meaningful role in predicting student outcomes. The stacked bar plot of "Tuition fees up to date" vs. Target clearly shows this relationship.

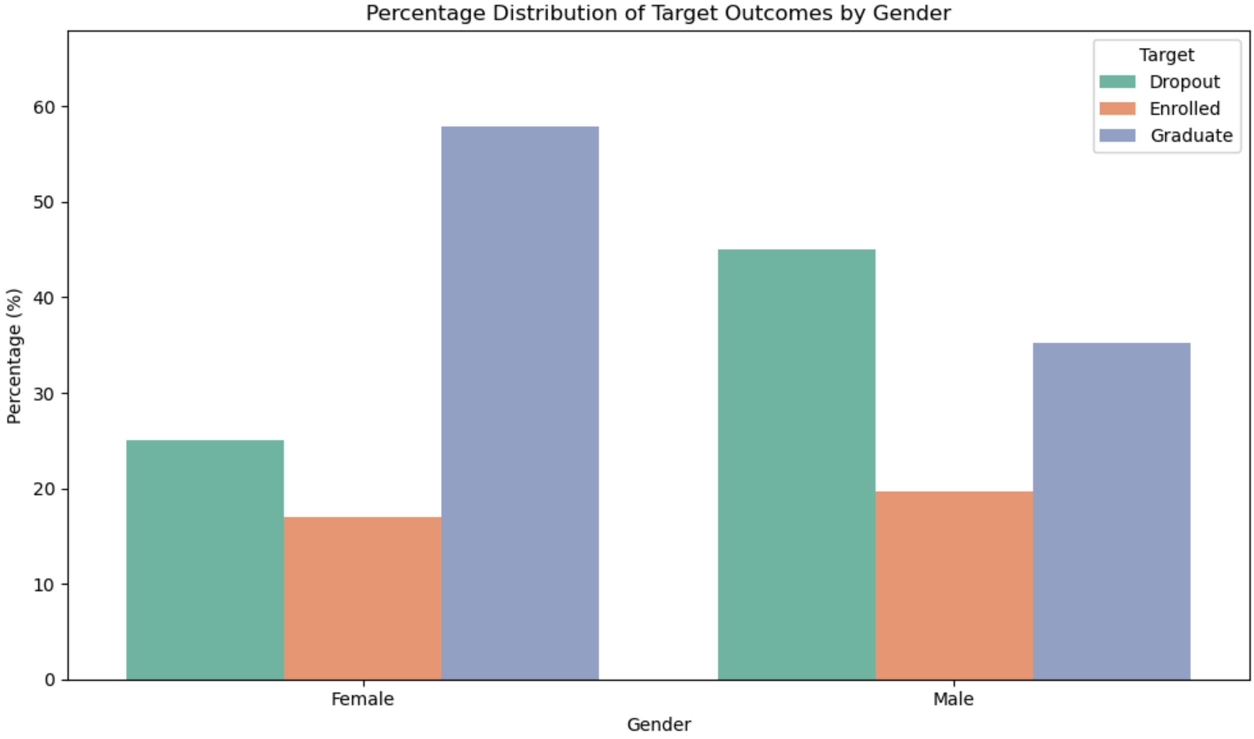
Students who are up to date with their tuition payments are more likely to achieve favorable outcomes, while those with outstanding fees have a higher dropout rate.

This supports the idea that financial stability is an important factor in student success.

