

**Московский государственный технический  
университет им. Н.Э. Баумана**

Отчёт по лабораторной работе №1 по курсу «Технологии машинного  
обучения».

«Разведочный анализ данных. Исследование и визуализация данных».

Выполнил:  
Анцифров Н. С.

Проверил:  
Гапанюк Ю.Е.

Подпись и дата:

Подпись и дата:

Москва, 2022 г.

# 1. Задание лабораторной работы

- Выбрать набор данных (датасет).
- Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных, например из Scikit-learn.
- Для лабораторных работ не рекомендуется выбирать датасеты большого размера.
- Создать ноутбук, который содержит следующие разделы: текстовое описание выбранного Вами набора данных, основные характеристики датасета, визуальное исследование датасета, информация о корреляции признаков.
- Сформировать отчет и разместить его в своем репозитории на GitHub.

## 2. Ячейки Jupyter-ноутбука

### 2.1. Текстовое описание датасета

В качестве датасета (набора данных) будем использовать набор данных, содержащий данные для распознавания вин. Данный набор доступен по адресу: [https://scikit-learn.org/stable/datasets/toy\\_dataset.html#wine-recognition-dataset](https://scikit-learn.org/stable/datasets/toy_dataset.html#wine-recognition-dataset)

Набор данных не содержит пропусков в данных.

Набор данных имеет следующие атрибуты:

- Alcohol - алкоголь
- Malic acid - яблочная кислота
- Ash - зола
- Alcalinity of ash - щелочность
- Magnesium - магний
- Total phenols - количество фенолов
- Flavanoids - флавоноиды
- Nonflavanoid phenols - нефлаваноидные фенолы
- Proanthocyanins - проантоцианы
- Color intensity - интенсивность цвета
- Hue - оттенок
- OD280/OD315 of diluted wines - OD280/OD315 разбавленных вин
- Proline - пролин

#### 2.1.1. Импорт библиотек

Импортируем библиотеки:

```
[1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
from sklearn.datasets import *
import warnings
warnings.filterwarnings('ignore')
```

#### 2.1.2. Загрузка данных

Загрузим набор данных, содержащий информацию для распознавания вин:

```
[2]: wine = load_wine()
```

Проверим загрузку:

```
[3]: type(wine)
```

```
[3]: sklearn.utils.Bunch
```

```
[4]: wine['target_names']
```

```
[4]: array(['class_0', 'class_1', 'class_2'], dtype='<U7')
```

```
[5]: wine['feature_names']
```

```
[5]: ['alcohol',  
      'malic_acid',  
      'ash',  
      'alcalinity_of_ash',  
      'magnesium',  
      'total_phenols',  
      'flavanoids',  
      'nonflavanoid_phenols',  
      'proanthocyanins',  
      'color_intensity',  
      'hue',  
      'od280/od315_of_diluted_wines',  
      'proline']
```

Преобразуем набор данных в Pandas Dataframe:

```
[6]: data_wine = pd.DataFrame(data= np.c_[wine['data'], wine['target']],  
                              columns= wine['feature_names'] + ['target'])
```

```
[7]: data_wine
```

```
[7]:
```

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	\
0	14.23	1.71	2.43	15.6	127.0	2.80	
1	13.20	1.78	2.14	11.2	100.0	2.65	
2	13.16	2.36	2.67	18.6	101.0	2.80	
3	14.37	1.95	2.50	16.8	113.0	3.85	
4	13.24	2.59	2.87	21.0	118.0	2.80	
..	...	...	...	...	...	...	
173	13.71	5.65	2.45	20.5	95.0	1.68	
174	13.40	3.91	2.48	23.0	102.0	1.80	
175	13.27	4.28	2.26	20.0	120.0	1.59	
176	13.17	2.59	2.37	20.0	120.0	1.65	
177	14.13	4.10	2.74	24.5	96.0	2.05	

	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	\
0	3.06	0.28	2.29	5.64	1.04	
1	2.76	0.26	1.28	4.38	1.05	
2	3.24	0.30	2.81	5.68	1.03	
3	3.49	0.24	2.18	7.80	0.86	
4	2.69	0.39	1.82	4.32	1.04	
..	...	...	...	...	...	
173	0.61	0.52	1.06	7.70	0.64	
174	0.75	0.43	1.41	7.30	0.70	
175	0.69	0.43	1.35	10.20	0.59	

176	0.68	0.53	1.46	9.30	0.60
177	0.76	0.56	1.35	9.20	0.61

	od280/od315_of_diluted_wines	proline	target
0	3.92	1065.0	0.0
1	3.40	1050.0	0.0
2	3.17	1185.0	0.0
3	3.45	1480.0	0.0
4	2.93	735.0	0.0
..	...	...	...
173	1.74	740.0	2.0
174	1.56	750.0	2.0
175	1.56	835.0	2.0
176	1.62	840.0	2.0
177	1.60	560.0	2.0

[178 rows x 14 columns]

## 2.2. Основные характеристики датасета

Выведем первые 5 строк датасета:

```
[8]: data_wine.head()
```

```
[8]:   alcohol  malic_acid  ash  alcalinity_of_ash  magnesium  total_phenols  \
0    14.23         1.71  2.43                15.6      127.0           2.80
1    13.20         1.78  2.14                11.2      100.0           2.65
2    13.16         2.36  2.67                18.6      101.0           2.80
3    14.37         1.95  2.50                16.8      113.0           3.85
4    13.24         2.59  2.87                21.0      118.0           2.80
```

	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue
0	3.06	0.28	2.29	5.64	1.04
1	2.76	0.26	1.28	4.38	1.05
2	3.24	0.30	2.81	5.68	1.03
3	3.49	0.24	2.18	7.80	0.86
4	2.69	0.39	1.82	4.32	1.04

	od280/od315_of_diluted_wines	proline	target
0	3.92	1065.0	0.0
1	3.40	1050.0	0.0
2	3.17	1185.0	0.0
3	3.45	1480.0	0.0
4	2.93	735.0	0.0

