A Mini Project Report on

Text Extractor using Amazon AWS Services

Submitted in partial fulfillment of the requirements

of the degree of Bachelor in Engineering

Third Year of Engineering In
Computer Engineering

By

Aditya Sawant	TE4	45
Swapnil Sawant	TE4	46
Chinmay Shinde	TE4	50

Under the guidance of
Prof. Krupa Chotai

Department of Computer Engineering

Shah and Anchor Kutchhi Engineering College

Chembur, Mumbai-400088

University of Mumbai

2022-2023

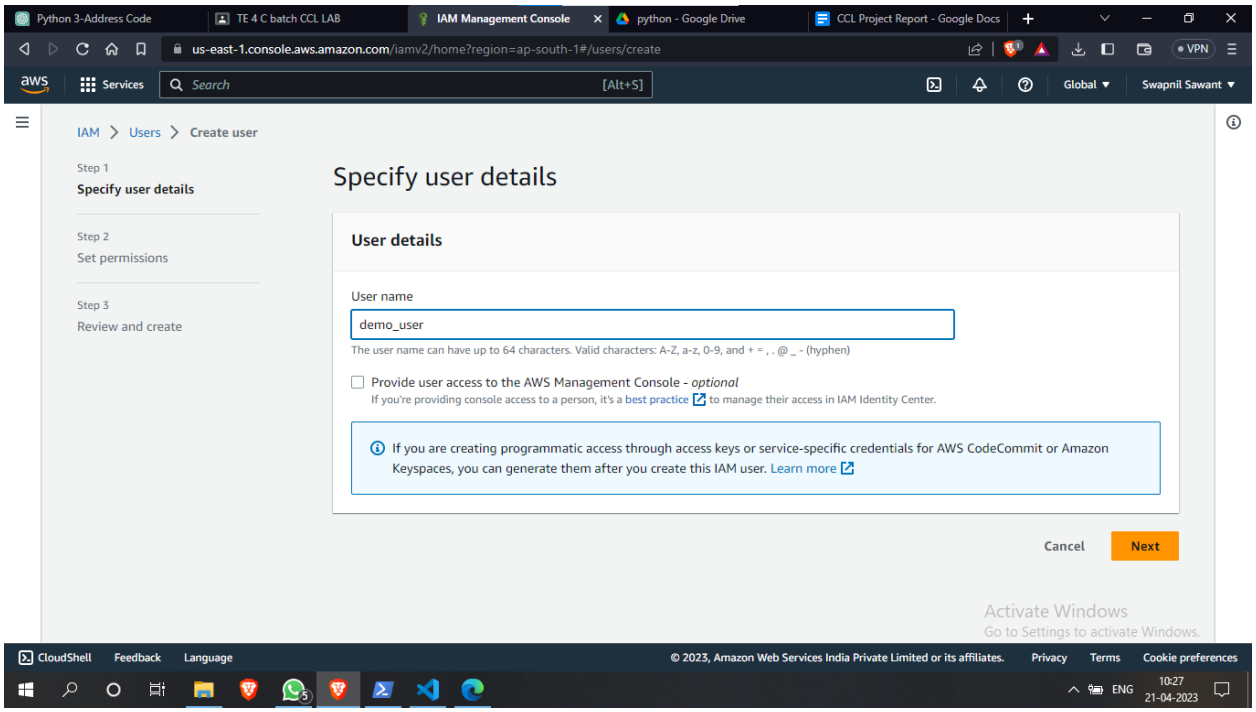
Aim :

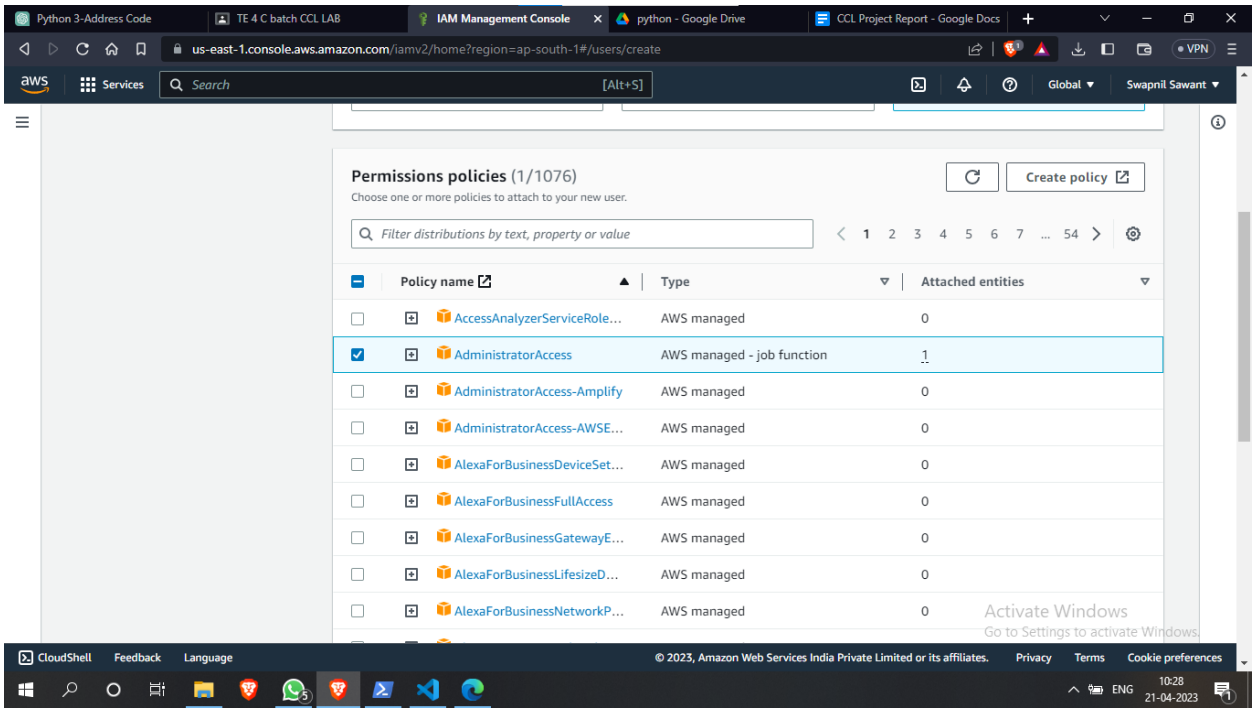
To extract the text from the image by using the AWS IAM, Textract, and AWS CLI service.