Success Rates by Gender and Outcome

In [8]:
```
# Calculate counts by Gender and Target without modifying original data
counts = data.groupby(['Gender', 'Target']).size().reset_index(name='count')

# Calculate total per Gender for percentage calculation
totals = counts.groupby('Gender')['count'].transform('sum')
counts['percentage'] = counts['count'] / totals * 100

# Map Gender codes to labels for clarity (adjust if needed)
gender_map = {1: 'Male', 0: 'Female'}
counts['Gender_label'] = counts['Gender'].map(gender_map)
print(counts)

# Plot using seaborn
plt.figure(figsize=(10, 6))
sns.barplot(
    data=counts,
    x='Gender_label',
    y='percentage',
    hue='Target',
    palette='Set2'
)

plt.ylabel('Percentage (%)')
plt.xlabel('Gender')
plt.title('Percentage Distribution of Target Outcomes by Gender')
plt.legend(title='Target')
plt.ylim(0, counts['percentage'].max() + 10)
plt.tight_layout()
plt.show()
```

```
   Gender    Target  count  percentage Gender_label
0       0   Dropout    720   25.104603       Female
1       0  Enrolled    487   16.980474       Female
2       0  Graduate   1661   57.914923       Female
3       1   Dropout    701   45.051414         Male
4       1  Enrolled    307   19.730077         Male
5       1  Graduate    548   35.218509         Male
```



Female students graduate at a significantly higher rate (58%) than male students (35%), while males have a much higher dropout rate (45% vs. 25%). This stark contrast highlights a gender gap in academic success.

4. Data Cleaning (if necessary, otherwise leave empty but justified)

According to the data source, lots of data cleaning and preperation was performed.
Also they removed any unexplainable outliers.

1. Incorrect values
   There seems to be no incorrect values

2. Missing values
   according to the initial analysis there are no missing values.

3. Justified feature reduction or type conversion
   There is no need to remove features.

4. Save the - if necessary - merged and integrated files from various sources into `korr.csv`
   Not necessary - No changes were made.

5. Data Preparation

1. Different pipelines per algorithm (SVM, DT, RF, kNN, Logistic Regression) including splitting into train-validation sets with stratification

In [9]:
```
# Split into train/validation with stratification
X_train, X_test, y_train, y_test = train_test_split(
    X, Y, test_size=0.3, random_state=my_random_state, stratify=Y)
```

In [10]:
```
# kNN
knn_pipeline = Pipeline(
    steps=[
        ('scaler', RobustScaler()),
        ('knn', KNeighborsClassifier())
    ]
)
```

In [11]:
```
# SVM
svm_pipeline = Pipeline(
    steps=[
        ("scaler", RobustScaler()),
        ("svm", SVC(random_state=my_random_state))
    ]
)
```

In [12]:
```
# RF
rf_pipeline = Pipeline(
    steps=[
        ('rf', RandomForestClassifier(random_state=my_random_state))
    ]
)
```

In [13]:
```
# LogisticRegression
logreg_pipeline = Pipeline(
    steps=[
        ("scaler", RobustScaler()),
        ("logreg", LogisticRegression(random_state=my_random_state))
```

```
    ]
)
```

In [14]:
```
# DT
dt_pipeline = Pipeline(
    steps=[
        ('dt', DecisionTreeClassifier(random_state=my_random_state))
    ]
)
```

6. Comparison of Classification Algorithms

1. Hyperparameter optimization using GridSearch per algorithm with `cv=5` + time < 15 min

kNN GridSearch

In [15]:
```
knn_param_grid = [{'knn__n_neighbors': range(1, 30),
                   'knn__weights': ['uniform', 'distance'],
                   'knn__metric': ['euclidean', 'manhattan'],
                   'knn__algorithm': ['ball_tree', 'kd_tree', 'brute'],
                   'knn__leaf_size': [1, 5, 10, 20, 30, 40, 50]}]

knn_grid_search = GridSearchCV(
    # using accuracy as scoring metric instead of f1_score since f1 didn't work
    estimator=knn_pipeline, param_grid=knn_param_grid, cv=5, scoring='accuracy', n_jobs=-1)

knn_grid_search.fit(X_train, y_train)

print(knn_grid_search.best_params_)
print(knn_grid_search.best_score_)
```

```
{'knn__algorithm': 'ball_tree', 'knn__leaf_size': 1, 'knn__metric': 'manhattan', 'knn__n_neighbors': 21, 'knn__weights': 'uniform'}
0.713825108134869
```