Определим размер датасета:

```
[9]: data_wine.shape
```

```
[9]: (178, 14)
```

В датасете 178 строк и 14 столбцов. Определим названия столбцов и их тип:

```
[10]: data_wine.columns
```

```
[10]: Index(['alcohol', 'malic_acid', 'ash', 'alcalinity_of_ash', 'magnesium',
          'total_phenols', 'flavanoids', 'nonflavanoid_phenols',
          'proanthocyanins', 'color_intensity', 'hue',
          'od280/od315_of_diluted_wines', 'proline', 'target'],
          dtype='object')
```

```
[11]: data_wine.dtypes
```

```
[11]: alcohol                float64
      malic_acid             float64
      ash                   float64
      alcalinity_of_ash      float64
      magnesium              float64
      total_phenols          float64
      flavanoids              float64
      nonflavanoid_phenols   float64
      proanthocyanins         float64
      color_intensity         float64
      hue                    float64
      od280/od315_of_diluted_wines float64
      proline                 float64
      target                  float64
      dtype: object
```

Проверим наличие пустых значений:

```
[12]: for col in data_wine.columns:
      temp_null_count = data_wine[data_wine[col].isnull()].shape[0]
      print('{} - {}'.format(col, temp_null_count))
```

```
alcohol - 0
malic_acid - 0
ash - 0
alcalinity_of_ash - 0
magnesium - 0
total_phenols - 0
flavanoids - 0
nonflavanoid_phenols - 0
proanthocyanins - 0
color_intensity - 0
hue - 0
od280/od315_of_diluted_wines - 0
proline - 0
target - 0
```

Видим, что пустых значений в датасете нет.

Основные статистические характеристики набора данных:

```
[13]: data_wine.describe()
```

```
[13]:
```

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	\
count	178.000000	178.000000	178.000000	178.000000	178.000000	
mean	13.000618	2.336348	2.366517	19.494944	99.741573	
std	0.811827	1.117146	0.274344	3.339564	14.282484	
min	11.030000	0.740000	1.360000	10.600000	70.000000	

25%	12.362500	1.602500	2.210000	17.200000	88.000000
50%	13.050000	1.865000	2.360000	19.500000	98.000000
75%	13.677500	3.082500	2.557500	21.500000	107.000000
max	14.830000	5.800000	3.230000	30.000000	162.000000

	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	\
count	178.000000	178.000000	178.000000	178.000000	
mean	2.295112	2.029270	0.361854	1.590899	
std	0.625851	0.998859	0.124453	0.572359	
min	0.980000	0.340000	0.130000	0.410000	
25%	1.742500	1.205000	0.270000	1.250000	
50%	2.355000	2.135000	0.340000	1.555000	
75%	2.800000	2.875000	0.437500	1.950000	
max	3.880000	5.080000	0.660000	3.580000	

	color_intensity	hue	od280/od315_of_diluted_wines	proline	\
count	178.000000	178.000000	178.000000	178.000000	
mean	5.058090	0.957449	2.611685	746.893258	
std	2.318286	0.228572	0.709990	314.907474	
min	1.280000	0.480000	1.270000	278.000000	
25%	3.220000	0.782500	1.937500	500.500000	
50%	4.690000	0.965000	2.780000	673.500000	
75%	6.200000	1.120000	3.170000	985.000000	
max	13.000000	1.710000	4.000000	1680.000000	

	target
count	178.000000
mean	0.938202
std	0.775035
min	0.000000
25%	0.000000
50%	1.000000
75%	2.000000
max	2.000000

Определим уникальные значения для целевого признака (сорт вина):

```
[14]: data_wine['target'].unique()
```

```
[14]: array([0., 1., 2.])
```

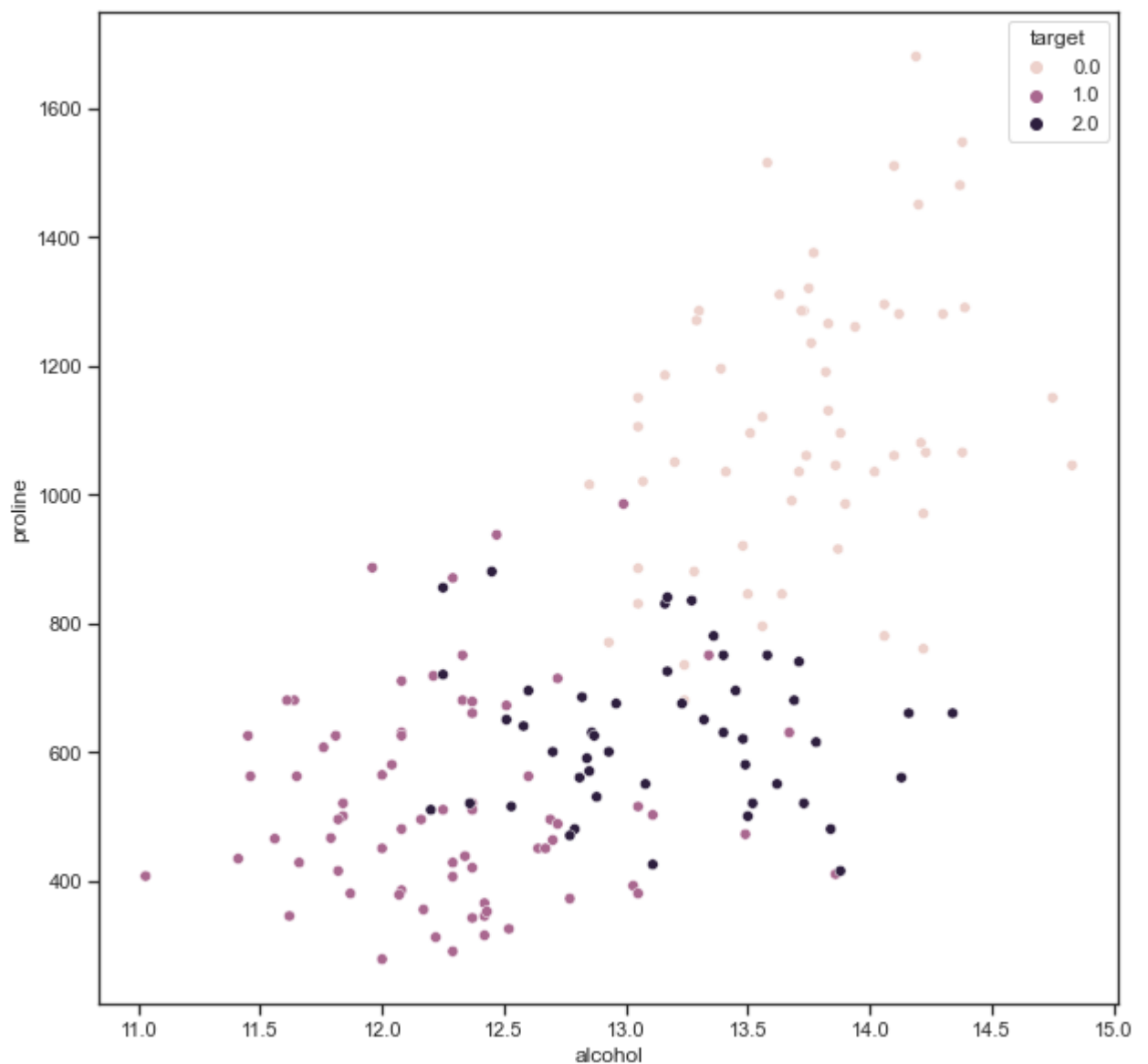
Целевой признак содержит только три значения (три сорта).

