|  |
| --- |
| https://lh6.googleusercontent.com/QcftzNtI05T0Y6fjdSh1Rr2rt8oqZ1IvnLvbn1jLJ7CCyteVir3k-xBLv4SL1wAgWJsRhmmJSR0UW-RP63_GQenE4vVWv05BRoZTsmIcBccVTnfxwmsnNMvjg599x9SqZd8E3dkd |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования  **«МИРЭА - Российский технологический университет»**  **РТУ МИРЭА** |

Институт информационных технологий (ИТ)

Кафедра прикладной математики (ПМ)

|  |  |
| --- | --- |
| **ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ №7-8** | |
| **по дисциплине** | |
| «Технологии и инструментарий анализа больших данных» | |
| Выполнил студент группы ИВБО-04-19 | Кузьников М.С. |
|  |  |
| Принял | Трифанова О.А. |

Практическая работа выполнена «\_\_» \_\_\_\_\_\_\_ 2022 г.

(подпись студента)

Зачтено «\_\_» 2022 г.

**СОДЕРЖАНИЕ**

1. [ПОСТАНОВКА ЗАДАЧИ 3](#_Toc116463437)

[2. ЗАДАНИЕ №1 4](#_Toc116463438)

[3. ЗАДАНИЕ №2 5](#_Toc116463439)

[4. ЗАДАНИЕ №3 6](#_Toc116463440)

[5. ЗАДАНИЕ №4 7](#_Toc116463441)

[6. ЗАДАНИЕ №5 10](#_Toc116463442)

[7. ЗАДАНИЕ №6 13](#_Toc116463443)

[8. ЗАДАНИЕ №7 15](#_Toc116463444)

[ЗАКЛЮЧЕНИЕ 17](#_Toc116463445)

[СПИСОК ЛИТЕРАТУРЫ 18](#_Toc116463446)

# ПОСТАНОВКА ЗАДАЧИ

**Цель работы:** изучить корреляцию и линейную регрессию

**Задание:**

1. Определить два вектора, представляющие собой число автомобилей, припаркованных в течении ***«5»*** рабочих дней у бизнес-центра на уличной стоянке и в подземном гараже.

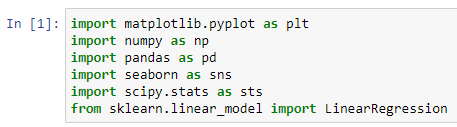
|  |  |  |
| --- | --- | --- |
| **День** | **Улица** | **Гараж** |
| Понедельник | 80 | 100 |
| Вторник | 98 | 82 |
| Среда | 75 | 105 |
| Четверг | 91 | 89 |
| Пятница | 78 | 102 |

Найти и интерпретировать корреляцию между переменными ***«Улица»*** и ***«Гараж»*** (подсчитать корреляцию по Пирсону).

1. Построить диаграмму рассеяния
2. Загрузить данные из файла ***«bitcoin.csv»***
3. Скрыть последние ***«14»*** дней
4. Предсказать стоимость криптовалюты за последние ***«14»*** дней с помощью линейной регрессии
5. Вывести угол наклона и ***«y»***-перехват. Построить диаграмму
6. Загрузить данные из файла ***«housePrice.csv»***
7. Произвести предобработку
8. Реализовать линейную регрессию вручную, без использования библиотеки
9. Вывести угол наклона и ***«y»***-перехват. Построить диаграмму

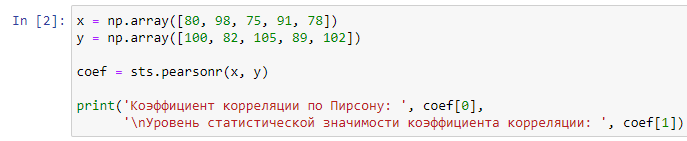
## ЗАДАНИЕ №1

Подключаю необходимые библиотеки для работы ***(Рисунок 1)***.



