



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.04.01 Информатика и вычислительная техника

МАГИСТЕРСКАЯ ПРОГРАММА 09.04.01/05 Современные интеллектуальные
программно-аппаратные комплексы.

О Т Ч Е Т

по лабораторной работе № 1

Дисциплина: Методы интерпретации и визуализации данных в СППР

Студент

ИУ6-31М

(Группа)

(Подпись, дата)

И.С. Марчук

(И.О. Фамилия)

Преподаватель

(Подпись, дата)

М.А. Захаров

(И.О. Фамилия)

Москва, 2024

Цель: изучить библиотеку matplotlib языка программирования python на примере Pie Charts, Box Plots, Scatter Plots, Bubble Plots.

Ход работы

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

# Чтение файла CSV
df = pd.read_csv('/content/diabetes_dataset.csv')

# Вывод первых нескольких строк таблицы
df.head()
```

```
Out[ ]:   year  gender  age  location  race:AfricanAmerican  race:Asian  race:Caucasian  race:
```

0	2020	Female	32.0	Alabama	0	0	0
1	2015	Female	29.0	Alabama	0	1	0
2	2015	Male	18.0	Alabama	0	0	0
3	2015	Male	41.0	Alabama	0	0	1
4	2016	Female	52.0	Alabama	1	0	0



Вывести информацию про

- первых 5 пациентов,
- всех, кроме последних 2х
- с 3 по 15 включительно

```
In [ ]: # 1. Первые 5 пациентов
first_5 = df.iloc[:5]

# 2. Все, кроме последних 2
except_last_2 = df.iloc[:-2]

# 3. С 3 по 15 включительно
from_3_to_15 = df.iloc[2:15]

# Отображаем результаты как DataFrame
from IPython.display import display

print("Первые 5 пациентов:")
display(first_5)

print("\nВсе, кроме последних 2 пациентов:")
display(except_last_2)

print("\nПациенты с 3 по 15 включительно:")
display(from_3_to_15)
```

Первые 5 пациентов:

	year	gender	age	location	race:AfricanAmerican	race:Asian	race:Caucasian	race:Hi
0	2020	Female	32.0	Alabama	0	0	0	
1	2015	Female	29.0	Alabama	0	1	0	
2	2015	Male	18.0	Alabama	0	0	0	
3	2015	Male	41.0	Alabama	0	0	1	
4	2016	Female	52.0	Alabama	1	0	0	

< >

Все, кроме последних 2 пациентов:

	year	gender	age	location	race:AfricanAmerican	race:Asian	race:Caucasian	ra
0	2020	Female	32.0	Alabama	0	0	0	
1	2015	Female	29.0	Alabama	0	1	0	
2	2015	Male	18.0	Alabama	0	0	0	
3	2015	Male	41.0	Alabama	0	0	1	
4	2016	Female	52.0	Alabama	1	0	0	
...
99993	2016	Male	58.0	Wyoming	1	0	0	
99994	2016	Male	62.0	Wyoming	0	1	0	
99995	2018	Female	33.0	Wyoming	0	0	0	
99996	2016	Female	80.0	Wyoming	0	1	0	
99997	2018	Male	46.0	Wyoming	0	1	0	

99998 rows × 16 columns

< >

Пациенты с 3 по 15 включительно:

	year	gender	age	location	race:AfricanAmerican	race:Asian	race:Caucasian	race:Hispanic
2	2015	Male	18.0	Alabama	0	0	0	0
3	2015	Male	41.0	Alabama	0	0	1	0
4	2016	Female	52.0	Alabama	1	0	0	0
5	2016	Male	66.0	Alabama	0	0	1	0
6	2015	Female	49.0	Alabama	0	0	1	0
7	2016	Female	15.0	Alabama	0	0	0	0
8	2016	Male	51.0	Alabama	1	0	0	0
9	2015	Male	42.0	Alabama	0	0	1	0
10	2016	Male	15.0	Alabama	1	0	0	0
11	2016	Female	53.0	Alabama	0	0	1	0
12	2015	Female	3.0	Alabama	1	0	0	0
13	2016	Female	40.0	Alabama	0	0	0	0
14	2016	Female	64.0	Alabama	0	0	0	0

Проанализировать на предмет наличия особенностей пациентов, таких как

1. Средний возраст, максимальный и минимальный
2. Узнать количество женщин и мужчин
3. Количество женщин, которым больше 18 лет
4. Количество мужчин, которые не курят.

