

**Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра обчислювальної техніки**

Лабораторна робота №4
з дисципліни
«Аналіз даних з використанням мови Python»

Виконав:
студент групи ПІ-04
Пащенко Дмитро Олексійович

Перевірила:
Тимофєєва Ю. С.

Київ 2022

Код програми та результат виконання

```
In [16]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
In [5]: data = pd.read_csv('diamonds.csv')
data
```

```
Out[5]:
```

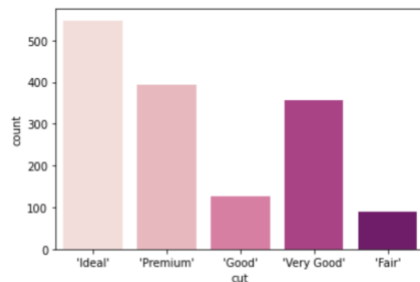
	Unnamed: 0	carat	cut	color	clarity	depth	table	price	x	y	z
0	1	0.23	'Ideal'	'E'	'SI2'	61.5	55.0	326	3.95	3.98	2.43
1	2	0.21	'Premium'	'E'	'SI1'	59.8	61.0	326	3.89	3.84	2.31
2	3	0.23	'Good'	'E'	'VS1'	56.9	65.0	327	4.05	4.07	2.31
3	4	0.29	'Premium'	'I'	'VS2'	62.4	58.0	334	4.20	4.23	2.63
4	5	0.31	'Good'	'J'	'SI2'	63.3	58.0	335	4.34	4.35	2.75
...
1509	1510	0.81	'Very Good'	'G'	'VS2'	63.1	58.0	2994	5.88	5.84	3.70
1510	1511	1.24	'Premium'	'J'	'I1'	61.9	55.0	2994	6.92	6.85	4.26
1511	1512	0.81	'Premium'	'G'	'VS2'	62.0	58.0	2994	5.95	5.92	3.68
1512	1513	0.81	'Premium'	'D'	'SI2'	61.7	58.0	2994	5.97	5.93	3.67
1513	1514	0.73	'Ideal'	'D'	'SI1'	61.4	56.0	2995	5.78	5.82	3.56

1. Побудувати стовпчикові діаграми, на яких відобразити:

а) кількість діамантів кожного з класів якості;

```
In [83]: sns.countplot(data = data, x = 'cut', palette = 'RdPu')
#data['cut'].value_counts().plot(kind = 'bar'); #або barh
```

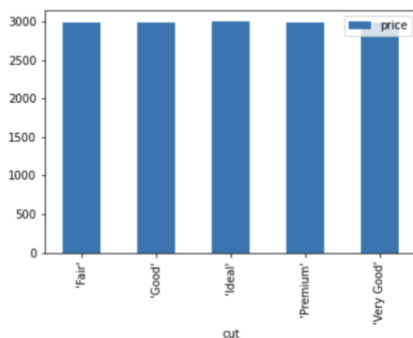
```
Out[83]: <AxesSubplot:xlabel='cut', ylabel='count'>
```



б) максимальну ціну діамантів кожного класу якості;

```
In [87]: max = pd.pivot_table(data, values = 'price', index = 'cut', aggfunc = np.max)
max.plot(kind = 'bar')
```

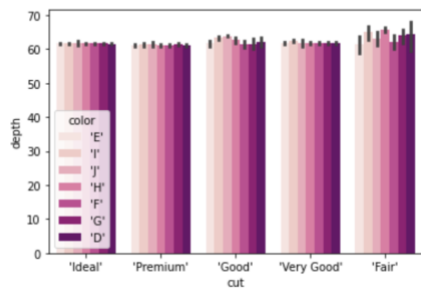
```
Out[87]: <AxesSubplot:xlabel='cut'>
```



в) середню глибину діамантів різного класу якості з різною якістю кольору.

```
In [30]: sns.barplot(data = data, x = 'cut', y = 'depth', hue = 'color', palette = 'RdPu')
```

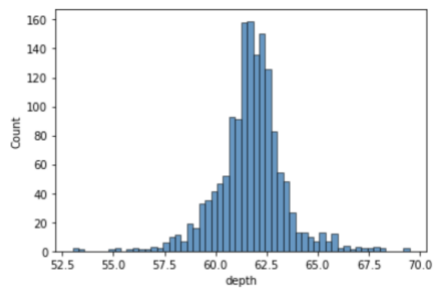
```
Out[30]: <AxesSubplot:xlabel='cut', ylabel='depth'>
```



2. Побудувати гістограму глибини діамантів у відсотках (depth), загальну і для кожного класу якості.

