

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
им. Н.Э. Баумана

Факультет «Информатика и системы управления»  
Кафедра «Систем обработки информации и управления»

ОТЧЕТ

Лабораторная работа № 1  
по дисциплине «Методы машинного обучения»

Тема: «Создание "истории о данных"»

ИСПОЛНИТЕЛЬ:

группа ИУ5-25

Алексеев А С  
ФИО

подпись

"\_\_" \_\_\_\_\_ 2024 г.

ПРЕПОДАВАТЕЛЬ:

Гапанюк Ю Е  
ФИО

подпись

"\_\_" \_\_\_\_\_ 2024 г.

Москва - 2024

---

## Задание

- Выбрать набор данных (датасет).
  - Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:
    1. История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.
    2. На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.
    3. Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.
    4. Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.
  - История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.
- Сформировать отчет и разместить его в своем репозитории на github.

Импорт библиотек

```
In [1]:import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
from pandas.plotting import parallel_coordinates
from math import pi
import squarify
```

Загрузка датасета

```
In [2]:df = pd.read_csv("winequalityN.csv")
print(df)
```

	type	fixed acidity	volatile acidity	citric acid	residual sugar \
0	white	7.0	0.270	0.36	20.7
1	white	6.3	0.300	0.34	1.6
2	white	8.1	0.280	0.40	6.9
3	white	7.2	0.230	0.32	8.5
4	white	7.2	0.230	0.32	8.5
...	...	...	...	...	...
6492	red	6.2	0.600	0.08	2.0
6493	red	5.9	0.550	0.10	2.2
6494	red	6.3	0.510	0.13	2.3
6495	red	5.9	0.645	0.12	2.0
6496	red	6.0	0.310	0.47	3.6

	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH \
0	0.045	45.0	170.0	1.00100	3.00
1	0.049	14.0	132.0	0.99400	3.30
2	0.050	30.0	97.0	0.99510	3.26
3	0.058	47.0	186.0	0.99560	3.19
4	0.058	47.0	186.0	0.99560	3.19
...	...	...	...	...	...
6492	0.090	32.0	44.0	0.99490	3.45
6493	0.062	39.0	51.0	0.99512	3.52
6494	0.076	29.0	40.0	0.99574	3.42
6495	0.075	32.0	44.0	0.99547	3.57
6496	0.067	18.0	42.0	0.99549	3.39

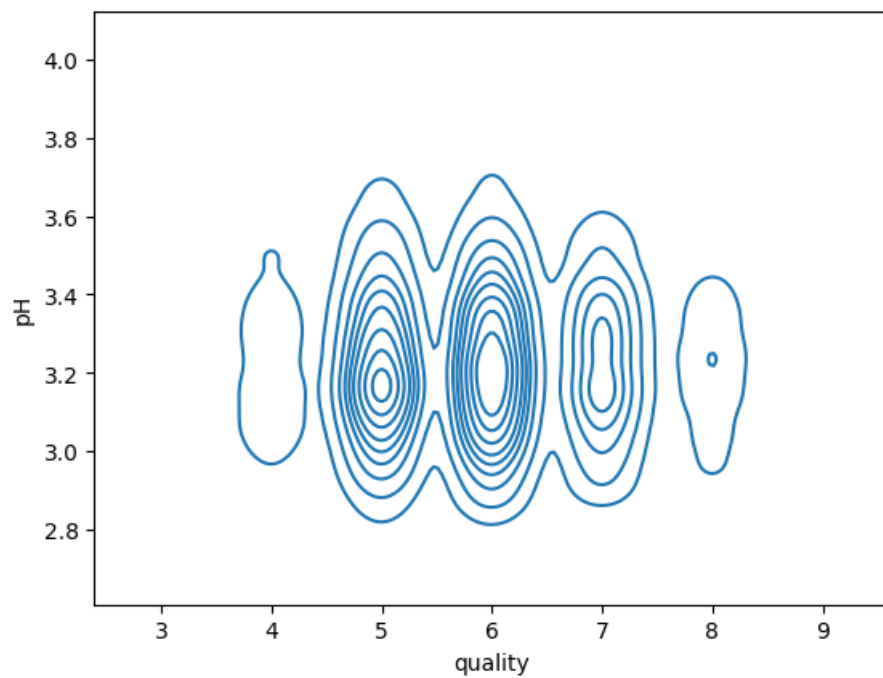
	sulphates	alcohol	quality
0	0.45	8.8	6
1	0.49	9.5	6
2	0.44	10.1	6
3	0.40	9.9	6
4	0.40	9.9	6
...	...	...	...
6492	0.58	10.5	5
6493	NaN	11.2	6
6494	0.75	11.0	6
6495	0.71	10.2	5
6496	0.66	11.0	6