*Рисунок 1 – Листинг программы*

Нахожу корреляцию между переменными ***«Улица»*** и ***«Гараж» (Рисунок 2.1 – 2.2).***

**

*Рисунок 2.1 – Листинг программы*

*C:\Users\Димасик\Desktop\Screens\Screenshot_3.png*

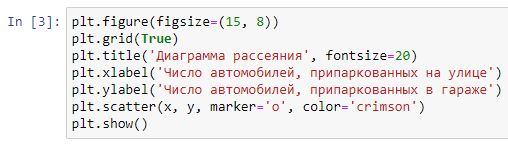
*Рисунок 2.2 – Вывод программы*

На основе полученных результатов делаю следующие выводы:

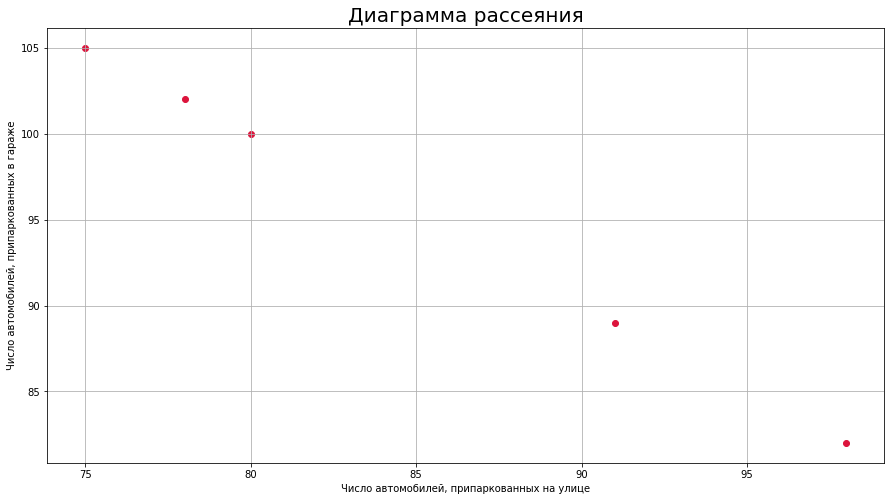
1. Коэффициент корреляции по Пирсону ***«r»*** равен ***«-1»***, что означает сильную отрицательную корреляцию.
2. Уровень статистической значимости коэффициента корреляции ***«p»*** равен ***«0»***, что означает существование статистически значимой корреляции между двумя переменными.

## ЗАДАНИЕ №2

Строю диаграмму рассеяния ***(Рисунок 3.1 – 3.2)***.

