

РК1 Домрачева Ксения ИУ5-65Б

Вариант 4

```
In [14]: import pandas as pd
import seaborn as sns
```

```
In [3]: df = pd.read_csv('heart.csv')
```

```
In [4]: df
```

Out[4]:

	age	sex	cp	trestbps	chol	fbs	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
...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

1025 rows × 14 columns

```
In [5]: df.head()
```

Out[5]:

	age	sex	cp	trestbps	chol	fbs	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

```
In [6]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   age         1025 non-null   int64
```

```
1 sex      1025 non-null int64
2 cp       1025 non-null int64
3 trestbps 1025 non-null int64
4 chol     1025 non-null int64
5 fbs      1025 non-null int64
6 restecg  1025 non-null int64
7 thalach  1025 non-null int64
8 exang     1025 non-null int64
9 oldpeak  1025 non-null float64
10 slope   1025 non-null int64
11 ca      1025 non-null int64
12 thal    1025 non-null int64
13 target  1025 non-null int64
dtypes: float64(1), int64(13)
memory usage: 112.2 KB
```

Пропущенных значений нет:

```
In [8]: df.isnull().sum()
```

```
Out[8]: 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
```

Корреляционная матрица

```
In [11]: corr = df.corr()
corr.style.background_gradient(cmap='coolwarm')
```

Matplotlib is building the font cache; this may take a moment.

```
Out[11]:
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak
age	1.000000	-0.103240	-0.071966	0.271121	0.219823	0.121243	-0.132696	-0.390227	0.088163	0.208137
sex	-0.103240	1.000000	-0.041119	-0.078974	-0.198258	0.027200	-0.055117	-0.049365	0.139157	0.084687
cp	-0.071966	-0.041119	1.000000	0.038177	-0.081641	0.079294	0.043581	0.306839	-0.401513	-0.174733
trestbps	0.271121	-0.078974	0.038177	1.000000	0.127977	0.181767	-0.123794	-0.039264	0.061197	0.187434
chol	0.219823	-0.198258	-0.081641	0.127977	1.000000	0.026917	-0.147410	-0.021772	0.067382	0.064880
fbs	0.121243	0.027200	0.079294	0.181767	0.026917	1.000000	-0.104051	-0.008866	0.049261	0.010859
restecg	-0.132696	-0.055117	0.043581	-0.123794	-0.147410	-0.104051	1.000000	0.048411	-0.065606	-0.050114
thalach	-0.390227	-0.049365	0.306839	-0.039264	-0.021772	-0.008866	0.048411	1.000000	-0.380281	-0.349796
exang	0.088163	0.139157	-0.401513	0.061197	0.067382	0.049261	-0.065606	-0.380281	1.000000	0.310844
oldpeak	0.208137	0.084687	-0.174733	0.187434	0.064880	0.010859	-0.050114	-0.349796	0.310844	1.000000
slope	-0.169105	-0.026666	0.131633	-0.120445	-0.014248	-0.061902	0.086086	0.395308	-0.267335	-0.575181

ca	0.271551	0.111729	-0.176206	0.104554	0.074259	0.137156	-0.078072	-0.207888	0.107849	0.22181
thal	0.072297	0.198424	-0.163341	0.059276	0.100244	-0.042177	-0.020504	-0.098068	0.197201	0.20267
target	-0.229324	-0.279501	0.434854	-0.138772	-0.099966	-0.041164	0.134468	0.422895	-0.438029	-0.43844

Модель построить можно, так как между целевым и некоторыми другими признаками есть корреляционные зависимости. Целевое значение коррелирует с признаками "Тип боли в груди" (0,43), "Максимальная ЧСС" ("thalach" 0,42), "Стенокардия, вызванная физической нагрузкой" ("exang" -0,43), "Депрессия ST, вызванная физической нагрузкой по сравнению с отдыхом" ("oldpeak" -0,43). Эти признаки окажут наибольший вклад в модель. Некоторые значения, с наименьшей корреляцией возможно стоит удалить из модели.

```
In [15]: sns.pairplot(df, hue="target")
```

```
Out[15]: <seaborn.axisgrid.PairGrid at 0x119df8e1780>
```



```
In [ ]:
```