[6497 rows x 13 columns]

DENSITY

```
In [3]:sns.kdeplot(x=df['quality'],y=df['pH'])
```

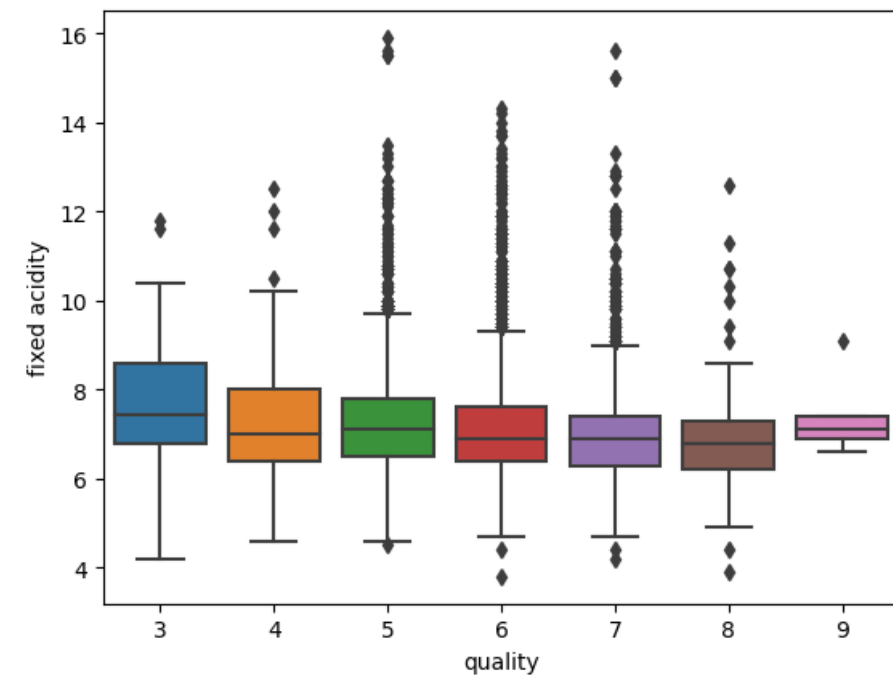
Out[3]:<AxesSubplot: xlabel='quality', ylabel='pH'>



Boxplot

In [4]:sns.boxplot(x=df['quality'],y=df['fixed acidity'])

