

ГУАП

КАФЕДРА №43

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
ПРЕПОДАВАТЕЛЬ

ст. преподаватель

должность, уч. степень, звание

подпись, дата

М.Д. Поляк

инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ №1

ЗНАКОМСТВО С JUPYTER NOTEBOOK И GOOGLE COLAB

по дисциплине: ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР.

4232

подпись, дата

А.В. Коновалова

инициалы, фамилия

Санкт-Петербург
2025

1. Цель работы

Знакомство со средами Jupyter Notebook и Google Colaboratory, а также библиотеками Pandas и matplotlib.

2. Задание

Задание 1

Откройте Jupyter-ноутбук `jupyter_assignment.ipynb` в этом репозитории. Скопируйте путь в адресной строке браузера. Перейдите в Google Colab, в меню выберите "Файл" -> "Открыть ноутбук", в открывшемся окне слева выбрать "GitHub", затем:

- вставить в поле для поиска скопированный URL;
- поставить галочку "Показывать личные хранилища" ("Include private repos");
- и нажать на иконку с лупой. При необходимости разрешить Colab доступ к аккаунту GitHub, если откроется новое окно с таким приглашением. Среди результатов поиска выбрать `jupyter_assignment.ipynb` и приступить к выполнению задания.

Задание 2

Откройте в Google Colab Jupyter-ноутбук `matplotlib_assignment.ipynb`, ознакомьтесь с его содержимым и выполните задание.

Задание 3

Откройте в Google Colab Jupyter-ноутбук `pandas_assignment.ipynb`, ознакомьтесь с его содержимым и выполните задание..

3. Выполнение задания

Задание 1:

Часть 1:

```
### BEGIN YOUR CODE
#I have read through the Introduction and Overview notebook
READ_INTRODUCTION = True

#I understand (at a high level) what Jupyter notebooks are and how to read and
#interact with them (or I have been in touch with the course instructor to ask for
help)
LEARNED_ABOUT_JUPYTER = True

#I've created (or already have) a Google account and can access Google
#Colaboratory under my own account
ACCESS_COLABORATORY = True

#I've created a GitHub account
CREATED_GITHUB_ACCOUNT = True
github_username = 'istoki0bespokoystva'

#My info
my_name = 'Anastasia'
### END YOUR CODE
```

Часть 2:

```
### BEGIN YOUR CODE
print('Hello, Anastasia')
### END YOUR CODE
```

Hello, Anastasia

Объявление функций:

```
def greet(name):
    ### BEGIN YOUR CODE
    return 'Hello, ' + str(name) + '!';
    ### END YOUR CODE
```

Задание 2:

```
import numpy as np
from matplotlib import pyplot as plt
```

```
### BEGIN YOUR CODE
Student_ID = 3
### END YOUR CODE
```

Функция:

$$y = tg(a * x^2 + b * x + c)$$

Рисунок 1 - Функция

Часть 2:

```
def my_function(x,a,b,c):
    ### BEGIN YOUR CODE
    return np.tan(a * x ** 2 + b * x + c)
    ### END YOUR CODE
```

```
### BEGIN YOUR CODE
import numpy as np
a = 0.1
b = 2
c = 1.5
x = np.linspace(-5, 5, 100)
```

```
### BEGIN YOUR CODE
import matplotlib.pyplot as plt
```

```

y = my_function(x,a,b,c)

plt.figure(figsize=(10, 6))
plt.plot(x, y, label='y = tan(ax^2 + bx + c)', color='darkviolet')

plt.title('График функции y = tan(ax^2 + bx + c)')
plt.xlabel('x')
plt.ylabel('y')

plt.grid()
plt.legend()
plt.show()

### END YOUR CODE

