МІНІСТЕРСТВО ОСВІТИ І НАУКИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»



**Лабораторна робота №1**

**З дисципліни «Візуалізація даних»**

***Виконав:*** *студент групи КН-310*

*Бурак Марко*

***Викладач:***

*Бойко Н. І.*

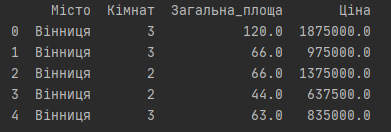
Львів – 2020

**Тема:** Аналіз даних та статистичне виведення

**Хід Роботи**

1. Завантажуємо дані та досліджуємо їх:

df = pd.read\_csv('flats.csv',decimal=",")

****

*Рис. 1 Інформація про дані*

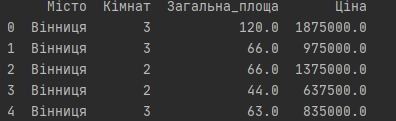
Перевіримо розмірність:

print(df.shape[0])

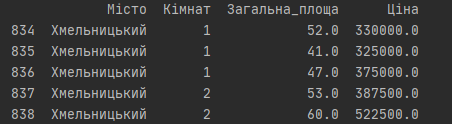
print(int(df.size/df.shape[0]))



1. Переглянемо перші п’ять та останні шість рядків з наших даних



*Рис. 2.1 перші п’ять рядків з наших даних*



*Рис. 2.2 останні шість рядків з наших даних*

1. Дізнаємося, яка кількість квартир продається у кожному місті згідно до нашого датасету



*Рис. 3 Кількість квартир у кожному місті*

1. Побудуємо стовпчикову діаграму для кількості кімнат

# кількість квартир за кімнатами

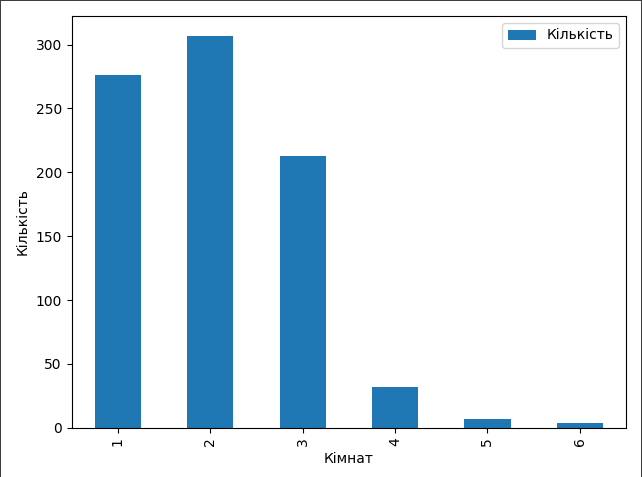
sums = df.groupby('Кімнат').size().reset\_index(name='Кількість')

sums.set\_index('Кімнат', inplace=True)

ax = sums.plot(kind='bar')

ax.set\_ylabel("Кількість")

plt.show()



*Рис. 4 Стовпчикова діаграма для кількості кімнат*

1. Побудуємо стовпчикову діаграму для змінної площа

# кількість загальних площ

sums2 = df.groupby('Загальна\_площа').size().reset\_index(name='Кількість')

x\_new = [x for x in range(0,201,50)]

x\_ax = sums2['Загальна\_площа'].tolist()

y\_ax = sums2['Кількість'].tolist()

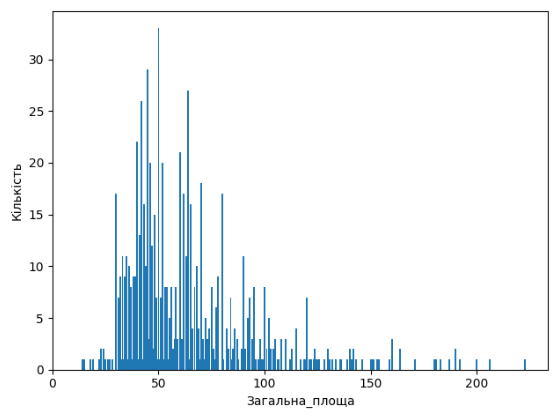
plt.bar(x\_ax,y\_ax)

plt.xlabel('Загальна\_площа')

plt.ylabel('Кількість')

plt.xticks(x\_new)

plt.show()



