

```

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

# 1. Загрузка данных
df = pd.read_csv("HousingData.csv")

# 2. Проверка пропусков
print("Количество пропусков по колонкам:")
print(df.isnull().sum())

# 3. Удаление строк с пропущенными значениями
df_clean = df.dropna()
print("\nФорма набора после удаления пропусков:", df_clean.shape)

# 4. Корреляционный анализ
corr_matrix = df_clean.corr(numeric_only=True)

# 5. Визуализация корреляционной матрицы
plt.figure(figsize=(12, 10))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f",
            square=True)
plt.title("Корреляционная матрица признаков Boston Housing")
plt.tight_layout()
plt.show()

# 6. Диаграмма рассеяния для пары признаков: LSTAT и MEDV
plt.figure(figsize=(8, 6))
sns.scatterplot(data=df_clean, x='LSTAT', y='MEDV', color='teal')
plt.title("Диаграмма рассеяния: LSTAT vs MEDV")
plt.xlabel("Процент малоимущих (LSTAT)")
plt.ylabel("Стоимость жилья (MEDV)")
plt.grid(True)
plt.tight_layout()
plt.show()

```

Количество пропусков по колонкам:

CRIM	20
ZN	20
INDUS	20
CHAS	20
NOX	0
RM	0
AGE	20
DIS	0
RAD	0
TAX	0
PTRATIO	0
B	0
LSTAT	20

MEDV 0  
dtype: int64

Форма набора после удаления пропусков: (394, 14)

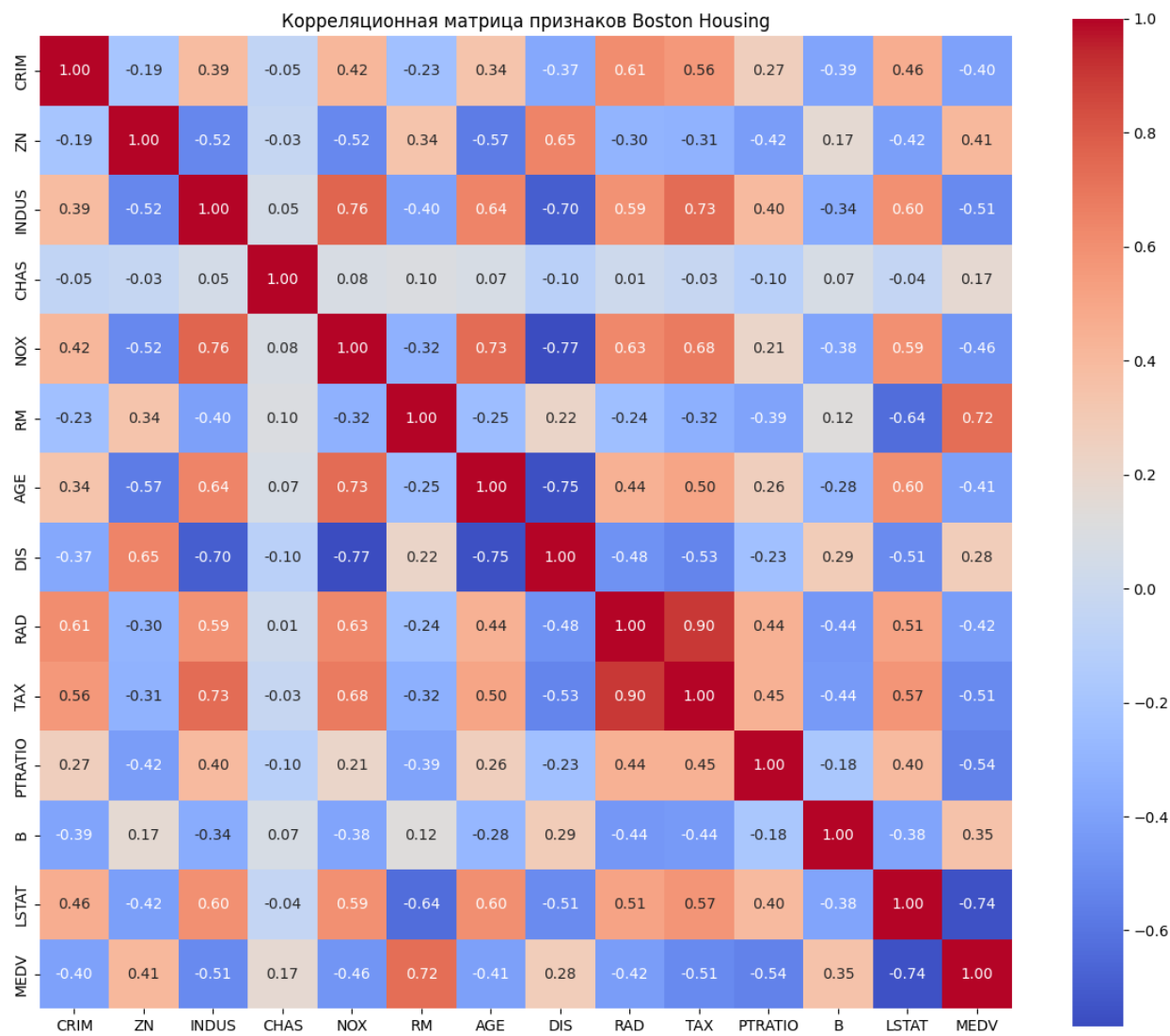


Диаграмма рассеяния: LSTAT vs MEDV