**

*Рисунок 3.1 – Листинг программы*



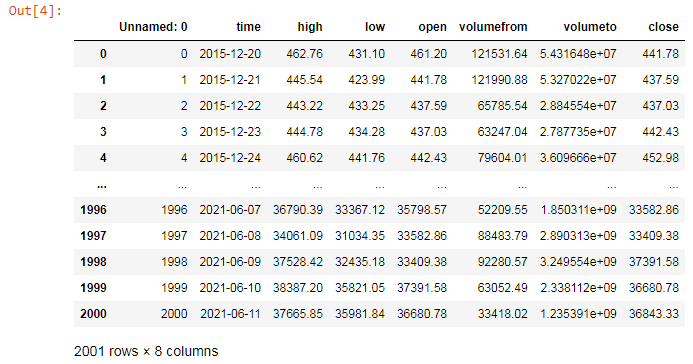
*Рисунок 3.2 – Вывод программы*

## ЗАДАНИЕ №3

Загружаю данные из файла ***«bitcoin.csv» (Рисунок 4.1 – 4.2).***

C:\Users\Димасик\Desktop\Screens\Screenshot_6.png

*Рисунок 4.1 – Листинг программы*

**

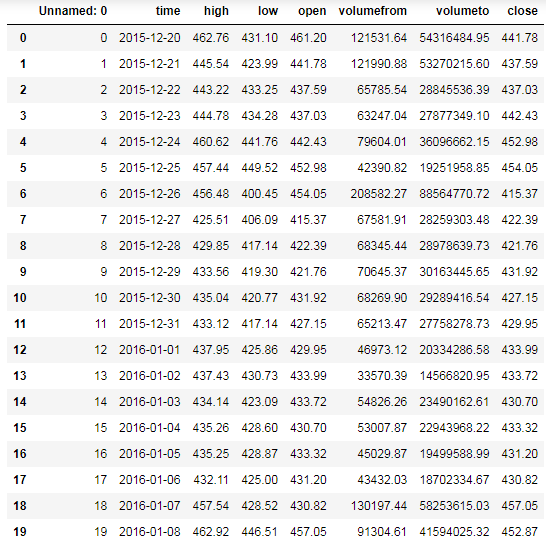
*Рисунок 4.2 – Вывод программы*

## ЗАДАНИЕ №4

Скрываю последние ***«14»*** дней ***(Рисунок 5.1 – 5.6).***

*C:\Users\Димасик\Desktop\Screens\Screenshot_8.png*

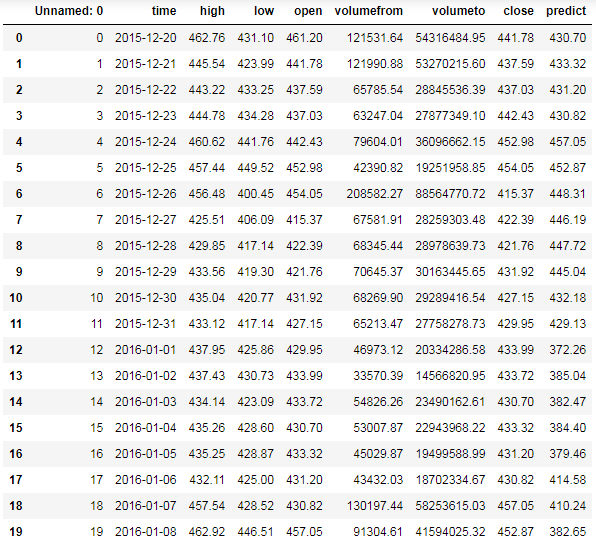
*Рисунок 5.1 – Листинг программы*

**

*Рисунок 5.2 – Вывод программы*