## 2.3. Визуальное исследование датасета

Диаграмма рассеяния - распределение двух столбцов данных и отображение визуальной зависимости между ними:

```
[15]: fig, ax = plt.subplots(figsize=(10,10))
      sns.scatterplot(ax=ax, x='alcohol', y='proline', hue='target', data=data_wine)
```

```
[15]: <AxesSubplot:xlabel='alcohol', ylabel='proline'>
```

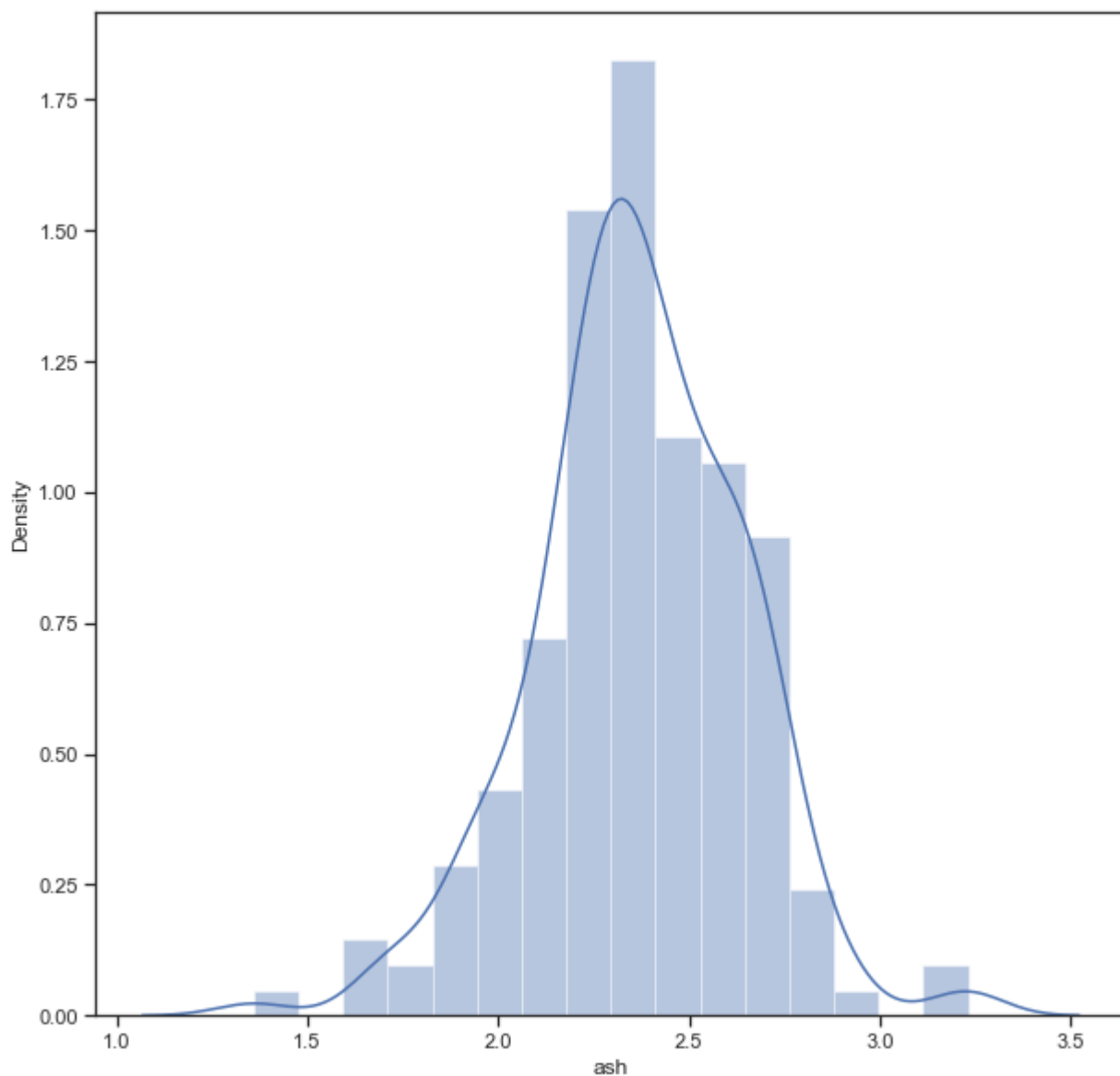


Из диаграммы можно сделать частичный вывод о том, что чем больше алкоголя в напитке, тем больше в нём пролина. Причём также наблюдается зависимость между 3 сортами напитка (на диаграмме разница по цвету).