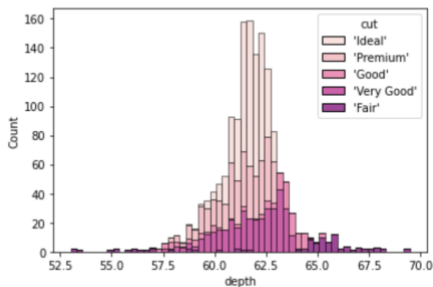
```
In [94]: sns.histplot(data = data, x = 'depth', palette = 'RdPu') # bins задає кількість стовпців  
#plt.hist(data = data, x = 'depth')
```

```
Out[94]: <AxesSubplot:xlabel='depth', ylabel='Count'>
```



```
In [96]: sns.histplot(data = data, x = 'depth', hue = 'cut', multiple = 'stack', palette = 'RdPu')
```

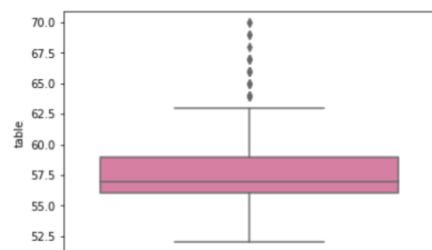
```
Out[96]: <AxesSubplot:xlabel='depth', ylabel='Count'>
```



3. Побудувати діаграму розмаху параметру table (загальну і в залежності від якості кольору), визначити чи присутні викиди.

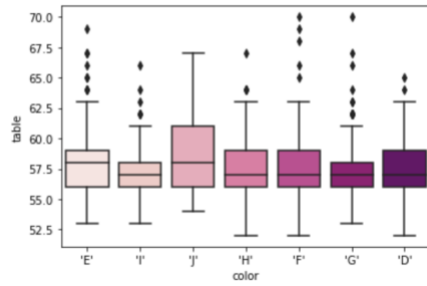
```
In [105]: sns.boxplot(data = data, y = 'table', palette = 'RdPu')
```

```
Out[105]: <AxesSubplot:ylabel='table'>
```



```
In [106]: sns.boxplot(data = data, y = 'table', x = 'color', palette = 'RdPu')
```

```
Out[106]: <AxesSubplot:xlabel='color', ylabel='table'>
```



4. За допомогою діаграм розсіювання зробити висновки щодо залежності між:

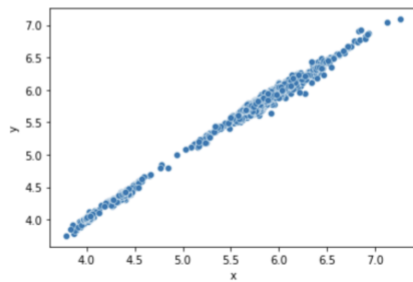
а) довжиною і шириною;

б) глибиною у % і глибиною у мм.

Порахувати коефіцієнт кореляції за допомогою відповідних функцій.

```
In [107]: sns.scatterplot(data = data, x = 'x', y = 'y')
```

```
Out[107]: <AxesSubplot:xlabel='x', ylabel='y'>
```

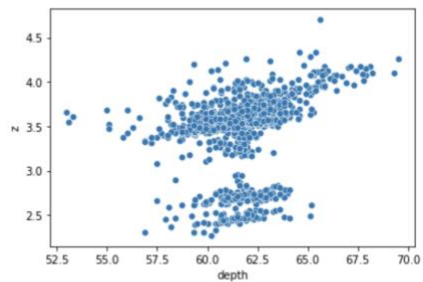


```
In [75]: np.corrcoef(data['x'], data['y'])[0, 1] * 100
```

```
Out[75]: 99.60042811801728
```

```
In [109]: sns.scatterplot(data = data, x = 'depth', y = 'z')
```

```
Out[109]: <AxesSubplot:xlabel='depth', ylabel='z'>
```



```
In [76]: np.corrcoef(data['depth'], data['z'])[0, 1] * 100
```

```
Out[76]: 22.79556791011829
```