



Министерство науки и высшего образования Российской Федерации  
Калужский филиал  
федерального государственного бюджетного  
образовательного учреждения высшего образования  
«Московский государственный технический университет имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(КФ МГТУ им. Н.Э. Баумана)

**ФАКУЛЬТЕТ ИУК «Информатика и управление»**

**КАФЕДРА ИУК4 «Программное обеспечение ЭВМ, информационные технологии»**

## **ЛАБОРАТОРНАЯ РАБОТА №1**

**«Обработка данных средствами библиотек NumPy и Pandas.  
Визуализация данных»**

**ДИСЦИПЛИНА: «Технологии анализа данных»**

Выполнил: студент гр. ИУК4-82Б \_\_\_\_\_ ( Карельский М.К. )  
(Подпись)

Проверил: \_\_\_\_\_ ( Ерохин И.И. )  
(Подпись)

Дата сдачи (защиты):

Результаты сдачи (защиты):

- Балльная оценка:

- Оценка:

Калуга, 2024

**Цель:** формирование практических навыков работы с библиотеками numpy и pandas, а также применения средств визуализации данных языка Python.

**Задачи:**

1. Ознакомиться с функциональными возможностями библиотек numpy и pandas.
2. Изучить средства визуализации языка Python.

**Вариант 5**

Считать данные из CSV файла в структуру DataFrame. Добавить в DataFrame еще один столбец, в котором содержится количество символов в строке идентификатора продукта. Определить возрастную группу покупателя (мужчина), который приобрел больше всего товаров (по стоимости). Построить график зависимости потраченных средств покупателей (среднее значение) от возрастной группы. Вывести результаты для 5 покупателей с самыми большими расходами и 5 с самыми маленькими. Построить трехмерную поверхность – цилиндр (параметры цилиндра задать произвольно).

**Листинг:**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

frame = pd.read_csv("data.csv")
frame = frame[["User ID", "Product ID", "Gender", "Age", "Purchase"]]
frame["ID_symbols"] = frame["Product_ID"].str.len()
print(frame.head())

grouped = frame[frame.Gender == "M"][["User ID",
"Purchase"]].groupby(["User_ID"], as_index=False).sum()
id = grouped[grouped.Purchase == grouped.Purchase.max()].iloc[0]["User_ID"]
print(id)
print(frame[frame.User_ID == id].iloc[0]["Age"])

grouped = frame[["Age", "Purchase"]].groupby(["Age"], as_index=False).mean()
grouped.plot(x="Age", y="Purchase")
plt.savefig("age_purchase.pdf")

grouped = frame[["User_ID", "Purchase"]].groupby(["User_ID"],
as_index=False).sum()
print(grouped.nlargest(5, "Purchase"))
print(grouped.nsmallest(5, "Purchase"))

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

def data_for_cylinder_along_z(center_x,center_y,radius,height_z):
```

```

z = np.linspace(0, height_z, 50)
theta = np.linspace(0, 2*np.pi, 50)
theta_grid, z_grid=np.meshgrid(theta, z)
x_grid = radius*np.cos(theta_grid) + center_x
y_grid = radius*np.sin(theta_grid) + center_y
return x_grid,y_grid,z_grid

Xc,Yc,Zc = data_for_cylinder_along_z(0.2,0.2,0.05,0.1)
ax.plot_surface(Xc, Yc, Zc, alpha=0.5)

plt.show()

```

## Результат:

	User_ID	Product_ID	Gender	Age	Purchase	ID_symbols
0	1000001	P00069042	F	0-17	8370	9
1	1000001	P00248942	F	0-17	15200	9
2	1000001	P00087842	F	0-17	1422	9
3	1000001	P00085442	F	0-17	1057	9
4	1000002	P00285442	M	55+	7969	9

Рис. 1. Дополнительный столбец