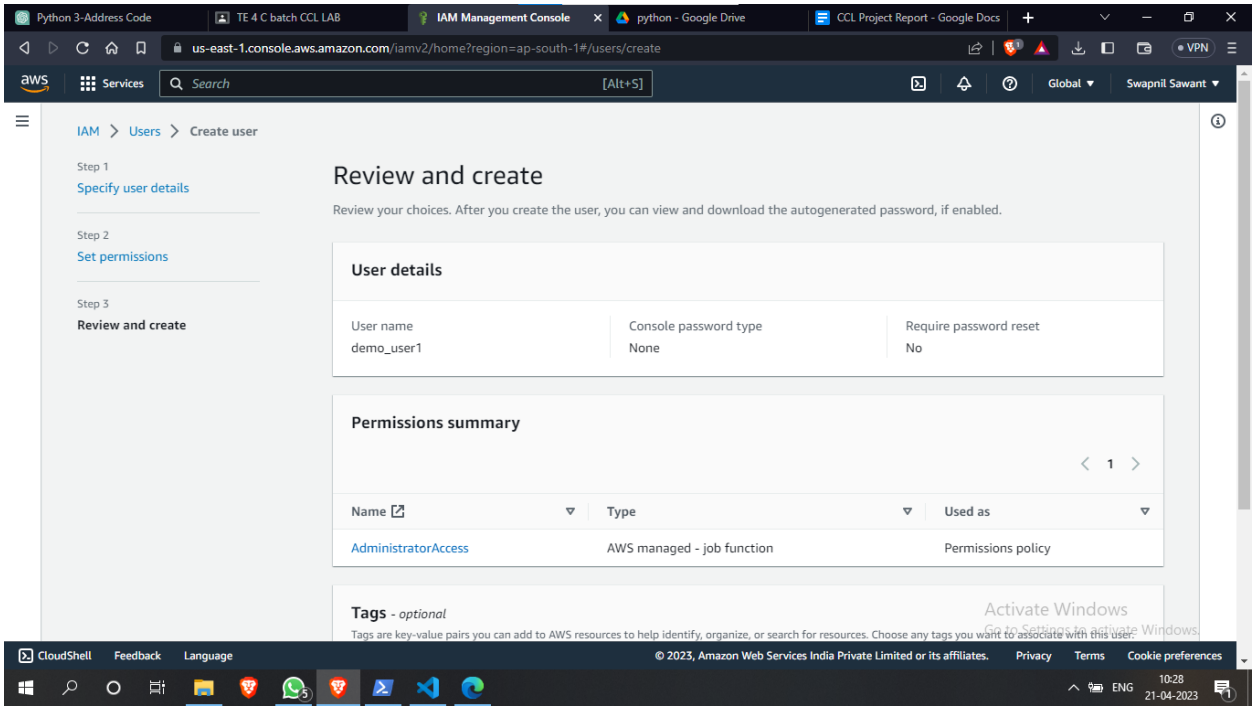
Problem Statement :

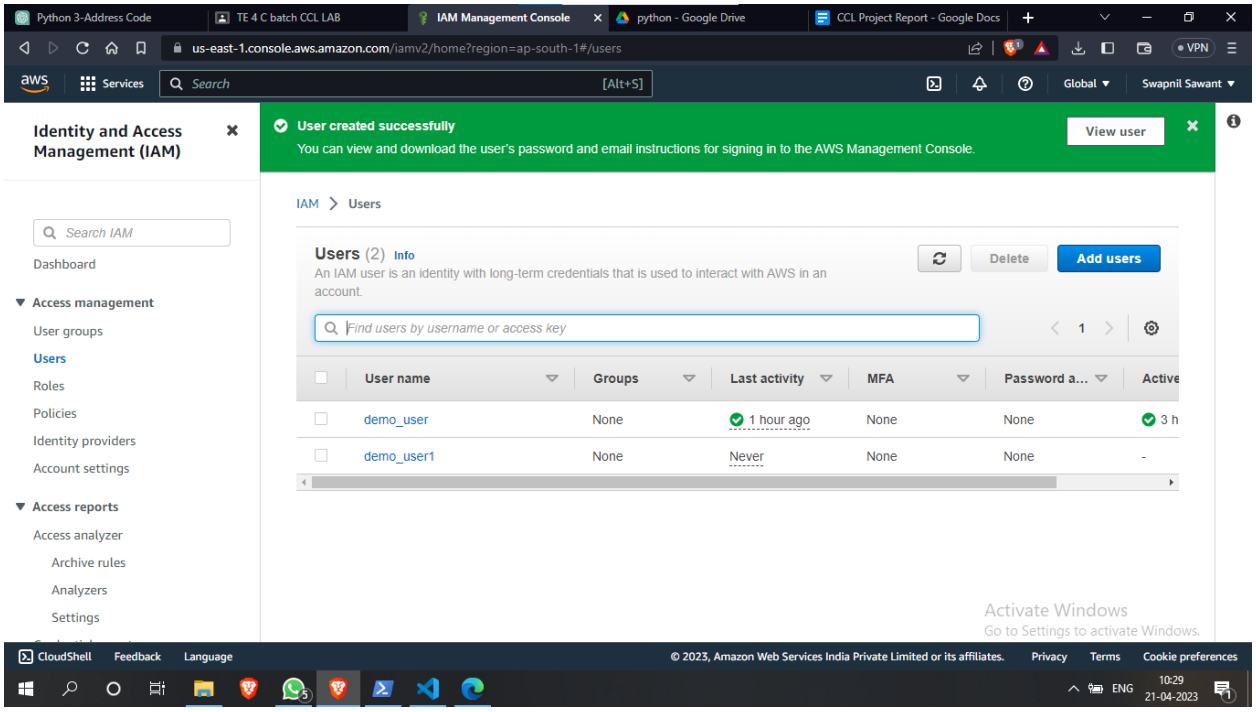
The goal of the project "Text Extractor using Amazon AWS Services" is to create a system that can automatically extract text from photos using the services provided by AWS IAM, Textract, and AWS CLI. The purpose of this project is to give businesses and organizations that need to handle a lot of photos with text an effective and affordable option. The software must be able to correctly recognize text included inside pictures, extract the content, and store it in a machine-readable format for future use. Optimising the system's efficiency to handle high numbers of photos, maintaining the confidentiality and privacy of the extracted data, and assuring the accuracy of text extraction are the project's main problems. The accomplishment of this project's objectives

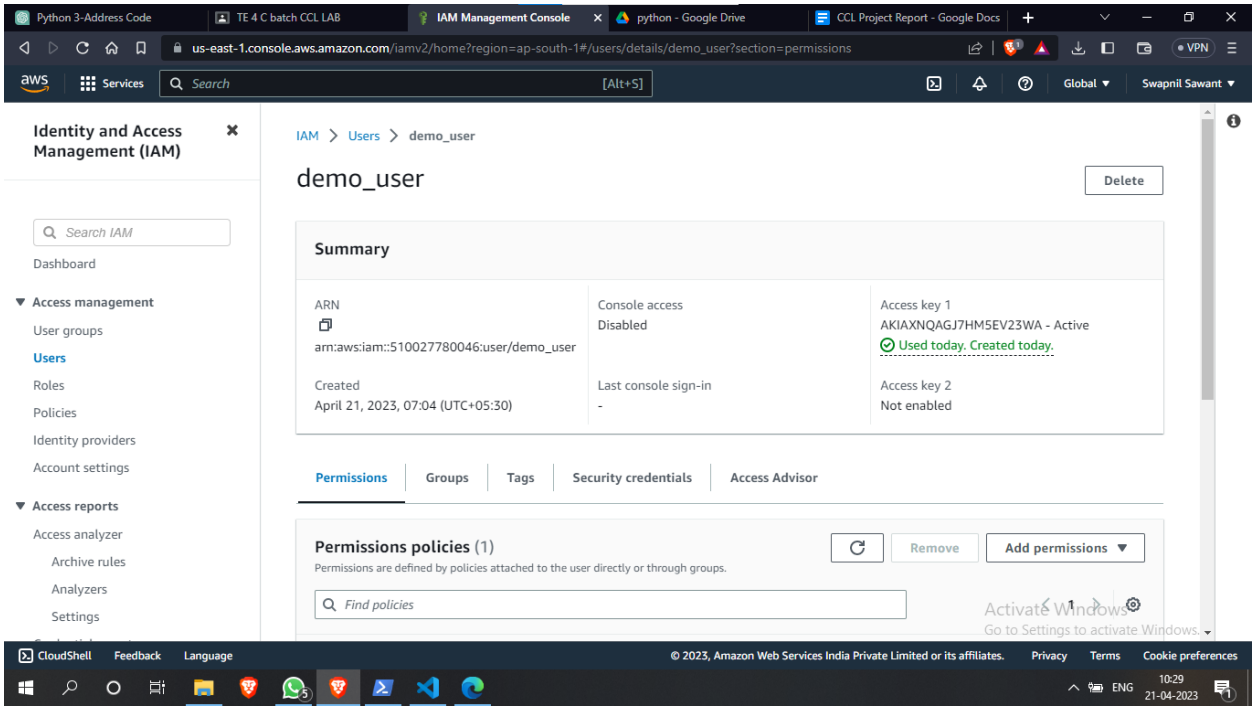
Creating IAM user(demo_user) :



