In [1]:
```
# Data wrangling
import numpy as np
import pandas as pd
from tqdm import tqdm
import matplotlib.pyplot as plt
import seaborn as sns

# Data Visualization
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots

# Data pre-processing
from sklearn.preprocessing import StandardScaler

# Data splitting
from sklearn.model_selection import train_test_split

# Evaluation metrics
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from sklearn.metrics import roc_curve, auc
from sklearn.preprocessing import StandardScaler, OneHotEncoder
```

In [2]:
```
# Read dataset from csv file
data =  pd.read_csv(r"C:\Users\gabro\OneDrive\Desktop\NCI springboard\machine learn

# Correcting typographical error for nationality
data.rename(columns = {'Nacionality':'Nationality'}, inplace = True)
data.head()
```

Out[2]:

	Marital status	Application mode	Application order	Course	Daytime/evening attendance\t	Previous qualification	Previous qualification (grade)
0	1	17	5	171	1	1	122.0
1	1	15	1	9254	1	1	160.0
2	1	1	5	9070	1	1	122.0
3	1	17	2	9773	1	1	122.0
4	2	39	1	8014	0	1	100.0

5 rows × 37 columns

In [3]:
```
# shape of data
data.shape
```

Out[3]: (4424, 37)

In [4]:
```
data.tail()
```

Out[4]:

	Marital status	Application mode	Application order	Course	Daytime/evening attendance\t	Previous qualification	Previo qualificati (grac
4419	1	1	6	9773	1	1	12
4420	1	1	2	9773	1	1	12
4421	1	1	1	9500	1	1	15
4422	1	1	1	9147	1	1	18
4423	1	10	1	9773	1	1	15

5 rows × 37 columns

In [5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 37 columns):
 #   Column                                          Non-Null Count  Dtype  
---  ------                                          --------------  -----  
 0   Marital status                                  4424 non-null   int64  
 1   Application mode                                4424 non-null   int64  
 2   Application order                               4424 non-null   int64  
 3   Course                                          4424 non-null   int64  
 4   Daytime/evening attendance                      4424 non-null   int64  
 5   Previous qualification                          4424 non-null   int64  
 6   Previous qualification (grade)                  4424 non-null   float64
 7   Nationality                                     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
```

In [6]:
```
data.nunique()
```

Out[6]:
```
Marital status                                     6
Application mode                                  18
Application order                                  8
Course                                            17
Daytime/evening attendance\t                       2
Previous qualification                            17
Previous qualification (grade)                   101
Nationality                                       21
Mother's qualification                            29
Father's qualification                            34
Mother's occupation                               32
Father's occupation                               46
Admission grade                                  620
Displaced                                          2
Educational special needs                          2
Debtor                                             2
Tuition fees up to date                            2
Gender                                             2
Scholarship holder                                 2
Age at enrollment                                 46
International                                      2
Curricular units 1st sem (credited)               21
Curricular units 1st sem (enrolled)               23
Curricular units 1st sem (evaluations)            35
Curricular units 1st sem (approved)               23
Curricular units 1st sem (grade)                 797
Curricular units 1st sem (without evaluations)    11
Curricular units 2nd sem (credited)               19
Curricular units 2nd sem (enrolled)               22
Curricular units 2nd sem (evaluations)            30
Curricular units 2nd sem (approved)               20
Curricular units 2nd sem (grade)                 782
Curricular units 2nd sem (without evaluations)    10
Unemployment rate                                 10
Inflation rate                                     9
GDP                                               10
Target                                             3
dtype: int64
```

In [7]:
```
data.isnull().sum()
```

Out[7]:
```
Marital status                                  0
Application mode                                0
Application order                               0
Course                                          0
Daytime/evening attendance\t                    0
Previous qualification                          0
Previous qualification (grade)                  0
Nationality                                     0
Mother's qualification                          0
Father's qualification                          0
Mother's occupation                             0
Father's occupation                             0
Admission grade                                 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
```

In [8]:
```
# Remove duplicates
data = data.drop_duplicates()
```

In [9]:
```
#Checking for null values
data.isnull().any().sum()
```

Out[9]:
```
0
```

Data Reduction

In [11]:
```
threshold = 0.5  # Example threshold
data = data.loc[:, data.isnull().mean() < threshold]
```

```
# Display dataset shape after reduction
data.shape
```

Out[11]: (4424, 37)

Exploratory Data Analysis

In [13]:
```
data.describe(include='all').T
```

Out[13]:

	count	unique	top	freq	mean	std	min	25%
Marital status	4424.0	NaN	NaN	NaN	1.178571	0.605747	1.0	1.0
Application mode	4424.0	NaN	NaN	NaN	18.669078	17.484682	1.0	1.0
Application order	4424.0	NaN	NaN	NaN	1.727848	1.313793	0.0	1.0
Course	4424.0	NaN	NaN	NaN	8856.642631	2063.566416	33.0	9085.0
Daytime/evening attendance\t	4424.0	NaN	NaN	NaN	0.890823	0.311897	0.0	1.0
Previous qualification	4424.0	NaN	NaN	NaN	4.577758	10.216592	1.0	1.0
Previous qualification (grade)	4424.0	NaN	NaN	NaN	132.613314	13.188332	95.0	125.0
Nationality	4424.0	NaN	NaN	NaN	1.873192	6.914514	1.0	1.0
Mother's qualification	4424.0	NaN	NaN	NaN	19.561935	15.603186	1.0	2.0
Father's qualification	4424.0	NaN	NaN	NaN	22.275316	15.343108	1.0	3.0
Mother's occupation	4424.0	NaN	NaN	NaN	10.960895	26.418253	0.0	4.0
Father's occupation	4424.0	NaN	NaN	NaN	11.032324	25.26304	0.0	4.0
Admission grade	4424.0	NaN	NaN	NaN	126.978119	14.482001	95.0	117.9
Displaced	4424.0	NaN	NaN	NaN	0.548373	0.497711	0.0	0.0
Educational special needs	4424.0	NaN	NaN	NaN	0.011528	0.10676	0.0	0.0
Debtor	4424.0	NaN	NaN	NaN	0.113698	0.31748	0.0	0.0
Tuition fees up to date	4424.0	NaN	NaN	NaN	0.880651	0.324235	0.0	1.0
Gender	4424.0	NaN	NaN	NaN	0.351718	0.47756	0.0	0.0
Scholarship holder	4424.0	NaN	NaN	NaN	0.248418	0.432144	0.0	0.0
Age at enrollment	4424.0	NaN	NaN	NaN	23.265145	7.587816	17.0	19.0
International	4424.0	NaN	NaN	NaN	0.024864	0.155729	0.0	0.0

	count	unique	top	freq	mean	std	min	25%
Curricular units 1st sem (credited)	4424.0	NaN	NaN	NaN	0.709991	2.360507	0.0	0.0
Curricular units 1st sem (enrolled)	4424.0	NaN	NaN	NaN	6.27057	2.480178	0.0	5.0
Curricular units 1st sem (evaluations)	4424.0	NaN	NaN	NaN	8.299051	4.179106	0.0	6.0
Curricular units 1st sem (approved)	4424.0	NaN	NaN	NaN	4.7066	3.094238	0.0	3.0
Curricular units 1st sem (grade)	4424.0	NaN	NaN	NaN	10.640822	4.843663	0.0	11.0
Curricular units 1st sem (without evaluations)	4424.0	NaN	NaN	NaN	0.137658	0.69088	0.0	0.0
Curricular units 2nd sem (credited)	4424.0	NaN	NaN	NaN	0.541817	1.918546	0.0	0.0
Curricular units 2nd sem (enrolled)	4424.0	NaN	NaN	NaN	6.232143	2.195951	0.0	5.0
Curricular units 2nd sem (evaluations)	4424.0	NaN	NaN	NaN	8.063291	3.947951	0.0	6.0
Curricular units 2nd sem (approved)	4424.0	NaN	NaN	NaN	4.435805	3.014764	0.0	2.0
Curricular units 2nd sem (grade)	4424.0	NaN	NaN	NaN	10.230206	5.210808	0.0	10.75
Curricular units 2nd sem (without evaluations)	4424.0	NaN	NaN	NaN	0.150316	0.753774	0.0	0.0
Unemployment rate	4424.0	NaN	NaN	NaN	11.566139	2.66385	7.6	9.4
Inflation rate	4424.0	NaN	NaN	NaN	1.228029	1.382711	-0.8	0.3
GDP	4424.0	NaN	NaN	NaN	0.001969	2.269935	-4.06	-1.7
Target	4424	3	Graduate	2209	NaN	NaN	NaN	NaN

Before we do EDA, lets separate Numerical and categorical variables for easy analysis

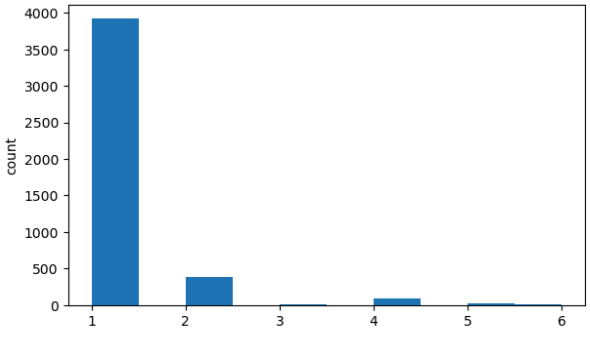
In [15]:
```
cat_cols=data.select_dtypes(include=['object']).columns
num_cols = data.select_dtypes(include=np.number).columns.tolist()
print("Categorical Variables:")
print(cat_cols)
print("Numerical Variables:")
print(num_cols)
```

