МІНІСТЕРСТВО ОСВІТИ І НАУКИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

Кафедра систем штучного інтелекту



Лабораторна робота №4

З курсу “Візуалізація даних”

Виконав:  
студент групи КН-310

Бурак Марко

Викладач:

Виклюк Я.І.

Львів – 2020

**Тема:** Незгруповані розподіли частот.

**Мета:** Побудова розподілу частот за емпіричними даними.

**Завдання:** Представити в табличній та графічній формі диференціальний та інтегральний розподіли частот за табличними даними:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3 | 5 | 4 | 4 | 2 | 6 | 3 | 5 | 4 | 4 |
| 4 | 2 | 3 | 5 | 5 | 3 | 4 | 4 | 3 | 5 |

Пояснення: кожен студент повинен до кожної цифри у клітинках добавити останню цифру порядкового номера у журналі, але з певними корективами: 307 – (+1), 308 – (+2), 309 – (-1), 310 – (-2).

Мій порядковий номер в журналі – 4, остання цифра – 4, група – КН-310.

Тому 4-2 = 2, додаю до кожного значення 2

**Імпортую бібліотеки для роботи з датасетом**

import pandas as pd

import matplotlib.pyplot as plt

**Додаю до кожного значення з датасету 2, як сказано за умовою задачі**

correct\_number = 4 - 2

**Функція програми**

def analys(data):

dataset1 = pd.DataFrame(data, columns=["Start\_data"])

dataset1["Start\_data"] = dataset1["Start\_data"] + correct\_number

dataset1["Variation\_series"] = dataset1["Start\_data"].sort\_values().reset\_index(drop=True)

dataset1\_table = pd.DataFrame(dataset1.value\_counts(["Variation\_series"], sort=False), columns=["Amount"])

summa = dataset1\_table["Amount"].sum()

dataset1\_table["Amount\_%"] = (dataset1\_table["Amount"] / summa \* 100)

dataset1\_table["Cumulative\_Sum"] = dataset1\_table["Amount"].cumsum()

dataset1\_table["Cumulative\_Sum\_%"] = dataset1\_table["Amount\_%"].cumsum()

print(dataset1\_table)

**У цій функції я формую датасет з даних Start data, Variation\_series, Amount, Amount%,Cumulative\_Sum та Cumulative\_Sum\_%**

Start data це початкові дані,Variation\_series це варіаційні дані, які посортовані від найбільшого до найменшого також це поле виступає індексом, поле Amount відповідає за кількість значень у варіативному ряді.Amount%, відсоткове представлдення Amount.Cumulative\_Sum та Cumulative\_Sum\_% це кумулятивна, або ж накопичувальна сума та відсодкове представлення цієї суми , відповідно.

fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(5, 5))

dataset1\_table.plot(y="Amount", ax=axes[0, 0], kind="bar", title="Amount of rating")

dataset1\_table.plot(y="Amount\_%", ax=axes[0, 1], kind="bar", title="Amount\_% of rating")

dataset1\_table.plot(y="Cumulative\_Sum", ax=axes[1, 0], kind="bar", title="Cumulative\_Sum of rating")

dataset1\_table.plot(y="Cumulative\_Sum\_%", ax=axes[1, 1], kind="bar", title="Cumulative\_Sum\_% of rating")

fig.tight\_layout()

plt.show()

me = dataset1\_table.plot(y="Cumulative\_Sum\_%", kind="line")

dataset1\_table.plot(y="Amount\_%", kind="bar",ax=me)

# ax.dataset1\_table.plot(y="Amount", ax=axes[0, 0], kind="bar", title="Amount of rating")

# ax.dataset1\_table.plot(y="Cumulative\_Sum", ax=axes[1, 0], kind="line", title="Cumulative\_Sum of rating")

plt.show()

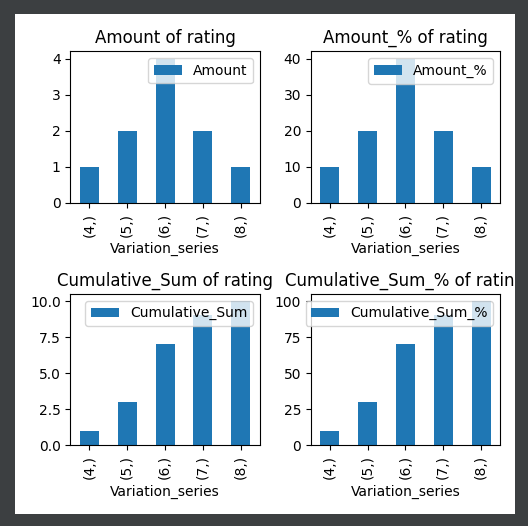
**Ця частина коду, відповідає за візуальне зображення графіків, після апробації, програма зображає 4 гістограми а також один графік який містить у собі гістограму та пряму.**