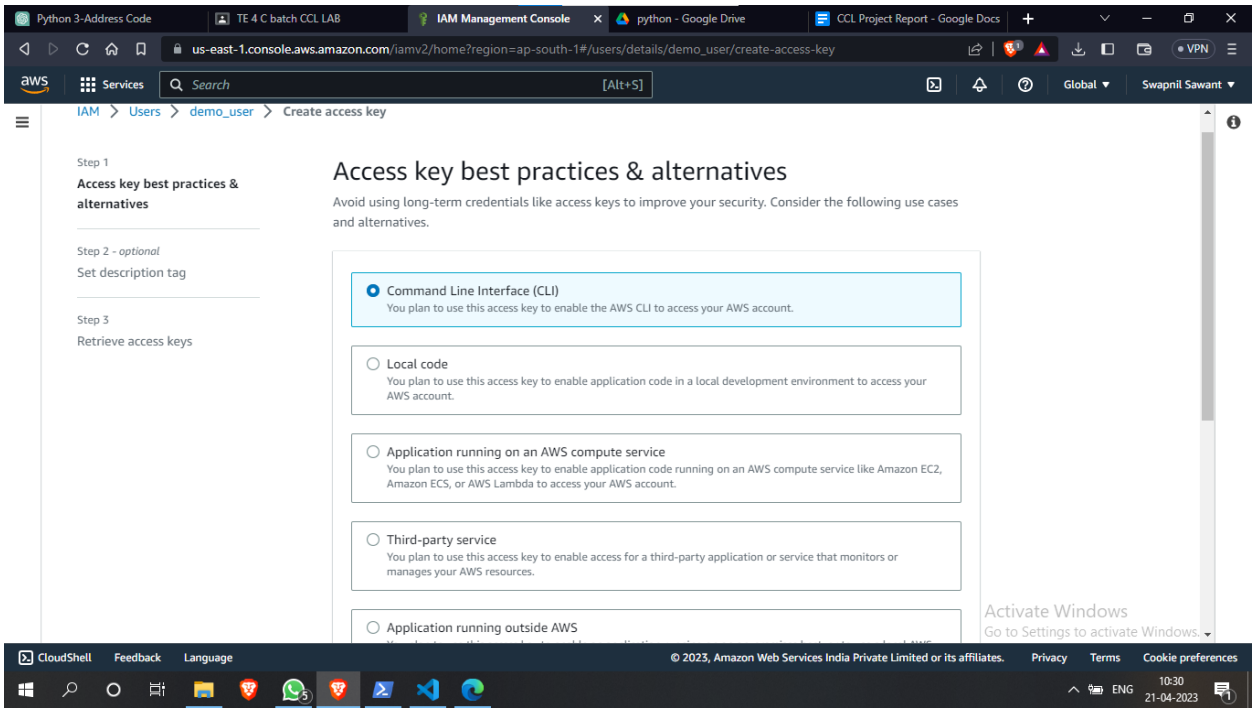


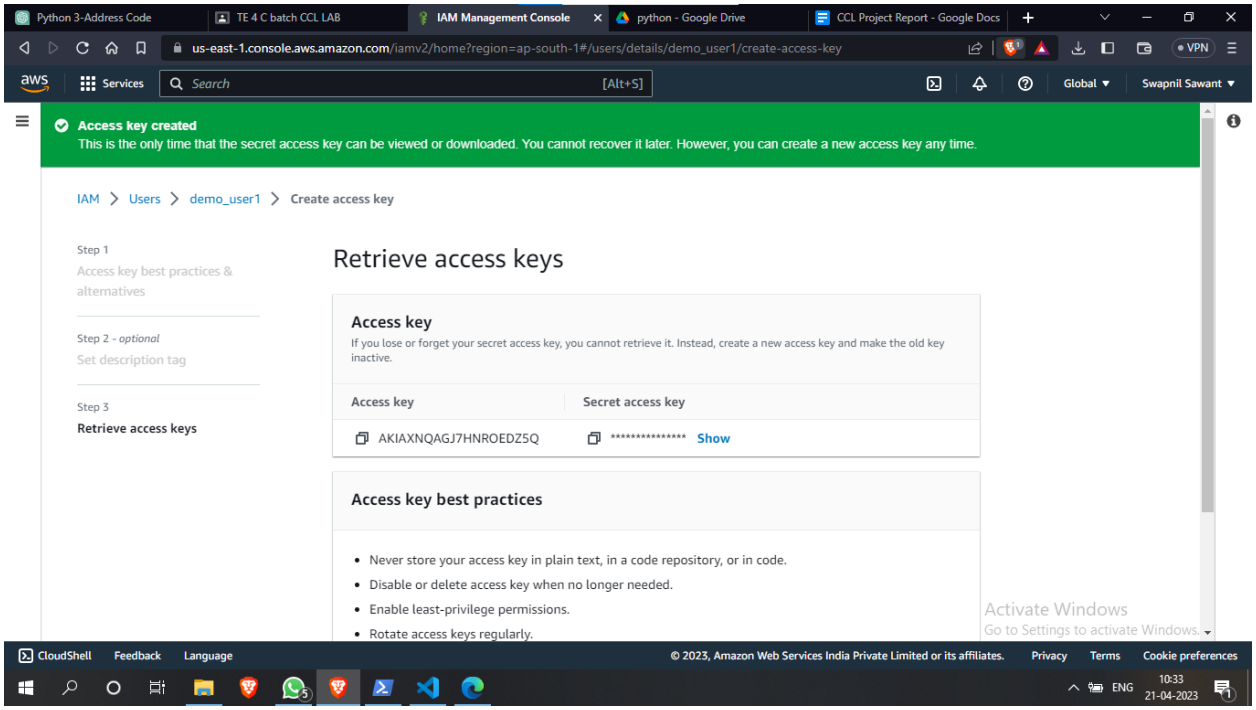

Python 3-Address Code
TE 4 C batch CCL LAB
IAM Management Console
python - Google Drive
CCL Project Report - Google Docs
us-east-1.console.aws.amazon.com/iamv2/home?region=ap-south-1#/users/create
Services
Search
[Alt+S]
Global
Swapnil Sawant
IAM > Users > Create user
Step 1
Specify user details
Step 2
Set permissions
Step 3
Review and create
Review and create
Review your choices. After you create the user, you can view and download the autogenerated password, if enabled.
User details
User name
demo_user1
Console password type
None
Require password reset
No
Permissions summary
Name
Type
Used as
AdministratorAccess
AWS managed - job function
Permissions policy
Tags - optional
Tags are key-value pairs you can add to AWS resources to help identify, organize, or search for resources. Choose any tags you want to associate with this user.
Activate Windows
Go to Settings to activate Windows.
CloudShell
Feedback
Language
© 2023, Amazon Web Services India Private Limited or its affiliates.
Privacy
Terms
Cookie preferences
ENG
10:28
21-04-2023