Гистограмма отображает плотность вероятности распределения данных:

```
[16]: fig, ax = plt.subplots(figsize=(10,10))
      sns.distplot(data_wine['ash'])
```

```
[16]: <AxesSubplot:xlabel='ash', ylabel='Density'>
```



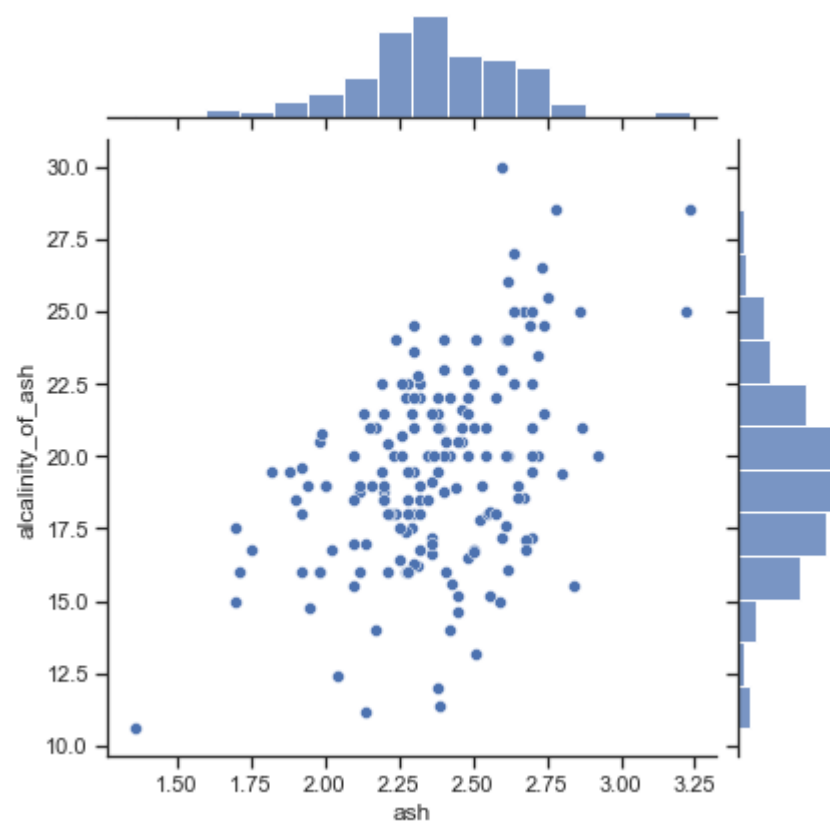
Видно распределение золы в напитках.

Комбинация гистограмм и диаграмм рассеивания выполняется с помощью `jointplot`:

```
[17]: sns.jointplot(x='ash', y='alcalinity_of_ash', data=data_wine)
```

```
[17]: <seaborn.axisgrid.JointGrid at 0x2257e6385e0>
```

