Московский государственный технический университет им. Н. Э. Баумана

Курс «Технологии машинного обучения»
Отчёт по лабораторной работе №1

Выполнил:	Проверил:
Файзуллин К. Х.	Гапанюк Ю.Е.
группа ИУ5-64Б	

Дата: 21.06.25 Дата:

Подпись: Подпись:

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
             21.0 6.238137 0.971880
,1 8.3014
                                                            37.86
                                         2401.0 2.109842
                                        496.0 2.802260 37.85
,2 7.2574
            52.0 8.288136 1.073446
            52.0 5.817352 1.073059
52.0 6.281853 1.081081
,3 5.6431
                                         558.0 2.547945 37.85
,4 3.8462
                                         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
```

[11]: •

.....

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	target
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422
20635	1.5603	25.0	5.045455	1.133333	845.0	2.560606	39.48	-121.09	0.781
20636	2.5568	18.0	6.114035	1.315789	356.0	3.122807	39.49	-121.21	0.771
20637	1.7000	17.0	5.205543	1.120092	1007.0	2.325635	39.43	-121.22	0.923
20638	1.8672	18.0	5.329513	1.171920	741.0	2.123209	39.43	-121.32	0.847
20639	2.3886	16.0	5.254717	1.162264	1387.0	2.616981	39.37	-121.24	0.894

20640 rows × 9 columns

,

,

,

...

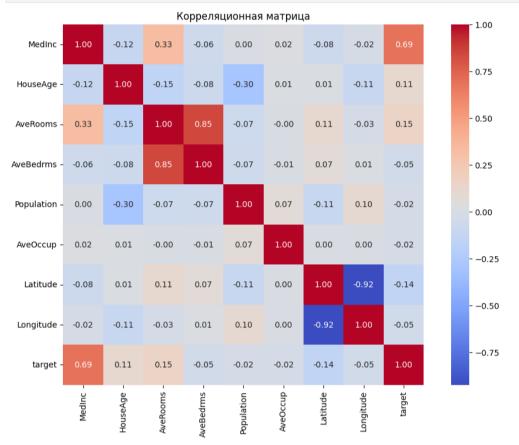
.

,

print(df.describe())



```
[]: 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()
```