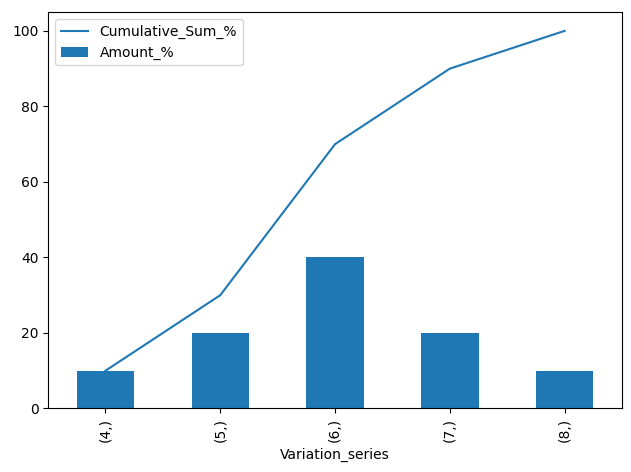
analys([3, 5, 4, 4, 2, 6, 3, 5, 4, 4])

analys([4, 2, 3, 5, 5, 3, 4, 4, 3, 5])Результат виконання:

**Після апробації, програма повернула ряд 4 5 5 6 6 6 6 7 7 8**

З першого графіку видно, що цей розподіл є схожим на нормальний розподіл.



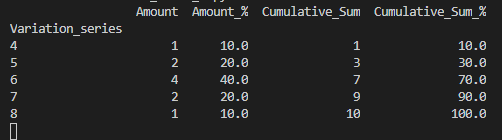


Графік деференціального та інтегрального розподілу частот

Деференціальні розподіли представляють значення частот окремо для кожного х.  
Інтегральні (накопичувальні, кумулятивні) це додавання усіх попередніх х частот.

Площо, яка знаходиться під диференціальним розподілом має сенс частоти.

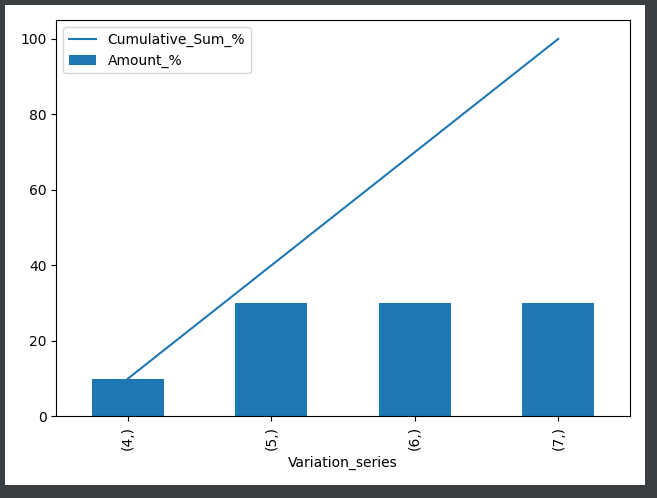
З графіка вище можна побачити, що значення з 4 до 5 включно становлять 30% від усієї вибірки, з 4 до 6 включно становлять 70% вибірки, а від 4 до 7 - 90%, і від 4 до 8 включно, 100%, тобто уся вибірка. Цей графік дуже зручний, щоб проводити таку аналітику.



Описані вище дані виведені у таблицю.

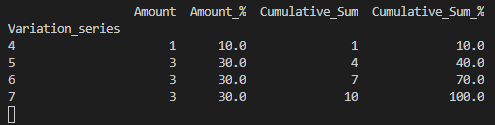


Повторюємо ті ж дії для другого датасету, тут бачимо, що у рейтингу є лише одна 4 і по три 5,6 та 7.



Графік деференціального та інтегрального розподілу частот

З графіка вище можна побачити, що значення з 4 до 5 включно становлять 40% від усієї вибірки, з 4 до 6 включно становлять 70% вибірки, а від 4 до 7 - 100%



Описані вище дані виведені у таблицю.

**Висновок:** Я навчився будувати та візуалізувати розподіл частот за емпіричними даними за допомогою бібліотеки Pandas мовою програмування Python.