|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **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

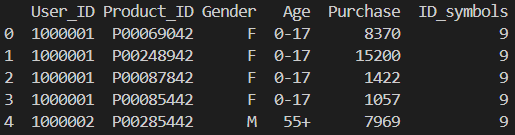
    return x\_grid,y\_grid,z\_grid

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()

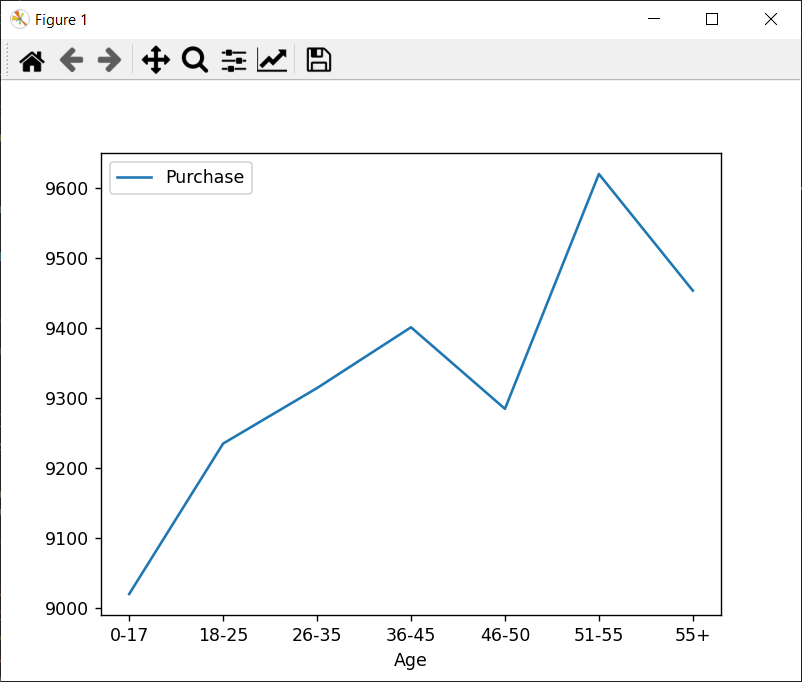
**Результат:**



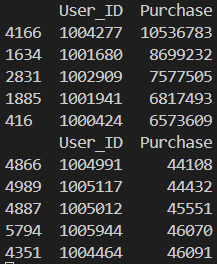
**Рис. 1.** Дополнительный столбец



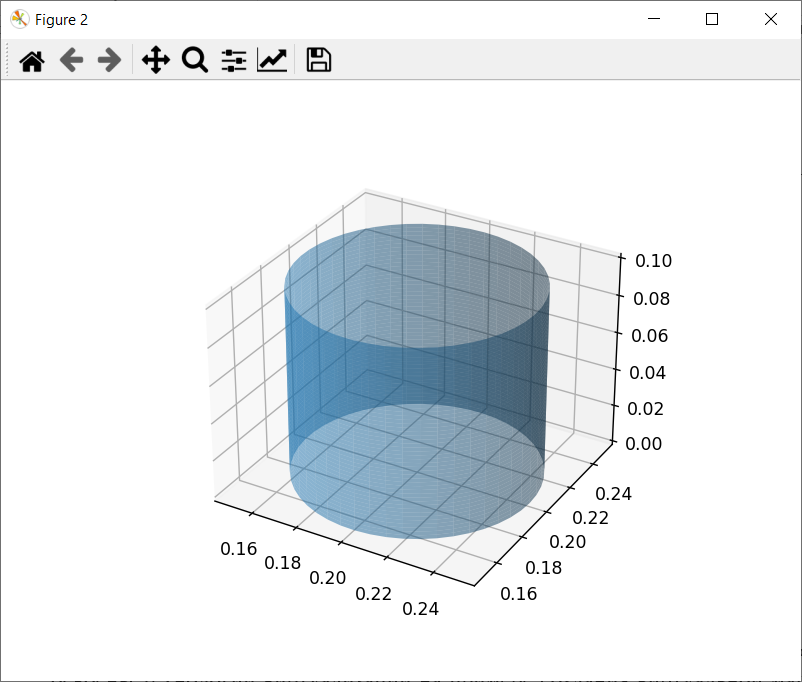
**Рис. 2.** Возрастная группа покупателя



**Рис. 3.** График



**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.