```

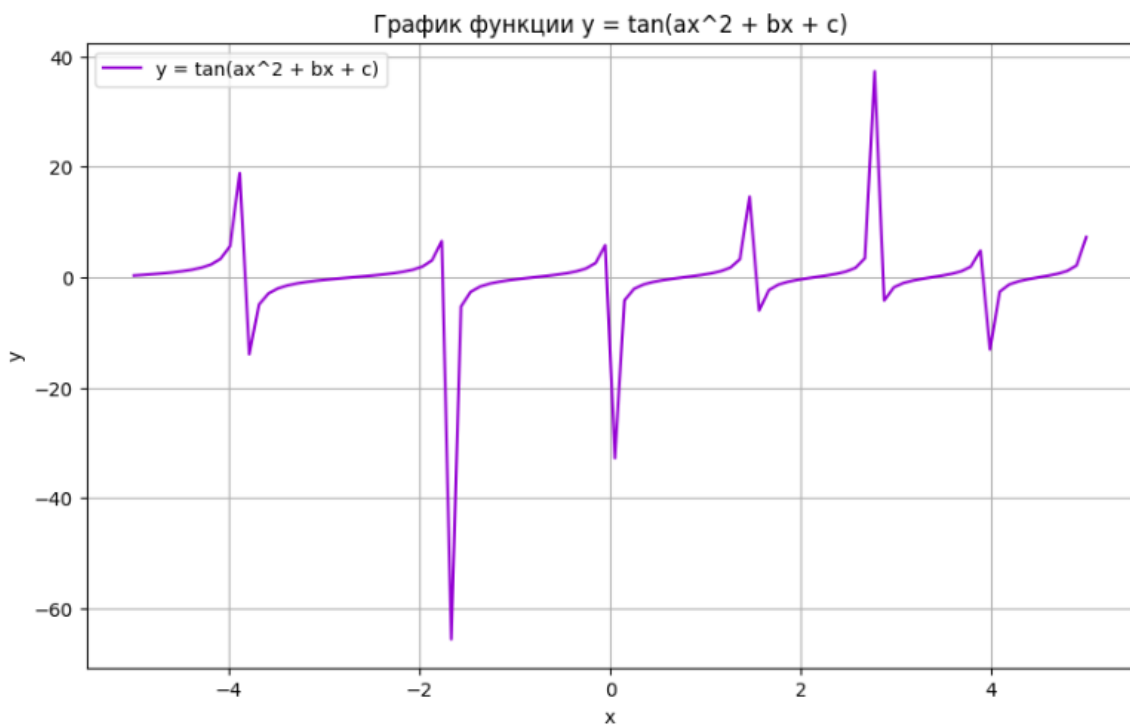


Рисунок 2 – График функции

Часть 3:

```

rng = np.random.RandomState(Student_ID)
vector_a = rng.choice(np.arange(100, dtype=np.int32), size=(1,5), replace=False)
vector_b = rng.choice(np.arange(100, dtype=np.int32), size=(5,1), replace=False)

### BEGIN YOUR CODE

dot_product = np.dot(vector_a, vector_b)

### END YOUR CODE

print(f'Вектор A: {vector_a}\nВектор B: {vector_b}\nСкалярное произведение <A, B>={dot_product}')

```

```

Вектор A: [[93 67  6 64 96]]
Вектор B: [[34]
 [25]
 [14]
 [85]
 [29]]
Скалярное произведение <A, B>=[[13145]]

```

Задание 3:

```
### BEGIN YOUR CODE
```

```
Student_ID = 3
```

```
### END YOUR CODE
```

Датасет 'Chipotle' доступен по следующей ссылке: <https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv>
В заданиях ниже, где нужно выбрать вопрос, всегда выбирайте вопрос № 1

```
### BEGIN YOUR CODE
```

```
!wget https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv
```

```
# !unzip ...
```

```
# !tar chipotle.tsv
```

```
# !gunzip ...
```

```
### END YOUR CODE
```

```
--2025-03-14 01:07:16-- https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.109.133, 185.199.
110.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.108.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 364975 (356K) [text/plain]
Saving to: 'chipotle.tsv'
```

```
chipotle.tsv      100%[=====>] 356.42K  --.-KB/s    in 0.04s
```

```
2025-03-14 01:07:16 (8.09 MB/s) - 'chipotle.tsv' saved [364975/364975]
```

```
import pandas as pd
```

```
df = pd.read_csv('chipotle.tsv', sep='\t')
```

```
df.columns = df.columns.str.lower()
```

```
# Place your code to read the dataset here. Don't forget to import all the
libraries you need!
```

```
columns = df.columns # Place your code here instead of '...'
```

```
print(columns)
```

```
Index(['order_id', 'quantity', 'item_name', 'choice_description',
       'item_price'],
      dtype='object')
```