```
Categorical Variables:
Index(['Target'], dtype='object')
Numerical Variables:
['Marital status', 'Application mode', 'Application order', 'Course', 'Daytime/evening attendance\t', 'Previous qualification', 'Previous qualification (grade)', 'Nationality', "Mother's qualification", "Father's qualification", "Mother's occupation", "Father's occupation", 'Admission grade', 'Displaced', 'Educational special needs', 'Debtor', 'Tuition fees up to date', 'Gender', 'Scholarship holder', 'Age at enrollment', 'International', 'Curricular units 1st sem (credited)', 'Curricular units 1st sem (enrolled)', 'Curricular units 1st sem (evaluations)', 'Curricular units 1st sem (approved)', 'Curricular units 1st sem (grade)', 'Curricular units 1st sem (without evaluations)', 'Curricular units 2nd sem (credited)', 'Curricular units 2nd sem (enrolled)', 'Curricular units 2nd sem (evaluations)', 'Curricular units 2nd sem (approved)', 'Curricular units 2nd sem (grade)', 'Curricular units 2nd sem (without evaluations)', 'Unemployment rate', 'Inflation rate', 'GDP']
```

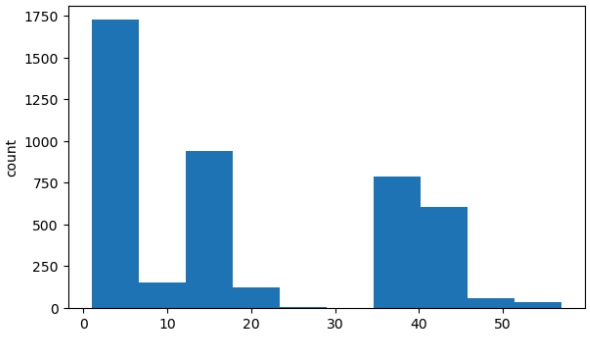