*Рис. 5 Стовпчикова діаграма для загальної площі*

1. На цьому графіку розподіл квартир, які продаються за загальною площею з кроком 50

# гістограма по загальній площі з кроком 50

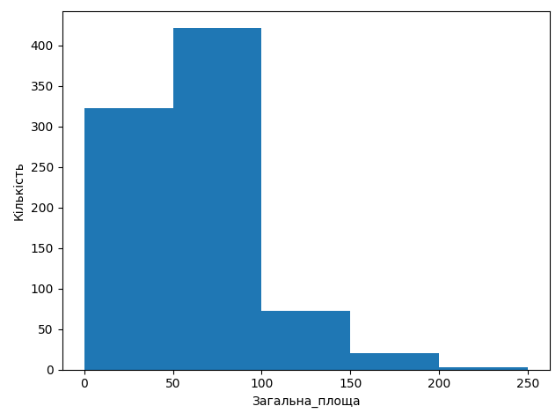
sums4 = df['Загальна\_площа']

ax = sums4.plot.hist(bins=[x for x in range(0,251,50)])

ax.set\_ylabel("Кількість")

ax.set\_xlabel('Загальна\_площа')

plt.show()



*Рис. 6 Розподіл квартир за площе*

1. Змінимо інтервал з 25 метрів квадратних до 50

# гістограма по загальній площі з кроком 25

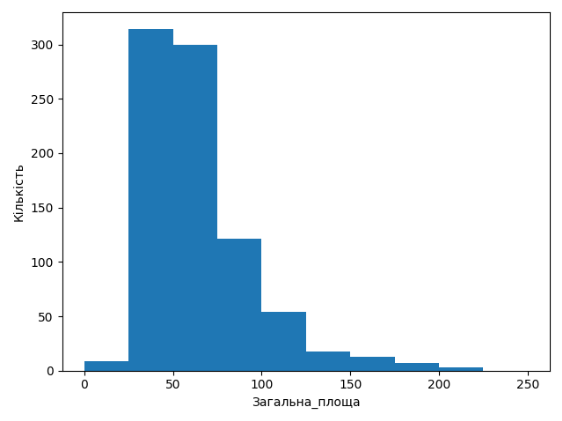
sums3 = df['Загальна\_площа']

ax = sums3.plot.hist(bins=[x for x in range(0,251,25)])

ax.set\_ylabel("Кількість")

ax.set\_xlabel('Загальна\_площа')

plt.show()



*Рис. 7 Розподіл квартир за площею*

1. Графік розсіювання. Побудуємо графік залежності ціни від загальної площі.

# scatter

x\_ax = df['Загальна\_площа']

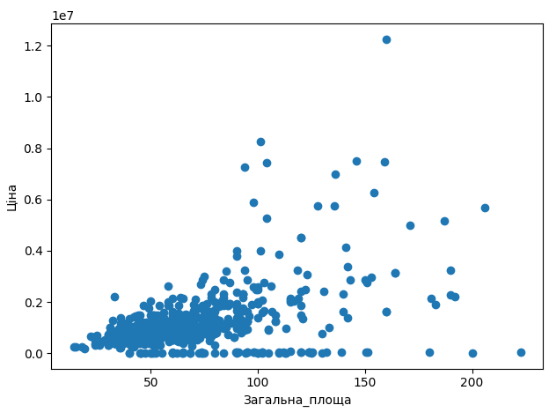
y\_ax = df['Ціна']

plt.scatter(x\_ax,y\_ax)

plt.xlabel('Загальна\_площа')

plt.ylabel('Ціна')

plt.show()