1. Какое блюдо (item_name) заказывали чаще всего?

```
answer1 = df['item_name'].value_counts().idxmax() # Place your code here instead of  
'...'
```

```
print(answer1)
```

Chicken Bowl

1. Сколько всего было заказов блюда, название которого сохранено в answer1?

```
answer2 = df['item_name'].value_counts()[answer1]  
print(answer2)
```

726

1. Какой доход получила сеть Chipotle Mexican Grill на заказах, попавших в датасет?

```
answer3 = df['item_price'] = df['item_price'].str.replace('$', '').astype(float)  
answer3 = df['item_price'].sum()  
print(answer3)
```

34500.16

1. Каков средний доход с одного заказа?

```
answer4 = df.groupby('order_id')['item_price'].sum().mean()  
print(answer4)
```

18.811428571428568

1. Сколько раз был заказан самый популярный напиток (Coke, Sprite, Mountain Dew и т.п.)?

```
drinks_df = df[df['item_name'].isin(['Canned Soda', 'Canned Soft Drink', '6 Pack  
Soft Drink'])]  
drink_counts = drinks_df.groupby('choice_description')['quantity'].sum()  
  
answer5 = drink_counts.max()  
print(answer5)
```

159

1. Какой суммарный доход принесли напитки в заказах вегетарианцев?

```
veggie_items = ['Veggie Bowl', 'Veggie Burrito', 'Veggie Salad Bowl']
veggie_orders = df[df['item_name'].isin(veggie_items)]['order_id'].unique()
veggie_orders_df = df[df['order_id'].isin(veggie_orders)]

drink_items = ['Canned Soda', 'Canned Soft Drink', '6 Pack Soft Drink', 'Bottled Water', 'Izze', 'Nantucket Nectar']

veggie_drinks_df =
veggie_orders_df[veggie_orders_df['item_name'].isin(drink_items)]

answer6 = (veggie_drinks_df['item_price'] * veggie_drinks_df['quantity']).sum()
print(answer6)
```

399.37

1. Сколько было сделано вегетарианских заказов? Заказ не считается вегетарианским, если в нем были не вегетарианские блюда.

```
veggie_items = ['Veggie Bowl', 'Veggie Burrito', 'Veggie Salad Bowl', 'Chips',
                'Chips and Guacamole',
                'Chips and Fresh Tomato Salsa', 'Chips and Tomatillo-Green Chili Salsa',
                'Chips and Tomatillo Red Chili Salsa', 'Side of Chips', 'Canned Soda',
                'Canned Soft Drink', '6 Pack Soft Drink', 'Bottled Water', 'Izze',
                'Nantucket Nectar']

answer7 = df.groupby('order_id')['item_name'].apply(lambda x: all(item in
veggie_items for item in x)).sum()
print(answer7)
```

73

1. Какой соус или дополнительный ингредиент по выбору (choice_description) чаще всего берут вместе с буррито с курицей (Chicken Burrito)?

```
chicken_burrito_df = df[df['item_name'] == 'Chicken Burrito']

answer8 = chicken_burrito_df['choice_description'].value_counts().idxmax()

print(answer8)

[Fresh Tomato Salsa, [Rice, Black Beans, Cheese, Sour Cream, Lettuce]]
```