EDA Univariate Analysis using Matplotlib and Seaborn libraries, Seaborn is also a python library built on top of Matplotlib that uses short lines of code to create and style statistical plots from Pandas and Numpy. n the below fig, a histogram and box plot is used to show the pattern of the variables, as some variables have skewness and outliers.

In [17]:
```
for col in num_cols:
    print(col)
    print('Skew :', round(data[col].skew(), 2))
    plt.figure(figsize = (15, 4))
    plt.subplot(1, 2, 1)
    data[col].hist(grid=False)
    plt.ylabel('count')
    plt.subplot(1, 2, 2)
    sns.boxplot(x=data[col])
    plt.show()
```

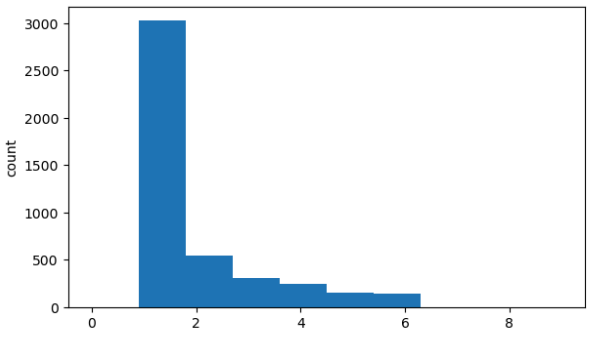
```
Marital status
Skew : 4.4
```



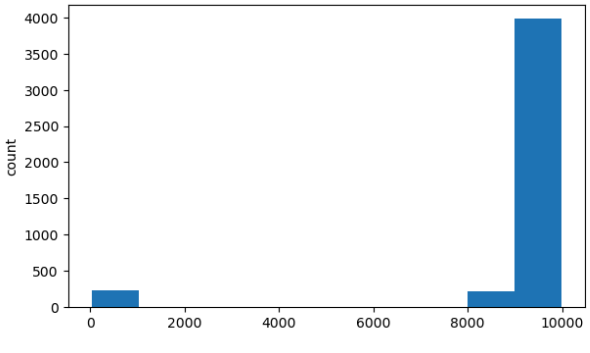
Application mode
Skew : 0.39



Application order
Skew : 1.88

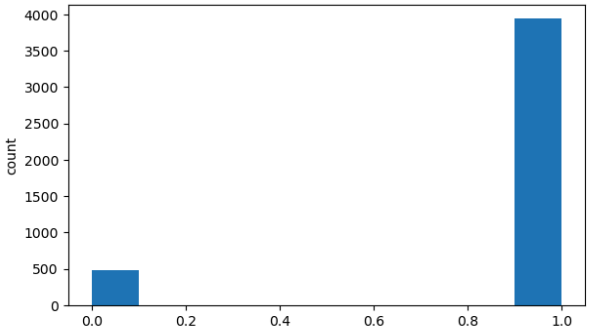


Course
Skew : -3.81

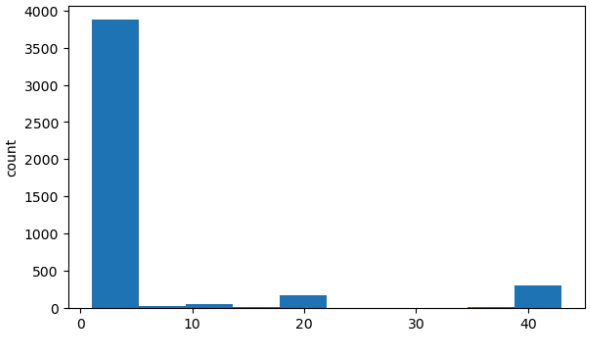


Daytime/evening attendance
Skew : -2.51

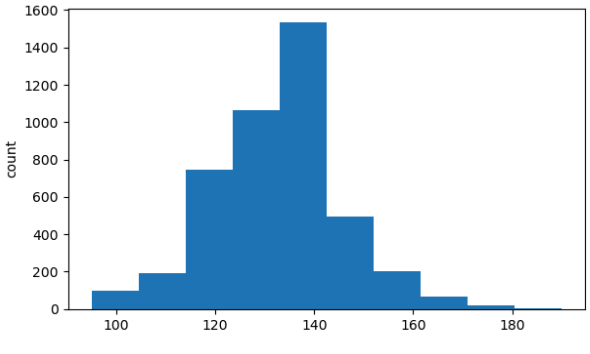
```
C:\Users\gabro\anaconda3\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarni
ng: Glyph 9 (   ) missing from current font.
  fig.canvas.print_figure(bytes_io, **kw)
```



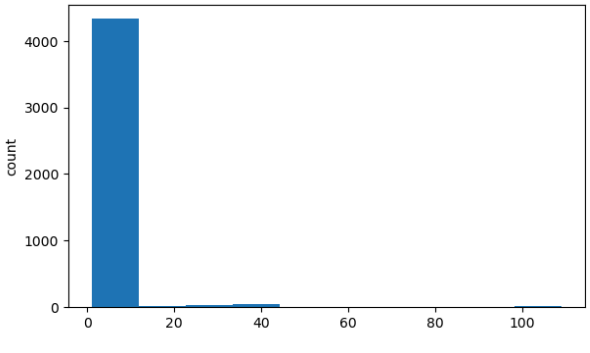
Previous qualification
Skew : 2.87



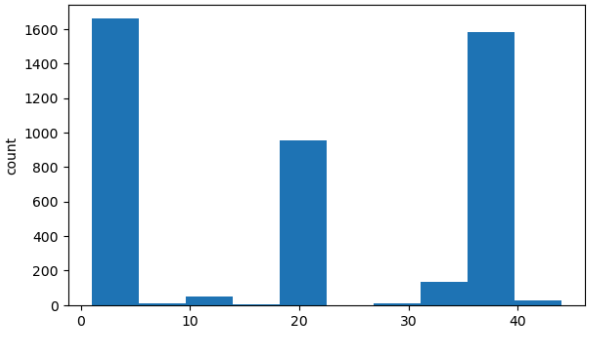
Previous qualification (grade)
