**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

**ЗВІТ**

до лабораторної роботи № 3 на тему

**«Аналіз розміру файлів»**

Студентки 3 курсу факультету комп'ютерних наук та кібернетики

групи ТТП-32

Сушко Діани

**Київ-2024**

**ЗМІСТ**

|  |  |
| --- | --- |
| Вступ ……………...…………………………………………. | 3 |
| Код програми …………………………………………………… | 4 |
| Виконання програми ………………………………………....... | 7 |
| Висновки ……………………...…………………………………. | 9 |

# **ВСТУП**

Проаналізувати частотну характеристику (частотний розподіл) розміру файлів у файловій системі на комп'ютері (залежність кількості файлів від їх розміру).

# **ПРОГРАМНИЙ КОД**

Main.py

import pandas as pd

import matplotlib.pyplot as plt

from matplotlib.widgets import Slider

file\_sizes = pd.read\_csv('file\_sizes.txt', header=None, names=['size'])

intervals = [0, 1024, 1024\*1024, 10\*1024\*1024, 100\*1024\*1024, 1024\*1024\*1024, float('inf')]

labels = ['<1KB', '1KB-1MB', '1MB-10MB', '10MB-100MB', '100MB-1GB', '>1GB']

file\_sizes['size\_interval'] = pd.cut(file\_sizes['size'], bins=intervals, labels=labels, right=False)

file\_size\_counts = file\_sizes['size\_interval'].value\_counts().sort\_index()

total\_sizes\_mb = file\_sizes.groupby('size\_interval')['size'].sum() / (1024 \* 1024)

total\_sizes\_gb = total\_sizes\_mb / 1024

fig, ax = plt.subplots(figsize=(10, 9))

bar\_plot = ax.bar(file\_size\_counts.index, file\_size\_counts.values, color='skyblue')

ax.set\_title(f'Гістограма кількості файлів за їх розміром (усього {**len**(file\_sizes)} файлів)')

ax.set\_xlabel('Розмір файлу (байт)')

ax.set\_ylabel('Кількість файлів')

ax.grid(axis='y')

ax.set\_xticklabels(labels=file\_size\_counts.index, rotation=0, ha='center')

for i, v in **enumerate**(file\_size\_counts.values):

    ax.text(i, v, str(v), ha='center', va='bottom', rotation=0)

ax\_slider = plt.axes([0.15, 0.01, 0.7, 0.03])

slider = Slider(ax\_slider, 'Кількість файлів', 25, **max**(file\_size\_counts), valinit=**max**(file\_size\_counts), valstep=25)

def **update**(val):

    max\_val = int(slider.val)

    ax.set\_ylim(0, max\_val)

    plt.draw()

slider.on\_changed(update)

conclusions = []

for interval, count in **zip**(labels, file\_size\_counts.values):

    percentage = count / **len**(file\_sizes) \* 100

    conclusion = f"Переважна більшість файлів ({percentage:.2f}%) має розміри у діапазоні {interval}. Загальний обсяг: {total\_sizes\_mb[interval]:.2f} MB ({total\_sizes\_gb[interval]:.2f} GB)"

    conclusions.append((interval, percentage, conclusion))

conclusions.sort(key=lambda x: x[1], reverse=True)

with **open**('conclusions.txt', 'w') as f:

    for \_, \_, conclusion in conclusions:

        f.write(conclusion + '\n')

plt.show()