```
In [ ]: # 1. Средний возраст, максимальный и минимальный
average_age = df['age'].mean()
max_age = df['age'].max()
min_age = df['age'].min()

print(f"Средний возраст: {average_age:.2f}")
print(f"Максимальный возраст: {max_age}")
print(f"Минимальный возраст: {min_age}")

# Визуализация распределения возраста
plt.figure(figsize=(8, 5))
plt.hist(df['age'], bins=20, color='skyblue', edgecolor='black')
plt.axvline(average_age, color='red', linestyle='--', label=f'Средний возраст: {average_age:.2f}')
plt.xlabel('Возраст')
plt.ylabel('Количество')
plt.title('Распределение возраста пациентов')
plt.legend()
plt.show()

# 2. Количество женщин и мужчин
gender_counts = df['gender'].value_counts()
print("\nКоличество мужчин и женщин:")
gender_counts
```

```

# Визуализация гендерного распределения
plt.figure(figsize=(5, 5))
gender_counts.plot(kind='pie', autopct='%1.1f%%', startangle=90, colors=['lightcoral', 'lightblue'])
plt.title('Распределение мужчин и женщин')
plt.ylabel('')
plt.show()

# 3. Количество женщин, которым больше 18 лет
women_above_18 = df[(df['gender'] == 'Female') & (df['age'] > 18)].shape[0]
print(f"\nКоличество женщин старше 18 лет: {women_above_18}")

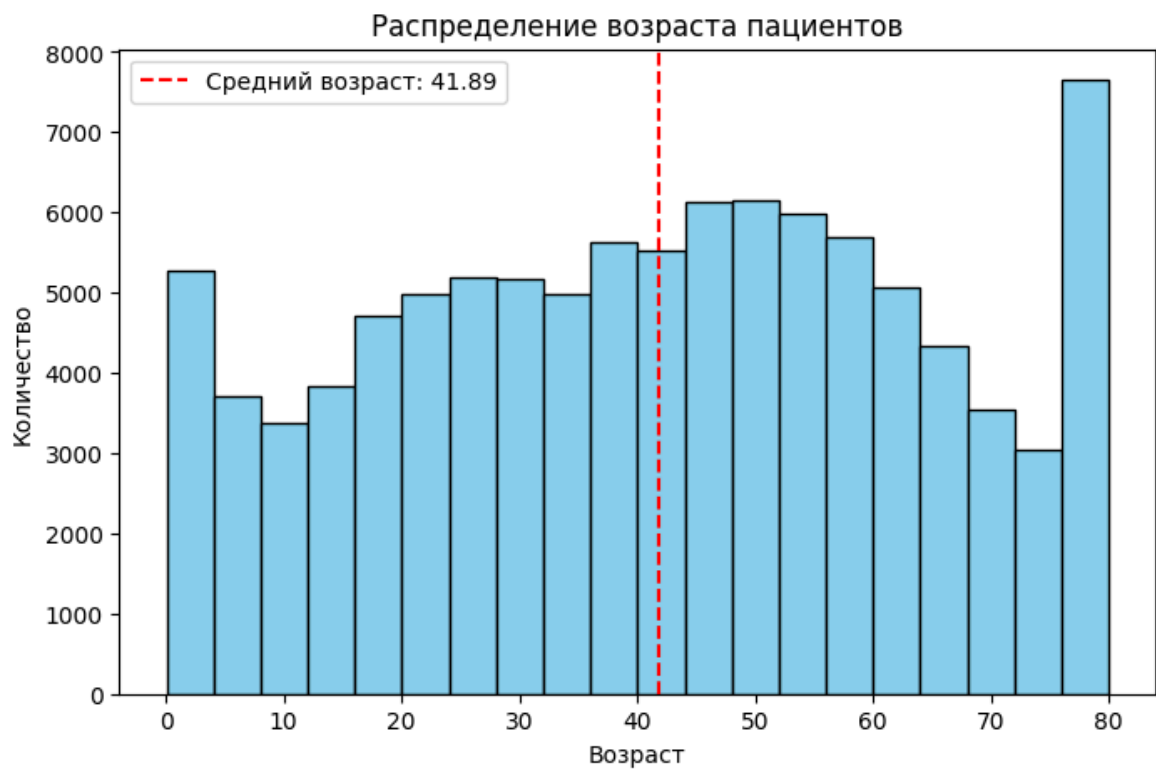
# 4. Количество мужчин, которые не курят
# Предполагаем, что "не курят" указано в столбце smoking_history как "never"
non_smoking_men = df[(df['gender'] == 'Male') & (df['smoking_history'] == 'never')].shape[0]
print(f"\nКоличество мужчин, которые не курят: {non_smoking_men}")

```

Средний возраст: 41.89

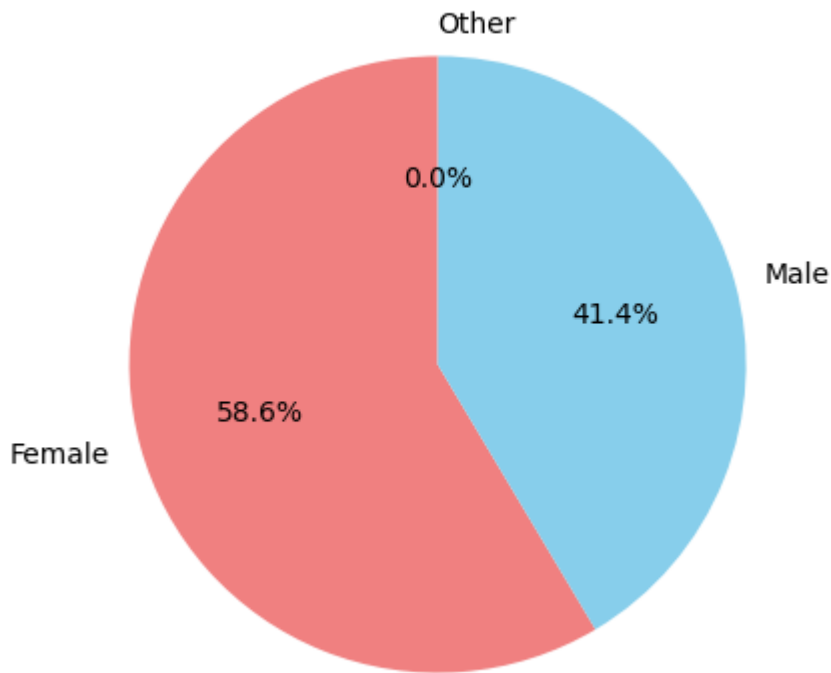
Максимальный возраст: 80.0

Минимальный возраст: 0.08



Количество мужчин и женщин:

Распределение мужчин и женщин



Количество женщин старше 18 лет: 49412

Количество мужчин, которые не курят: 12223

Вывести процент диабетиков в каждом штате

