```
### РК1 ИУ5-65Б Бенц Ян
### Задача №1. 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 студентов группы ИУ5-65Б, ИУ5И-65Б - для набора данных
построить "парные диаграммы".
```

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
```

```python
from sklearn.datasets import load_wine
raw = load_wine()
```

```python
raw.feature_names
```

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

```python
raw.target_names
```

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

```python
data = pd.DataFrame(data= np.c_[raw['data'], raw['target']],
                     columns= raw['feature_names'] + ['wine classes'])
```

```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 14 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   alcohol                       178 non-null    float64
 1   malic_acid                    178 non-null    float64
```

```
 2   ash                             178 non-null    float64
 3   alcalinity_of_ash               178 non-null    float64
 4   magnesium                       178 non-null    float64
 5   total_phenols                   178 non-null    float64
 6   flavanoids                      178 non-null    float64
 7   nonflavanoid_phenols            178 non-null    float64
 8   proanthocyanins                 178 non-null    float64
 9   color_intensity                 178 non-null    float64
 10  hue                             178 non-null    float64
 11  od280/od315_of_diluted_wines  178 non-null    float64
 12  proline                         178 non-null    float64
 13  wine classes                    178 non-null    float64
dtypes: float64(14)
memory usage: 19.6 KB
```

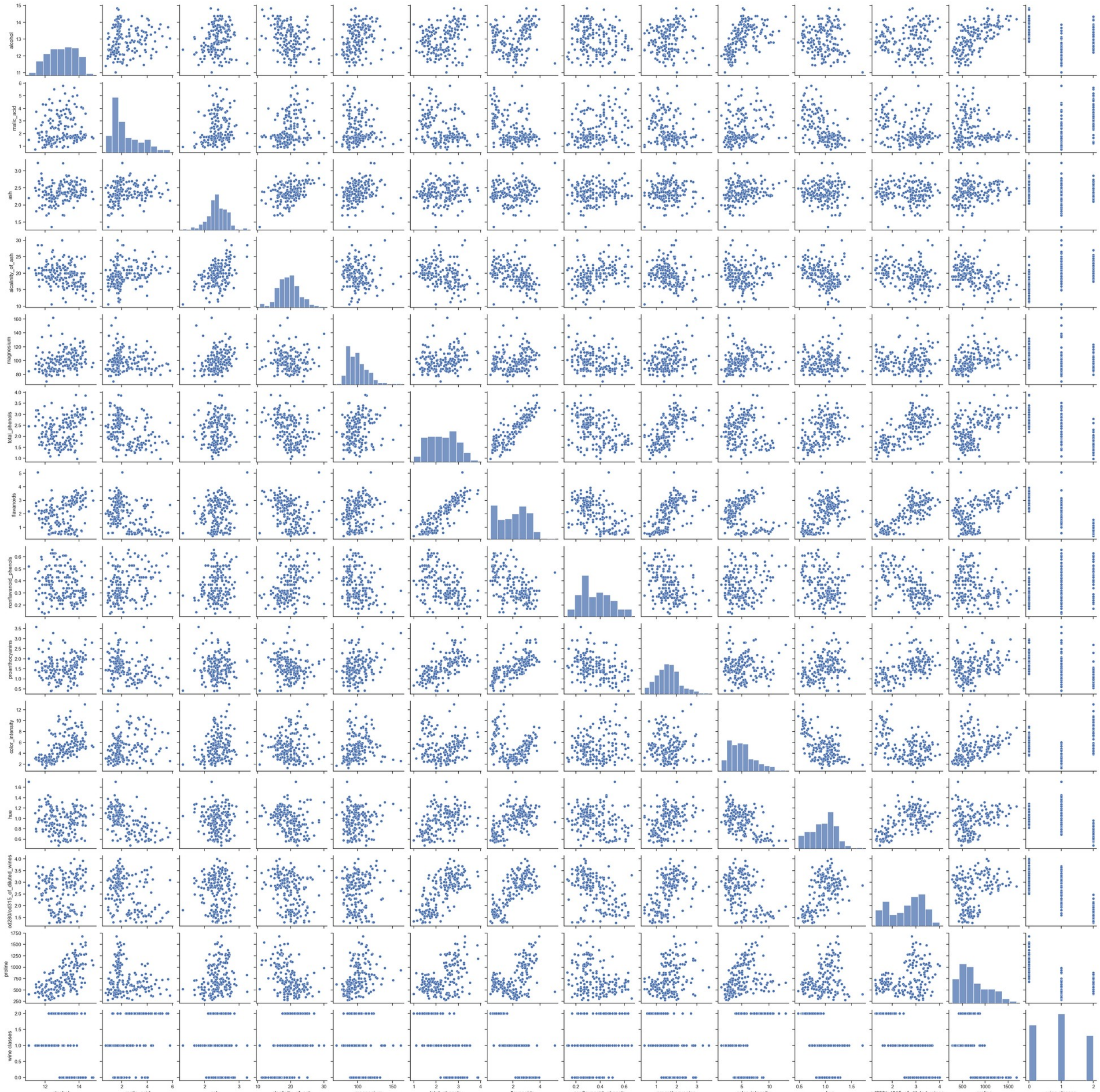
```
### датасет не содержит пропусков