```

ID: 1004277
Age: 36-45

```

Рис. 2. Возрастная группа покупателя

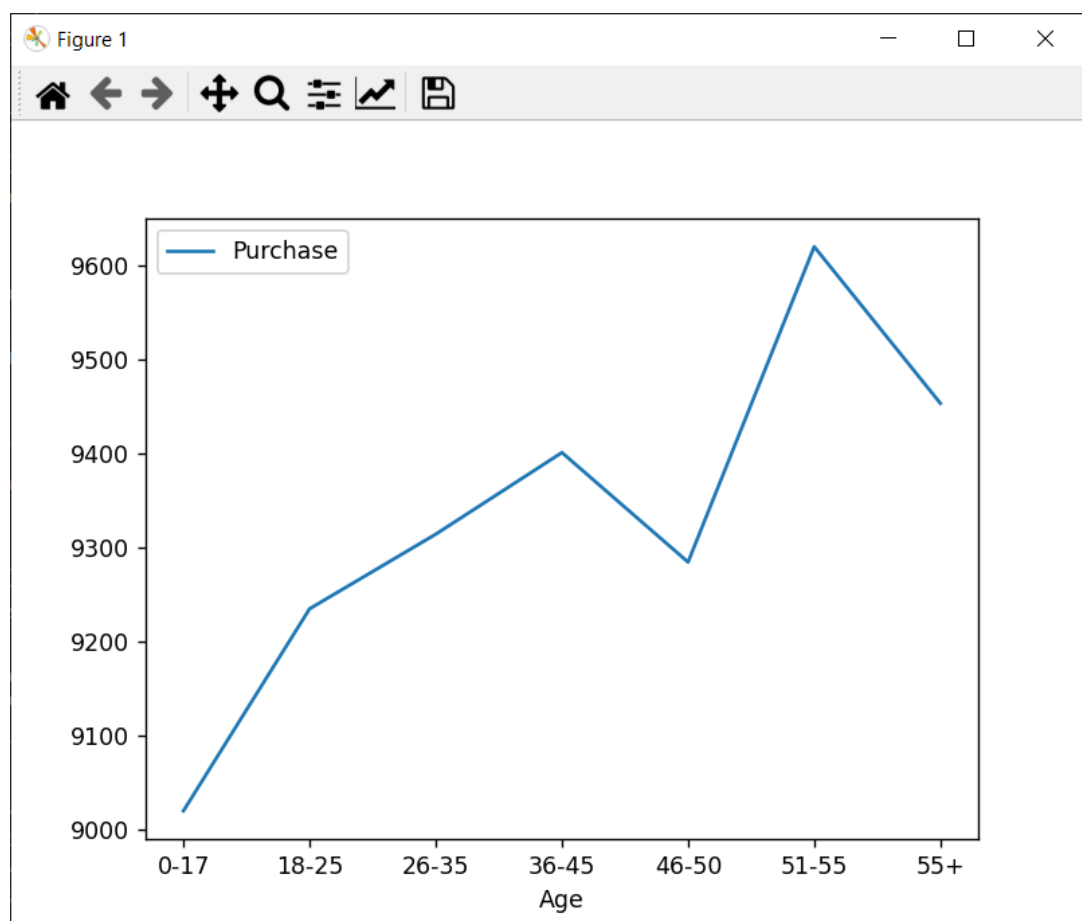
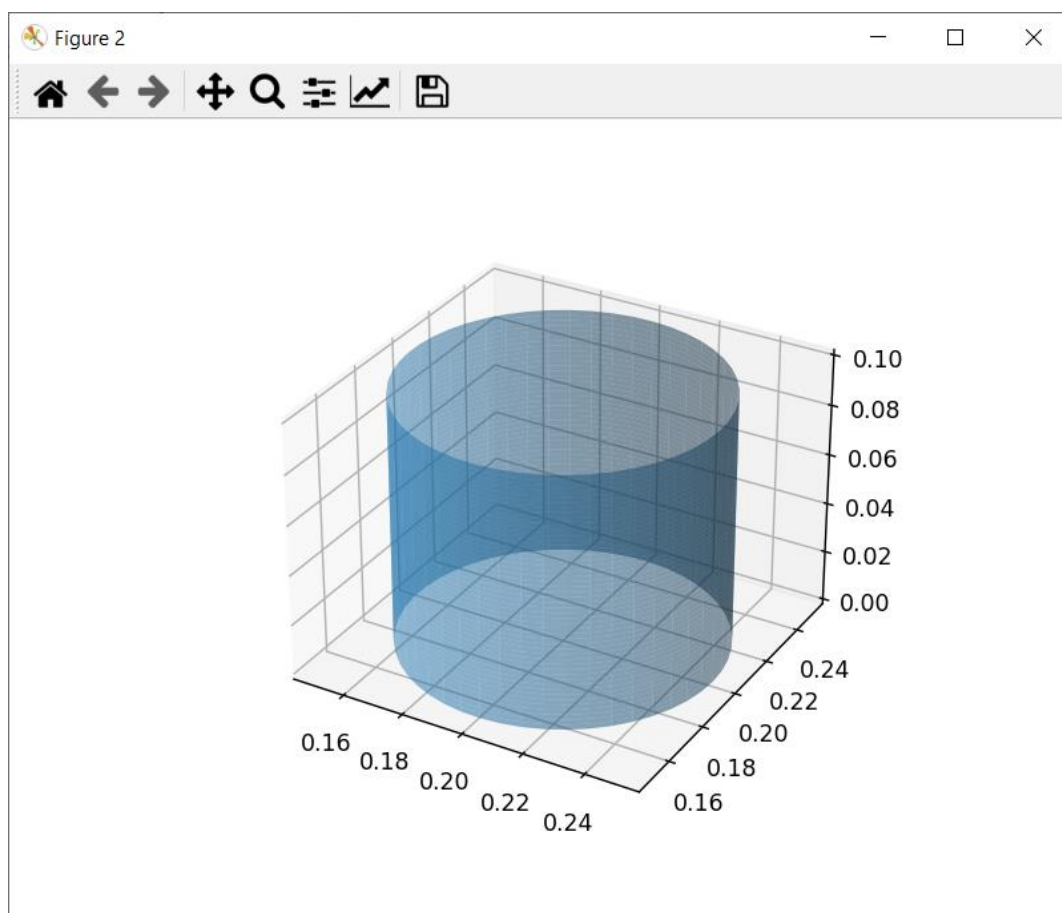


Рис. 3. График

	User_ID	Purchase
4166	1004277	10536783
1634	1001680	8699232
2831	1002909	7577505
1885	1001941	6817493
416	1000424	6573609
	User_ID	Purchase
4866	1004991	44108
4989	1005117	44432
4887	1005012	45551
5794	1005944	46070
4351	1004464	46091

**Рис. 4.** Самый большие и маленькие расходы



**Рис. 5.** Трехмерная поверхность

**Вывод:** в ходе выполнения лабораторной работы были получен практические навыки работы с библиотеками numpy и pandas, а также применения средств визуализации данных языка Python.