Python 3-Address Code
TE 4 C batch CCL LAB
IAM Management Console
python - Google Drive
CCL Project Report - Google Docs
us-east-1.console.aws.amazon.com/iamv2/home?region=ap-south-1#/users
Services
Search
[Alt+S]
Global
Swapnil Sawant
Identity and Access Management (IAM)
User created successfully
You can view and download the user's password and email instructions for signing in to the AWS Management Console.
View user
IAM > Users
Search IAM
Dashboard
Access management
User groups
Users
Roles
Policies
Identity providers
Account settings
Access reports
Access analyzer
Archive rules
Analyzers
Settings
Users (2) Info
An IAM user is an identity with long-term credentials that is used to interact with AWS in an account.
Delete
Add users
Find users by username or access key
User name
Groups
Last activity
MFA
Password a...
demo_user
None
1 hour ago
None
None
demo_user1
None
Never
None
None
Activate Windows
Go to Settings to activate Windows.
CloudShell
Feedback
Language
© 2023, Amazon Web Services India Private Limited or its affiliates.
Privacy
Terms
Cookie preferences
ENG
10:29
21-04-2023

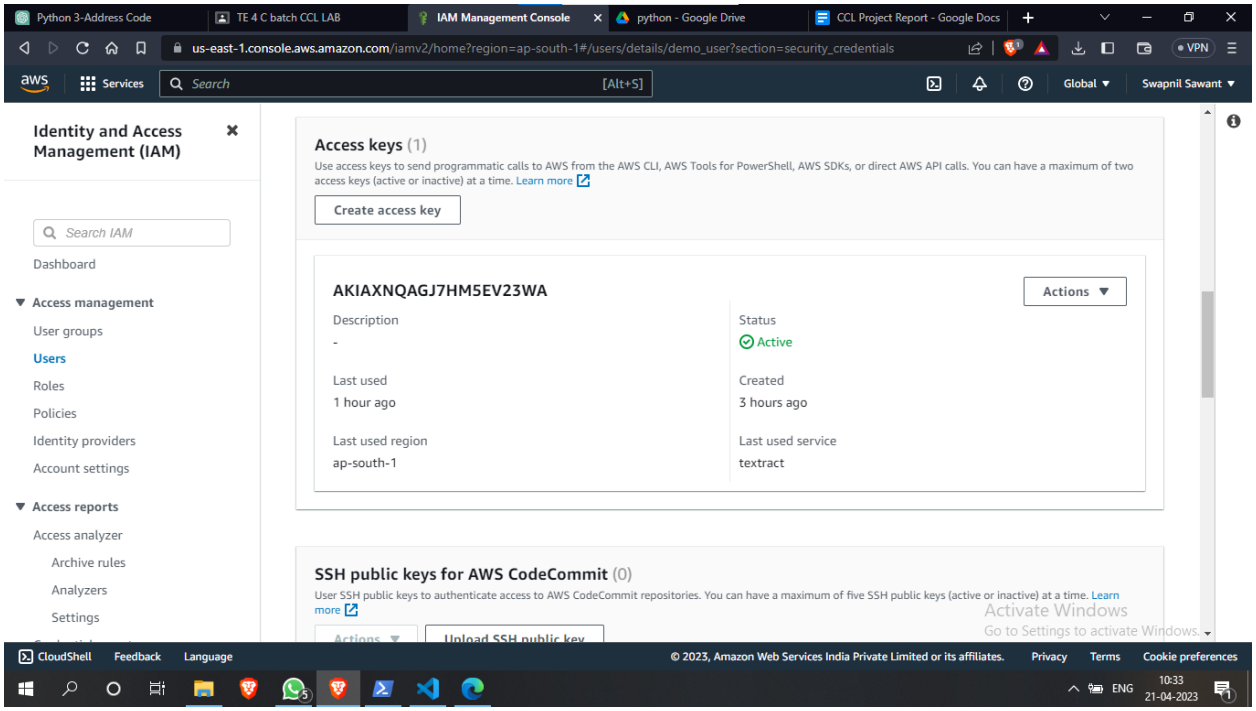


Give programmatic access to the user using the access key creation.