Out[4]:<AxesSubplot: xlabel='quality', ylabel='fixed acidity'>

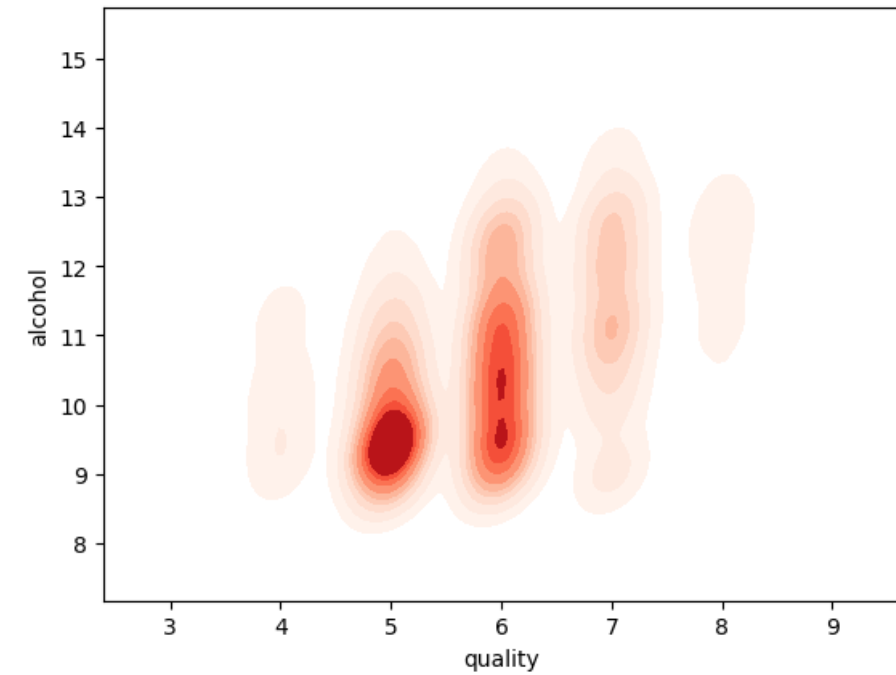


Density 2D

In [5]:sns.kdeplot(x=df['quality'],y=df['alcohol'], cmap="Reds", shade=True, bw\_adjust=1)

`shade` is now deprecated in favor of `fill`; setting `fill=True`.  
This will become an error in seaborn v0.14.0; please update your code.

"""Entry point for launching an IPython kernel.  
Out[5]:<AxesSubplot:xlabel='quality', ylabel='alcohol'>



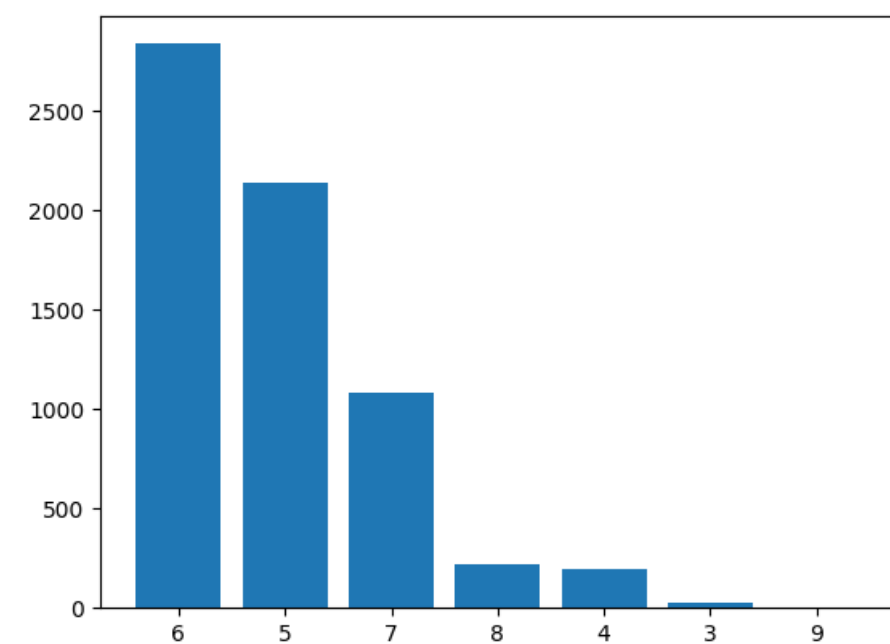
Treemap

```
In [6]: numtype = []
        white = 0
        red = 0
        for i in df['type']:
            if i == 'white':
                white += 1
            else:
                red += 1
        numtype.append(white)
        numtype.append(red)
        squarify.plot(sizes=numtype, label=['white','red'], alpha=.8)
        plt.axis('off')
        plt.show()
```



Quick start

```
In [7]: height = df['quality'].value_counts().to_frame()
        bars = df['quality'].unique()
        y_pos = np.arange(len(bars))
        plt.bar(y_pos, height['quality'])
        plt.xticks(y_pos, bars)
        plt.show()
```



Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js