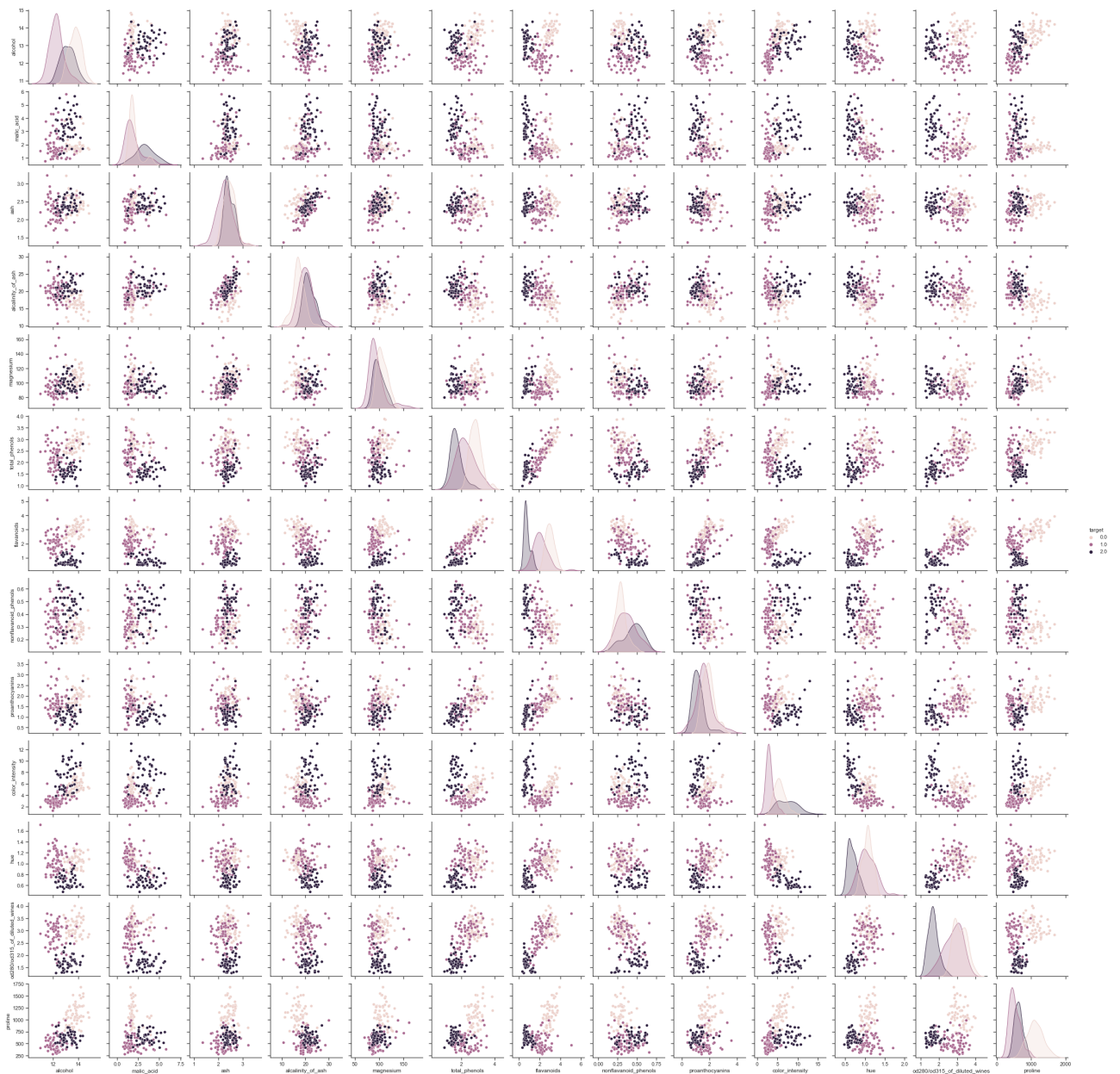




Данные можно представить в виде парных диаграмм - матрицы графиков:

```
[18]: sns.pairplot(data_wine, hue="target")
```

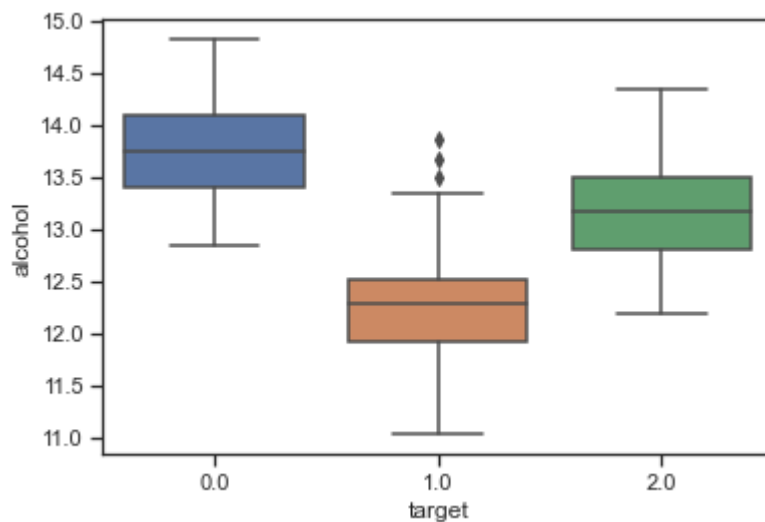
```
[18]: <seaborn.axisgrid.PairGrid at 0x2257edec0d0>
```



Отображение в виде “Ящика с усами”:

```
[19]: sns.boxplot(x='target', y='alcohol', data=data_wine)
```

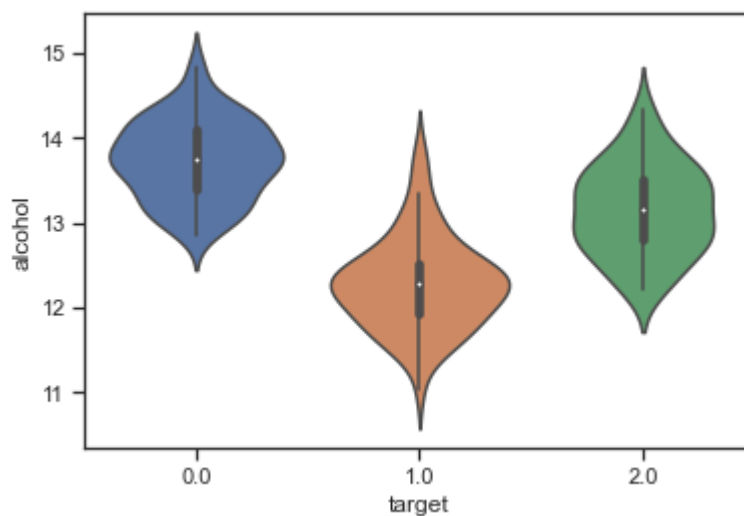
```
[19]: <AxesSubplot:xlabel='target', ylabel='alcohol'>
```



Он показывает количество алкоголя напитков в зависимости от сортов.  
Violin Plot дополнительно показывает распределение плотности:

```
[20]: sns.violinplot(x='target', y='alcohol', data=data_wine)
```

```
[20]: <AxesSubplot:xlabel='target', ylabel='alcohol'>
```



## 2.4. Информация о корреляции признаков

Проверка корреляции помогает найти корреляции с целевым признаком (информативные для машинного обучения), а также выявить линейно независимые нецелевые признаки.

Построим корреляционную матрицу:

```
[21]: data_wine.corr()
```

```
[21]:
```

	alcohol	malic_acid	ash	\
alcohol	1.000000	0.094397	0.211545	
malic_acid	0.094397	1.000000	0.164045	

ash	0.211545	0.164045	1.000000
alcalinity_of_ash	-0.310235	0.288500	0.443367
magnesium	0.270798	-0.054575	0.286587
total_phenols	0.289101	-0.335167	0.128980
flavanoids	0.236815	-0.411007	0.115077
nonflavanoid_phenols	-0.155929	0.292977	0.186230
proanthocyanins	0.136698	-0.220746	0.009652
color_intensity	0.546364	0.248985	0.258887
hue	-0.071747	-0.561296	-0.074667
od280/od315_of_diluted_wines	0.072343	-0.368710	0.003911
proline	0.643720	-0.192011	0.223626
target	-0.328222	0.437776	-0.049643

	alcalinity_of_ash	magnesium	total_phenols \
alcohol	-0.310235	0.270798	0.289101
malic_acid	0.288500	-0.054575	-0.335167
ash	0.443367	0.286587	0.128980
alcalinity_of_ash	1.000000	-0.083333	-0.321113
magnesium	-0.083333	1.000000	0.214401
total_phenols	-0.321113	0.214401	1.000000
flavanoids	-0.351370	0.195784	0.864564
nonflavanoid_phenols	0.361922	-0.256294	-0.449935
proanthocyanins	-0.197327	0.236441	0.612413
color_intensity	0.018732	0.199950	-0.055136
hue	-0.273955	0.055398	0.433681
od280/od315_of_diluted_wines	-0.276769	0.066004	0.699949
proline	-0.440597	0.393351	0.498115
target	0.517859	-0.209179	-0.719163

