**Министерство науки высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»**

**(Университет ИТМО)**

**ОТЧЕТ**

по лабораторной работе №1

Тема: «Коэффициент корреляции Пирсона»

**Выполнил:**

студент группы К42112с

Дарбинян В.М.

**Проверил:**

Мусаев А.А.

Санкт-Петербург

2022 г.

1. **Задача**

Используя коэффициент корреляции Пирсона, постройте матрицу корреляций для выбранных котировок за январь, февраль, …, декабрь. Сделайте выводы о исследуемом процессе исходя из изменения матриц корреляций

1. **Получение котировок**

Чтобы получить актуальные данные за определенный промежуток времени, стоит воспользоваться сайтом <https://www.finam.ru/>. На этом сайте можно гибко настроить параметры, которые нужны нам для анализа данных. Параметры можно увидеть на рисунке 1.

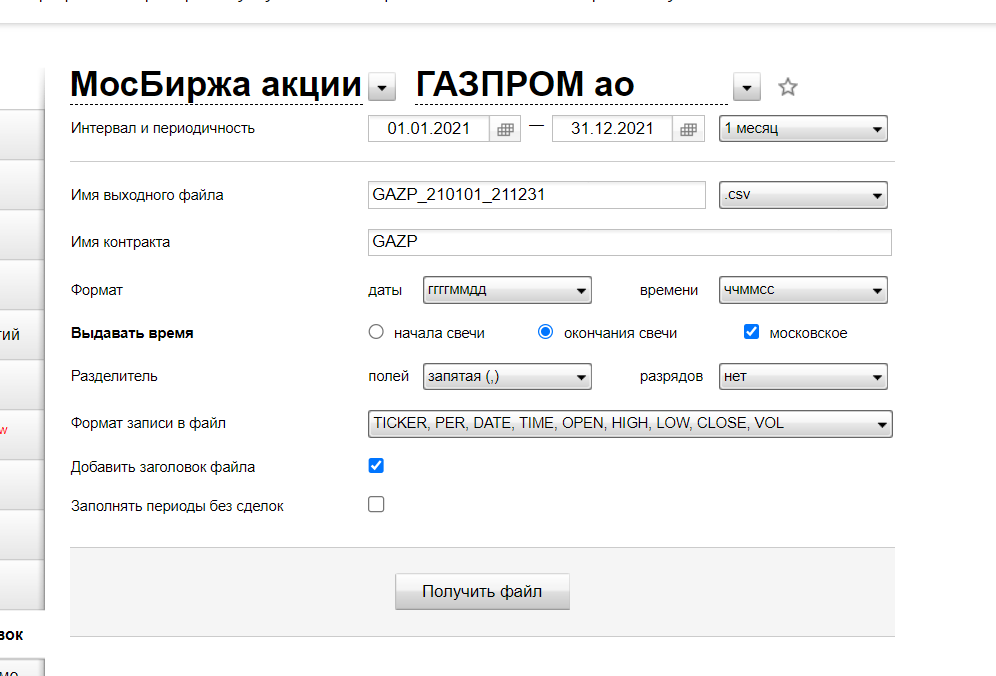


Рисунок 1. Сайт

1. **Открыть CSV файл в Python**

Самым простым вариантом для решения данной задачи, является использование Jupyter Notebook, в нем уже предустановлены самые популярные библиотеки, которыми мы, собственно, и должны воспользоваться.

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

Для чтения CSV файла, были написаны следующие строки кода:  
pd.set\_option('display.max\_rows', None)

pd.set\_option('display.max\_columns', None)

pd.set\_option('display.max\_colwidth', None)

file = 'GAZP\_210701\_220701.csv'

table = pd.read\_csv(file)

print(table)

После чего, получаем такую таблицу:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2. Данные котировки за год

1. **Коэффициент корреляции Пирсона**

Коэффициент характеризует наличие только линейной связи между признаками, обозначаемыми, как правило, символами X и Y. Формула расчёта коэффициента корреляции построена таким образом, что, если связь между признаками имеет линейный характер, коэффициент Пирсона точно устанавливает тесноту этой связи.

Метод **corr().round(3)** в определенном объекте, показывает корреляцию между параметрами, сразу округляя их на 3 значащих цифры.

Вывод коэффициента между данными:

**5. Графическое изображение**

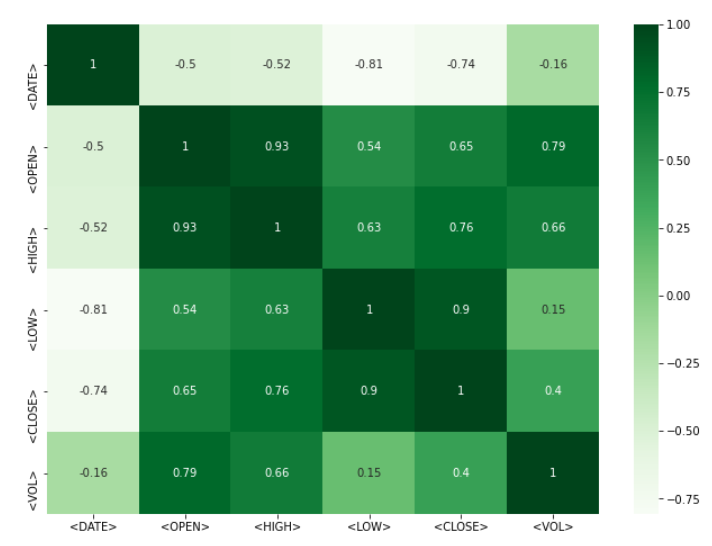
По итогу, после написания следующих строк кода  
corr = table.corr()

plt.figure(figsize=(11,8))

sns.heatmap(corr, cmap="Greens",annot=True)

plt.show()

Получаем следующую матрицу:



**6. Вывод**

При решение данной работы, мы научились парсить данные из CSV файла, освоили метод корреляции Пирсона, используя библиотеку pandas, а так же смогли вывести графическое изображение наших данных.

Если же говорить, что мы видим в данной матрице, не трудно заметить, что у нас чрезвычайно низкая корреляция DATE и LOW и это неудивительно, ведь в рассматриваемой выборке были взяты даты с 1 июля 2021 года по 1 июля 2022 года. Таким образом, матрица показывает, что цены упали кратно, особенно это видно в графе LOW.