```
In [ ]: # Отбор пациентов с диабетом
diabetic_patients = df[df['diabetes'] == 1]

# Подсчет больных диабетом по штатам
state_counts = diabetic_patients['location'].value_counts()

# Перевод в проценты
state_percentage = (state_counts / state_counts.sum()) * 100

print("Процентное распределение больных диабетом по штатам:")
print(state_percentage)

# график
plt.figure(figsize=(8, 15))
bars = state_percentage.plot(kind='barh', color='teal', edgecolor='black')
plt.xlabel('Процент больных диабетом (%)')
plt.ylabel('Штат')
plt.title('Процентное распределение больных диабетом по штатам')
plt.gca().invert_yaxis()

# Добавляем процентные значения на каждый столбец
for index, value in enumerate(state_percentage):
    plt.text(value+0.05, index, f'{value:.2f}%', va='center') # Расположение те

plt.show()
```

Процентное распределение больных диабетом по штатам:

location	
Delaware	2.352941
Kansas	2.341176
Illinois	2.294118
Montana	2.282353
Rhode Island	2.258824
Kentucky	2.258824
Hawaii	2.188235
New Mexico	2.164706
Mississippi	2.164706
Maryland	2.164706
Connecticut	2.117647
District of Columbia	2.094118
North Dakota	2.082353
Florida	2.070588
Nebraska	2.058824
Alabama	2.058824
Arkansas	2.035294
Michigan	2.035294
Alaska	2.035294
Minnesota	2.023529
Georgia	2.023529
Pennsylvania	2.000000
Maine	2.000000
Nevada	1.988235
Louisiana	1.964706
Oregon	1.964706
Indiana	1.964706
Missouri	1.964706
New Jersey	1.917647
Colorado	1.917647
Iowa	1.905882
California	1.894118
North Carolina	1.882353
South Carolina	1.870588
Massachusetts	1.870588
South Dakota	1.858824
Ohio	1.823529
New Hampshire	1.776471
Idaho	1.752941
Tennessee	1.717647
Oklahoma	1.670588
Arizona	1.647059
New York	1.647059
United States	1.541176
Vermont	1.470588
Virginia	1.470588
Washington	1.423529
Utah	1.411765
Guam	1.294118
Texas	1.282353
West Virginia	1.258824
Puerto Rico	1.223529
Virgin Islands	0.847059
Wyoming	0.376471
Wisconsin	0.294118

Name: count, dtype: float64

Процентное распределение больных диабетом по штатам