get\_file\_sizes.sh

*#!/bin/bash*

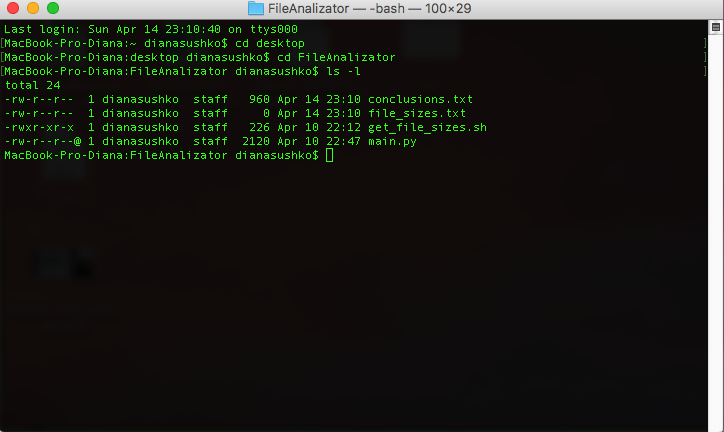
**read** -p "Enter the directory path: " dir\_path

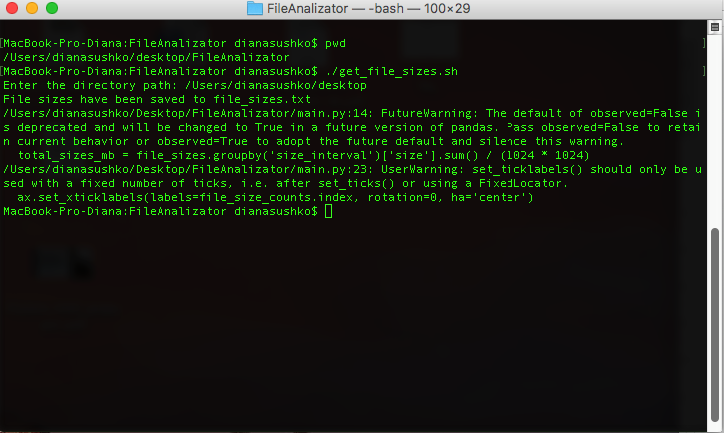
**find** "$dir\_path" -type f -exec ls -l {} + | **grep** -v '^total' | **awk** '{print $5}' > file\_sizes.txt

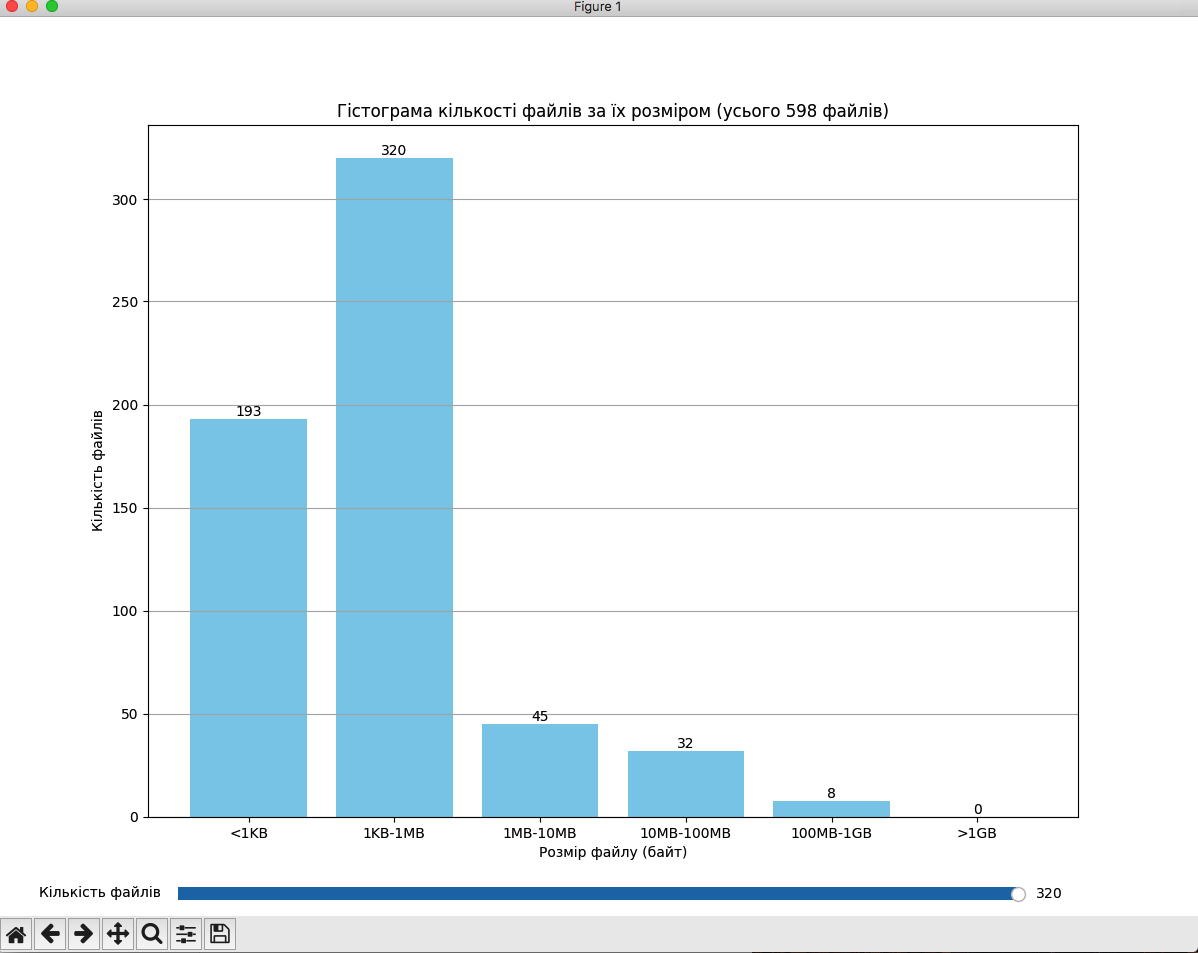
**echo** "File sizes have been saved to file\_sizes.txt"