*C:\Users\Димасик\Desktop\Screens\Screenshot_10.png*

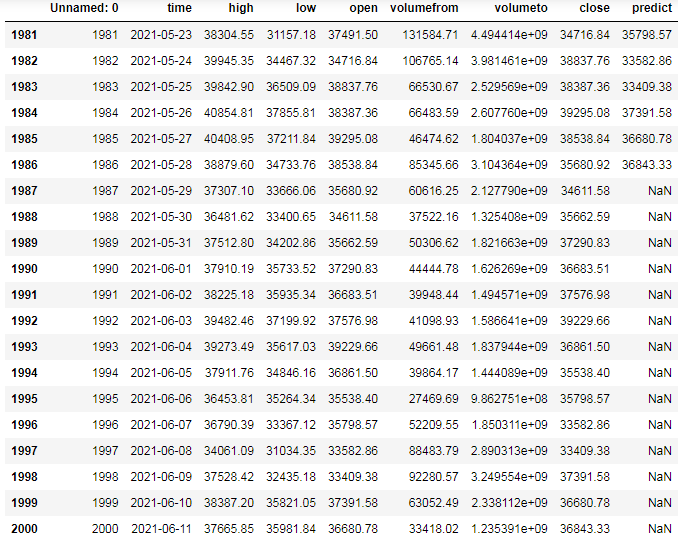
*Рисунок 5.3 – Листинг программы*



*Рисунок 5.4 – Вывод программы*

C:\Users\Димасик\Desktop\Screens\Screenshot_12.png

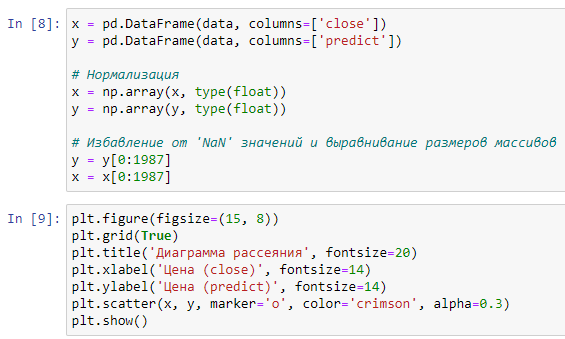
*Рисунок 5.5 – Листинг программы*



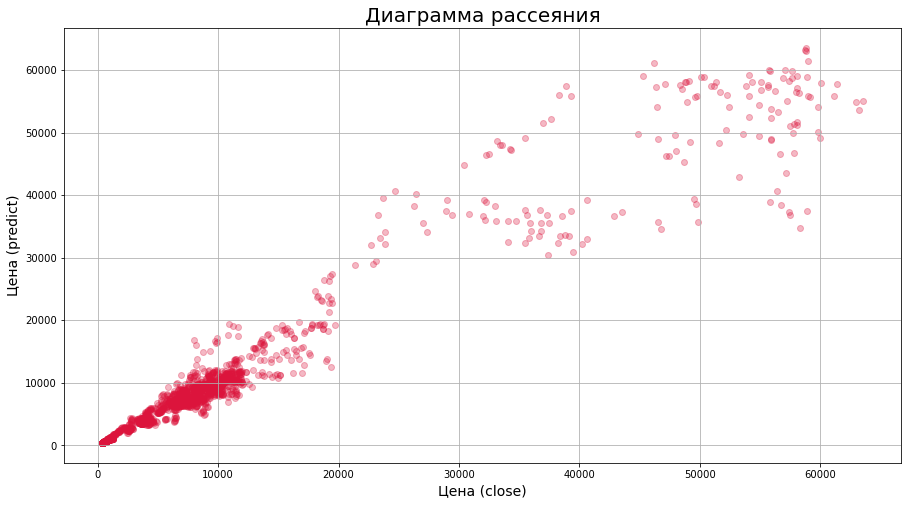
*Рисунок 5.6 – Вывод программы*

## ЗАДАНИЕ №5

Строю диаграмму рассеяния ***(Рисунок 6.1 – 6.2).***

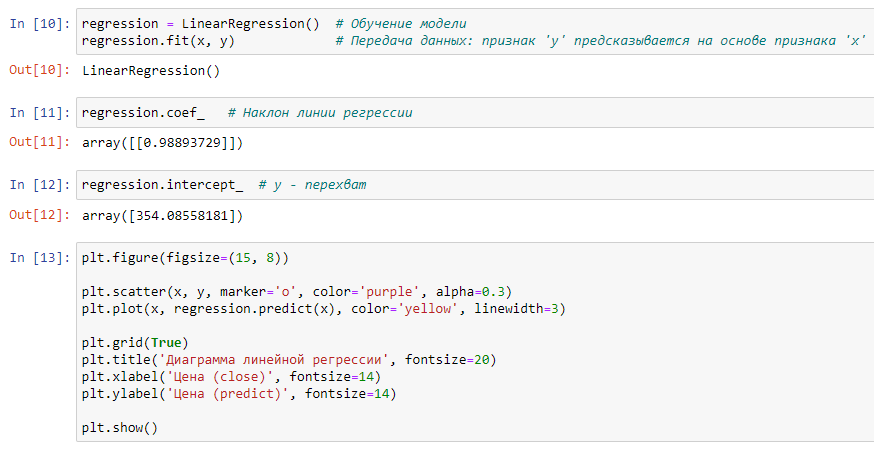
**

*Рисунок 6.1 – Листинг программы*

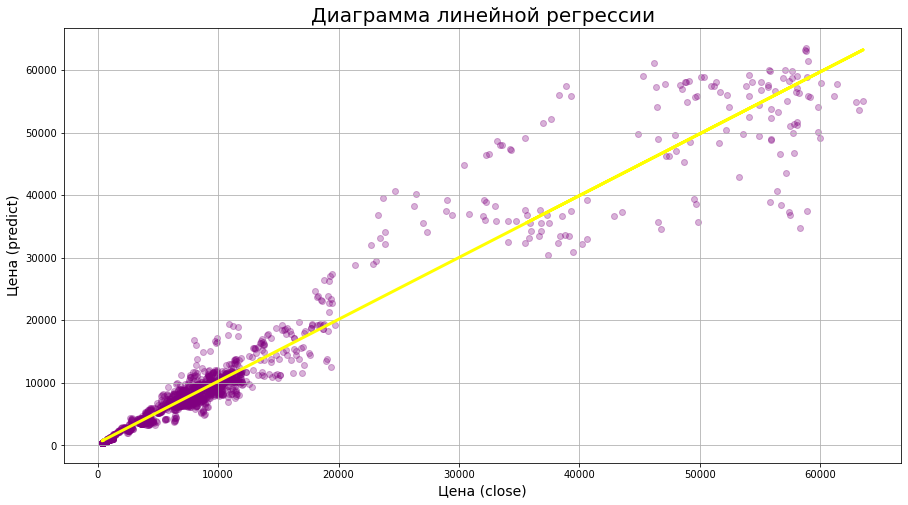


