**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра информационных систем**

отчет

**по лабораторной работе №1**

**по дисциплине «Машинное обучение»**

Тема: Исследование набора данных

|  |  |  |
| --- | --- | --- |
| Студент гр. |  | Громова В.С. |
| Преподаватель |  | Татчина Я.А. |

Санкт-Петербург

2024

**Цель работы.**

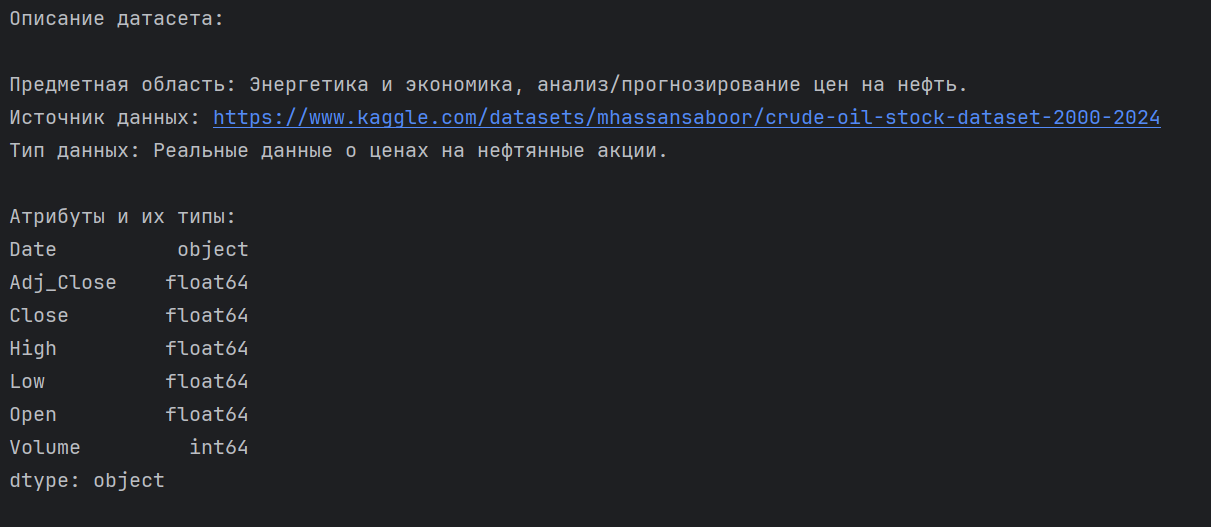
Выбор набора данных для дальнейшего исследования алгоритмов кластеризации и классификации. Получение практических навыков работы с инструментами анализа и визуализации данных.

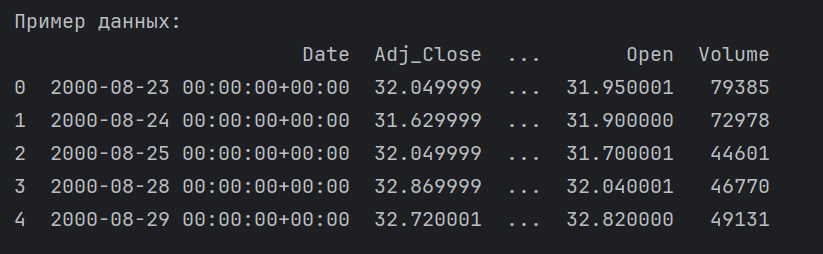
**Постановка задачи.**

1. Создать Jupyter Notebook
2. Выбрать исследуемый набор данных из предложенных источников
3. Описать выбранный набор данных, а именно:
   1. Предметную область, источник данных, характер данных;
   2. Атрибуты, их тип, что они обозначают;
   3. Задачу анализу.
4. Для каждого атрибута:
   1. Определить среднее значение, СКО;
   2. Построить гистограмму распределения значений, определить наличие выбросов;
   3. Определить наличие пропущенных значений и их количество;
   4. Предложить вариант обработки пропущенных значений.
5. Определить:
   1. Высоко коррелированные атрибуты, характер корреляции;
   2. Не коррелированные атрибуты;
   3. Графики рассеивания.
6. Проанализировать полученные результаты.

