



**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 «Системы обработки информации и управления»**

**Отчет по рубежному контролю №1
по дисциплине «Технология мультимедиа»
Вариант №3**

**Выполнила:
студентка группы ИУ5-63Б
Ваксина И. Р.
«18» апреля 2021г.**

**Проверил:
Гапанюк Ю. Е.
«18» апреля 2021г.**

2021 г.

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.

Дополнительное требование по группам:

Для студентов групп ИУ5-63Б, ИУ5Ц-83Б - для произвольной колонки данных построить график "Ящик с усами (boxplot)".

Набор данных:

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html#sklearn.datasets.load_wine

Рубежный контроль №1

Ваксина Ия ИУ5 63Б 3 Вариант

Импорт библиотек

```
B [5]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.datasets import load_wine
%matplotlib inline
sns.set(style="ticks")
```

```
B [14]: A = load_wine()
data = pd.DataFrame(A.data, columns=A.feature_names)
data['123'] = A.target
```

```
B [16]: # Первые пять строк датасета
data.head()
```

```
Out[16]:
```

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	od280/od315_1
0	14.23	1.71	2.43		15.6	127.0	2.80	3.06	0.28	2.29	5.64	1.04
1	13.20	1.78	2.14		11.2	100.0	2.65	2.76	0.26	1.28	4.38	1.05
2	13.16	2.36	2.67		18.6	101.0	2.80	3.24	0.30	2.81	5.68	1.03
3	14.37	1.95	2.50		16.8	113.0	3.85	3.49	0.24	2.18	7.80	0.86
4	13.24	2.59	2.87		21.0	118.0	2.80	2.69	0.39	1.82	4.32	1.04

```
B [17]: # Размер датасета
data.shape
```

```
Out[17]: (178, 14)
```

```
B [18]: # Количество нулевых элементов
data.isnull().sum()
```

```
Out[18]: 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
123                            0
dtype: int64
```

```
B [19]: # Колонки и их типы данных
data.dtypes
```

```
Out[19]: 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
123                            int32
dtype: object
```

```
B [21]: 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   123                   178 non-null   int32
dtypes: float64(13), int32(1)
memory usage: 18.9 KB
```

```
B [22]: data.describe()
```

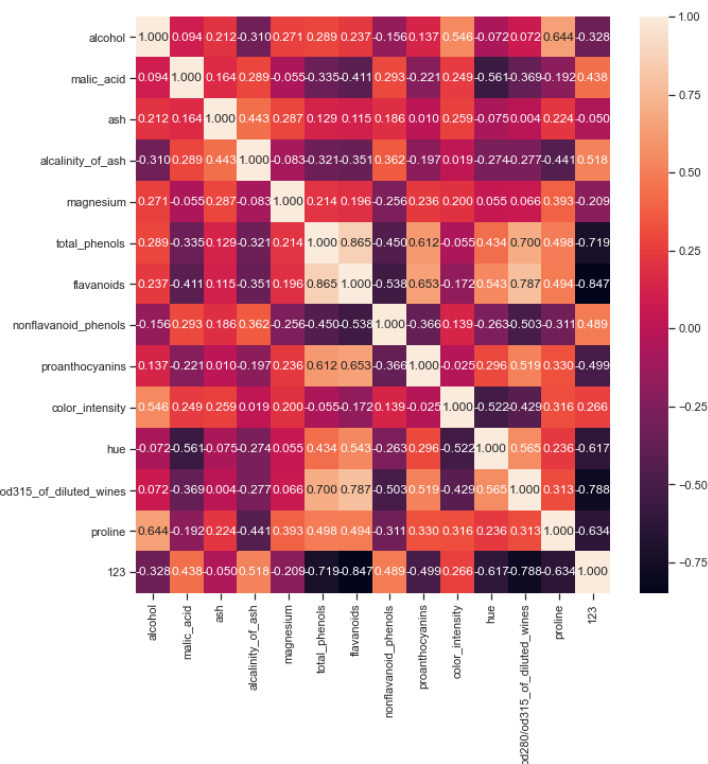
```
Out[22]:
```

	agnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	od280/od315_of_diluted_wines	proline	123
'8.000000'	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000
'9.741573'	2.295112	2.029270	0.361854	1.590899	5.058090	2.318286	0.957449	2.611685	746.893258	0.938202
'14.282484'	0.625851	0.998859	0.124453	0.572359	2.318286	0.228572	0.709990	314.907474	0.775035	
'0.000000'	0.980000	0.340000	0.130000	0.410000	1.280000	0.480000	1.270000	278.000000	0.000000	
'8.000000'	1.742500	1.205000	0.270000	1.250000	3.220000	0.782500	1.937500	500.500000	0.000000	
'8.000000'	2.355000	2.135000	0.340000	1.555000	4.690000	0.965000	2.780000	673.500000	1.000000	
'0.000000'	2.800000	2.875000	0.437500	1.950000	6.200000	1.120000	3.170000	985.000000	2.000000	
'2.000000'	3.880000	5.080000	0.660000	3.580000	13.000000	1.710000	4.000000	1680.000000	2.000000	

```
B [24]: corr_matrix = data.corr()
```

```
B [30]: corr_matrix['123']  
plt.figure(figsize=(10,10))  
sns.heatmap(corr_matrix, annot=True, fmt='.3f')
```

```
Out[30]: <AxesSubplot:>
```



"Ящик с усами (boxplot)"

```
B [32]: sns.boxplot(x=data['hue'])
```

```
Out[32]: <AxesSubplot:xlabel='hue'>
```