SVM GridSearch

In [16]:
```
from sklearn.experimental import enable_halving_search_cv
from sklearn.model_selection import HalvingGridSearchCV
# we use HalvingGridSearchCV for SVM to speed up the search process

svm_param_grid = {
    'svm__C': [0.001, 0.01, 0.1, 1, 10, 100, 1000],
    'svm__gamma': [1, 0.1, 0.01, 0.001, 0.0001, 0.00001],
    'svm__kernel': ['linear', 'rbf', 'poly', 'sigmoid'],
    'svm__degree': [1, 2, 3, 4],
    'svm__coef0': [0.0, 0.1, 0.5, 1]
}

svm_grid_search = HalvingGridSearchCV(svm_pipeline, svm_param_grid,
                                      cv=5, scoring='accuracy', n_jobs=-1) # enable multi-processing with n_jobs=-1

svm_grid_search.fit(X_train, y_train)

print(svm_grid_search.best_params_)
print(svm_grid_search.best_score_)
```

```
{'svm__C': 0.1, 'svm__coef0': 1, 'svm__degree': 3, 'svm__gamma': 0.01, 'svm__kernel': 'linear'}
0.7654592507742565
```

Random Forrest GridSearch

In [17]:
```
rf_param_grid = {
    'rf__n_estimators': [50, 75, 100, 150, 200],
    'rf__criterion': ['gini', 'entropy'],
    'rf__max_depth': range(1, 15),
    'rf__min_samples_leaf': range(1, 10),
    'rf__max_features': ['sqrt', 'log2'],
    'rf__bootstrap': [True, False]
}
rf_grid_search = GridSearchCV(rf_pipeline, rf_param_grid,
                              cv=5, scoring='accuracy', n_jobs=-1)

rf_grid_search.fit(X_train, y_train)

print(rf_grid_search.best_params_)
print(rf_grid_search.best_score_)
```

```
{'rf__bootstrap': False, 'rf__criterion': 'entropy', 'rf__max_depth': 11, 'rf__max_features': 'sqrt', 'rf__min_samples_leaf': 1, 'rf__n_estimat
ors': 150}
0.7706753869404346
```

Logistic Regression GridSearch

In [18]:
```
logreg_param_grid = [
    {
        "logreg__solver": ["liblinear"],
        "logreg__penalty": ["l1", "l2"],
        "logreg__fit_intercept": [True, False],
        "logreg__max_iter": [5000],
        "logreg__C": [0.01, 0.1, 1, 10, 100],
        "logreg__class_weight": [None, "balanced"]
    },
    {
        "logreg__solver": ["saga"],
        "logreg__penalty": ["l1", "l2"],
```

```python
        "logreg__fit_intercept": [True, False],
        "logreg__max_iter": [5000],
        "logreg__C": [0.01, 0.1, 1, 10, 100],
        "logreg__class_weight": [None, "balanced"]
    }
]

logreg_grid_search = GridSearchCV(
    logreg_pipeline, logreg_param_grid, cv=5, scoring='accuracy',  n_jobs=-1)

logreg_grid_search.fit(X_train, y_train)

print(logreg_grid_search.best_params_)
print(logreg_grid_search.best_score_)
```

```
{'logreg__C': 0.1, 'logreg__class_weight': None, 'logreg__fit_intercept': True, 'logreg__max_iter': 5000, 'logreg__penalty': 'l2', 'logreg__sol
ver': 'saga'}
0.7664729793110636
```

Decision Tree GridSearch

In [19]:
```python
dt_param_grid = [{'dt__criterion': ['gini', 'entropy'],
                  'dt__splitter': ['best', 'random'],
                  'dt__max_depth': range(1, 20),
                  'dt__min_samples_leaf': range(1, 20),
                  'dt__max_features': [None, 'sqrt', 'log2'],
                  'dt__min_samples_split': [2, 5, 10]}]

dt_grid_search = GridSearchCV(dt_pipeline, dt_param_grid,
                              cv=5, scoring='accuracy', n_jobs=-1)

dt_grid_search.fit(X_train, y_train)

print(dt_grid_search.best_params_)
print(dt_grid_search.best_score_)
```

```
{'dt__criterion': 'gini', 'dt__max_depth': 8, 'dt__max_features': None, 'dt__min_samples_leaf': 12, 'dt__min_samples_split': 2, 'dt__splitter':
'random'}
0.751942258585648
```

