

**Московский государственный технический
университет им. Н. Э. Баумана**

Курс «Технологии машинного обучения»

Отчёт по лабораторной работе №1

Выполнил:
Файзуллин К. Х.
группа ИУ5-64Б

Проверил:
Гапанюк Ю.Е.

Дата: 21.06.25

Дата:

Подпись:

Подпись:

Москва, 2025 г.

Цель лабораторной работы: изучение различных методов визуализация данных.

Задание:

1. Выбрать набор данных (датасет).
2. Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных.

Мой датасет: fetch_california_housing

Ход выполнения:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import fetch_california_housing

# Загрузка датасета
california_housing = fetch_california_housing()

# Доступ к данным и целевым значениям
X = california_housing.data
y = california_housing.target
```

```
import pandas as pd
from sklearn.datasets import fetch_california_housing

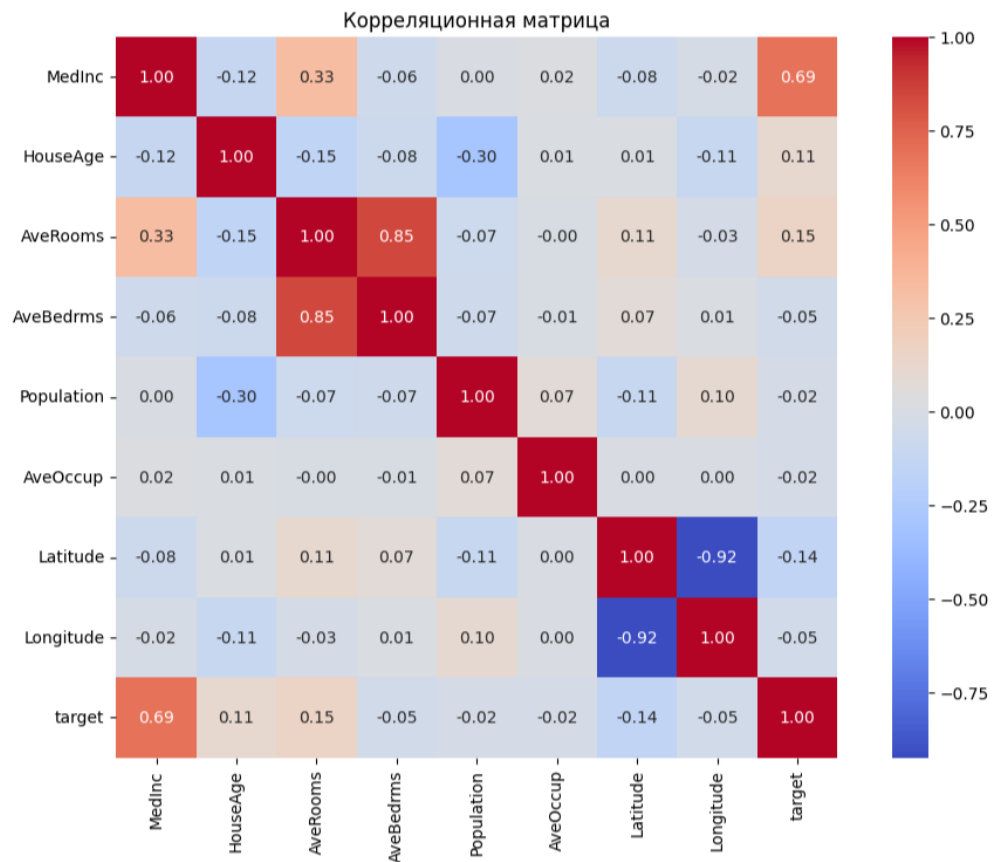
# Загрузка набора данных
california_housing = fetch_california_housing()

# Преобразование в DataFrame
df = pd.DataFrame(data=california_housing.data, columns=california_housing.feature_names)
df['target'] = california_housing.target # Добавить целевую переменную

# Просмотр первых строк датафрейма
print(df.head())
```

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	\
,0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	
,1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	
,2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	
,3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	
,4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	
,								
,	Longitude	target						
,0	-122.23	4.526						
,1	-122.22	3.585						
,2	-122.24	3.521						
,3	-122.25	3.413						
,4	-122.25	3.422						


```
[ ]: plt.figure(figsize=(12, 8))
correlation = df.corr()
sns.heatmap(correlation, annot=True, fmt=".2f", cmap='coolwarm', square=True)
plt.title('Корреляционная матрица')
plt.show()
```



```
[ ]: plt.figure(figsize=(10, 6))
sns.scatterplot(x='MedInc', y='target', data=df)
plt.title('Зависимость между средним доходом и целевым значением')
plt.show()
```