*Рис. 8 графік залежності ціни від загальної площі*

1. Порівняємо розподіл цін по містах

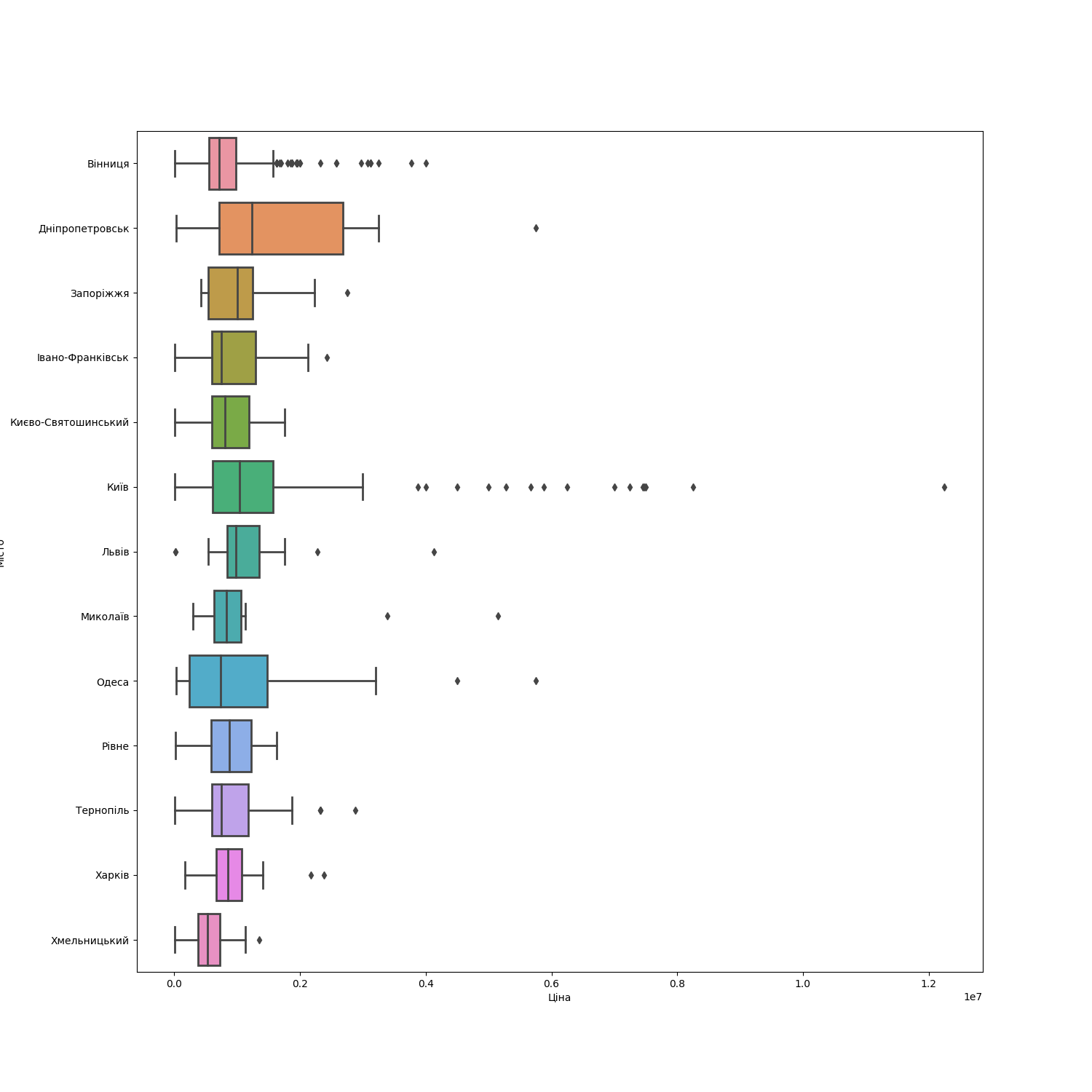
# box

plt.figure(figsize=(15, 15))

box=sns.boxplot(y='Місто', x="Ціна",orient='h', data=df, linewidth=2)

plt.savefig('foo.png')

plt.show()



*Рис. 9 Діаграма залежності ціни від міста*

**Висновок:** В дані лабораторній роботі проаналізував та візуалізував дані з файлу “flats.csv”, а також навчився основаб мови програмування Python і бібліотек “ matplotlib” та “ pandas”.