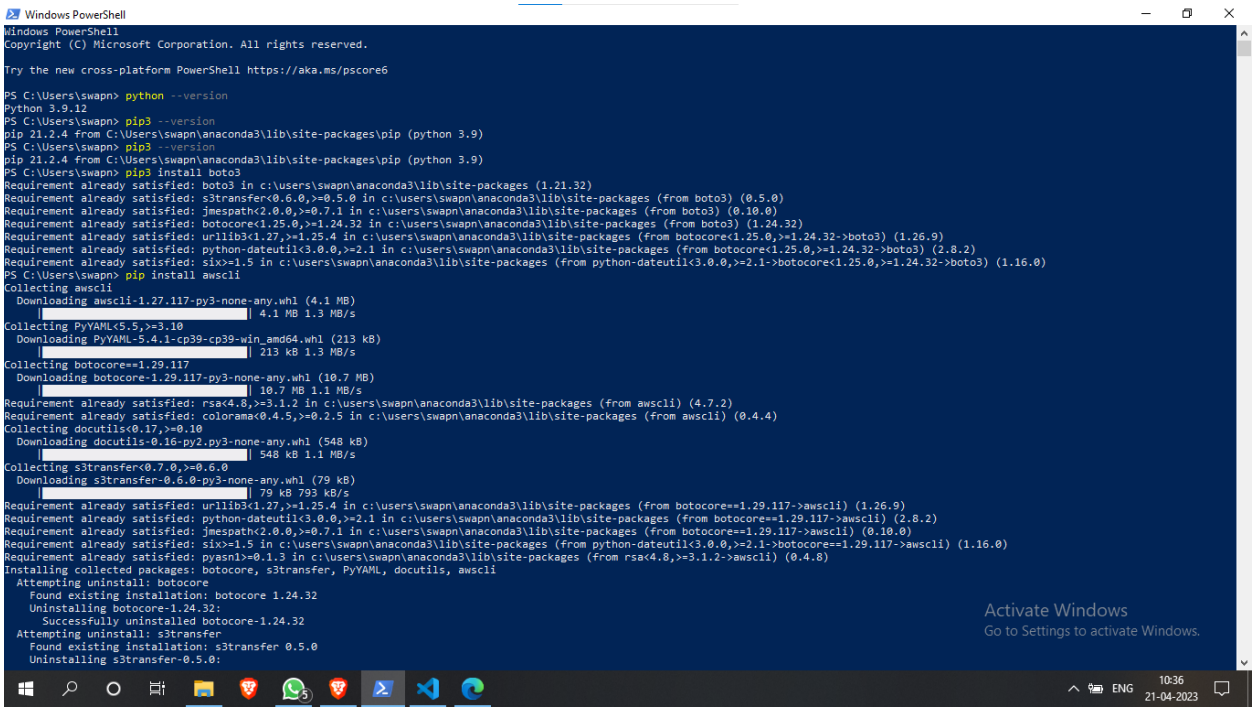


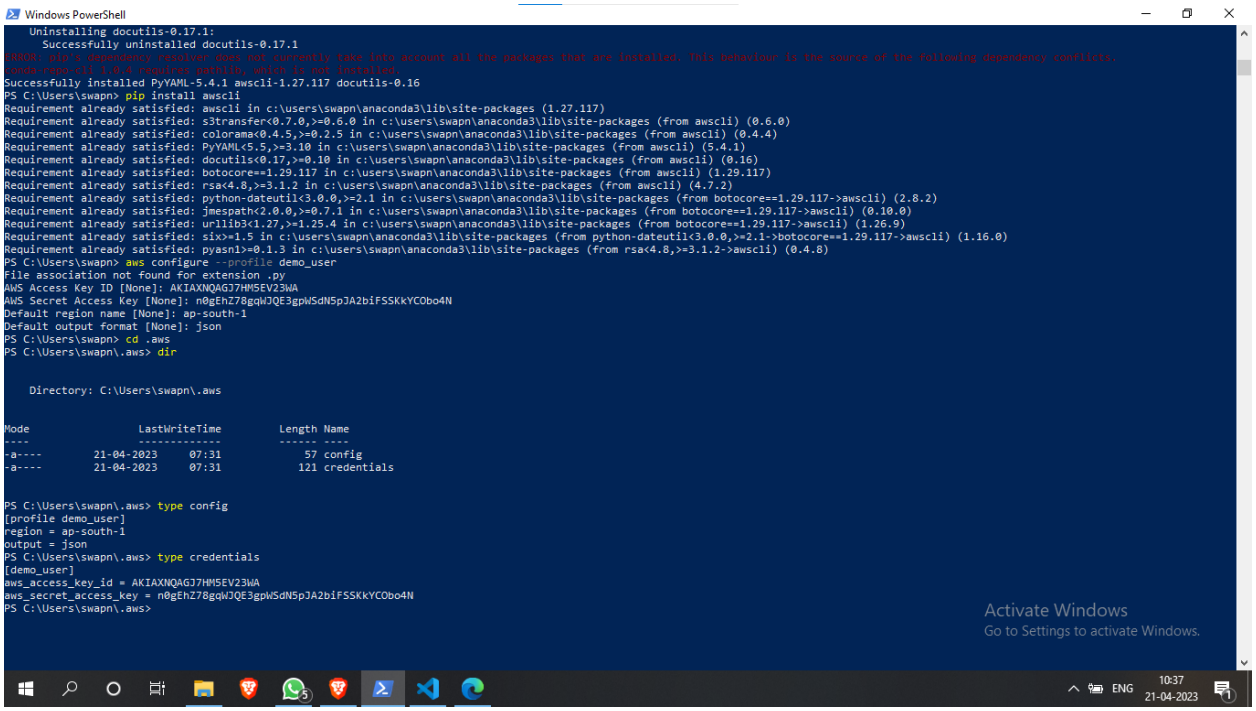
Note down the access key and secret access key in the file.



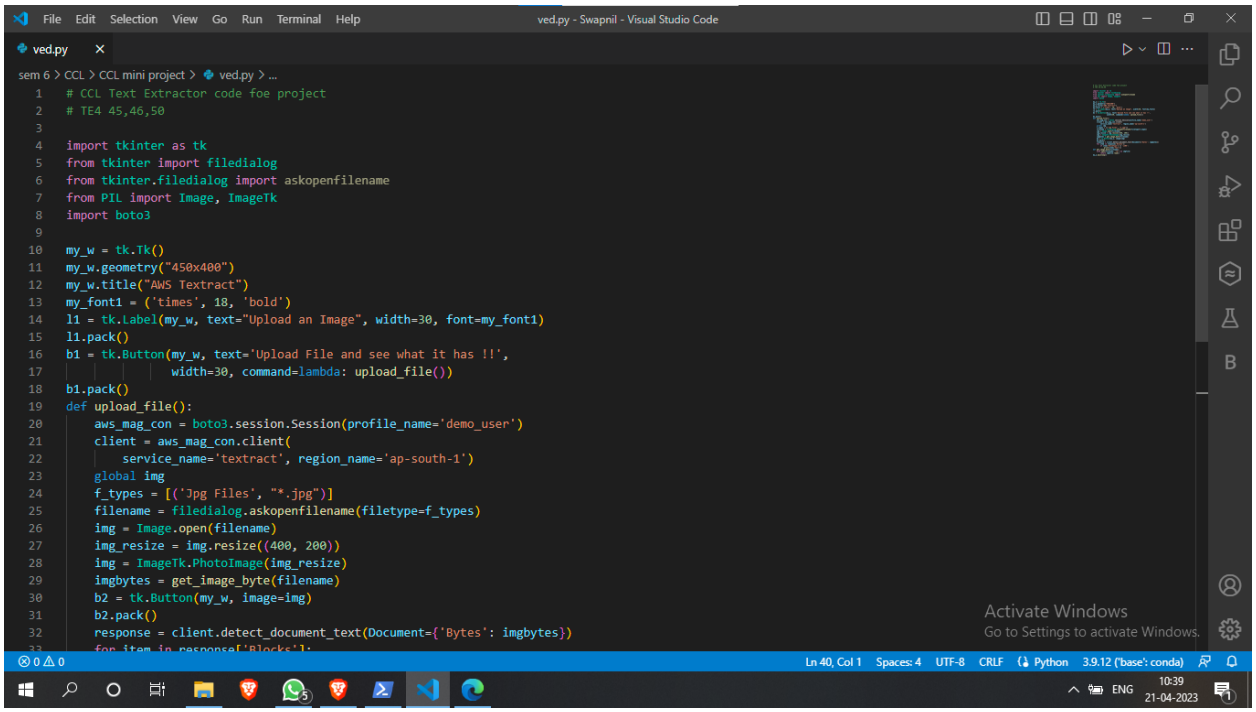
Configure the user for command line in powershell:

Download all the dependencies needed.

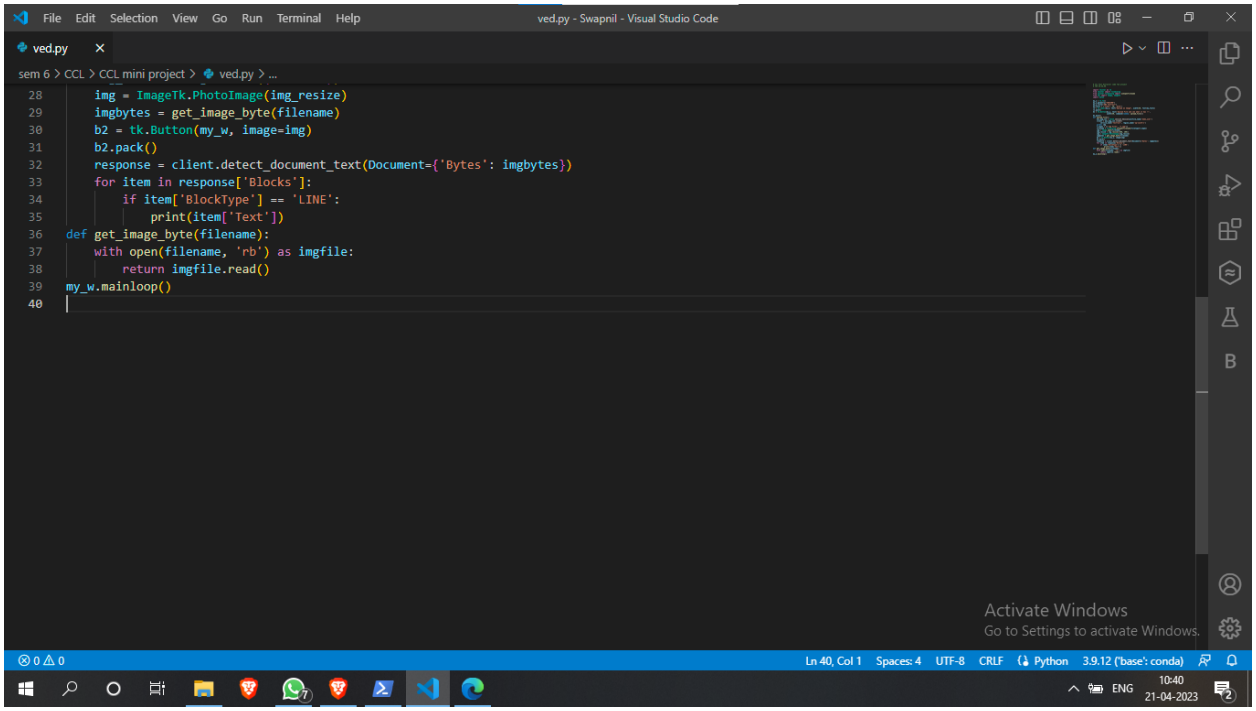




Write the code for the text extractor :

