|  |  |
| --- | --- |
| Gerb-BMSTU_01 | Министерство науки и высшего образования Российской Федерации  Федеральное государственное бюджетное образовательное учреждение  высшего образования  «Московский государственный технический университет  имени Н.Э. Баумана  (национальный исследовательский университет)»  (МГТУ им. Н.Э. Баумана) |

|  |  |
| --- | --- |
| ФАКУЛЬТЕТ | Информатика и системы управления (ИУ) |

|  |  |
| --- | --- |
| КАФЕДРА | Система обработки информации и управления |

|  |  |
| --- | --- |
| дисциплина | Методы машинного обучения |

|  |
| --- |
| отчет по лабораторной работе № 1 |

|  |
| --- |
| Создание "истории о данных" (Data Storytelling) |
| *название лабораторной работы* |

|  |  |
| --- | --- |
| Группа | ИУ5-14М |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Студент | 14.03.2022 |  |  |  | Молева А. А. |
|  | *дата выполнения работы* |  | *подпись* |  | *фамилия, и.о.* |

|  |  |  |  |
| --- | --- | --- | --- |
| Преподаватель |  |  | Гапанюк Ю. Е. |
|  | *подпись* |  | *фамилия, и.о.* |

Москва, 2022 г.

# **Цель работы**

Цель лабораторной работы: изучение различных методов визуализация данных и создание истории на основе данных.

Краткое описание. Построение графиков, помогающих понять структуру данных, и их интерпретация.

# **Задание**

* Выбрать набор данных (датасет).
* Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:
  + История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.
  + На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.
  + Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.
  + Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.
  + История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.
* Сформировать отчет и разместить его в своем репозитории на github.

# **Текст программы**

dataset = load\_boston()

import pandas as pd

X = pd.DataFrame(dataset['data'], columns=dataset['feature\_names'])

y = pd.Series(dataset['target'], name='target')

from sklearn.datasets import fetch\_california\_housing

dataset = fetch\_california\_housing(return\_X\_y=True, as\_frame=True)

X, y = dataset

X.info()

X.describe()

import pandas as pd

housing = pd.concat([X, y], axis=1)

from matplotlib import pyplot as plt

#skatter

housing.plot(kind='scatter', x='Longitude', y='Latitude', alpha=0.4, s=X.Population/100, label='population', figsize=(10,7), c=housing.MedHouseVal,

             cmap=plt.get\_cmap("jet"), colorbar=True)

plt.legend()

#парные диаграммы

import seaborn as sns

sns.pairplot(housing)

# Violin plot

fig, ax = plt.subplots(9, 1, figsize=(10, 35))

for idx, column in enumerate(housing.columns):

    sns.violinplot(x=housing[column], ax=ax[idx])

# Histograms

fig, ax = plt.subplots(9, 1, figsize=(10, 35))

for idx, column in enumerate(housing.columns):

    sns.distplot(x=housing[column], ax=ax[idx], axlabel=column)

# Boxplot

fig, ax = plt.subplots(9, 1, figsize=(10, 35))

for idx, column in enumerate(housing.columns):

    sns.boxplot(y=housing[column], ax=ax[idx])

housing.corr()

# Вывод значений в ячейках

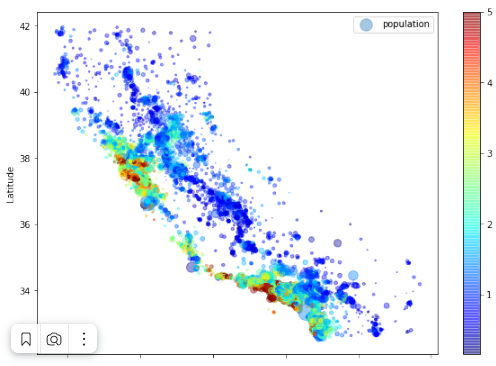
plt.figure(figsize=(10,10))

sns.heatmap(housing.corr(), annot=True, fmt='.3f')

