МІНІСТЕРСТВО ОСВІТИ І НАУКИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»



**Лабораторна робота №2**

**З дисципліни «Візуалізація даних»**

***Виконав:*** *студент групи КН-310*

*Бурак Марко*

***Викладач:***

*Бойко Н. І.*

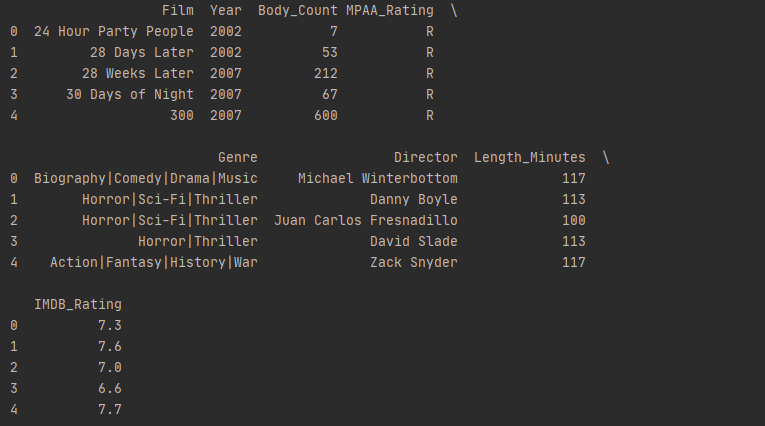
Львів – 2020

**Тема:** Аналіз даних та статистичне виведення

**Хід Роботи**

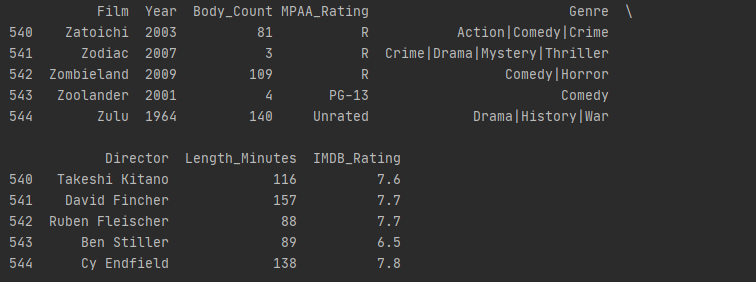
1. Завантажуємо дані та досліджуємо їх:

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

****

*Рис. 1 Інформація про дані*

1. Переглянемо останні рядки з наших даних



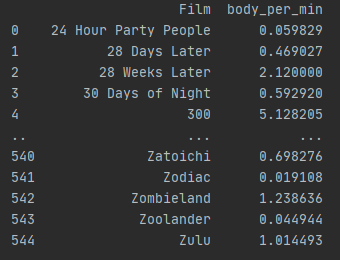
*Рис. 2.2 останні рядки з наших даних*

1. Дізнаємося, яка кількість смертей за хвилину у кожному фільмі

for i in range(len(df)):

body\_per\_min.append(df.iloc[i,2]/(df.iloc[i,6]))

df['body\_per\_min'] = body\_per\_min



*Рис. 3 Кількість смертей за хвилину у кожному місті*

1. Побудуємо стовпчикову діаграму для кількості смертей

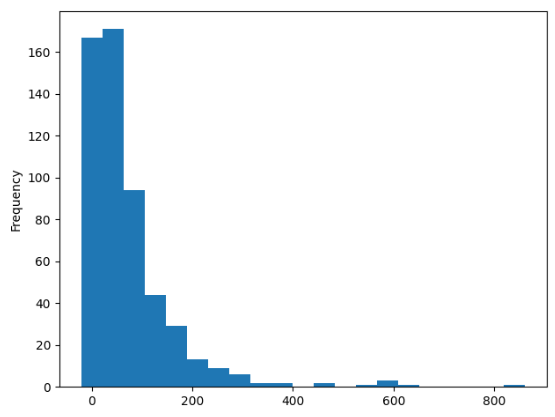
# hist count bodies

sums3 = df['Body\_Count']

sums3.plot.hist(bins=[x for x in range(-21,875,42)])

# plt.hist(df["Body\_Count"], bins=20)

plt.show()



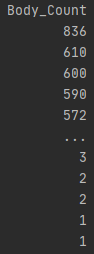
*Рис. 4 Стовпчикова діаграма для кількості смертей*

1. Посортуємо дані по смертях, та смертях за хв.