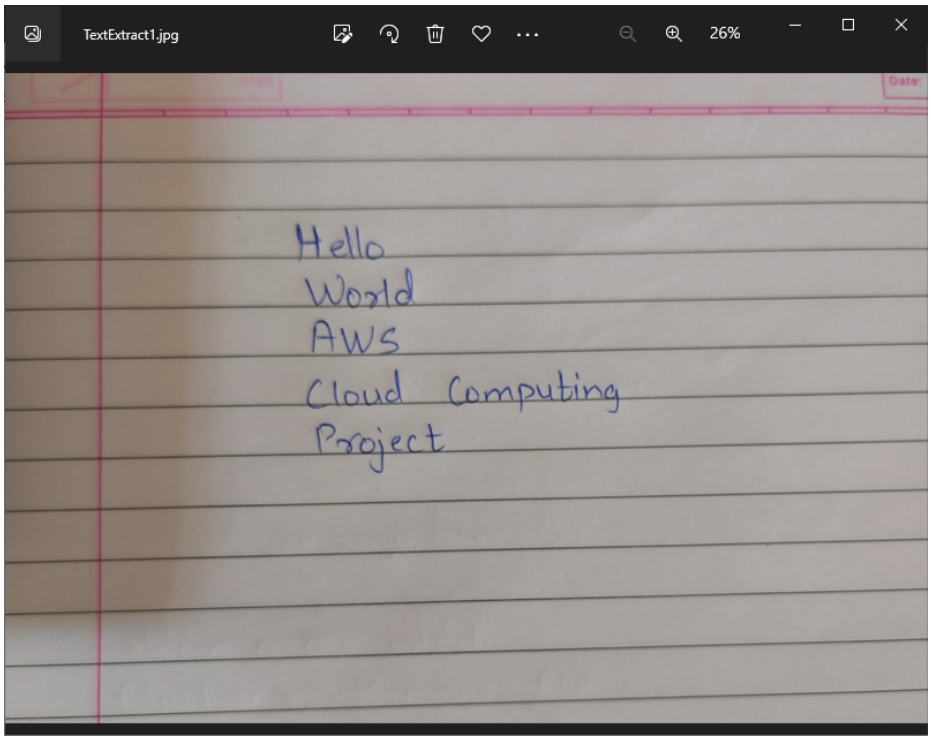


```
# CCL Text Extractor code foe project
# TE4 45,46,50

import tkinter as tk
from tkinter import filedialog
from tkinter.filedialog import askopenfilename
from PIL import Image, ImageTk
import boto3

my_w = tk.Tk()
my_w.geometry("450x400")
my_w.title("AWS Textract")
my_font1 = ('times', 18, 'bold')
l1 = tk.Label(my_w, text="Upload an Image", width=30, font=my_font1)
l1.pack()
b1 = tk.Button(my_w, text='Upload File and see what it has !!',
               width=30, command=lambda: upload_file())
b1.pack()
def upload_file():
    aws_mag_con = boto3.session.Session(profile_name='demo_user')
    client = aws_mag_con.client(
        service_name='textract', region_name='ap-south-1')
    global img
    f_types = [('Jpg Files', "*.jpg")]
    filename = filedialog.askopenfilename(filetype=f_types)
    img = Image.open(filename)
    img_resize = img.resize((400, 200))
    img = ImageTk.PhotoImage(img_resize)
    imgbytes = get_image_byte(filename)
    b2 = tk.Button(my_w, image=img)
    b2.pack()
    response = client.detect_document_text(Document={'Bytes': imgbytes})
```