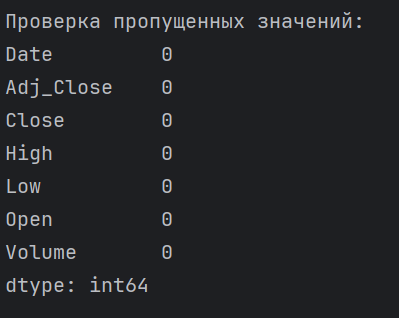
**Выполнение работы.**

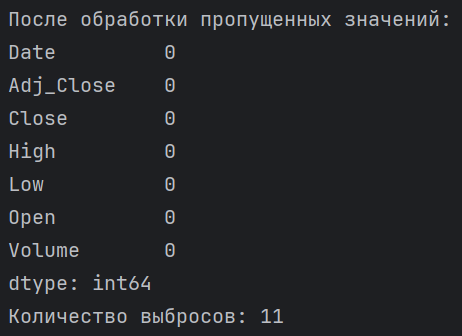
Для исследования был выбран набор данных для задач прогнозирования и классификации.



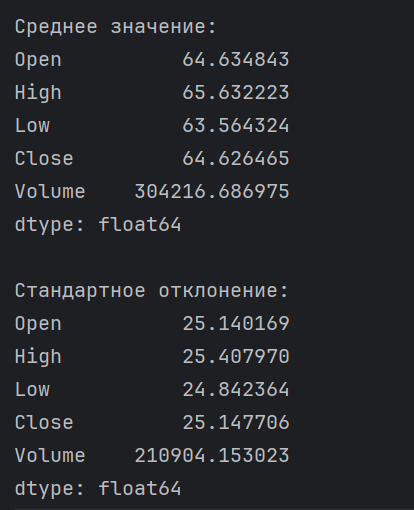


Была проведена проверка на пропущенные значения и выбросы

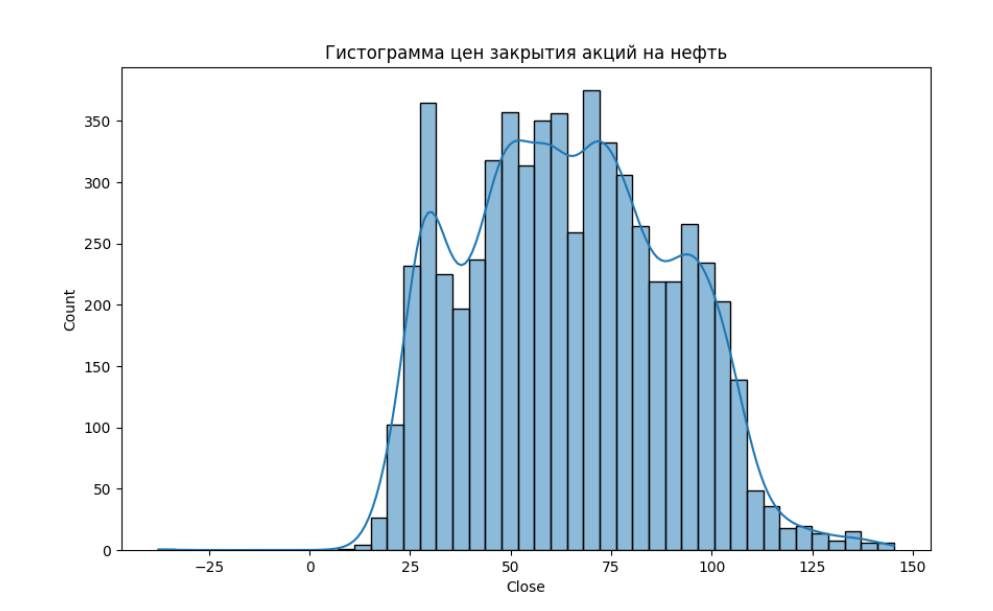


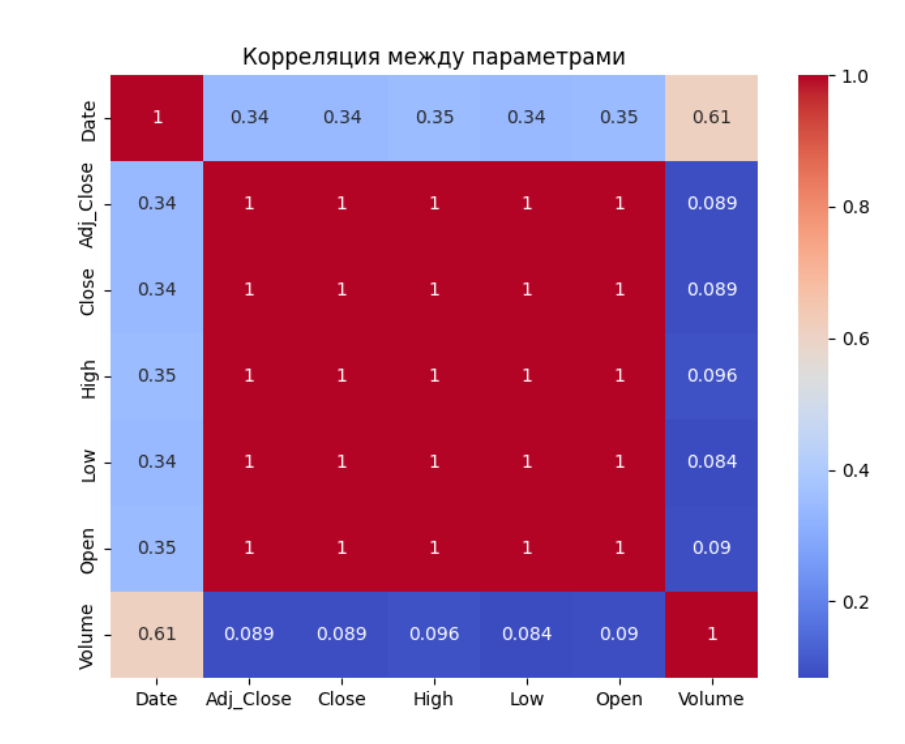


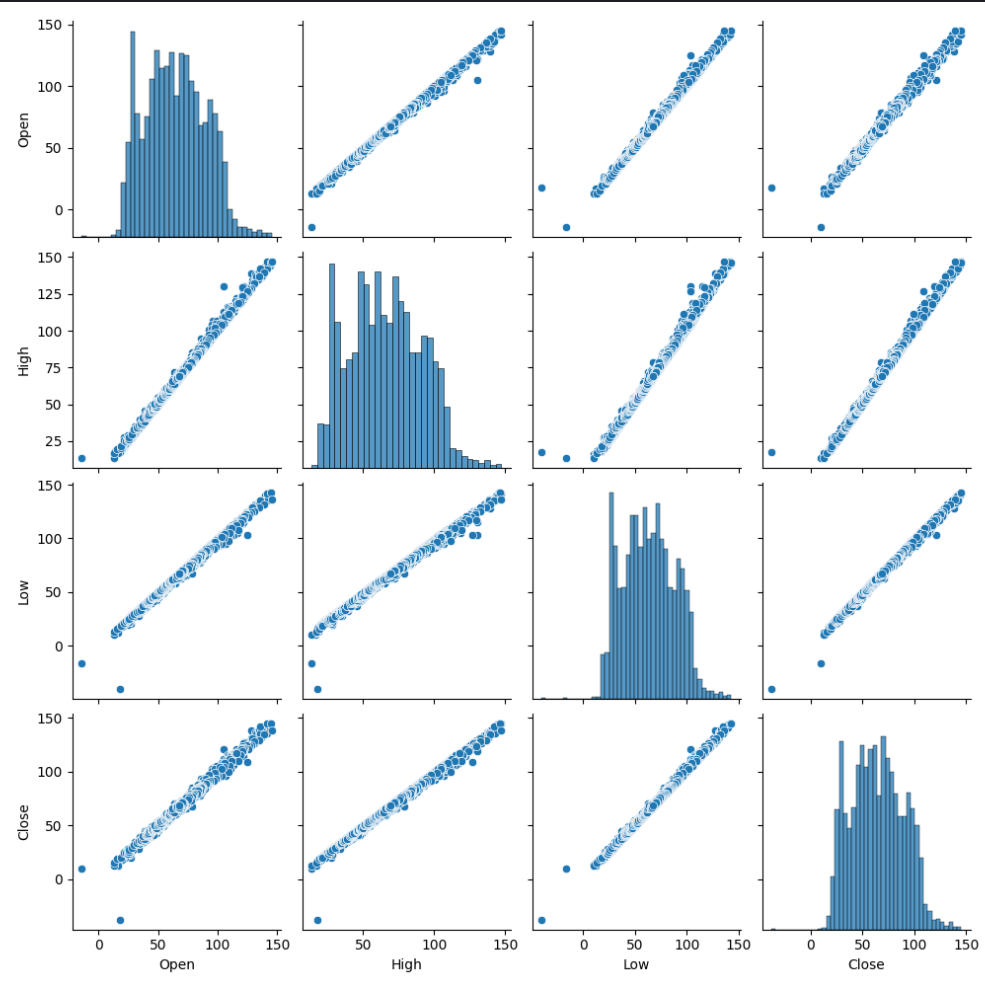
Вычислены средние значения и СКО для каждого параметра



С помощью средств pandas.df.describe() и seaborn.histplot() вычислены основные параметры признаков и построены гистограммы распределения соответственно.







**Выводы.**

В ходе работы были выбран набор данных для исследования. Получены практические навыки работы с анализом и визуализацией данных с помощью инструментов pandas, matplotlib, seaborn.