	flavanoids	nonflavanoid_phenols \
alcohol	0.236815	-0.155929
malic_acid	-0.411007	0.292977
ash	0.115077	0.186230
alcalinity_of_ash	-0.351370	0.361922
magnesium	0.195784	-0.256294
total_phenols	0.864564	-0.449935
flavanoids	1.000000	-0.537900
nonflavanoid_phenols	-0.537900	1.000000
proanthocyanins	0.652692	-0.365845
color_intensity	-0.172379	0.139057
hue	0.543479	-0.262640
od280/od315_of_diluted_wines	0.787194	-0.503270
proline	0.494193	-0.311385
target	-0.847498	0.489109

	proanthocyanins	color_intensity	hue \
alcohol	0.136698	0.546364	-0.071747
malic_acid	-0.220746	0.248985	-0.561296
ash	0.009652	0.258887	-0.074667
alcalinity_of_ash	-0.197327	0.018732	-0.273955
magnesium	0.236441	0.199950	0.055398
total_phenols	0.612413	-0.055136	0.433681
flavanoids	0.652692	-0.172379	0.543479

nonflavanoid_phenols	-0.365845	0.139057	-0.262640
proanthocyanins	1.000000	-0.025250	0.295544
color_intensity	-0.025250	1.000000	-0.521813
hue	0.295544	-0.521813	1.000000
od280/od315_of_diluted_wines	0.519067	-0.428815	0.565468
proline	0.330417	0.316100	0.236183
target	-0.499130	0.265668	-0.617369

	od280/od315_of_diluted_wines	proline	target
alcohol	0.072343	0.643720	-0.328222
malic_acid	-0.368710	-0.192011	0.437776
ash	0.003911	0.223626	-0.049643
alcalinity_of_ash	-0.276769	-0.440597	0.517859
magnesium	0.066004	0.393351	-0.209179
total_phenols	0.699949	0.498115	-0.719163
flavanoids	0.787194	0.494193	-0.847498
nonflavanoid_phenols	-0.503270	-0.311385	0.489109
proanthocyanins	0.519067	0.330417	-0.499130
color_intensity	-0.428815	0.316100	0.265668
hue	0.565468	0.236183	-0.617369
od280/od315_of_diluted_wines	1.000000	0.312761	-0.788230
proline	0.312761	1.000000	-0.633717
target	-0.788230	-0.633717	1.000000

Можно сделать следующие выводы:

- Целевой признак наиболее сильно коррелирует OD280/OD315 разбавленных вин (-0.78), количеством фенолов (-0.72) и флаваноидами (-0.85) - эти признаки очень важны для модели
- Целевой признак частично коррелирует с нефлаваноидными фенолами (0.49) и проантоцианами (0.5) и щелочностью (0.52) - эти признаки также можно оставить в модели
- Целевой признак слабо коррелирует с алкоголем (-0.33), золой (-0.05), магнием (-0.21) и интенсивностью света (0.27). Такие признаки стоит исключить из модели, так как они ухудшат её качество.

Выше была построена матрица корреляции по Пирсону, но также можно построить матрицы по критерию Кендалла и Спирмена, но разница в значениях будет невелика:

```
[22]: data_wine.corr(method='kendall')
```

```
[22]:
```

	alcohol	malic_acid	ash	\
alcohol	1.000000	0.093844	0.170154	
malic_acid	0.093844	1.000000	0.158178	
ash	0.170154	0.158178	1.000000	
alcalinity_of_ash	-0.212978	0.210119	0.258352	
magnesium	0.250506	0.050869	0.254246	
total_phenols	0.209099	-0.174929	0.089855	
flavanoids	0.191087	-0.211918	0.049474	
nonflavanoid_phenols	-0.109554	0.175129	0.098937	
proanthocyanins	0.133526	-0.168714	0.018240	
color_intensity	0.434353	0.195607	0.187786	
hue	-0.021717	-0.388707	-0.037234	
od280/od315_of_diluted_wines	0.061513	-0.162909	-0.006341	
proline	0.449387	-0.044660	0.171574	
target	-0.238984	0.247494	-0.038085	

	alcalinity_of_ash	magnesium	total_phenols	\
alcohol	-0.212978	0.250506	0.209099	
malic_acid	0.210119	0.050869	-0.174929	
ash	0.258352	0.254246	0.089855	
alcalinity_of_ash	1.000000	-0.121005	-0.256669	
magnesium	-0.121005	1.000000	0.172195	
total_phenols	-0.256669	0.172195	1.000000	
flavanoids	-0.309865	0.161603	0.701999	
nonflavanoid_phenols	0.278091	-0.158361	-0.310443	
proanthocyanins	-0.171404	0.117871	0.466517	
color_intensity	-0.057281	0.241781	0.028264	
hue	-0.239210	0.023760	0.289210	
od280/od315_of_diluted_wines	-0.226253	0.034307	0.478267	
proline	-0.313218	0.343016	0.280203	
target	0.449402	-0.184992	-0.590404	

	flavanoids	nonflavanoid_phenols	\
alcohol	0.191087	-0.109554	
malic_acid	-0.211918	0.175129	
ash	0.049474	0.098937	
alcalinity_of_ash	-0.309865	0.278091	
magnesium	0.161603	-0.158361	
total_phenols	0.701999	-0.310443	
flavanoids	1.000000	-0.378099	
nonflavanoid_phenols	-0.378099	1.000000	
proanthocyanins	0.534615	-0.269189	
color_intensity	0.028674	0.036065	
hue	0.354372	-0.179755	
od280/od315_of_diluted_wines	0.520448	-0.363787	
proline	0.263661	-0.174108	
target	-0.725255	0.379234	