2. Discussion of the optimized final models

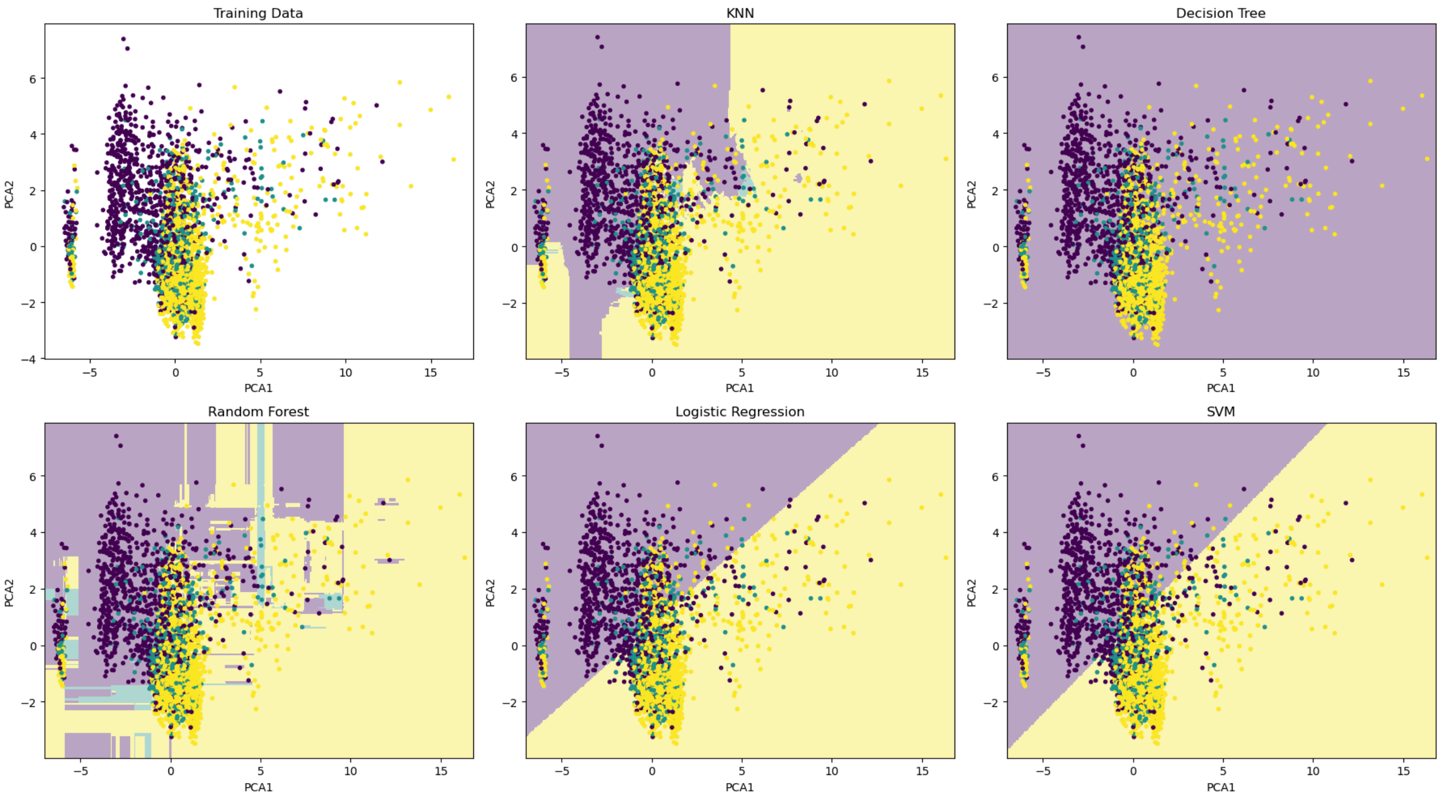
In [20]:
```python
models = {
    'KNN': knn_grid_search,
    'SVM': svm_grid_search,
    'RF': rf_grid_search,
    'Logistic Regression': logreg_grid_search,
    'Decision Tree': dt_grid_search
}
```

