

Рубежный контроль №1

Койбаев Тамерлан ИУ5-61Б

Задание.

Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель. Для пары произвольных колонок данных построить график "Диаграмма рассеяния"

```
#Импорт библиотек
import pandas as pd
from sklearn.preprocessing import LabelEncoder
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

В качестве датасета возьмём данные о болезнях сердца

```
from google.colab import drive
drive.mount('/content/drive')
data = pd.read_csv('/content/drive/My Drive/heart.csv', sep=",")

Mounted at /content/drive
```

```
data.head()
```

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Next steps: [View recommended plots](#)

Расшифровка полей:

- age
- sex
- chest pain type (4 values)
- resting blood pressure
- serum cholestoral in mg/dl
- fasting blood sugar > 120 mg/dl
- resting electrocardiographic results (values 0,1,2)
- maximum heart rate achieved
- exercise induced angina
- oldpeak = ST depression induced by exercise relative to rest the slope of the peak exercise ST segment
- number of major vessels (0-3) colored by flourosopy
- thal: 0 = normal; 1 = fixed defect; 2 = reversable defect

```
data.shape

(1025, 14)
```

Целевым признаком будет наличие сердечных заболеваний у пациента (поле "target")

## Проверка типов данных и наличие пропусков

```
#Проверка типов
data.dtypes
```

```
age          int64
sex          int64
cp           int64
trestbps     int64
chol         int64
fbs          int64
restecg      int64
thalach      int64
exang        int64
oldpeak      float64
slope        int64
ca           int64
thal         int64
target       int64
dtype: object
```

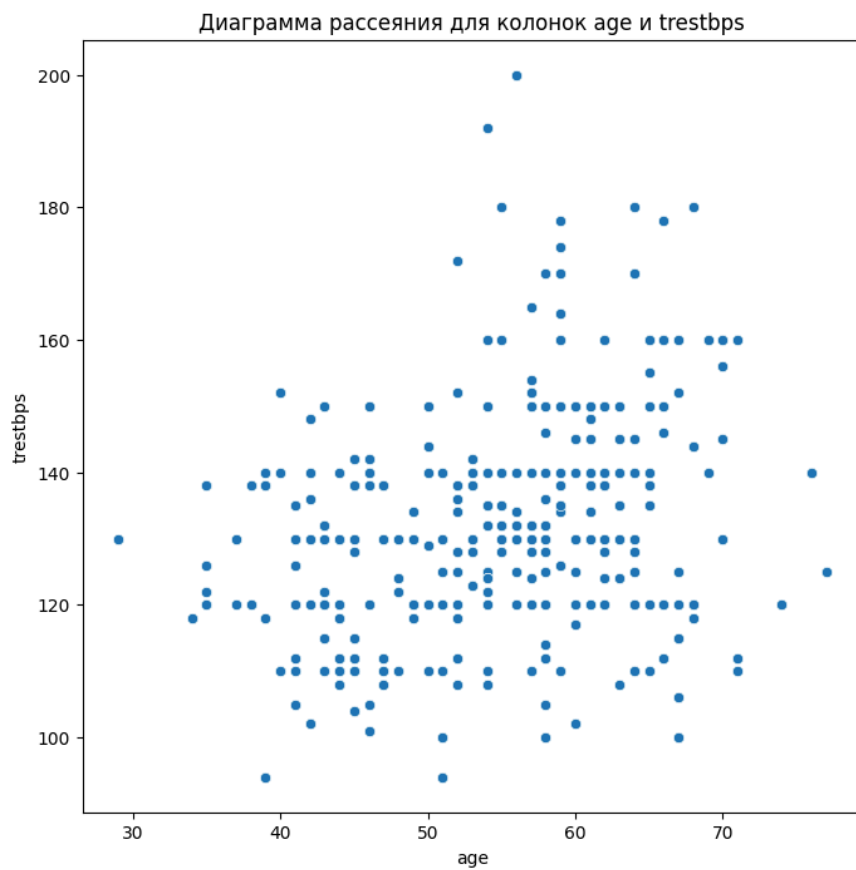
```
data.isnull().sum()
```

```
age          0
sex          0
cp           0
trestbps     0
chol         0
fbs          0
restecg      0
thalach      0
exang        0
oldpeak      0
slope        0
ca           0
thal         0
target       0
dtype: int64
```

## ▼ Диаграмма рассеяния

```
fig, ax = plt.subplots(figsize=(8,8))
ax.set_title("Диаграмма рассеяния для колонок age и trestbps")
sns.scatterplot(ax=ax, x='age', y='trestbps', data=data)
```

```
<Axes: title={'center': 'Диаграмма рассеяния для колонок age и trestbps'},  
xlabel='age', ylabel='trestbps'>
```



```
sns.pairplot(data)
```