	proanthocyanins	color_intensity	hue	\
alcohol	0.133526	0.434353	-0.021717	
malic_acid	-0.168714	0.195607	-0.388707	
ash	0.018240	0.187786	-0.037234	
alcalinity_of_ash	-0.171404	-0.057281	-0.239210	
magnesium	0.117871	0.241781	0.023760	
total_phenols	0.466517	0.028264	0.289210	
flavanoids	0.534615	0.028674	0.354372	
nonflavanoid_phenols	-0.269189	0.036065	-0.179755	
proanthocyanins	1.000000	-0.014962	0.231071	
color_intensity	-0.014962	1.000000	-0.291561	
hue	0.231071	-0.291561	1.000000	
od280/od315_of_diluted_wines	0.369104	-0.206046	0.324678	
proline	0.204172	0.316632	0.143508	
target	-0.450225	0.065124	-0.479229	

	od280/od315_of_diluted_wines	proline	target
alcohol	0.061513	0.449387	-0.238984
malic_acid	-0.162909	-0.044660	0.247494
ash	-0.006341	0.171574	-0.038085

alcalinity_of_ash	-0.226253	-0.313218	0.449402
magnesium	0.034307	0.343016	-0.184992
total_phenols	0.478267	0.280203	-0.590404
flavanoids	0.520448	0.263661	-0.725255
nonflavanoid_phenols	-0.363787	-0.174108	0.379234
proanthocyanins	0.369104	0.204172	-0.450225
color_intensity	-0.206046	0.316632	0.065124
hue	0.324678	0.143508	-0.479229
od280/od315_of_diluted_wines	1.000000	0.151559	-0.607572
proline	0.151559	1.000000	-0.406260
target	-0.607572	-0.406260	1.000000

```
[23]: data_wine.corr(method='spearman')
```

```
[23]:
```

	alcohol	malic_acid	ash \
alcohol	1.000000	0.140430	0.243722
malic_acid	0.140430	1.000000	0.230674
ash	0.243722	0.230674	1.000000
alcalinity_of_ash	-0.306598	0.304069	0.366374
magnesium	0.365503	0.080188	0.361488
total_phenols	0.310920	-0.280225	0.132193
flavanoids	0.294740	-0.325202	0.078796
nonflavanoid_phenols	-0.162207	0.255236	0.145583
proanthocyanins	0.192734	-0.244825	0.024384
color_intensity	0.635425	0.290307	0.283047
hue	-0.024203	-0.560265	-0.050183
od280/od315_of_diluted_wines	0.103050	-0.255185	-0.007500
proline	0.633580	-0.057466	0.253163
target	-0.354167	0.346913	-0.053988

	alcalinity_of_ash	magnesium	total_phenols \
alcohol	-0.306598	0.365503	0.310920
malic_acid	0.304069	0.080188	-0.280225
ash	0.366374	0.361488	0.132193
alcalinity_of_ash	1.000000	-0.169558	-0.376657
magnesium	-0.169558	1.000000	0.246417
total_phenols	-0.376657	0.246417	1.000000
flavanoids	-0.443770	0.233167	0.879404
nonflavanoid_phenols	0.389390	-0.236786	-0.448013
proanthocyanins	-0.253695	0.173647	0.666689
color_intensity	-0.073776	0.357029	0.011162
hue	-0.352507	0.036095	0.439457
od280/od315_of_diluted_wines	-0.325890	0.056963	0.687207
proline	-0.456090	0.507575	0.419470
target	0.569792	-0.250498	-0.726544

	flavanoids	nonflavanoid_phenols \
alcohol	0.294740	-0.162207
malic_acid	-0.325202	0.255236
ash	0.078796	0.145583
alcalinity_of_ash	-0.443770	0.389390
magnesium	0.233167	-0.236786
total_phenols	0.879404	-0.448013

flavanoids	1.000000	-0.543897
nonflavanoid_phenols	-0.543897	1.000000
proanthocyanins	0.730322	-0.384629
color_intensity	-0.042910	0.059639
hue	0.535430	-0.267813
od280/od315_of_diluted_wines	0.741533	-0.494950
proline	0.429904	-0.270112
target	-0.854908	0.474205

	proanthocyanins	color_intensity	hue \
alcohol	0.192734	0.635425	-0.024203
malic_acid	-0.244825	0.290307	-0.560265
ash	0.024384	0.283047	-0.050183
alcalinity_of_ash	-0.253695	-0.073776	-0.352507
magnesium	0.173647	0.357029	0.036095
total_phenols	0.666689	0.011162	0.439457
flavanoids	0.730322	-0.042910	0.535430
nonflavanoid_phenols	-0.384629	0.059639	-0.267813
proanthocyanins	1.000000	-0.030947	0.342795
color_intensity	-0.030947	1.000000	-0.418522
hue	0.342795	-0.418522	1.000000
od280/od315_of_diluted_wines	0.554031	-0.317516	0.485454
proline	0.308249	0.457096	0.207740
target	-0.570648	0.131170	-0.616570