Decision surface

In [21]:
```python
# --- Plot decision surface without altering the model ---
def plot_decision_surface(model, X, Y, ax=None, title=None):
    h = .05
    x_min, x_max = X.iloc[:, 0].min() - .5, X.iloc[:, 0].max() + .5
    y_min, y_max = X.iloc[:, 1].min() - .5, X.iloc[:, 1].max() + .5
    xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
                         np.arange(y_min, y_max, h))

    # Convert mesh to DataFrame with same column names
    X_mesh = pd.DataFrame(np.c_[xx.ravel(), yy.ravel()], columns=X.columns)
    Z = model.predict(X_mesh).reshape(xx.shape)

    if ax is None:
        plt.figure()
        ax = plt.gca()
    ax.set_xlim(xx.min(), xx.max())
    ax.set_ylim(yy.min(), yy.max())
    ax.pcolormesh(xx, yy, Z, alpha=0.35, shading='auto')
    ax.scatter(X.iloc[:, 0], X.iloc[:, 1], c=Y, marker='.')

    ax.set_xlabel(X.columns[0])
    ax.set_ylabel(X.columns[1])
    ax.set_title(title if title else str(model))


# --- Encode target labels ---
le = LabelEncoder()
Y_encoded = le.fit_transform(y_train)

# --- Reduce features to 2D with PCA (once for plotting only) ---
pca_pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('pca', PCA(n_components=2))
])
X_2d_array = pca_pipeline.fit_transform(X_train)
X_2d = pd.DataFrame(X_2d_array, columns=['PCA1', 'PCA2'])

# --- Extract best parameter sets from GridSearchCV results (do NOT reuse models) ---
model_specs = [
    ("KNN", KNeighborsClassifier, knn_grid_search.best_params_),
    ("Decision Tree", DecisionTreeClassifier, dt_grid_search.best_params_),
    ("Random Forest", RandomForestClassifier, rf_grid_search.best_params_),
    ("Logistic Regression", LogisticRegression, logreg_grid_search.best_params_),
    ("SVM", SVC, svm_grid_search.best_params_)
]
```

```python
# --- Create cloned estimators for PCA-transformed plotting only ---
plotting_models = []
for name, cls, best_params in model_specs:
    # Strip prefixes like 'clf__' or 'dt__' from param grid keys
    clean_params = {k.split('__')[-1]: v for k, v in best_params.items()}
    estimator = cls(**clean_params)
    # Clone the PCA pipeline and append classifier
    plot_pipeline = Pipeline([
        ('clf', estimator)
    ])
    print(name, clean_params, estimator)
    # Fit on the original full dataset, but only for the PCA visualization
    plot_pipeline.fit(X_2d, Y_encoded)
    print(plot_pipeline)
    plotting_models.append((name, plot_pipeline))


# --- Plotting ---
fig, axes = plt.subplots(2, 3, figsize=(18, 10))
axes = axes.flatten()

# First plot: data
axes[0].scatter(X_2d.iloc[:, 0], X_2d.iloc[:, 1], c=Y_encoded, marker='.')
axes[0].set_title("Training Data")
axes[0].set_xlabel("PCA1")
axes[0].set_ylabel("PCA2")

# Remaining plots
for i, (name, model) in enumerate(plotting_models, start=1):
    # Reuse transformed PCA data for input; model includes PCA inside too
    plot_decision_surface(model, X_2d, Y_encoded, ax=axes[i], title=name)

# Hide unused axes
for j in range(len(plotting_models) + 1, len(axes)):
    axes[j].axis('off')

plt.tight_layout()
plt.show()
```

```
KNN {'algorithm': 'ball_tree', 'leaf_size': 1, 'metric': 'manhattan', 'n_neighbors': 21, 'weights': 'uniform'} KNeighborsClassifier(algorithm
='ball_tree', leaf_size=1, metric='manhattan',
                     n_neighbors=21)
Pipeline(steps=[('clf',
                 KNeighborsClassifier(algorithm='ball_tree', leaf_size=1,
                                      metric='manhattan', n_neighbors=21))])
Decision Tree {'criterion': 'gini', 'max_depth': 8, 'max_features': None, 'min_samples_leaf': 12, 'min_samples_split': 2, 'splitter': 'random'}
DecisionTreeClassifier(max_depth=8, min_samples_leaf=12, splitter='random')
Pipeline(steps=[('clf',
                 DecisionTreeClassifier(max_depth=8, min_samples_leaf=12,
                                        splitter='random'))])
Random Forest {'bootstrap': False, 'criterion': 'entropy', 'max_depth': 11, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'n_estimators': 150}
RandomForestClassifier(bootstrap=False, criterion='entropy', max_depth=11,
                       n_estimators=150)
Pipeline(steps=[('clf',
                 RandomForestClassifier(bootstrap=False, criterion='entropy',
                                        max_depth=11, n_estimators=150))])
Logistic Regression {'C': 0.1, 'class_weight': None, 'fit_intercept': True, 'max_iter': 5000, 'penalty': 'l2', 'solver': 'saga'} LogisticRegres
sion(C=0.1, max_iter=5000, solver='saga')
Pipeline(steps=[('clf',
                 LogisticRegression(C=0.1, max_iter=5000, solver='saga'))])
SVM {'C': 0.1, 'coef0': 1, 'degree': 3, 'gamma': 0.01, 'kernel': 'linear'} SVC(C=0.1, coef0=1, gamma=0.01, kernel='linear')
Pipeline(steps=[('clf', SVC(C=0.1, coef0=1, gamma=0.01, kernel='linear'))])
```