*Рисунок 6.2 – Вывод программы*

Строю диаграмму линейной регрессии ***(Рисунок 7.1 – 7.2).***

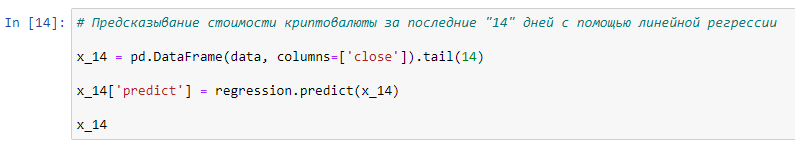


*Рисунок 7.1 – Листинг программы*

**

*Рисунок 7.2 – Вывод программы*

Предсказываю стоимость криптовалюты за последние ***«14»*** дней с помощью линейной регрессии ***(Рисунок 8.1 – 8.2).***



*Рисунок 8.1 – Листинг программы*

**

*Рисунок 8.2 – Вывод программы*

## ЗАДАНИЕ №6

Загружаю данные из файла ***«housePrice.csv»*** ***(Рисунок 9.1 – 9.2) .***

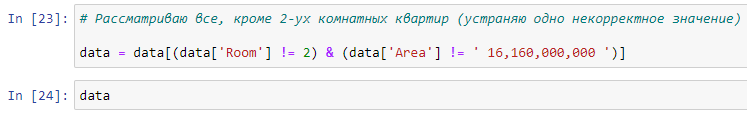
*C:\Users\Димасик\Desktop\Screens\Screenshot_20.png*