Skew : 0.31

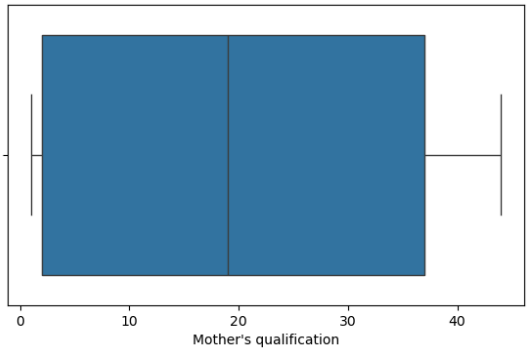


Nationality
Skew : 10.7

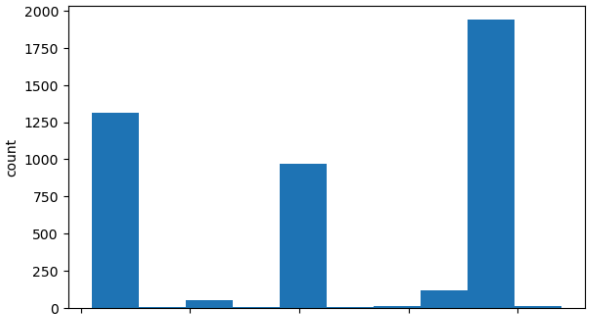


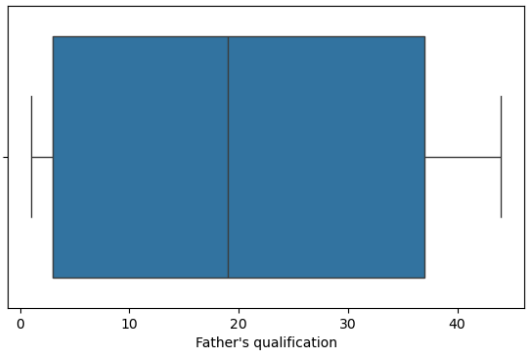
Mother's qualification
Skew : 0.0



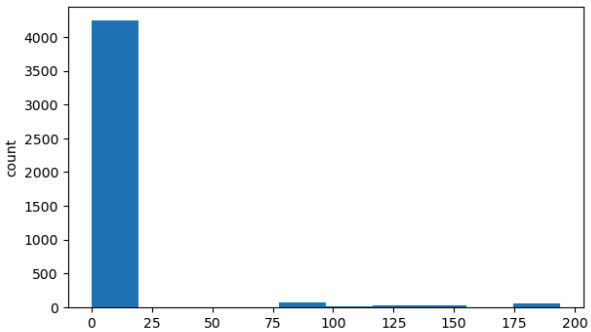


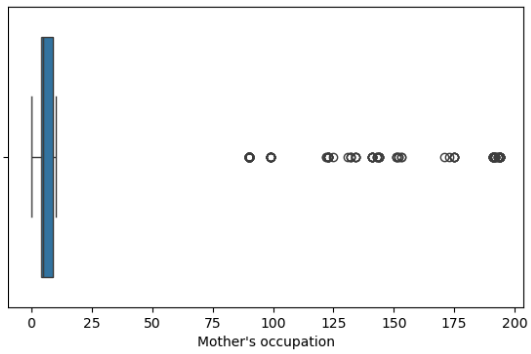
Father's qualification
Skew : -0.3



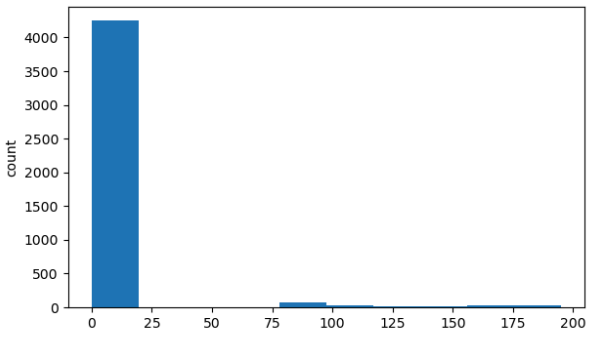


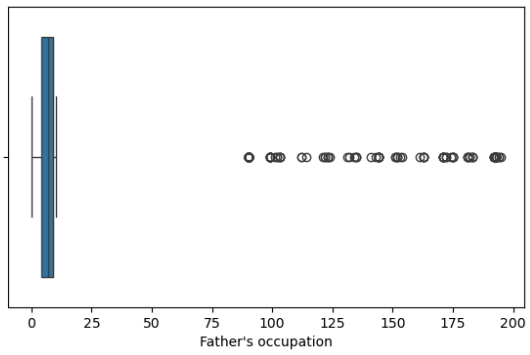
Mother's occupation
Skew : 5.34



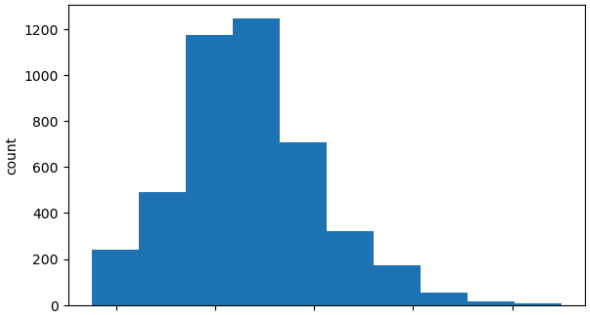


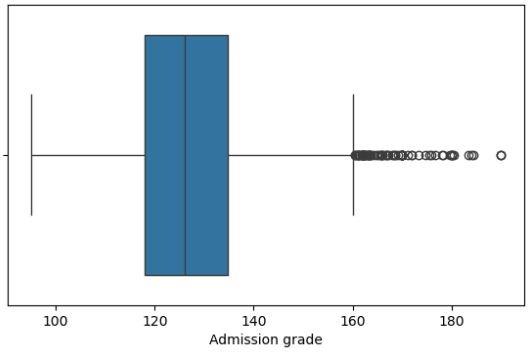
Father's occupation
Skew : 5.4





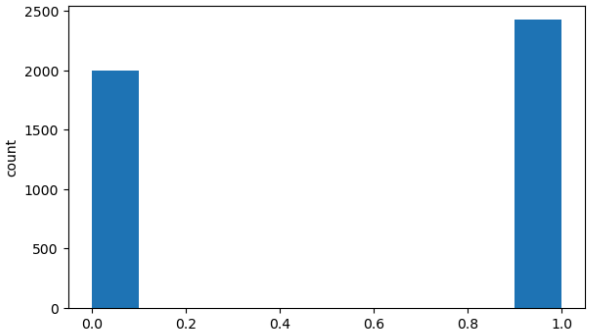
Admission grade
Skew : 0.53

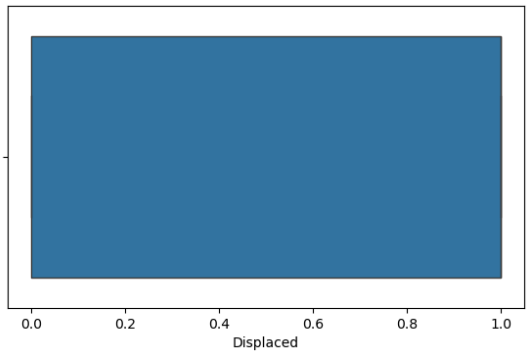




Displaced

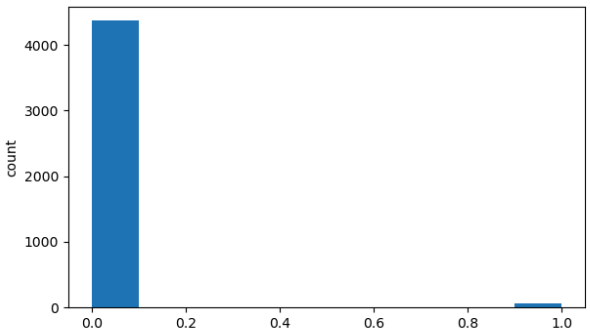
Skew : -0.19

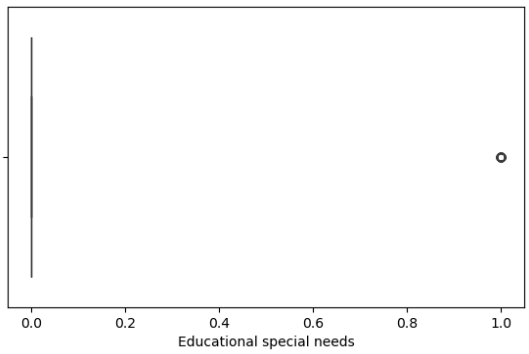




Educational special needs

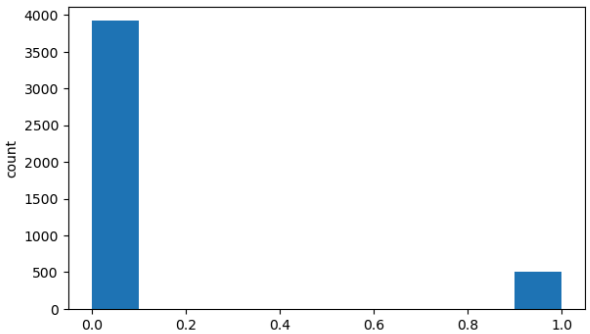
Skew : 9.15

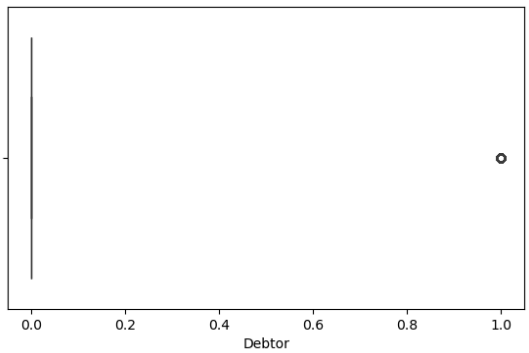




Debtor

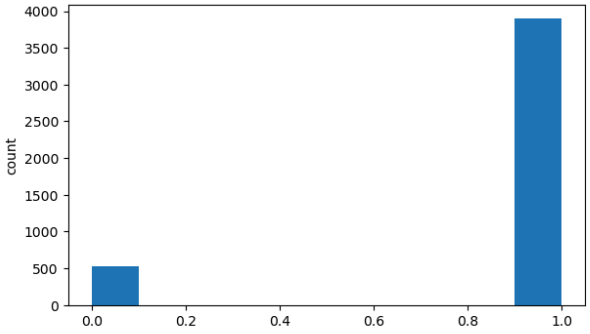
Skew : 2.43

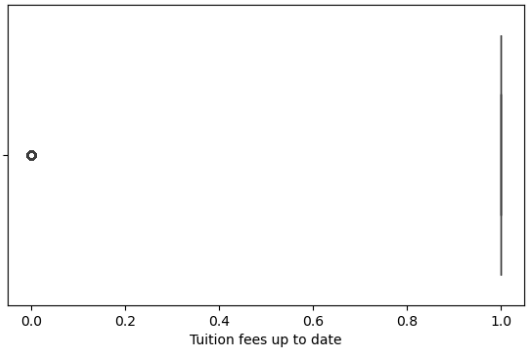




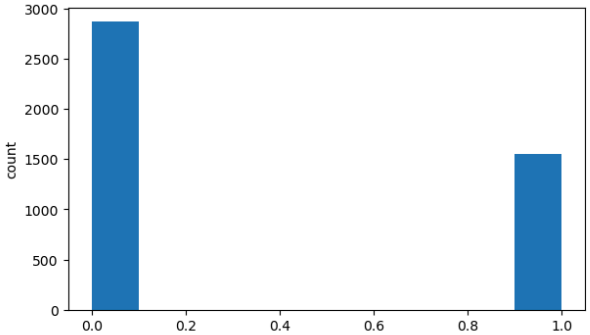
Tuition fees up to date

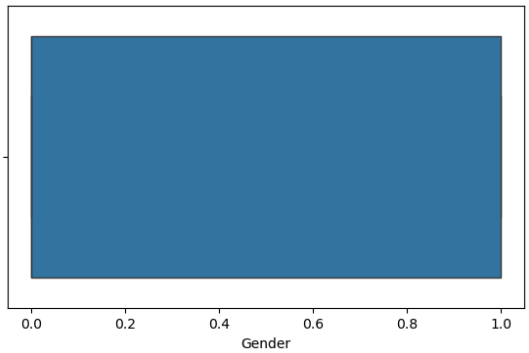
Skew : -2.35



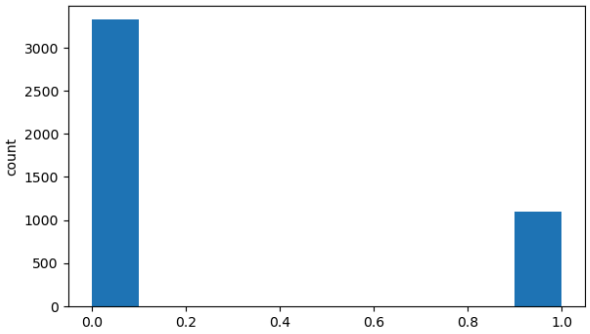


