Вадим Пичурин РТ5-61Б

Вариант 14, задание 2, датасет 6

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

# Загрузка данных
df = pd.read_csv("data/Admission_Predict_Ver1.1.csv")

# Очистка названий столбцов от лишних пробелов
df.columns = df.columns.str.strip()

# Создание копии датасета
df_demo = df.copy()

# Вводим пропуски в 5% строк для 'GRE Score' и 'Research'
np.random.seed(42)
missing_indices = np.random.choice(df_demo.index, size=int(0.05 *
len(df_demo)), replace=False)
df_demo.loc[missing_indices, 'GRE Score'] = np.nan
df_demo.loc[missing_indices, 'Research'] = np.nan

# Обработка пропусков
# Количественный признак — среднее значение
df_demo['GRE Score'].fillna(df_demo['GRE Score'].mean(), inplace=True)

# Категориальный признак — мода
df_demo['Research'].fillna(df_demo['Research'].mode()[0],
inplace=True)

# Выбор признаков для ML (исключим Serial No. и целевую переменную)
features_for_ml = df_demo.drop(columns=['Serial No.', 'Chance of
Admit']).columns.tolist()

print("Выбранные признаки для ML-модели:", features_for_ml)

# Построение jointplot для 'CGPA' и 'Chance of Admit'
sns.set(style="white", color_codes=True)
plot = sns.jointplot(data=df_demo, x='CGPA', y='Chance of Admit',
kind='scatter')
plot.fig.suptitle("Jointplot: CGPA vs Chance of Admit", y=1.02)
plt.show()
```

```
Выбранные признаки для ML-модели: ['GRE Score', 'TOEFL Score',
'University Rating', 'SOP', 'LOR', 'CGPA', 'Research']
```

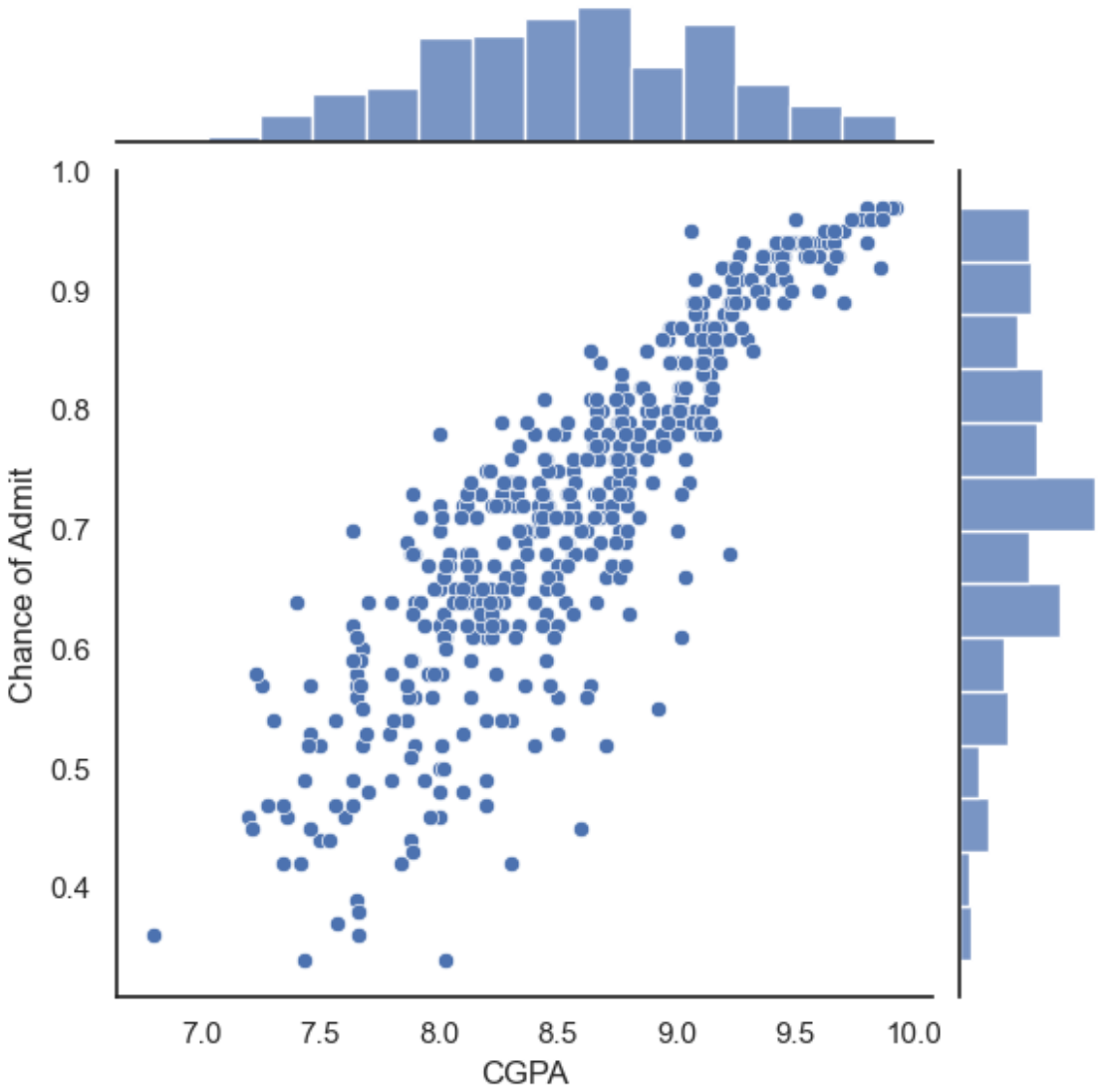
```
/var/folders/6l/m51c4vbn6_3g7bhxs2fnbn5h0000gn/T/
ipykernel_63037/911591534.py:23: FutureWarning: A value is trying to
be set on a copy of a DataFrame or Series through chained assignment
using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.


  df_demo['GRE Score'].fillna(df_demo['GRE Score'].mean(),
inplace=True)
/var/folders/6l/m51c4vbn6_3g7bhxs2fnbn5h0000gn/T/ipykernel_63037/91159
1534.py:26: FutureWarning: A value is trying to be set on a copy of a
DataFrame or Series through chained assignment using an inplace
method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.


  df_demo['Research'].fillna(df_demo['Research'].mode()[0],
inplace=True)
```


Jointplot: CGPA vs Chance of Admit
Chance of Admit
1.0
0.9
0.8
0.7
0.6
0.5
0.4
7.0
7.5
8.0
8.5
9.0
9.5
10.0
CGPA