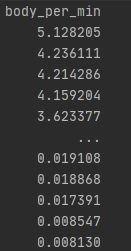
#sort

body\_sort = df.sort\_values(by = ['Body\_Count'],ascending=False)

body\_per\_min = df.sort\_values(by = ['body\_per\_min'],ascending=False)



*Рис. 5 Сортування за смертями*



*Рис. 6 Сортування за смертями за хв*

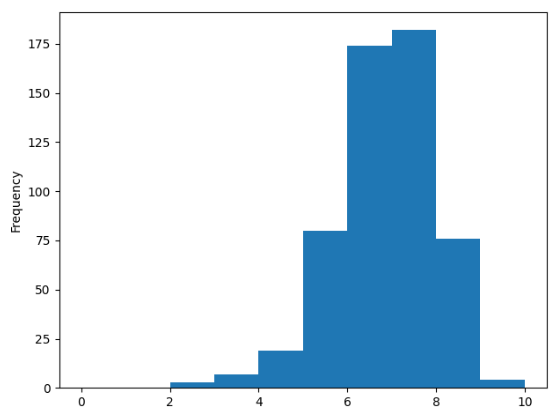
1. Побудуємо стовпчикову діаграму за рейтингом

# imdb simulation

rating = df['IMDB\_Rating']

rating.plot.hist(bins=[i for i in range(0,11,1)])

plt.show()



*Рис.7 Стовпчикова діаграма для рейтингу*

1. Середнє значення та середньоквадратичне за рейтингом

#середнє значення для змінної IMDBrating

imdb\_mean = df['IMDB\_Rating'].mean()

# середньоквадратичне відхилення для змінної IMDBrating

imdb\_sd = df['IMDB\_Rating'].std()



*Рис. 8 Середнє значення та середньоквадратичне за рейтингом*

1. Побудуємо нормальний розподіл за останніми значеннями

#нормальний розподіл

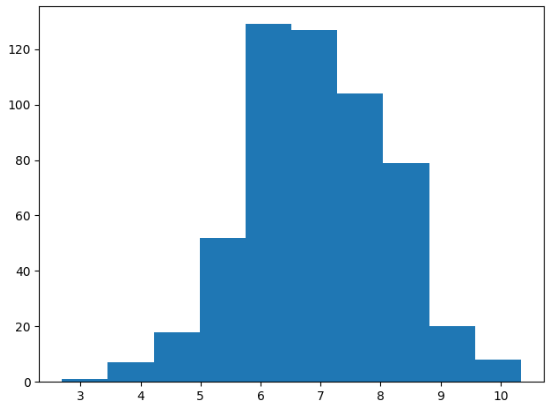
random.seed(900)

imdb\_simulation = np.random.normal(imdb\_mean, imdb\_sd, df.shape[0])

df['imdb\_simulation'] = imdb\_simulation

plt.hist(df['imdb\_simulation'], bins=10)

plt.show()



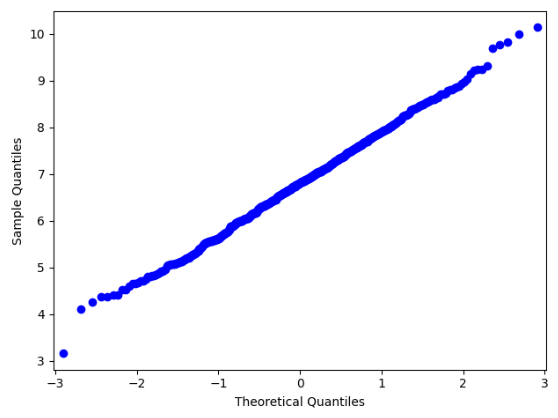
*Рис. 9 Нормальний розподіл*

1. Графік симуляції по рейтингу

#симуляція imdb\_simulation

sm.qqplot(df['imdb\_simulation'])

plt.show()



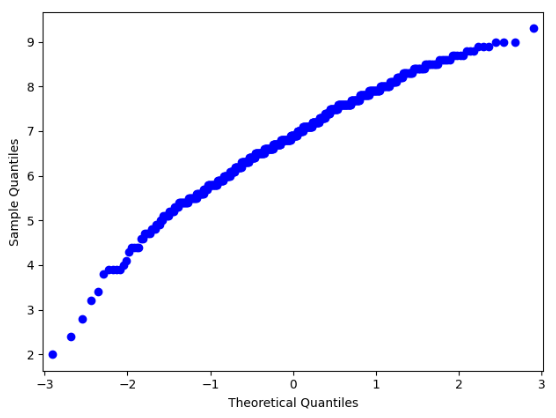
*Рис. 10 графік симуляції*

1. Реальний графік по рейтингу

#справжній рейтинг IMDB\_Rating:

sm.qqplot(df['IMDB\_Rating'])

plt.show()



*Рис. 9 Справжній графік по рейтингу*

**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”.