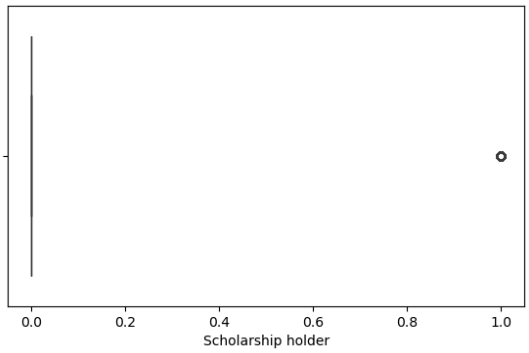
Gender
Skew : 0.62



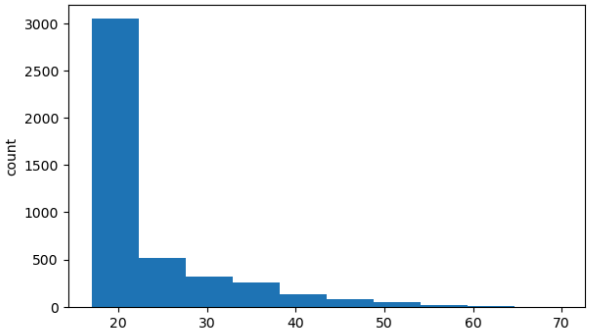


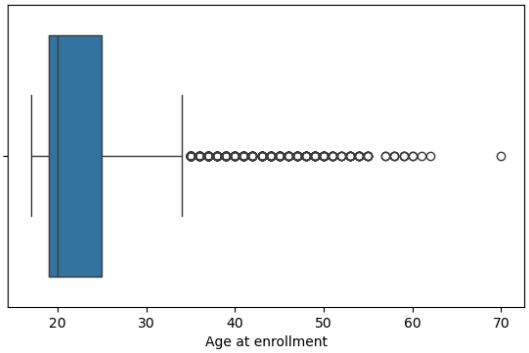
Scholarship holder
Skew : 1.16



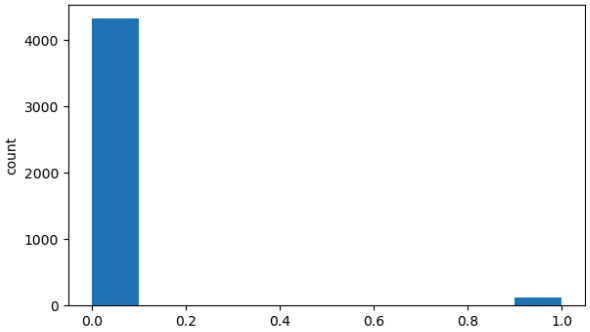


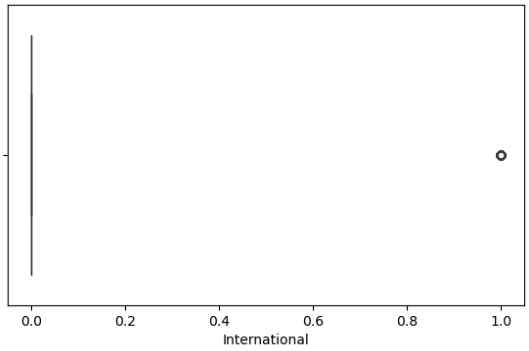
Age at enrollment
Skew : 2.05



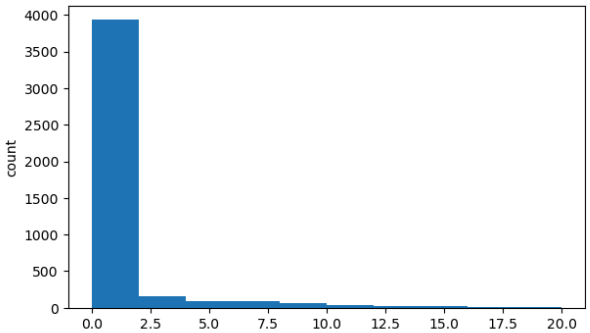


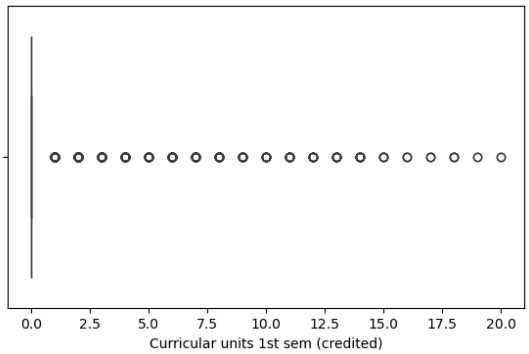
International
Skew : 6.1



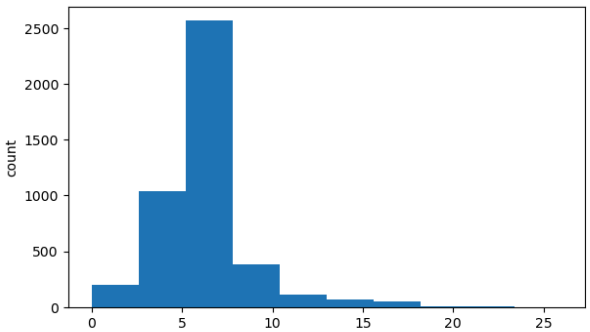


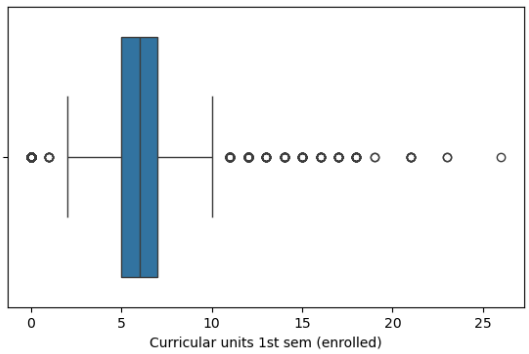
Curricular units 1st sem (credited)
Skew : 4.17



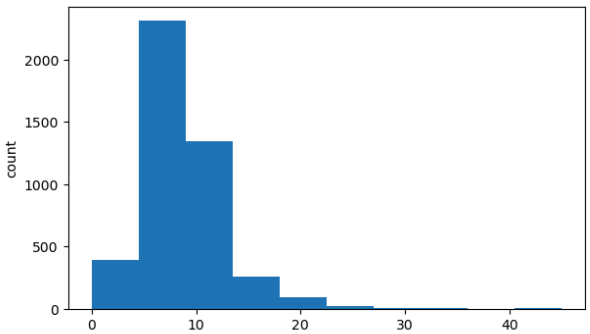


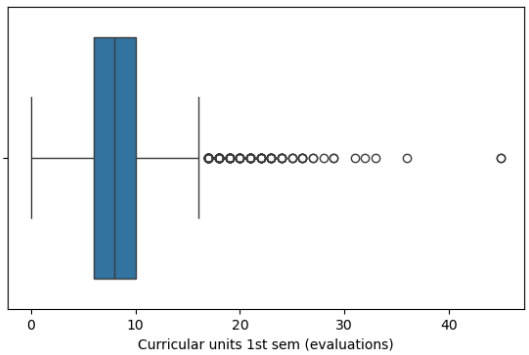
Curricular units 1st sem (enrolled)
Skew : 1.62



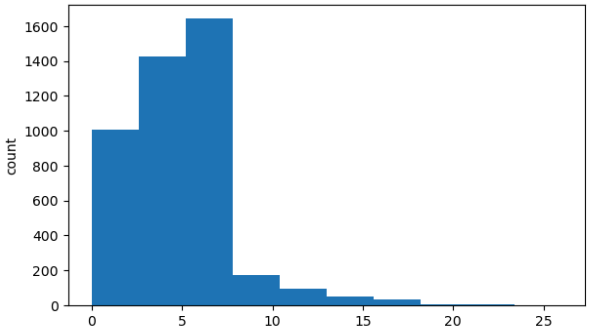


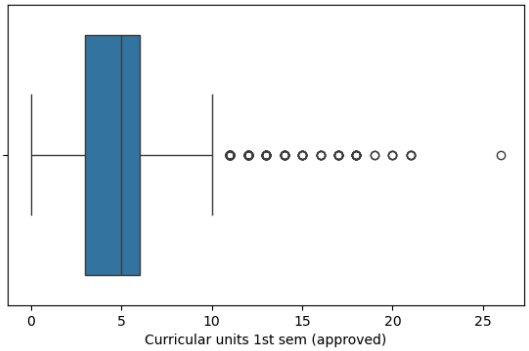
Curricular units 1st sem (evaluations)
Skew : 0.98



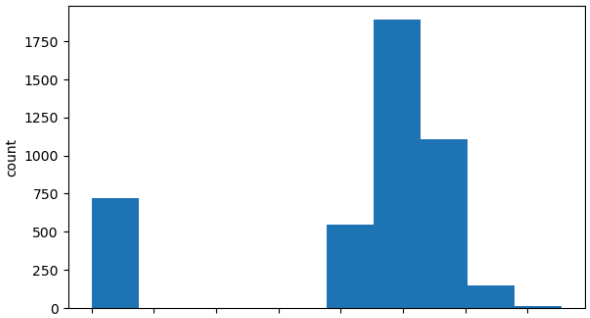


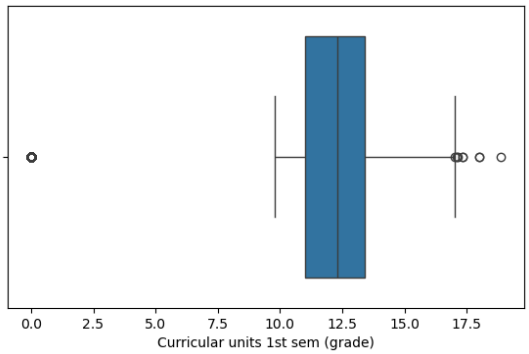
Curricular units 1st sem (approved)
Skew : 0.77



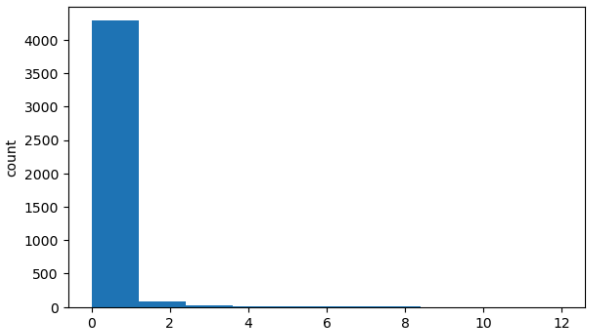


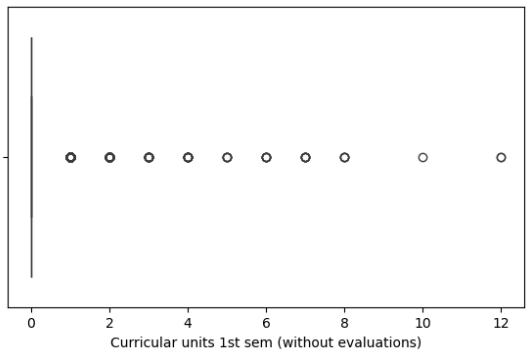
Curricular units 1st sem (grade)
Skew : -1.57



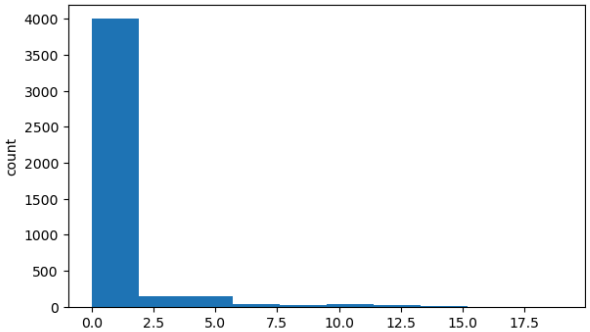


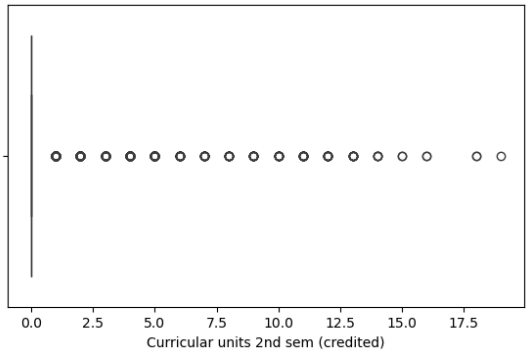
Curricular units 1st sem (without evaluations)
Skew : 8.21