Performance comparison table on validation data (Accuracy, F1, Precision, Recall, MCC)

In [22]:
```python
# Metrics storage
results = []

# Evaluate each model
for name, model in models.items():
    y_pred = model.predict(X_test)

    results.append({
        "Model": name,
        "Accuracy": accuracy_score(y_test, y_pred),
        "F1": f1_score(y_test, y_pred, average='weighted'),
        "Precision": precision_score(y_test, y_pred, average='weighted'),
        "Recall": recall_score(y_test, y_pred, average='weighted'),
        "MCC": matthews_corrcoef(y_test, y_pred)
    })

# Convert to DataFrame
results_df = pd.DataFrame(results).set_index("Model")
results_df = results_df.round(4)

# Display table
results_df
```

Out[22]:

Model	Accuracy	F1	Precision	Recall	MCC
KNN	0.7078	0.6798	0.6932	0.7078	0.5130
SVM	0.7741	0.7641	0.7699	0.7741	0.6271
RF	0.7922	0.7787	0.7817	0.7922	0.6558
Logistic Regression	0.7613	0.7457	0.7440	0.7613	0.6028
Decision Tree	0.7274	0.7204	0.7200	0.7274	0.5488

Looking at the table, we can conclude that the **Random Forest (RF)** model performs the best overall.

Achieving the highest values across most evaluation metrics—**accuracy (0.7922), F1 score (0.7787), precision (0.7817), recall (0.7922), and MCC (0.6558)**. This suggests that RF is the most reliable and balanced model for this classification task, effectively managing both false positives and false negatives compared to the others.

The **K-Nearest Neighbors (KNN)** model performs the poorest overall.

It has the **lowest scores in all key metrics**—accuracy (0.7078), F1 score (0.6798), precision (0.6932), recall (0.7078), and especially **MCC (0.5130)**, which indicates weaker overall predictive power and less balanced performance compared to the other models.