rk1

April 23, 2025

5-65 1 14

[1]:
```
#
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[3]:
```
#
df = pd.read_csv('Admission_Predict_Ver1.1.csv')
```

[4]:
```
#
print("                :")
print(df.head())
```

```
                :
   Serial No.  GRE Score  TOEFL Score  University Rating  SOP  LOR   CGPA  \
0           1        337          118                  4  4.5   4.5  9.65
1           2        324          107                  4  4.0   4.5  8.87
2           3        316          104                  3  3.0   3.5  8.00
3           4        322          110                  3  3.5   2.5  8.67
4           5        314          103                  2  2.0   3.0  8.21

   Research  Chance of Admit
0         1             0.92
1         1             0.76
2         1             0.72
3         1             0.80
4         0             0.65
```

[5]:
```
#              (               'LOR ')
df.rename(columns={'LOR ': 'LOR'}, inplace=True)

#
print("\n                  :")
print(df.isnull().sum())
```

                :

```
Serial No.           0
GRE Score            0
TOEFL Score          0
University Rating    0
SOP                  0
LOR                  0
CGPA                 0
Research             0
Chance of Admit      0
dtype: int64
```

[9]:
```
#      7%
np.random.seed(100)
miss_rate = 0.07

#
miss_idx_gre = df.sample(frac=miss_rate).index
miss_idx_un_r = df.sample(frac=miss_rate).index

df.loc[miss_idx_gre, 'GRE Score'] = np.nan  #
df.loc[miss_idx_un_r, 'University Rating'] = np.nan  #

#
print("\n                :")
print(df.isnull().sum())
```

```
              :
Serial No.            0
GRE Score            36
TOEFL Score           0
University Rating    36
SOP                   0
LOR                   0
CGPA                  0
Research              0
Chance of Admit       0
dtype: int64
```

1. (GRE Score)

[13]:
```
#
df['GRE Score'] = df['GRE Score'].fillna(df['GRE Score'].mean())
```

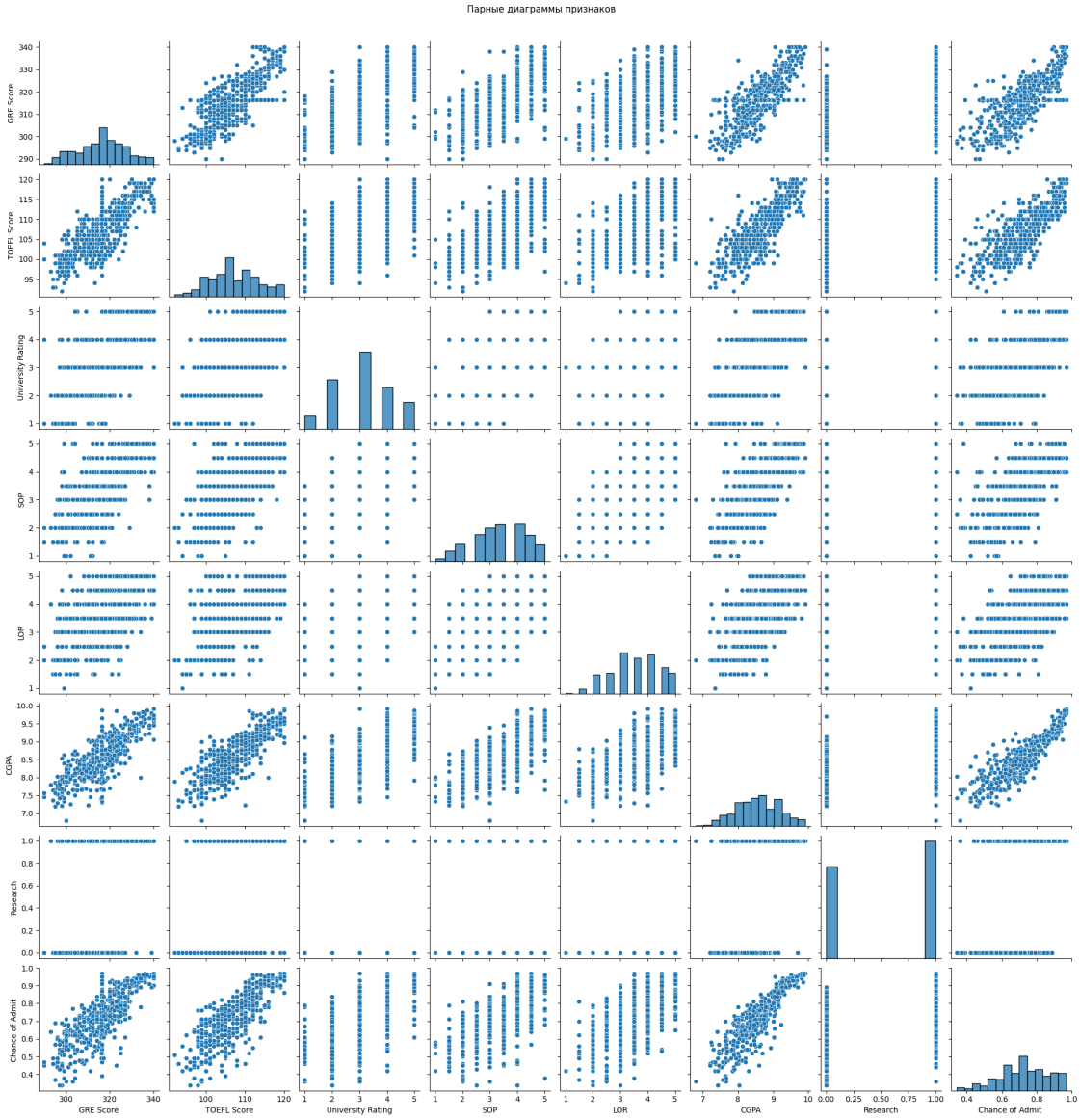
2. (University Rating)

[14]:
```
#
df['University Rating'] = df['University Rating'].fillna(df['University␣
↪Rating'].mode()[0])

#
print("\n               :")
print(df.isnull().sum())
```

```
               :
Serial No.            0
GRE Score             0
TOEFL Score           0
University Rating     0
SOP                   0
LOR                   0
CGPA                  0
Research              0
Chance of Admit       0
dtype: int64
```

[15]:
```
#
print("\n               :")
print(df.isnull().sum())

#
sns.pairplot(df.drop(columns=["Serial No."]))
plt.suptitle("               ", y=1.02)
plt.show()
```

```
               :
Serial No.            0
GRE Score             0
TOEFL Score           0
University Rating     0
SOP                   0
LOR                   0
CGPA                  0
Research              0
Chance of Admit       0
dtype: int64
```



:

```
['GRE Score', 'TOEFL Score', 'University Rating', 'SOP', 'LOR', 'CGPA',
'Research']
```

: GRE Score, TOEFL Score, University Rating, SOP, LOR, CGPA, Research

:

- , ‘Serial No.’ .
- ‘CGPA’, ‘GRE Score’, ‘TOEFL Score’ — .
- ‘University Rating’, ‘SOP’, ‘LOR’, ‘Research’ — .
- ‘Serial No.’ — , .