*Рисунок 9.1 – Листинг программы*

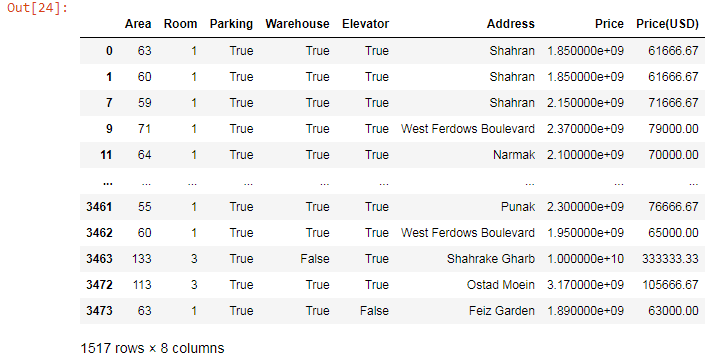


*Рисунок 9.2 – Вывод программы*

Провожу предобработку данных ***(Рисунок 10.1 – 10.2) .***

**

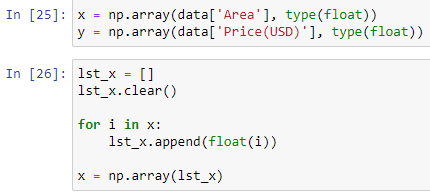
*Рисунок 10.1 – Листинг программы*

**

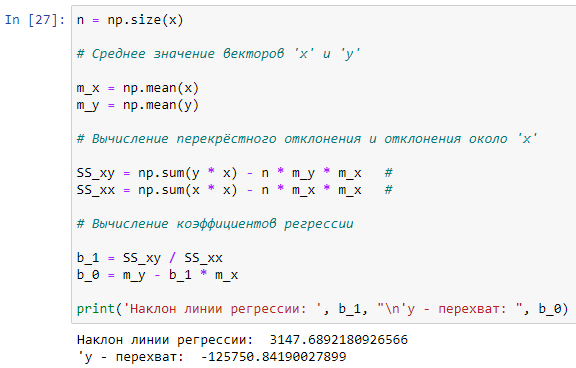
*Рисунок 10.2 – Вывод программы*

## ЗАДАНИЕ №7

Реализовываю линейную регрессию вручную, без использования библиотеки ***(Рисунок 11.1 – 11.2).***

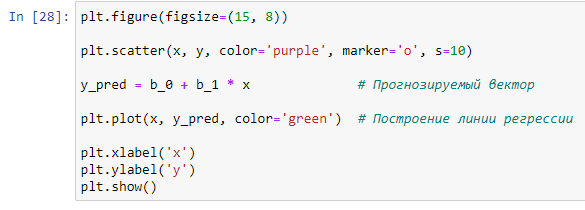
**

*Рисунок 11.1 – Листинг программы*

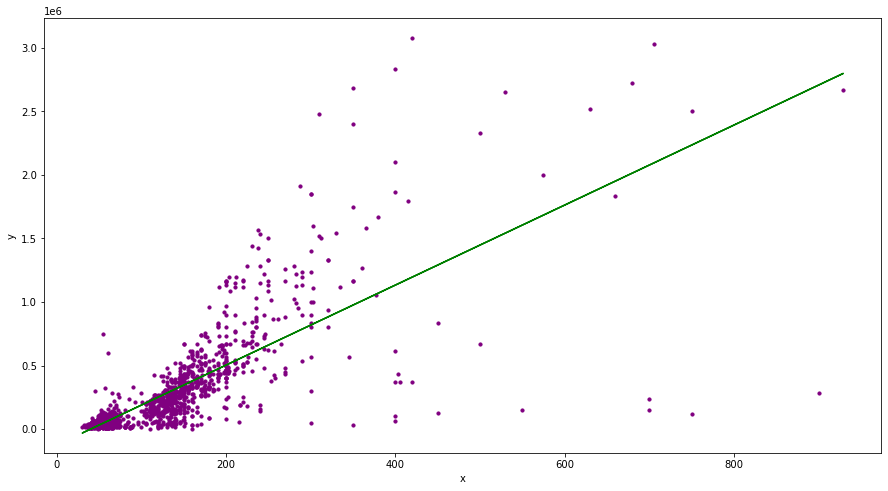


*Рисунок 11.2 – Листинг программы*

Строю диаграмму линейной регрессии ***(Рисунок 12.1 – 12.2).***

**

*Рисунок 12.1 – Листинг программы*



*Рисунок 12.2 – Вывод программы*

# ЗАКЛЮЧЕНИЕ

В результате выполнения практической работы были найдены коэффициенты корреляции, на основании которых были построены диаграммы рассеяния и диаграммы линейной регрессии как с использованием библиотеки, так и без неё вручную.

# СПИСОК ЛИТЕРАТУРЫ

1. [Документация по языку программирования ***«Python***](https://pydocs.ru/)***»*** [Электронный ресурс]: <https://pydocs.ru/>
2. Документация по ***«Jupyter Notebook»*** [Электронный ресурс]: <https://docs.jupyter.org/en/latest/>