1. Постройте гистограмму распределения общей стоимости заказов. Найти и отметить на графике средний чек и медианную стоимость заказа.

```
import matplotlib.pyplot as plt

order_totals = df.groupby('order_id')['item_price'].sum()

plt.figure(figsize=(10, 6))

plt.hist(order_totals, bins=50, color='skyblue', edgecolor='black')

mean_total = order_totals.mean()
median_total = order_totals.median()

plt.axvline(mean_total, color='red', linestyle='--', label=f'Средний чек:
${mean_total:.2f}')

plt.axvline(median_total, color='green', linestyle='--', label=f'Медиана:
${median_total:.2f}')

plt.title('Распределение общей стоимости заказов')
plt.xlabel('Общая стоимость заказа ($)')
plt.ylabel('Количество заказов')
plt.legend()
```

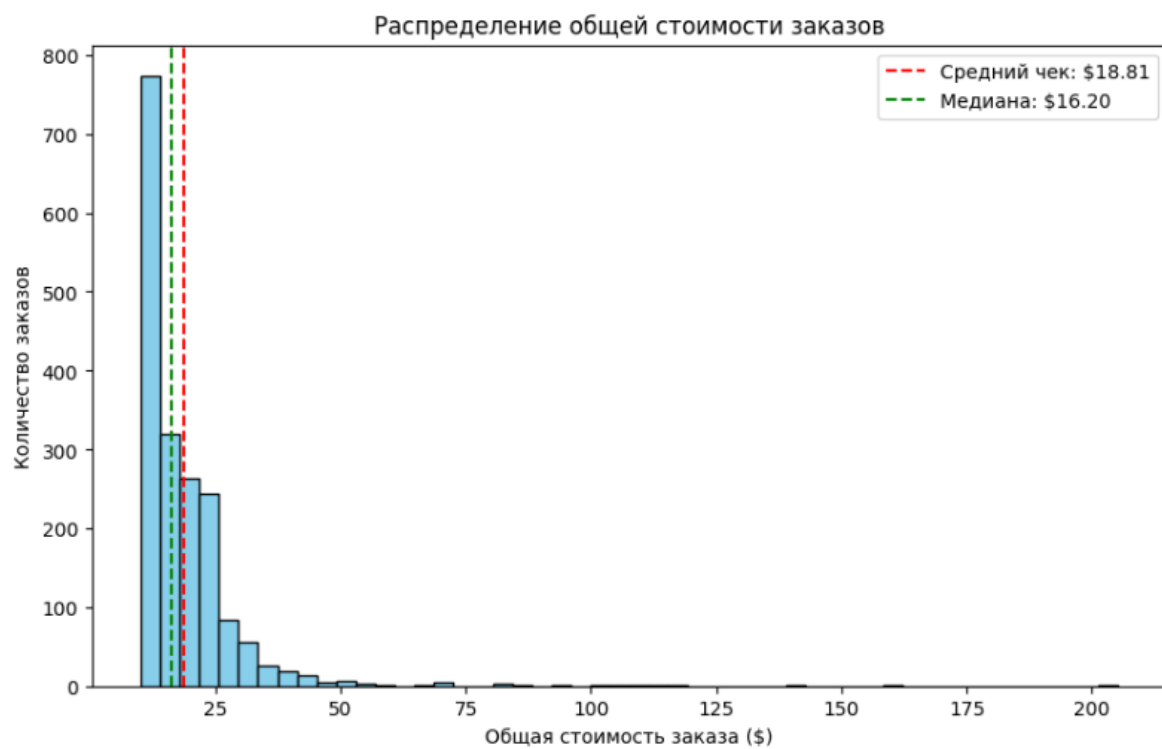



Рисунок 3 – Гистограмма распределения общей стоимости заказов

4. Вывод

В ходе выполнения данной лабораторной работы я познакомилась с современными инструментами для анализа данных и визуализации. Использование Jupyter Notebook и Google Colaboratory предоставляет множество преимуществ, таких как интерактивность, возможность работы в облаке и доступ к мощным вычислительным ресурсам без необходимости установки программного обеспечения на локальном компьютере. Также я получила практические навыки работы с библиотеками Pandas и Matplotlib, с помощью них выполнила различные задания и построила графики по итоговым результатам.