**python3** main.py

# **ВИКОНАННЯ КОДУ**







Результати виконання програмного коду main.py було збережено у файл conclusions.txt.

Усього було знайдено 598 файлів.

Переважна більшість файлів 53.51% має розміри у діапазоні 1KB-1MB, загальний обсяг: 52.24 MB (0.05 GB); 32.27% файлів має розміри у діапазоні <1KB, загальний обсяг: 0.04 MB (0.00 GB); 7.53% файлів має розміри у діапазоні 1MB-10MB, загальний обсяг: 103.31 MB (0.10 GB); 5.35% файлів має розміри у діапазоні 10MB-100MB, загальний обсяг: 1366.27 MB (1.33 GB); 1.34% файлів має розміри у діапазоні 100MB-1GB, загальний обсяг: 2405.01 MB (2.35 GB); 0.00% файлів має розміри у діапазоні >1GB, загальний обсяг: 0.00 MB (0.00 GB).

# **ВИСНОВКИ**

На основі виконаного аналізу розмірів файлів директорії /Users/dianasushko/desktop можна зробити такі висновки:

Усього було знайдено і проаналізовано 598 файлів.

Найбільш поширеними є файли з розмірами у діапазоні від 1KB до 1MB, що становлять 53.51% усіх файлів, тобто більшість. Загальний обсяг цих файлів становить 52.24 MB (0.05 GB).

Наступними йдуть файли розміром <1KB, що становлять 32.27% усіх файлів. Загальний обсяг таких файлив складає 0.04 MB (0.00 GB).

Файли у діапазоні від 1MB до 10MB зустрічаються в 7.53%, їх значно менше, і їх загальний обсяг становіть 103.31 MB (0.10 GB).

Ще меншу частку становлять файли з розмірами в діапазоні від 10MB до 100MB, що складає 5.35%. Їх загальний обсяг становить 1366.27 MB (1.33 GB).

Частка файлів з розмірами від 100MB до 1GB складає всього 1.34%, є найбильш ваговими (важкими) і займають 2405.01 MB (2.35 GB).

А вот файлів з розмірами понад 1GB в досліджуваній директорії взагалі немає, тобто їхній загальний обсяг становить 0.00 MB (0.00 GB).