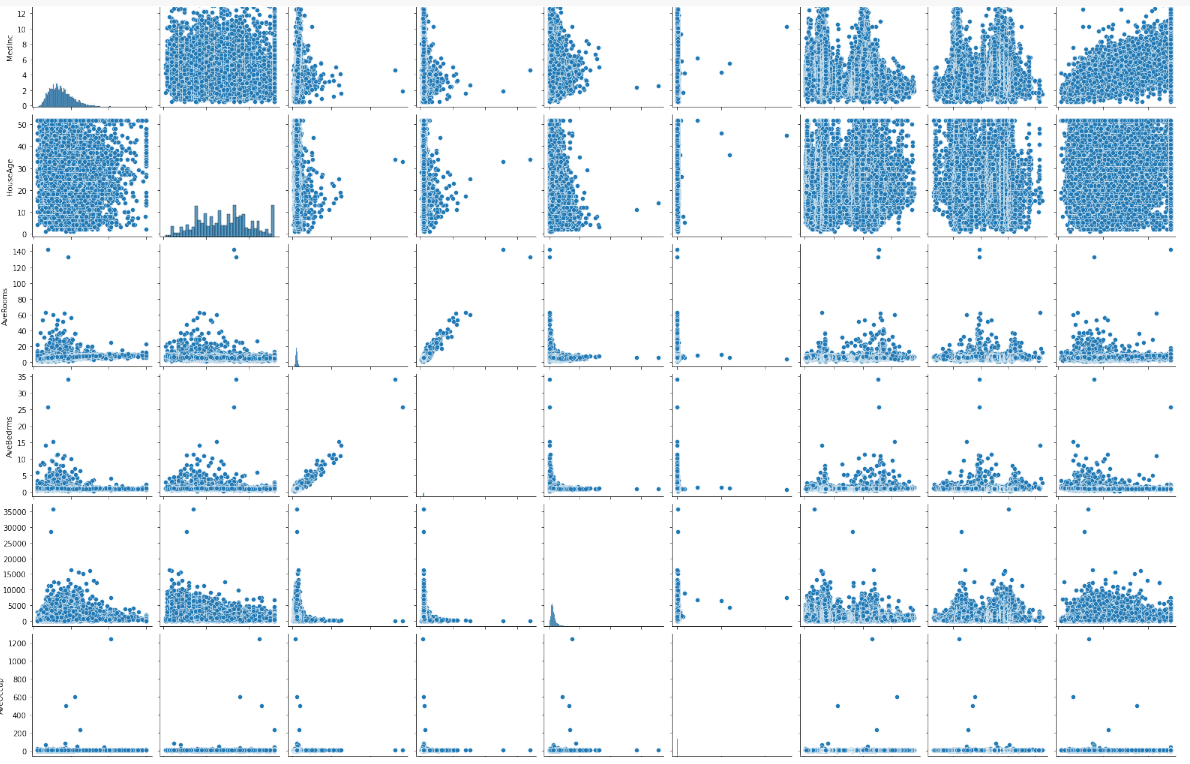
**Экранные формы**

****

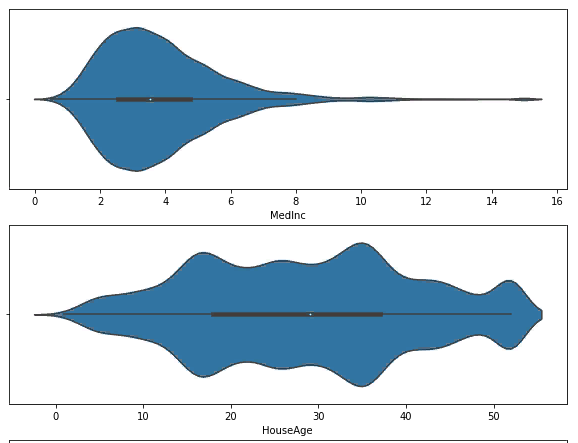
**Рисунок 1 – Датасет**

****

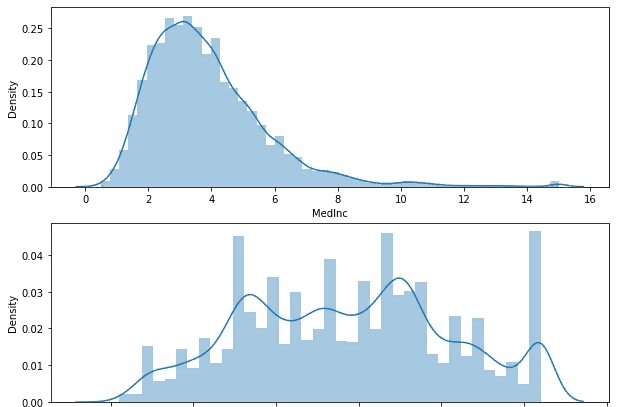
**Рисунок 2 – Диаграмма рассеяния**

****

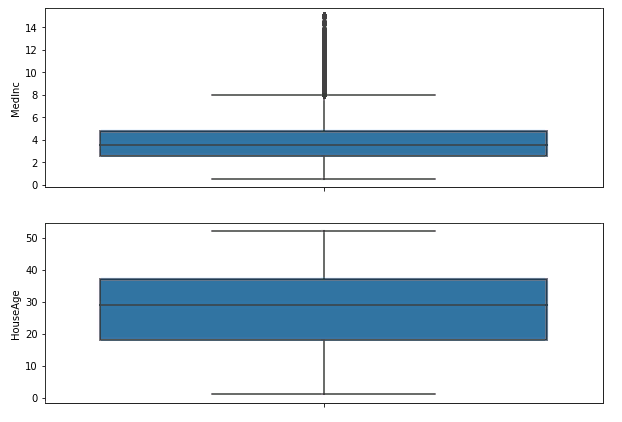
**Рисунок 3 – Парные диаграммы**

****

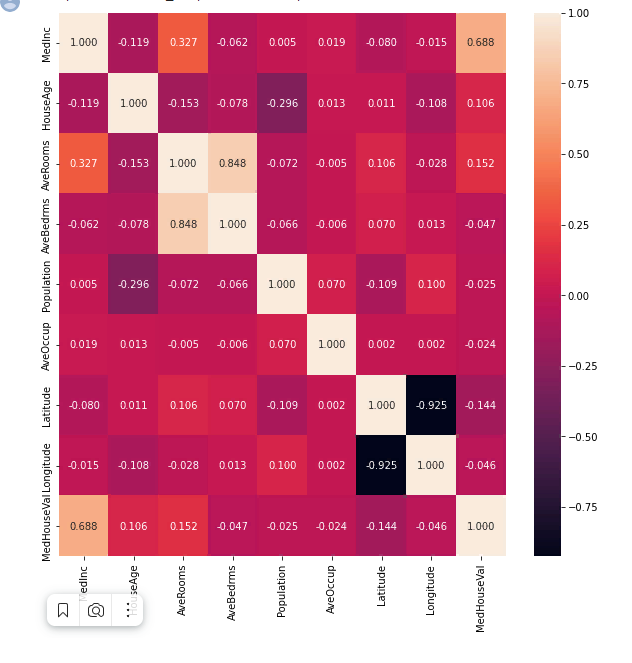
**Рисунок 4 – Скрипичная диаграмма**



**Рисунок 5 – Гистограммы**

****

**Рисунок 6 – Боксплот**

****

**Рисунок 7 – Тепловая карта**

**Выводы**

В результате проделанной работы была создана история о данных со следующими диаграммами: диаграмма расеяния, парные диаграммы, скрипичные диаграммы, гистограммы, боксплоты, тепловая карта.