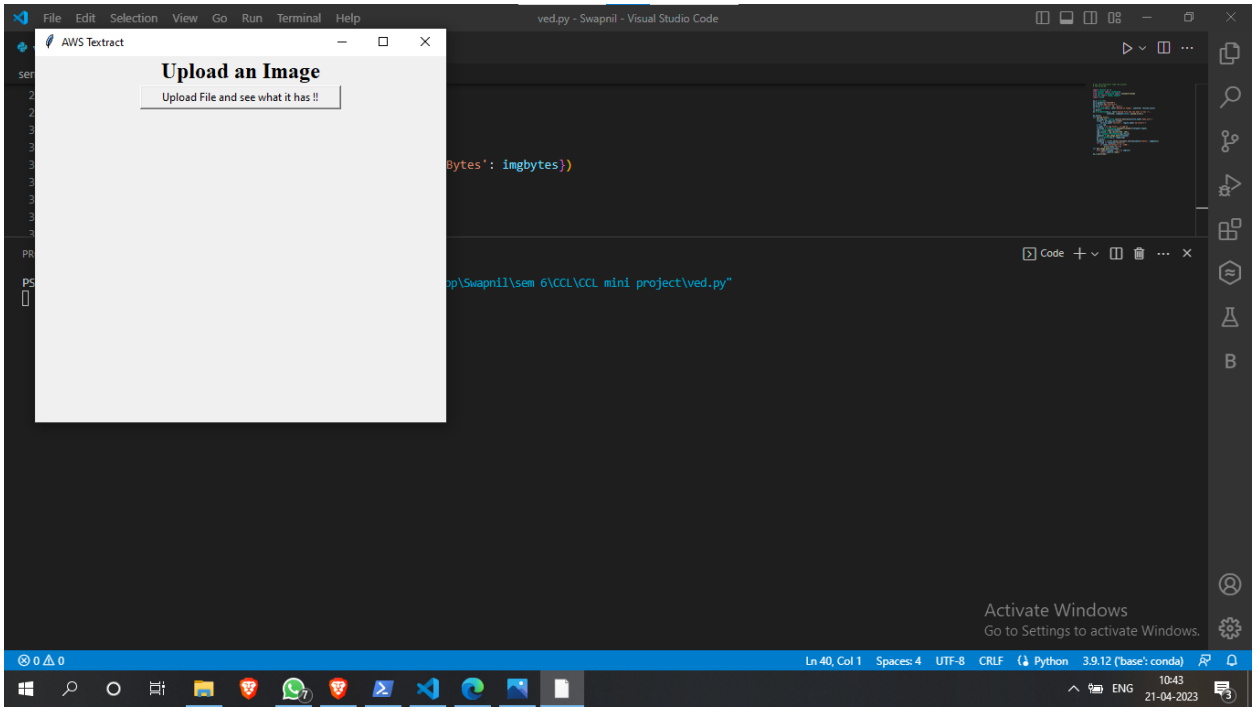


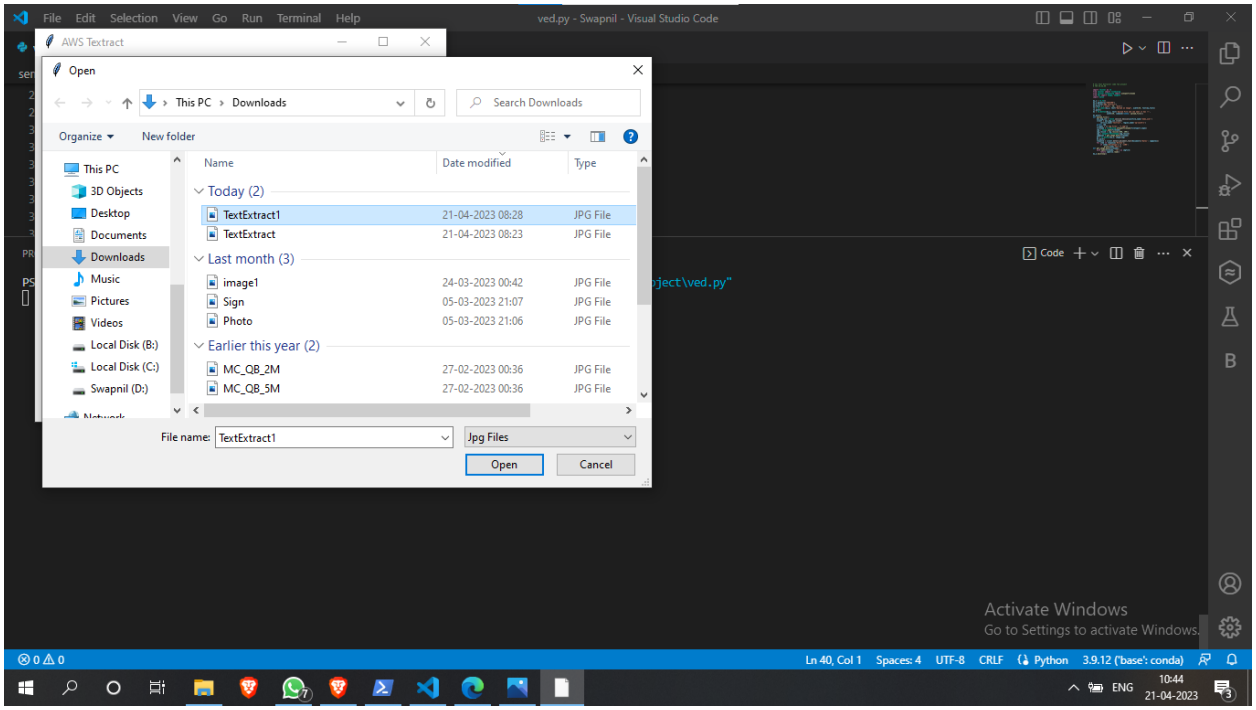
```
    img = ImageTk.PhotoImage(img_resize)
    imgbytes = get_image_byte(filename)
    b2 = tk.Button(my_w, image=img)
    b2.pack()
    response = client.detect_document_text(Document={'Bytes': imgbytes})
    for item in response['Blocks']:
        if item['BlockType'] == 'LINE':
            print(item['Text'])
def get_image_byte(filename):
    with open(filename, 'rb') as imgfile:
        return imgfile.read()
my_w.mainloop()
```

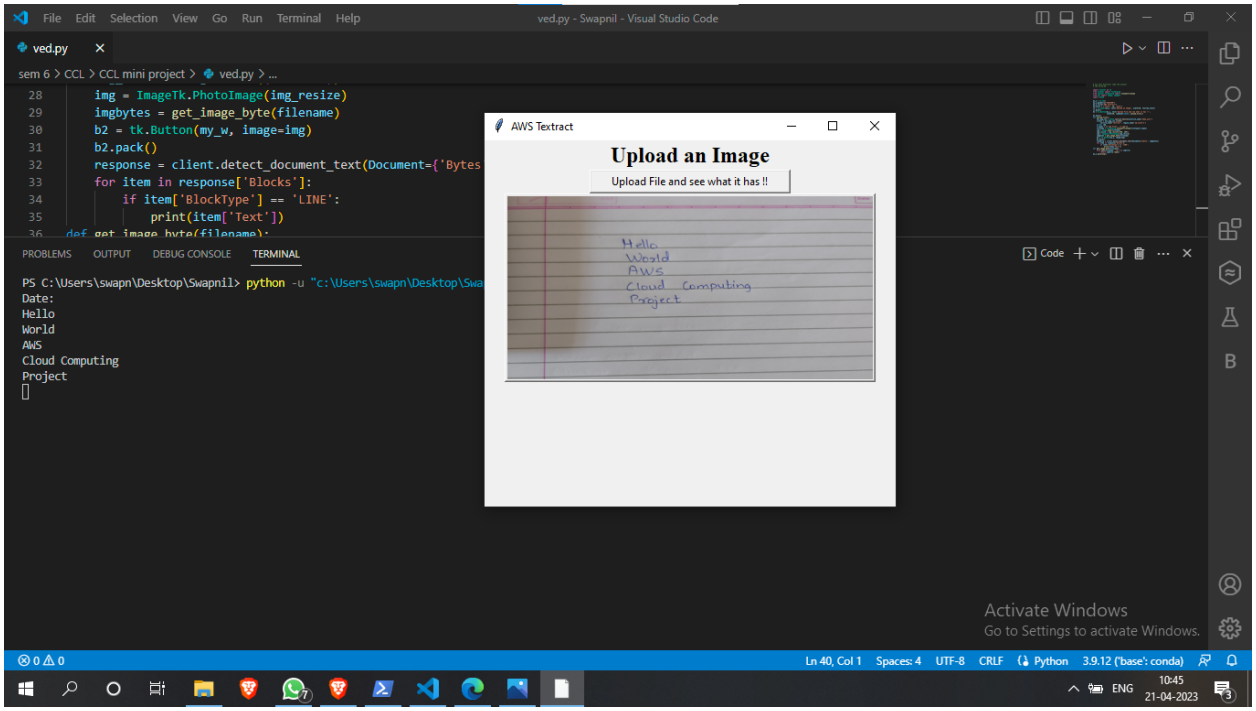
INPUT :