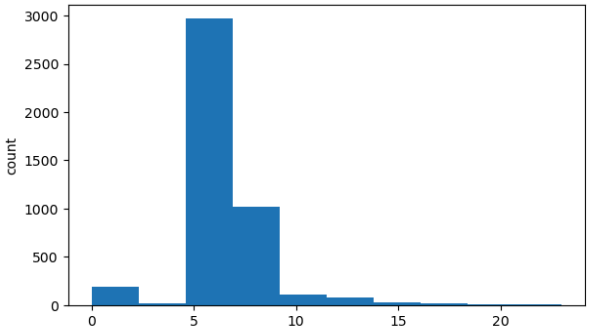


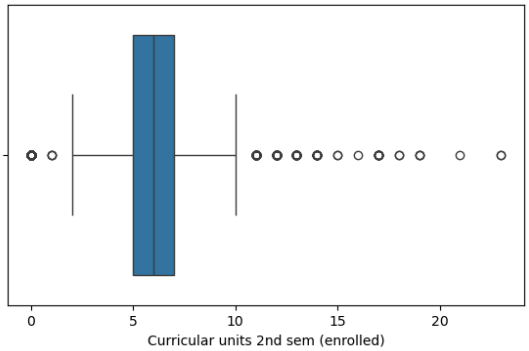
Curricular units 2nd sem (credited)
Skew : 4.63



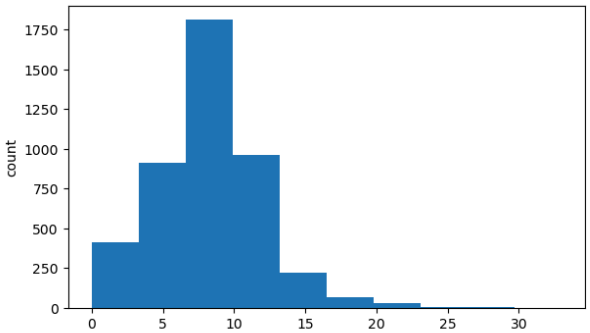


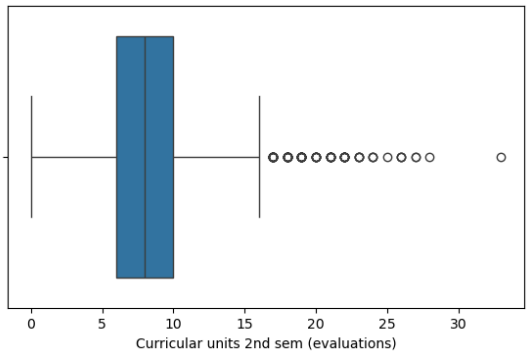
Curricular units 2nd sem (enrolled)
Skew : 0.79



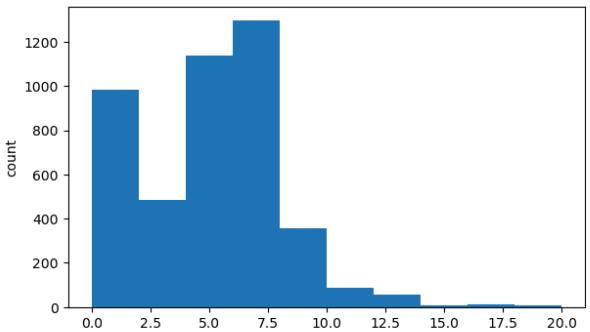


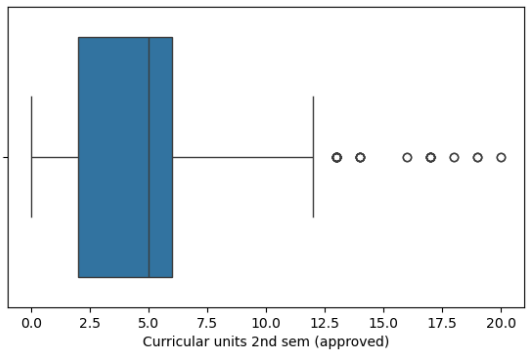
Curricular units 2nd sem (evaluations)
Skew : 0.34



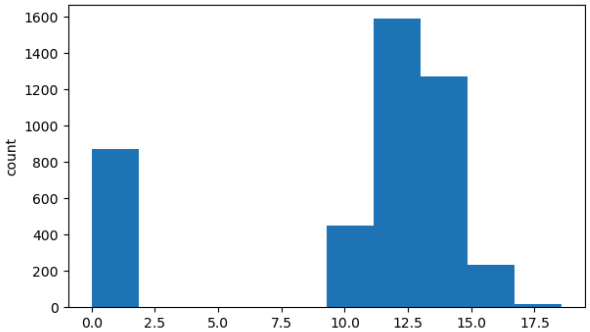


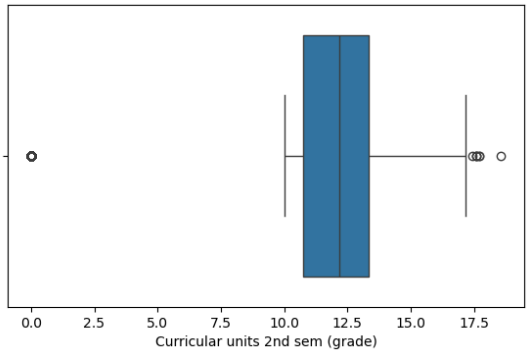
Curricular units 2nd sem (approved)
Skew : 0.31



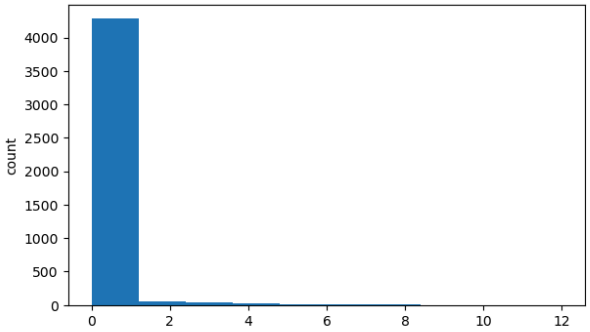


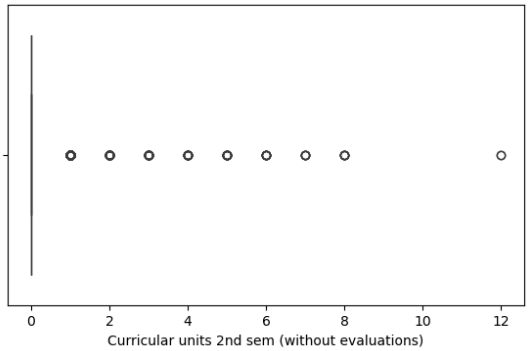
Curricular units 2nd sem (grade)
Skew : -1.31



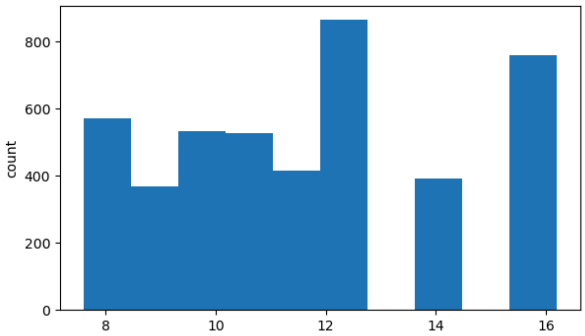


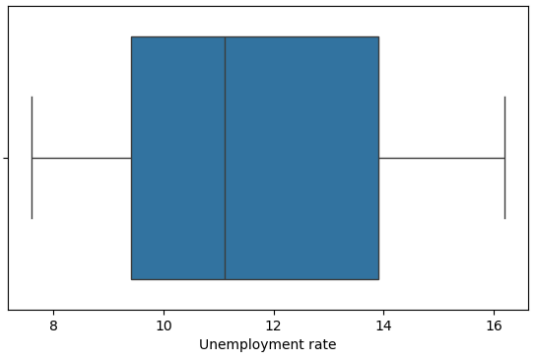
Curricular units 2nd sem (without evaluations)
Skew : 7.27



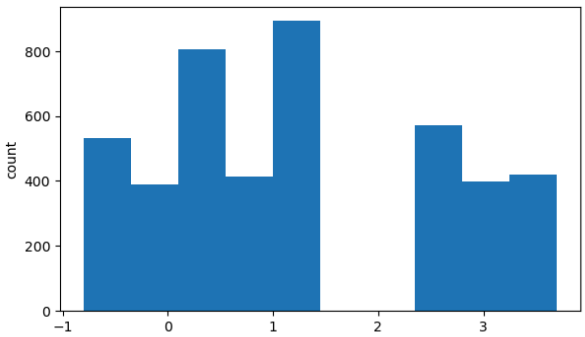


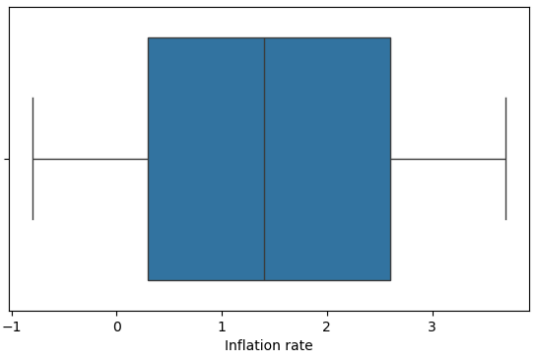
Unemployment rate
Skew : 0.21



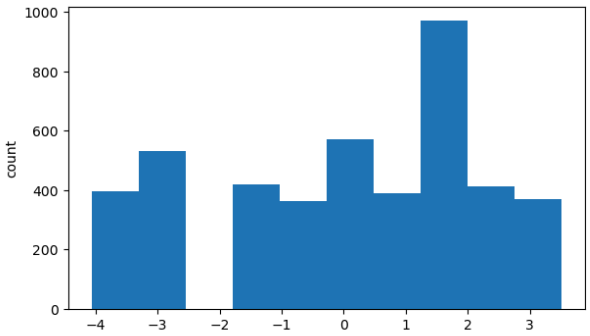


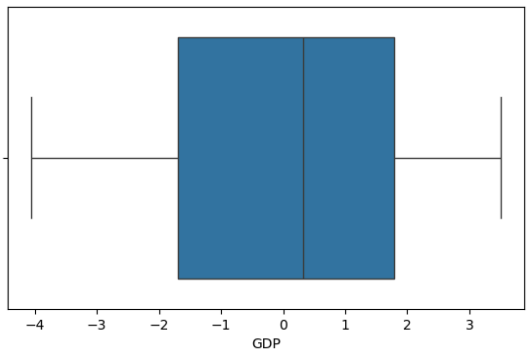
Inflation rate
Skew : 0.25





GDP
Skew : -0.39





In [18]:
```
# Identify numerical columns for outlier analysis
numerical_columns = data.select_dtypes(include=[np.number]).columns