data.describe()
```

```
          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

       wine classes
count    178.000000
mean       0.938202
std        0.775035
min        0.000000
25%        0.000000
50%        1.000000
75%        2.000000
max        2.000000
```

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

```
<seaborn.axisgrid.PairGrid at 0x20cdd1f9dc0>
```




```
sns.pairplot(data, hue="wine classes")
```

```
<seaborn.axisgrid.PairGrid at 0x20cdcca8dd0>
```


```
data.corr()

                               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
wine classes                 -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
wine classes                           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
```

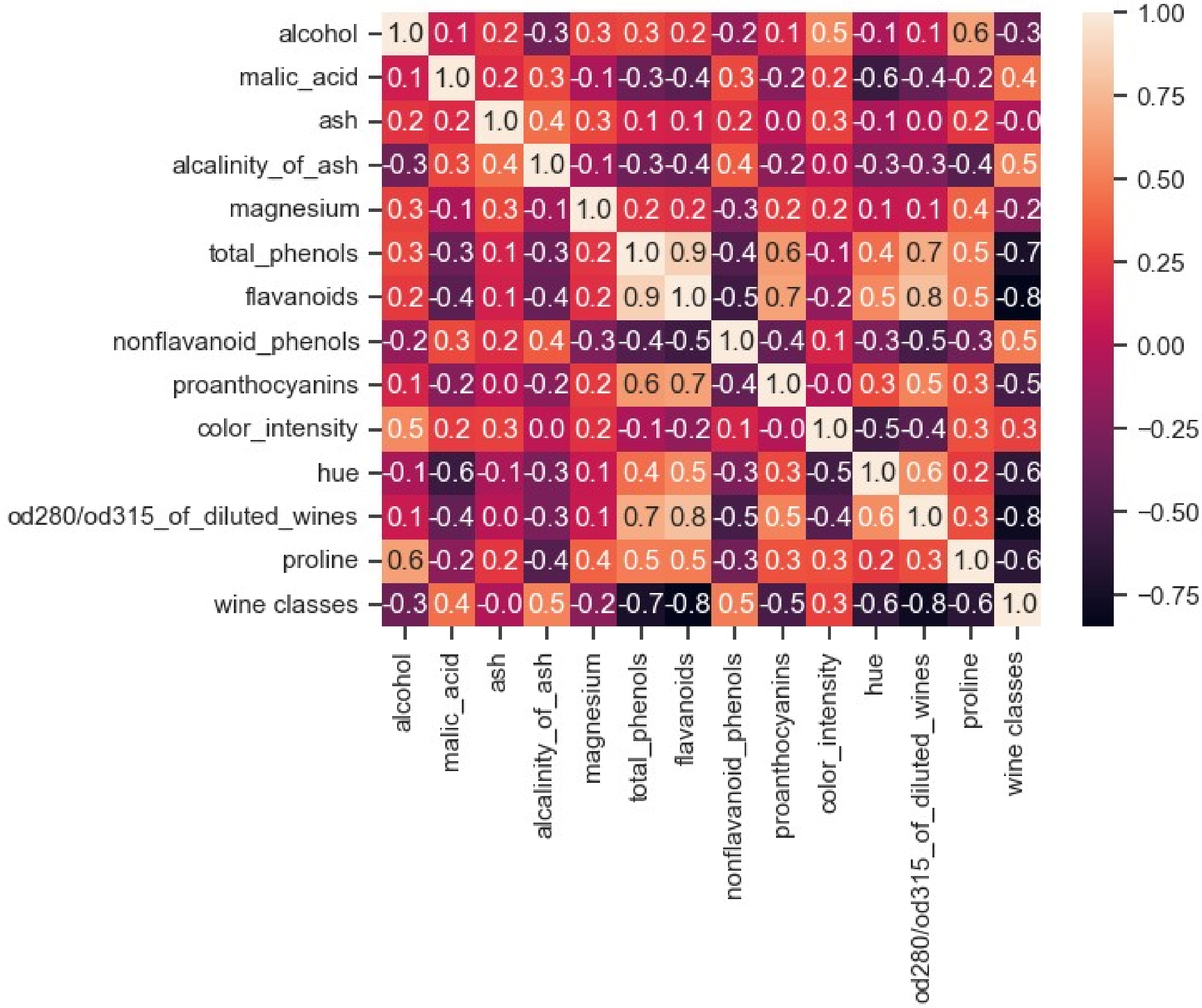
```
wine classes                        -0.847498            0.489109