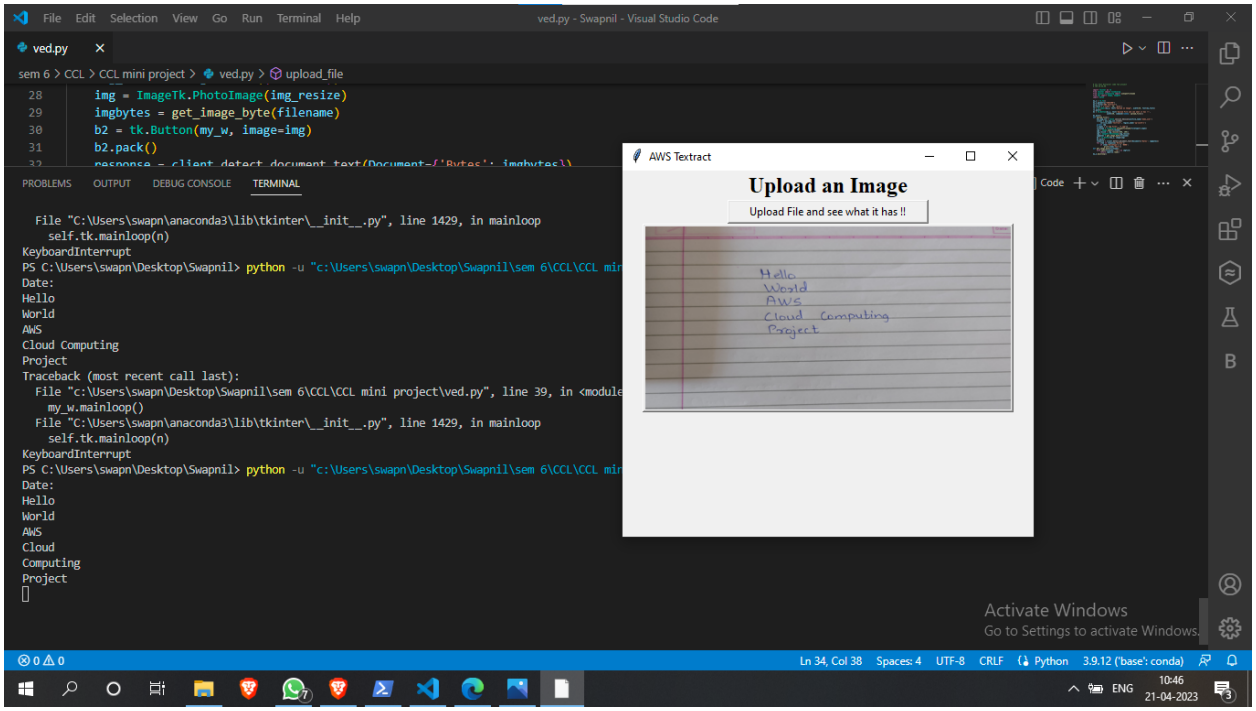
OUTPUT :

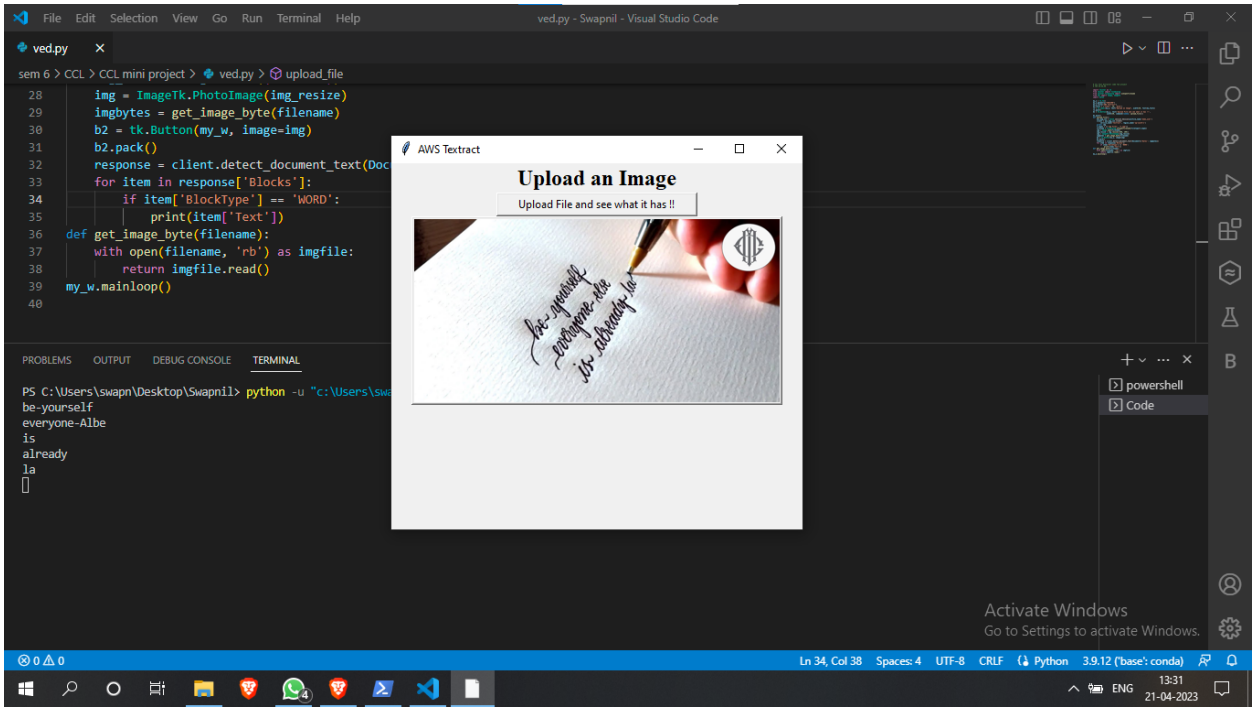



File Edit Selection View Go Run Terminal Help
ved.py - Swapnil - Visual Studio Code
AWS Textract
Open
This PC › Downloads
Search Downloads
Organize
New folder
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (B:)
Local Disk (C:)
Swapnil (D:)
Name
Date modified
Type
Today (2)
TextExtract1 21-04-2023 08:28 JPG File
TextExtract 21-04-2023 08:23 JPG File
Last month (3)
image1 24-03-2023 00:42 JPG File
Sign 05-03-2023 21:07 JPG File
Photo 05-03-2023 21:06 JPG File
Earlier this year (2)
MC_QB_2M 27-02-2023 00:36 JPG File
MC_QB_5M 27-02-2023 00:36 JPG File
File name: TextExtract1
Jpg Files
Open
Cancel
Activate Windows
Go to Settings to activate Windows.
Ln 40, Col 1 Spaces: 4 UTF-8 CRLF Python 3.9.12 ('base': conda)
ENG 10:44 21-04-2023


File Edit Selection View Go Run Terminal Help
ved.py - Swapnil - Visual Studio Code
ved.py
sem 6 > CCL > CCL mini project > ved.py > ...
img = ImageTk.PhotoImage(img_resize)
imgbytes = get_image_byte(filename)
b2 = tk.Button(my_w, image=img)
b2.pack()
response = client.detect_document_text(Document={'Bytes
for item in response['Blocks']:
if item['BlockType'] == 'LINE':
print(item['Text'])
def get_image_byte(filename):
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
PS C:\Users\swapn\Desktop\Swapnil> python -u "c:\Users\swapn\Desktop\Swa
Date:
Hello
World
AWS
Cloud Computing
Project
AWS Textract
Upload an Image
Upload File and see what it has !!
Hello
World
AWS
Cloud Computing
Project
Activate Windows
Go to Settings to activate Windows.
Ln 40, Col 1 Spaces: 4 UTF-8 CRLF Python 3.9.12 ('base': conda)
ENG 10:45 21-04-2023

To display word by word





Conclusion :

In conclusion, creating a Text Extractor with Amazon AWS Services is a worthwhile project that has the potential to greatly increase commercial and organizational productivity. The system can automatically extract text from photos and save it in a machine-readable format for subsequent processing by utilizing the AWS IAM, Textract, and AWS CLI services. The project has a number of difficulties, such as guaranteeing text extraction accuracy, enhancing system performance, and guaranteeing the confidentiality and privacy of extracted data. However, with proper design and execution, the project may offer considerable advantages to companies and organizations, simplifying their processes, lowering manual labor requirements, and enhancing overall effectiveness. This project is an excellent illustration of how contemporary technology may be utilized to address complicated issues and improve our way of life.

References :

https://aws.amazon.com/textract/

https://docs.aws.amazon.com/workspaces-web/latest/adminguide/getting-started-iam-user-access-keys.html

https://www.geeksforgeeks.org/how-to-extract-text-from-images-with-python/