# Calculate summary statistics
summary_stats = data[numerical_columns].describe()
```

```
# Calculate the IQR for each numerical column
Q1 = data[numerical_columns].quantile(0.25)
Q3 = data[numerical_columns].quantile(0.75)
IQR = Q3 - Q1

# Identify outliers using the IQR method
outliers = ((data[numerical_columns] < (Q1 - 1.5 * IQR)) | (data[numerical_columns]

# Display summary statistics and outliers
summary_stats, outliers
```

Out[18]: (

	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\t	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)	Nationality	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

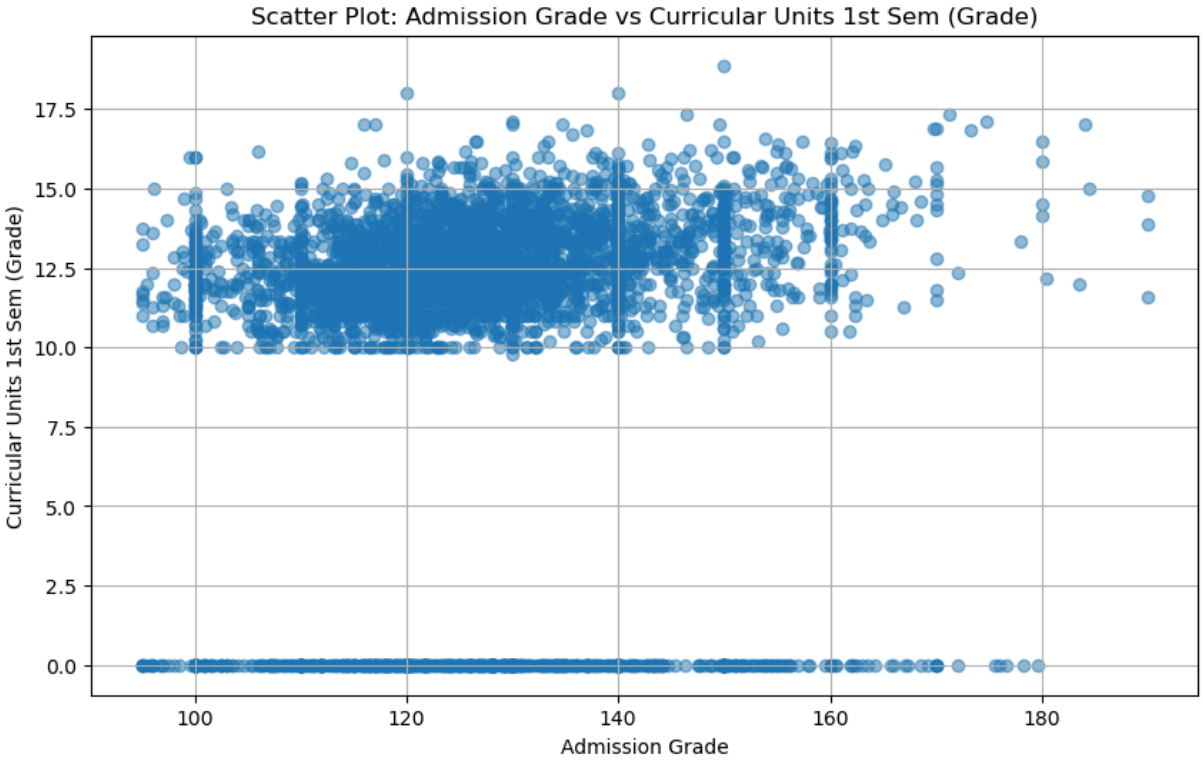
```
50%                                0.000000
75%                                0.000000
max                               19.000000