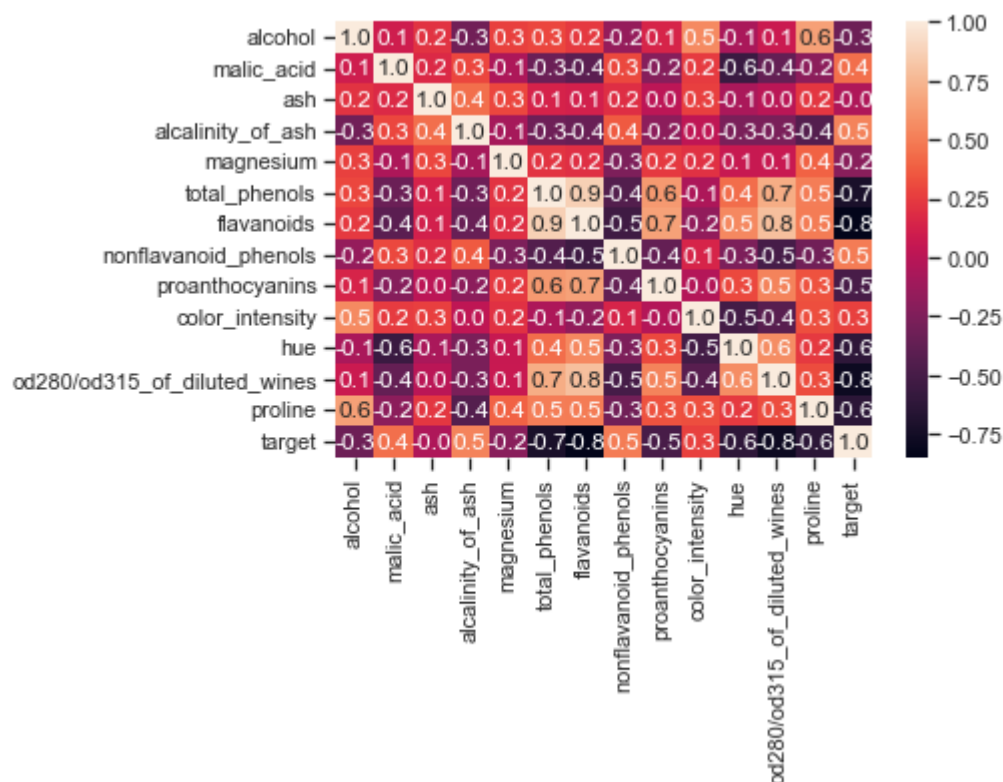
	od280/od315_of_diluted_wines	proline	target
alcohol	0.103050	0.633580	-0.354167
malic_acid	-0.255185	-0.057466	0.346913
ash	-0.007500	0.253163	-0.053988
alcalinity_of_ash	-0.325890	-0.456090	0.569792
magnesium	0.056963	0.507575	-0.250498
total_phenols	0.687207	0.419470	-0.726544
flavanoids	0.741533	0.429904	-0.854908
nonflavanoid_phenols	-0.494950	-0.270112	0.474205
proanthocyanins	0.554031	0.308249	-0.570648
color_intensity	-0.317516	0.457096	0.131170
hue	0.485454	0.207740	-0.616570
od280/od315_of_diluted_wines	1.000000	0.253266	-0.743787
proline	0.253266	1.000000	-0.576383
target	-0.743787	-0.576383	1.000000

Для визуализации корреляционных матриц используют тепловую карту:

```
[24]: sns.heatmap(data_wine.corr(), annot=True, fmt='.1f')
```

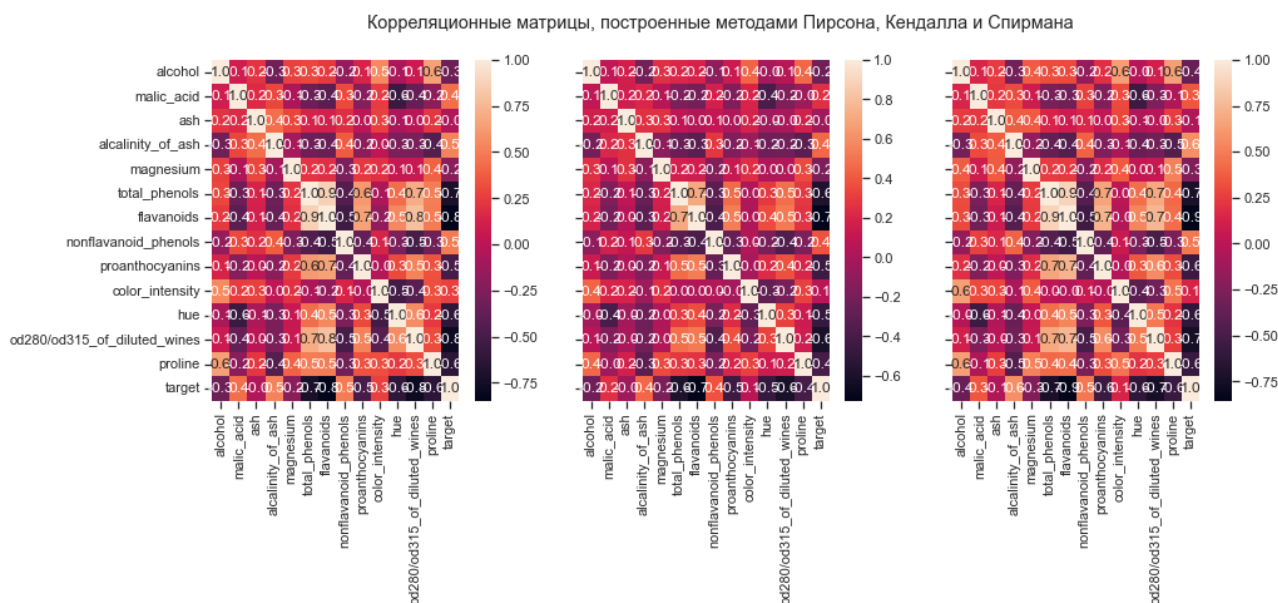
```
[24]: <AxesSubplot:>
```





```
[25]: fig, ax = plt.subplots(1, 3, sharex='col', sharey='row', figsize=(15,5))
sns.heatmap(data_wine.corr(method='pearson'), ax=ax[0], annot=True, fmt='.1f')
sns.heatmap(data_wine.corr(method='kendall'), ax=ax[1], annot=True, fmt='.1f')
sns.heatmap(data_wine.corr(method='spearman'), ax=ax[2], annot=True, fmt='.1f')
fig.suptitle('Корреляционные матрицы, построенные методами Пирсона, Кендалла и
↳Спирмана')
```

[25]: Text(0.5, 0.98, 'Корреляционные матрицы, построенные методами Пирсона, Кендалла и Спирмана')



Также можно вывести треугольную матрицу:

```
[26]: mask = np.zeros_like(data_wine.corr(), dtype=np.bool)
      # чтобы оставить нижнюю часть матрицы - mask[np.triu_indices_from(mask)] = True
      # чтобы оставить верхнюю часть матрицы - mask[np.tril_indices_from(mask)] = True
      mask[np.triu_indices_from(mask)] = True
      sns.heatmap(data_wine.corr(), mask=mask, annot=True, fmt='.1f')
```

[26]: <AxesSubplot:>