                              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
wine classes                        -0.499130         0.265668 -
0.617369

                              od280/od315_of_diluted_wines
proline  \
alcohol                                           0.072343  0.643720

malic_acid                                       -0.368710 -0.192011

ash                                               0.003911  0.223626

alcalinity_of_ash                                -0.276769 -0.440597

magnesium                                         0.066004  0.393351

total_phenols                                     0.699949  0.498115

flavanoids                                        0.787194  0.494193

nonflavanoid_phenols                             -0.503270 -0.311385
```

proanthocyanins	0.519067	0.330417
color_intensity	-0.428815	0.316100
hue	0.565468	0.236183
od280/od315_of_diluted_wines	1.000000	0.312761
proline	0.312761	1.000000
wine classes	-0.788230	-0.633717

```
                              wine classes
alcohol                          -0.328222
malic_acid                        0.437776
ash                              -0.049643
alcalinity_of_ash                 0.517859
magnesium                        -0.209179
total_phenols                    -0.719163
flavanoids                       -0.847498
nonflavanoid_phenols              0.489109
proanthocyanins                  -0.499130
color_intensity                   0.265668
hue                              -0.617369
od280/od315_of_diluted_wines     -0.788230
proline                          -0.633717
wine classes                      1.000000
```

```
sns.heatmap(data.corr(), annot=True, fmt='.1f')
```

```
<Axes: >
```




```
# На основании корреляционной матрицы можно сделать следующие выводы:
# 1. Целевой признак сильно коррелирует с количеством фенолов
(total_phenols = -0.7) и флавонидами (flavonoids = -0.8)
# 2. Целевой признак сильно коррелирует с od280/od315 = -0.8.
# Эти признаки стоит оставить в модели. Остальные признаки стоит
убрать из модели, так как скорее всего они ухудшат её качество.
# Таким образом, на основании проведённого корреляционного анализа,
можно строить модели машинного обучения.
```