       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],
Marital status                                505
```

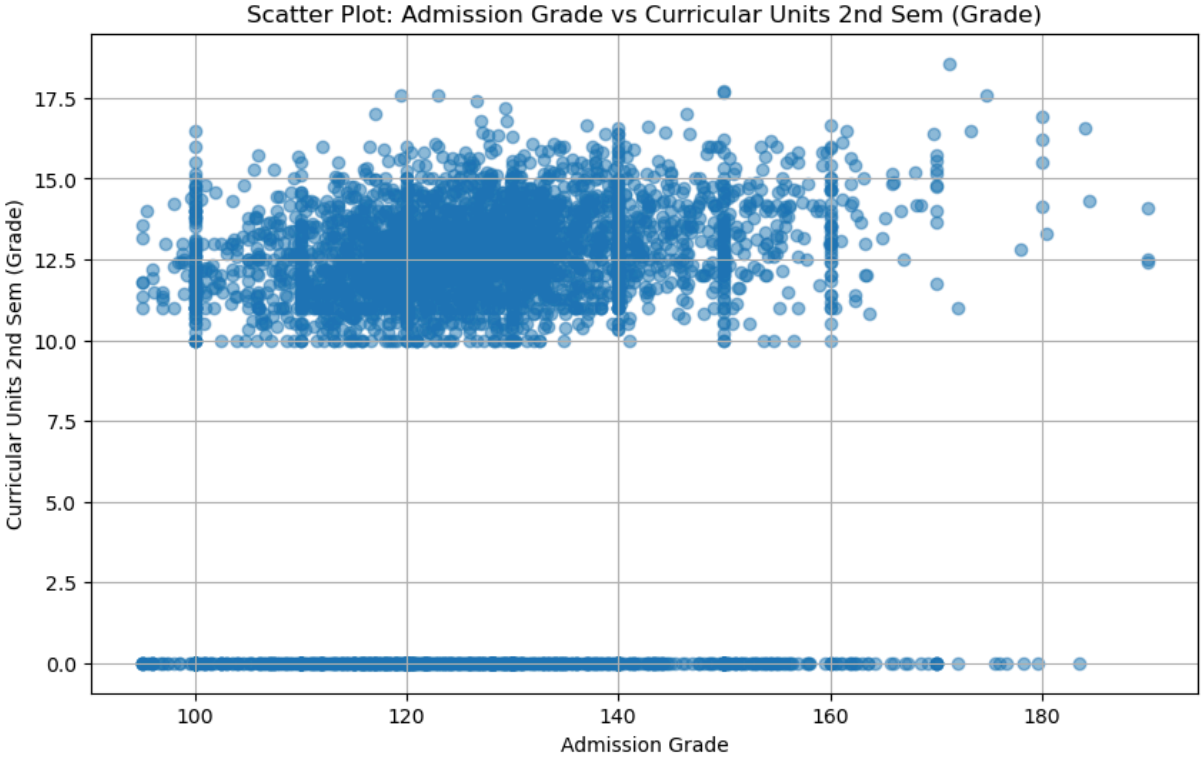
```
Application mode                                  0
Application order                               541
Course                                          442
Daytime/evening attendance\t                    483
Previous qualification                          707
Previous qualification (grade)                  179
Nationality                                     110
Mother's qualification                            0
Father's qualification                            0
Mother's occupation                             182
Father's occupation                             177
Admission grade                                  86
Displaced                                         0
Educational special needs                        51
Debtor                                          503
Tuition fees up to date                         528
Gender                                            0
Scholarship holder                             1099
Age at enrollment                               441
International                                   110
Curricular units 1st sem (credited)             577
Curricular units 1st sem (enrolled)             424
Curricular units 1st sem (evaluations)          158
Curricular units 1st sem (approved)             180
Curricular units 1st sem (grade)                726
Curricular units 1st sem (without evaluations)  294
Curricular units 2nd sem (credited)             530
Curricular units 2nd sem (enrolled)             369
Curricular units 2nd sem (evaluations)          109
Curricular units 2nd sem (approved)              44
Curricular units 2nd sem (grade)                877
Curricular units 2nd sem (without evaluations)  282
Unemployment rate                                 0
Inflation rate                                    0
GDP                                               0
dtype: int64)
```

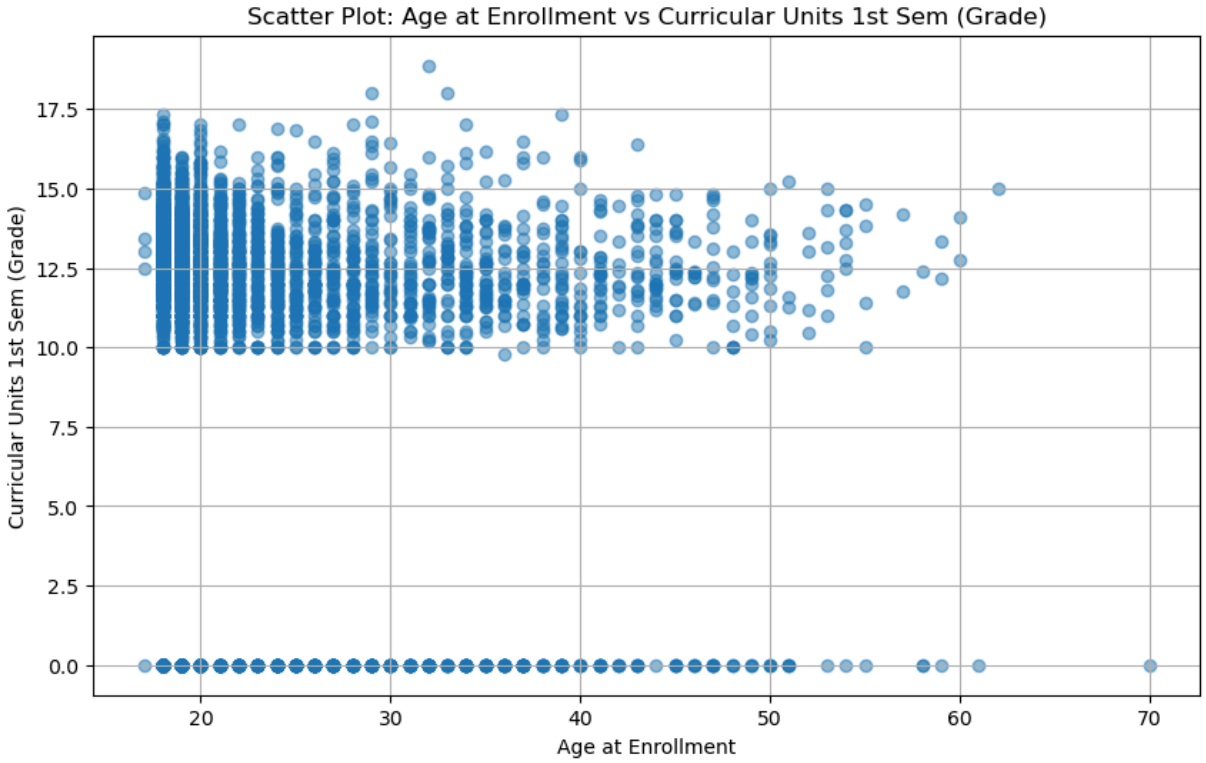
In [19]:

```python
# Scatter plot for Admission grade vs Curricular units 1st sem (grade)
plt.figure(figsize=(10, 6))
plt.scatter(data['Admission grade'], data['Curricular units 1st sem (grade)'], alph
plt.title('Scatter Plot: Admission Grade vs Curricular Units 1st Sem (Grade)')
plt.xlabel('Admission Grade')
plt.ylabel('Curricular Units 1st Sem (Grade)')
plt.grid(True)
plt.show()

# Scatter plot for Admission grade vs Curricular units 2nd sem (grade)
plt.figure(figsize=(10, 6))
plt.scatter(data['Admission grade'], data['Curricular units 2nd sem (grade)'], alph
plt.title('Scatter Plot: Admission Grade vs Curricular Units 2nd Sem (Grade)')
plt.xlabel('Admission Grade')
plt.ylabel('Curricular Units 2nd Sem (Grade)')
plt.grid(True)
plt.show()

# Scatter plot for Age at enrollment vs Curricular units 1st sem (grade)
```

```
plt.figure(figsize=(10, 6))
plt.scatter(data['Age at enrollment'], data['Curricular units 1st sem (grade)'], al
plt.title('Scatter Plot: Age at Enrollment vs Curricular Units 1st Sem (Grade)')
plt.xlabel('Age at Enrollment')
plt.ylabel('Curricular Units 1st Sem (Grade)')
plt.grid(True)
plt.show()
```







In [20]:
```
# Create a copy for exploratory data analysis
data_viz = data.copy()
print(data_viz.columns)
```

```
Index(['Marital status', 'Application mode', 'Application order', 'Course',
       'Daytime/evening attendance\t', 'Previous qualification',
       'Previous qualification (grade)', 'Nationality',
       'Mother's qualification', 'Father's qualification',
       'Mother's occupation', 'Father's occupation', 'Admission grade',
       'Displaced', 'Educational special needs', 'Debtor',
       'Tuition fees up to date', 'Gender', 'Scholarship holder',
       'Age at enrollment', 'International',
       'Curricular units 1st sem (credited)',
       'Curricular units 1st sem (enrolled)',
       'Curricular units 1st sem (evaluations)',
       'Curricular units 1st sem (approved)',
       'Curricular units 1st sem (grade)',
       'Curricular units 1st sem (without evaluations)',
       'Curricular units 2nd sem (credited)',
       'Curricular units 2nd sem (enrolled)',
       'Curricular units 2nd sem (evaluations)',
       'Curricular units 2nd sem (approved)',
       'Curricular units 2nd sem (grade)',
       'Curricular units 2nd sem (without evaluations)', 'Unemployment rate',
       'Inflation rate', 'GDP', 'Target'],
      dtype='object')
```

In [21]:
```
data['Target'].unique()
```

Out[21]:
```
array(['Dropout', 'Graduate', 'Enrolled'], dtype=object)
```

In [ ]:
```
# Distribution of Target feature
fig = px.pie(values= data_viz['Target'].value_counts(),
             names= data_viz['Target'].value_counts().index.to_list())

fig.update_traces(textposition='inside', textinfo='percent+label',
                  marker=dict(colors=['teal', 'goldenrod', 'slateblue']))

fig.update_layout(showlegend = False, height=400, width=800,
                  title='Distribution of Target')
fig.write_image('fig.svg', engine='kaleido')
fig.show('svg')
```

In [ ]:

In [ ]:
```
# Plotting the histogram
fig = px.histogram(data_viz, x='Age at enrollment', color='Target',
                   opacity=0.75, barmode='overlay',
                   width=800, height=500, color_discrete_sequence=px.colors.qualita

fig.update_layout(title='Age distribution of students')

# Showing the plot
fig.show()
```

In [ ]:

In [ ]:

In [ ]: