

AIT 580 - 001 - Project Assignment 04

Name: Akhil Arekatika

G01445921

Title: AWS

The screenshot shows a web browser window with multiple tabs open at the top. The active tab is 'awsacademy.instructure.com/courses/51818'. The page content is titled 'AWS Academy Learner Lab [51818]'. On the left, there's a sidebar with icons for Account, Dashboard, Courses, Calendar, Inbox, History, and Help. The main area features a large graphic of a classical building with columns and clouds above it, labeled 'aws academy'. To the right, there are three buttons: 'View Course Stream', 'View Course Calendar', and 'View Course Notifications'. Below these are sections for 'To Do' (Nothing for now) and 'Recent Feedback' (Nothing for now). At the bottom of the browser window, the taskbar shows various application icons.

The screenshot shows the AWS Academy Learner Lab Student Guide page. On the left, there is a sidebar with icons for Account, Dashboard, Courses, Calendar, Inbox, History, and Help. The main content area is titled "AWS Academy Learner Lab Student Guide". It contains three sections: "AWS Academy Learner Lab Compliance and Security" (with a "Complete All Items" button), "Learn how to effectively use the Academy Learner Lab", and "Module Knowledge Check" (100 pts | Score at least 70.0). Below these are sections for "AWS Academy Learner Lab" (with "Launch AWS Academy Learner Lab" button) and "AWS Academy Learner Lab Resources" (with "Demo - How to Access Learner Lab", "Demo - General Troubleshooting Tips", and "Demo - How to Launch Services through AWS Console"). The bottom right corner shows system status: ENG IN, WiFi, battery level, and the date/time 29-11-2023 19:33.

The screenshot shows the "Launch AWS Academy Learner Lab" page. The sidebar is identical to the previous screenshot. The main content area shows a terminal window with the command "eee_W_2558226@runweb104920:~" and a "Learner Lab" sidebar on the right. The "Learner Lab" sidebar contains links for "Environment Overview", "Environment Navigation", "Access the AWS Management Console", "Region restriction", "Service usage and other restrictions", "Using the terminal in the browser", "Running AWS CLI commands", "Using the AWS SDK for Python", "Preserving your budget", "Accessing EC2 Instances", "SSH Access to EC2 Instances", "SSH Access from Windows", and "SSH Access from a Mac". At the bottom, there are "Previous" and "Next" navigation buttons.

The screenshot shows the AWS Console Home page. At the top, there's a navigation bar with tabs like 'AWS' and 'Services'. Below it is a search bar and a user info section. The main area is divided into several cards:

- Recently visited**: Links to EC2, RDS, AWS Glue DataBrew, S3, and IAM.
- Welcome to AWS**: Includes sections for 'Getting started with AWS' (with a rocket icon), 'Training and certification' (with a graduation cap icon), and 'What's new with AWS?' (with a lightbulb icon).
- AWS Health**: Shows 0 open issues, 1 scheduled change, and 0 other notifications.
- Cost and usage**: Displays current month costs and top costs for the current month.
- Build a solution**: A card for starting building with simple wizards and automated workflows.

At the bottom, there's a footer with links for CloudShell, Feedback, and various AWS services like Lambda, S3, and CloudWatch.

The screenshot shows the AWS EC2 Dashboard. On the left, there's a sidebar with navigation links for EC2 Global View, Events, Console-to-Code, Instances, Images, and Elastic Block Store. The main content area is divided into several sections:

- Resources**: A summary of EC2 resources: Instances (running: 1), Auto Scaling Groups (0), Dedicated Hosts (0), Elastic IPs (0), Instances (1), Key pairs (2), Load balancers (0), Placement groups (0), Security groups (6), Snapshots (0), and Volumes (1).
- Launch instance**: A button to 'Launch instance' and an option to 'Migrate a server'.
- Service health**: A link to the AWS Health Dashboard.
- Zones**: A table showing Zone name and Zone ID.
- Account attributes**: Information about the Default VPC (vpc-08639426935079ac2) and Settings (Data protection and security, Zones, EC2 Serial Console, Default credit specification, Console experiments).
- Explore AWS**: Promotional cards for 'Up to 90% on EC2 with Spot Instances' and 'Enable Best Price-Performance with AWS Graviton2'.

The footer includes a URL for the Launch Instance Wizard and standard AWS footer links.

The screenshot shows the 'Launch an instance' wizard on the AWS EC2 service. The 'Name and tags' section has 'Name' set to 'ait_final_project'. The 'Software Image (AMI)' section shows 'Amazon Linux 2023 AMI 2023.2.2...' and 'Virtual server type (instance type)' set to 't2.micro'. A tooltip for 'Free tier' is visible, stating: 'Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance'. The 'Launch instance' button is highlighted.

The screenshot shows the 'Launch an instance' wizard. A note says 'We noticed that you didn't select a key pair. If you want to be able to connect to your instance it is recommended that you create one.' Two options are shown: 'Create new key pair' (selected) and 'Proceed without key pair'. The 'Key pair name' field contains 'researchpaper'. The 'Key pair type' section shows 'RSA' selected. A tooltip for 'RSA' says: 'RSA encrypted private and public key pair'. The 'Private key file format' section shows '.pem' selected. A note at the bottom says: 'When prompted, store the private key in a secure and accessible location on your computer. You will need it later to connect to your instance.' The 'Create key pair' button is highlighted.

The screenshot shows the AWS EC2 'Launch an instance' success page. At the top, there's a green success banner stating 'Successfully initiated launch of instance (i-09fcbd10593c0c43c)'. Below the banner, a 'Next Steps' section provides links to various AWS services: 'Create billing and free tier usage alerts', 'Connect to your instance', 'Connect an RDS database', and 'Create EBS snapshot policy'. The bottom of the page includes standard AWS navigation links like CloudShell, Feedback, and a footer with copyright information.

The screenshot shows the AWS EC2 Instances page. On the left, a sidebar lists various EC2 services: EC2 Dashboard, EC2 Global View, Events, Console-to-Code, Instances (selected), Instances Types, Launch Templates, Spot Requests, Savings Plans, Reserved Instances, Dedicated Hosts, Capacity Reservations, Images (AMIs, AMI Catalog), and Elastic Block Store. The main pane displays a table of instances. One instance, 'ait_final_project' (Instance ID: i-09fcbd10593c0c43c), is highlighted and shown in more detail below. The instance details include its Public IPv4 address (34.203.239.142), Instance state (Running), Instance type (t2.micro), Status check (Initializing), Alarm status (No alarms), Availability Zone (us-east-1b), and Public IPv6 address (172.31.32.228). The bottom of the page includes standard AWS navigation links like CloudShell, Feedback, and a footer with copyright information.

Connect to instance [Info](#)

Connect to your instance i-09fcbd10593c0c43c (ait_final_project) using any of these options

EC2 Instance Connect | Session Manager | **SSH client** | EC2 serial console

Instance ID
i-09fcbd10593c0c43c (ait_final_project)

1. Open an SSH client.
2. Locate your private key file. The key used to launch this instance is researchpaper.pem
3. Run this command, if necessary, to ensure your key is not publicly viewable.
 chmod 400 researchpaper.pem
4. Connect to your instance using its Public DNS:
 ec2-34-203-239-142.compute-1.amazonaws.com

Example:
ssh -i "researchpaper.pem" ec2-user@ec2-34-203-239-142.compute-1.amazonaws.com

Note: In most cases, the guessed user name is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.

```
ec2-user@ip-172-31-32-228:~ % + ~
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-34-203-239-142.compute-1.amazonaws.com' (ED25519) to the list of known hosts.
Bad permissions. Try removing permissions for user: NT AUTHORITY\Authenticated Users (S-1-5-11) on file E:/AKHIL/AIT580/researchpaper.pem.
Permissions for 'researchpaper.pem' are too open.
It is required that your private key files are NOT accessible by others.
This private key will be ignored.
Load key "researchpaper.pem": bad permissions
ec2-user@ec2-34-203-239-142.compute-1.amazonaws.com: Permission denied (publickey,gssapi-keyex,gssapi-with-mic).

E:\AKHIL\AIT580>chmod 400 researchpaper.pem
'chmod' is not recognized as an internal or external command,
operable program or batch file.

E:\AKHIL\AIT580>ssh -i "researchpaper.pem" ec2-user@ec2-34-203-239-142.compute-1.amazonaws.com
#_ _###_ Amazon Linux 2023
\_ _###_
 \###|
  #/ __ https://aws.amazon.com/linux/amazon-linux-2023
 V` '---'
 .-. /_/
 _/ /` ~
 [ec2-user@ip-172-31-32-228 ~]$
```

The screenshot shows the AWS S3 service page. On the left, a sidebar lists various options like Buckets, Access Grants, and Storage Lens. The main area displays an 'Account snapshot' with metrics: Total storage (24.6 KB), Object count (4), and Average object size (6.1 KB). A callout box 'Introducing Amazon Q' provides information about receiving guidance and troubleshooting tips. Below this, a table lists 'General purpose buckets' with one entry: 'my-ait-s3-bucket' (US East (N. Virginia) us-east-1, Creation date: October 23, 2023, 20:09:25 (UTC-04:00)).

The screenshot shows the 'Create S3 bucket' page. It features a section titled 'Block Public Access settings for this bucket'. Under this, several checkboxes are listed: 'Block all public access' (unchecked), 'Block public access to buckets and objects granted through new access control lists (ACLs)' (unchecked), 'Block public access to buckets and objects granted through any access control lists (ACLs)' (unchecked), 'Block public access to buckets and objects granted through new public bucket or access point policies' (checked), and 'Block public and cross-account access to buckets and objects through any public bucket or access point policies' (checked). A warning message states: 'Turning off block all public access might result in this bucket and the objects within becoming public. AWS recommends that you turn on block all public access, unless public access is required for specific and verified use cases such as static website hosting.' A checkbox at the bottom accepts this risk. The page includes standard AWS navigation and status bars.

AWS Region: US East (N. Virginia) us-east-1

Bucket type: General purpose

Bucket name: aitfinalproject

Object Ownership: ACLs disabled (recommended)

Success message: Successfully created bucket "aitfinalproject". To upload files and folders, or to configure additional bucket settings, choose View details.

Total storage: 24.6 KB

Object count: 4

Average object size: 6.1 KB

Name	AWS Region	Access	Creation date
aitfinalproject	US East (N. Virginia) us-east-1	Objects can be public	November 29, 2023, 19:51:00 (UTC-05:00)
my-ait-s3-bucket	US East (N. Virginia) us-east-1	Objects can be public	October 23, 2023, 20:09:25 (UTC-04:00)

Objects (0) [Info](#)
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[C](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

Find objects by prefix

Name	Type	Last modified	Size	Storage class
No objects				

You don't have any objects in this bucket.

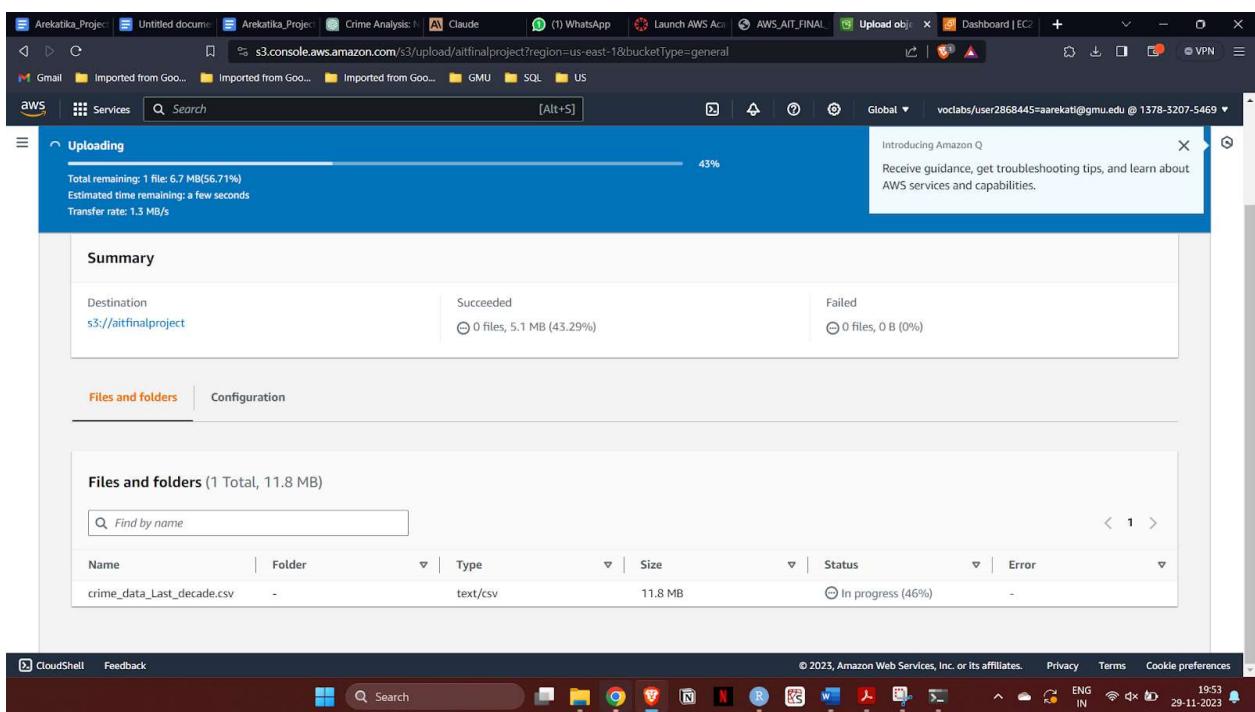
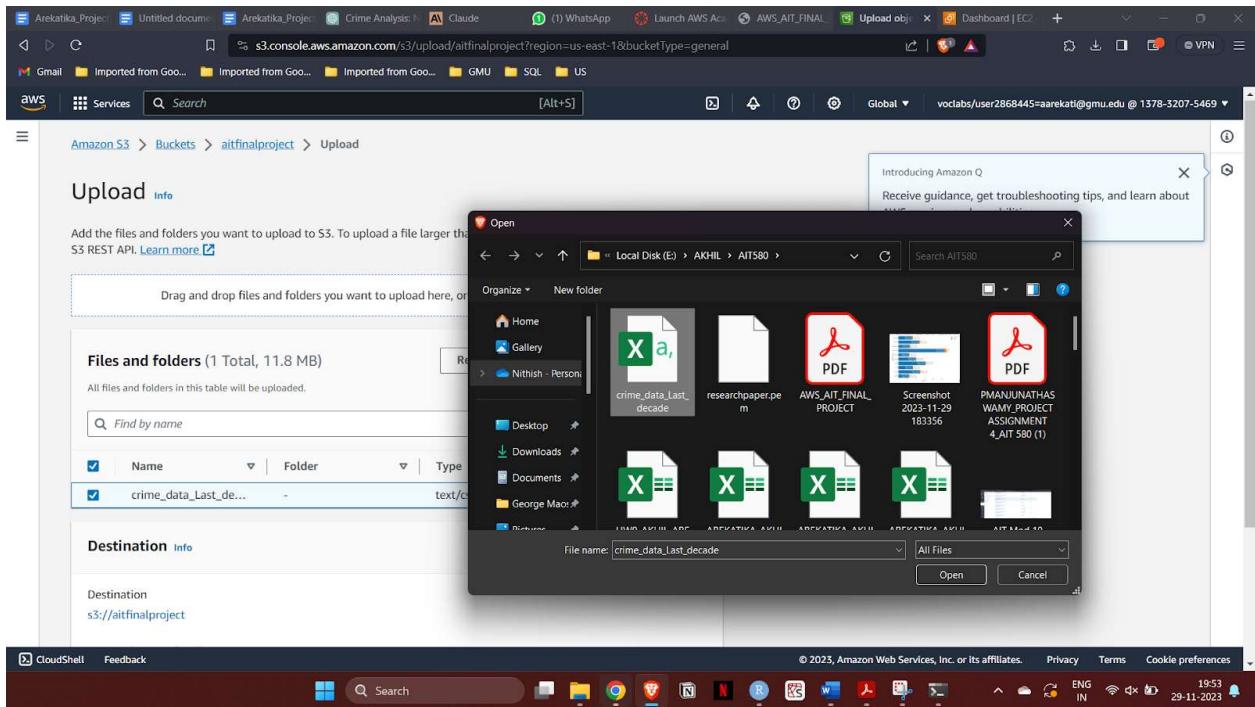
[Upload](#)

Upload [Info](#)
Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose [Add files](#) or [Add folder](#).

Files and folders (1 Total, 11.8 MB)				
All files and folders in this table will be uploaded.				
Find by name				
<input checked="" type="checkbox"/>	Name	Folder	Type	Size
<input checked="" type="checkbox"/>	crime_data_Last_de...	-	text/csv	11.8 MB

Destination [Info](#)
Destination
`s3://aitfinalproject`



Arekatika_Project Untitled document Arekatika_Proj Crime Analysis Claude (1) WhatsApp Launch AWS A... AWS_AIT_FINAL Upload object Dashboard | EC2 + - X

Gmail Imported from Goo... Imported from Goo... Imported from Goo... GMU SQL US

aws Services Search [Alt+S] Global v vodlabs/user2868445=aarekati@gmu.edu @ 1378-3207-5469

Upload succeeded View details below.

The information below will no longer be available after you navigate away from this page.

Summary

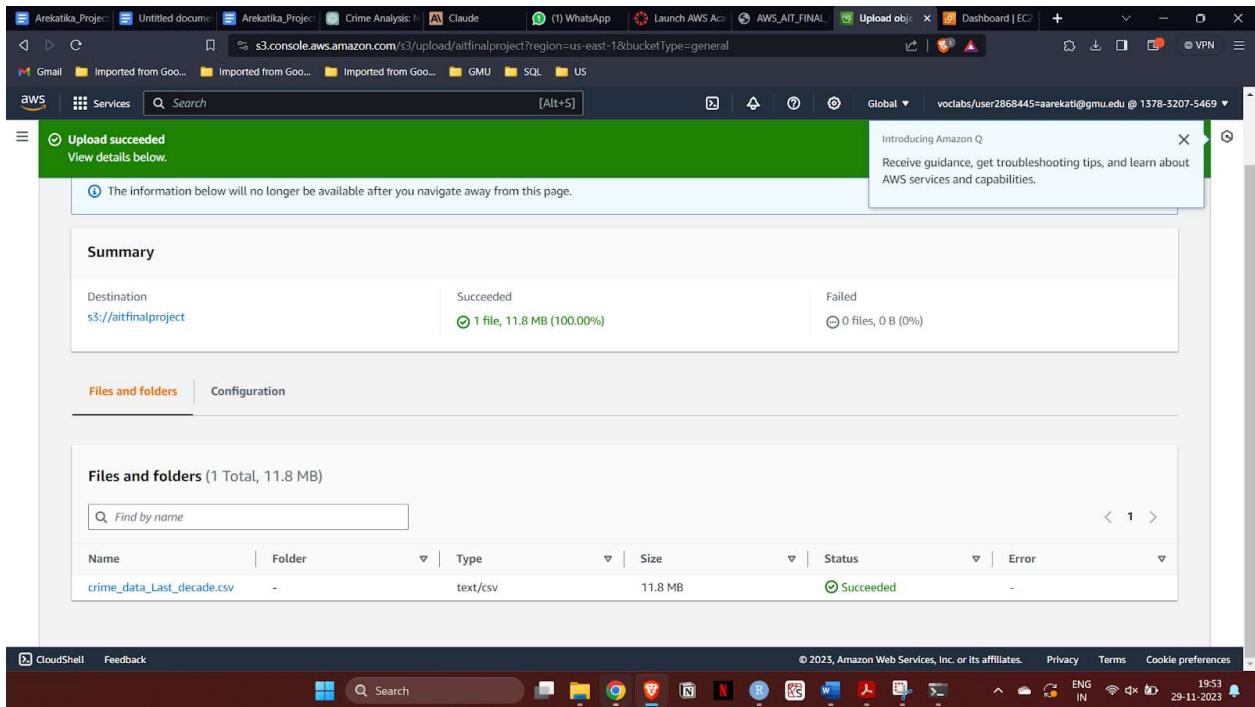
Destination	Succeeded	Failed
s3://aitfinalproject	1 file, 11.8 MB (100.00%)	0 files, 0 B (0%)

Files and folders Configuration

Files and folders (1 Total, 11.8 MB)

Find by name						
Name	Folder	Type	Size	Status	Error	
crime_data_Last_decade.csv	-	text/csv	11.8 MB	Succeeded	-	

CloudShell Feedback © 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences ENG IN 19:53 29-11-2023



Arekatika_Proj Untitled document Arekatika_Proj Crime Analysis Claude (1) WhatsApp Launch AWS A... AWS_AIT_FINAL Upload object crime_data Dashboard | E2 + - X

Gmail Imported from Goo... Imported from Goo... Imported from Goo... GMU SQL US

aws Services Search [Alt+S] Global v vodlabs/user2868445=aarekati@gmu.edu @ 1378-3207-5469

Amazon S3 Buckets aitfinalproject crime_data_Last_decade.csv

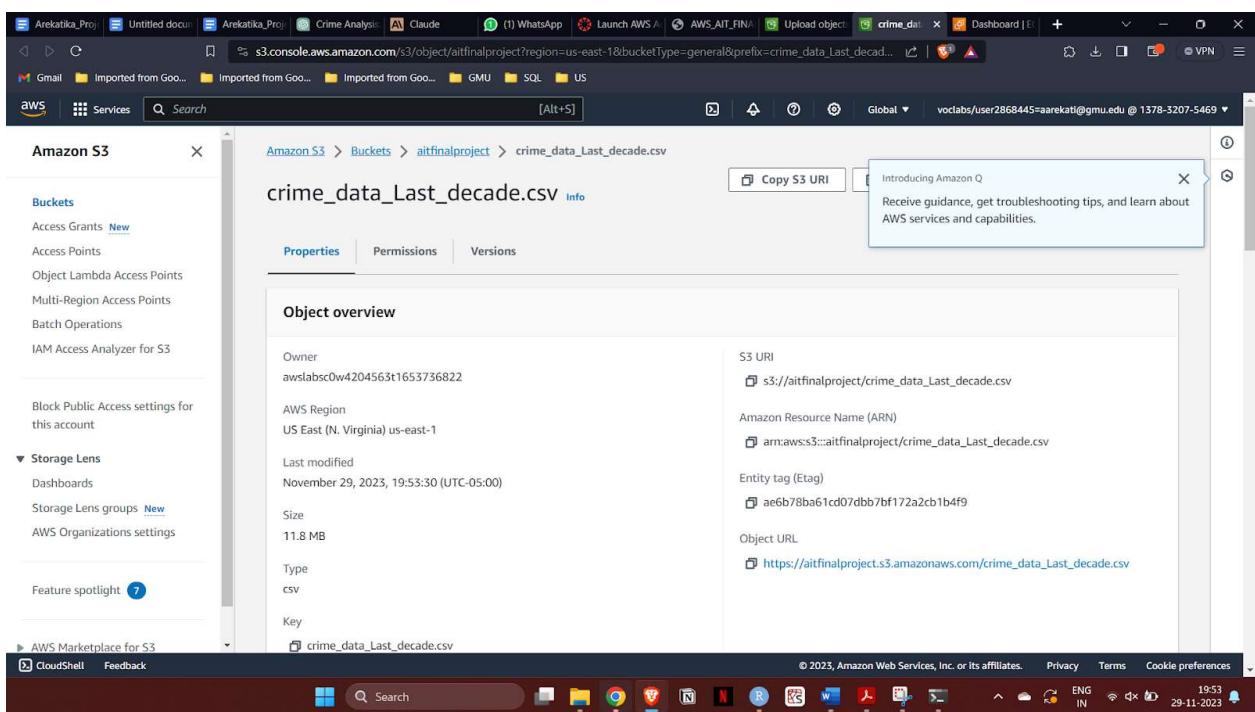
Copy S3 URI Introducing Amazon Q Receive guidance, get troubleshooting tips, and learn about AWS services and capabilities.

Properties Permissions Versions

Object overview

Owner	S3 URI
awslabsc0w4204563t1653736822	s3://aitfinalproject/crime_data_Last_decade.csv
AWS Region	Amazon Resource Name (ARN)
US East (N. Virginia) us-east-1	arn:aws:s3:::aitfinalproject/crime_data_Last_decade.csv
Last modified	Entity tag (Etag)
November 29, 2023, 19:53:30 (UTC-05:00)	ae6b78ba61cd07dbb7bf172a2cb1b4f9
Size	Object URL
11.8 MB	https://aitfinalproject.s3.amazonaws.com/crime_data_Last_decade.csv
Type	
csv	
Key	
crime_data_Last_decade.csv	

CloudShell Feedback © 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences ENG IN 19:53 29-11-2023



```
C:\Windows\System32\cmd.exe + Microsoft Windows [Version 10.0.23580.1000] (c) Microsoft Corporation. All rights reserved. E:\AKHIL\AIT580>scp usage: scp [-346ABCpqrv] [-c cipher] [-F ssh_config] [-i identity_file] [-J destination] [-l limit] [-o ssh_option] [-P port] [-S program] source ... target E:\AKHIL\AIT580>scp -i "E:\AKHIL\AIT580\researchpaper.pem" "E:\AKHIL\AIT580\Starbucks satisfactory survey.csv" ec2-user@34.203.239.142:files ssh: connect to host 34.203.0.239 port 22: Connection timed out lost connection E:\AKHIL\AIT580>scp -i "E:\AKHIL\AIT580\researchpaper.pem" "E:\AKHIL\AIT580\Starbucks satisfactory survey.csv" ec2-user@34.203.239.142:files The authenticity of host '34.203.239.142' (34.203.239.142) can't be established. ED25519 key fingerprint is SHA256:qQe6gvE0d1LAdMdFYK35w0/Uy/L/8A5L6EqPatqmD4. This host key is known by the following other names/addresses: C:\Users\DELL\.ssh\known_hosts:1: ec2-24-203-239-142.compute-1.amazonaws.com Are you sure you want to continue connecting (yes/no/[fingerprint])? Warning: Permanently added '34.203.239.142' (ED25519) to the list of known hosts. Starbucks satisfactory survey.csv 100% 27KB 846.6KB/s 00:00 E:\AKHIL\AIT580>scp -i "E:\AKHIL\AIT580\researchpaper.pem" "E:\AKHIL\AIT580\crime_data_Last_decade.csv" ec2-user@34.203.239.142:files crime_data_Last_decade.csv 100% 12MB 1.4MB/s 00:08 E:\AKHIL\AIT580>
```

```
C:\Windows\System32\cmd.e Microsoft Windows [Version 10.0.23580.1000] (c) Microsoft Corporation. All rights reserved. E:\AKHI ec2-user@ip-172-31-32-228~ usage: Microsoft Windows [Version 10.0.23580.1000] (c) Microsoft Corporation. All rights reserved. E:\AKHI E:\AKHIL\AIT580>ssh -i "researchpaper.pem" ec2-user@ec2-34-203-239-142.compute-1.amazonaws.com files ssh: connect to host ec2-34-203-239-142.compute-1.amazonaws.com port 22: Connection refused Amazon Linux 2023 E:\AKHI The auth https://aws.amazon.com/linux/amazon-linux-2023 ED25519 This host C:\ Are you Warning Starbuc Last login: Thu Nov 30 02:11:27 2023 from 18.206.107.29 Starbuc [ec2-user@ip-172-31-32-228 ~]$ ls files E:\AKHI [ec2-user@ip-172-31-32-228 ~]$ cd files [ec2-user@ip-172-31-32-228 files]$ ls 'Starbucks satisfactory survey.csv' E:\AKHI [ec2-user@ip-172-31-32-228 files]$ ls 'Starbucks satisfactory survey.csv' crime_data_Last_decade.csv [ec2-user@ip-172-31-32-228 files]$
```

ec2-user@ip-172-31-32-228:~

```
2021,"DECEMBER","CONDITIONAL RELEASE","SARATOGA","FEMALE",39,"CPCS 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SARATOGA","MALE",36,"IDENTITY THEFT 2"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",23,"BURGLARY 2ND, SUB 2"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",23,"C POS WEAPON 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",22,"SCSC 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",30,"SCSC 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",40,"MURDER 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",34,"TAMPERING EVIDENCE"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",44,"BURGLARY 2ND, SUB 2"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",49,"SEXUAL ABUSE 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",61,"ARSON 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",28,"BURGLARY 2ND, SUB 2"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",45,"SCSC 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",41,"ASSAULT 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",48,"SCSC 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",55,"SCSC 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",71,"SEX CONDUCT CHILD 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",46,"SCSC 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","FEMALE",35,"C POS WEAPON 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHENECTADY","MALE",30,"C POS WEAPON 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SCHOHARIE","MALE",39,"SCSC 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SENECA","MALE",25,"CPCS 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SENECA","MALE",50,"RAPE 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","STEUBEN","MALE",33,"CPCS 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","STEUBEN","MALE",46,"RAPE 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","STEUBEN","MALE",72,"RAPE 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","STEUBEN","MALE",40,"CPCS 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",30,"ROBBERY 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",55,"BURGLARY 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",49,"BURGLARY 1ST"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",49,"CPCS 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",60,"MURDER 2ND"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",60,"ASSAULT 1ST B FELONY"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",21,"BURGLARY 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",21,"YO DANG WEAPONS"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",43,"CPCS 4TH"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",44,"CPCS 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",55,"CPCS 3RD"
2021,"DECEMBER","CONDITIONAL RELEASE","SUFFOLK","MALE",34,"FAIL REGIS SEX 2ND"
```

```

ec2-user@ip-172-31-32-228:~ + 
Collecting pillow>=8
  Downloading Pillow-10.1.0-cp39-cp39-manylinux_2_28_x86_64.whl (3.6 MB)
    |████████| 3.6 MB 30.6 MB/s
Collecting cycler>=0.10
  Downloading cycler-0.12.1-py3-none-any.whl (8.3 kB)
Collecting importlib-resources>=3.2.0
  Downloading importlib_resources-6.1.1-py3-none-any.whl (33 kB)
Collecting fonttools>=4.22.0
  Downloading fonttools-4.45.1-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (4.6 MB)
    |████████| 4.6 MB 30.2 MB/s
Collecting pyparsing>=2.3.1
  Downloading pyparsing-3.1.1-py3-none-any.whl (103 kB)
    |████████| 103 kB 50.6 MB/s
Collecting kiwisolver>=1.3.1
  Downloading kiwisolver-1.4.5-cp39-cp39-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (1.6 MB)
    |████████| 1.6 MB 24.2 MB/s
Requirement already satisfied: python-dateutil>=2.7 in /home/ec2-user/.local/lib/python3.9/site-packages (from matplotlib) (2.8.2)
Collecting zipp>=3.1.0
  Downloading zipp-3.17.0-py3-none-any.whl (7.4 kB)
Requirement already satisfied: six>=1.5 in /usr/lib/python3.9/site-packages (from python-dateutil>=2.7>matplotlib) (1.15.0)
Installing collected packages: zipp, pyparsing, pillow, packaging, kiwisolver, importlib-resources, fonttools, cycler, contourpy, matplotlib
Successfully installed contourpy-1.2.0 cycler-0.12.1 fonttools-4.45.1 importlib-resources-6.1.1 kiwisolver-1.4.5 matplotlib-3.8.2 packaging-23.2 pillow-10.1
.0 pyparsing-3.1.1 zipp-3.17.0
[ec2-user@ip-172-31-32-228 files]$ python3 aitfinal.py
   Year      Month     Release_Type County_of_Commitment   Sex Age_at_Release Most_Serious_Crime
0  2013    JANUARY        PAROLE       ALBANY    MALE    34      BURGLARY 3RD
1  2013    JANUARY        PAROLE       ALBANY    MALE    35  SEXUAL ABUSE 1ST
2  2013    JANUARY        PAROLE       ALBANY    MALE    36      BURGLARY 3RD
3  2013    JANUARY        PAROLE       ALBANY    MALE    42      RAPE 1ST
4  2013    JANUARY        PAROLE       ALBANY    MALE    48      BURGLARY 2ND
...  ...  ...
184783  2022 DECEMBER MAXIMUM EXPIRATION SULLIVAN    MALE    23      BURGLARY 3RD
184784  2022 DECEMBER MAXIMUM EXPIRATION ULSTER    MALE    27  CONTRABAND 1ST
184785  2022 DECEMBER MAXIMUM EXPIRATION ULSTER    MALE    35      ASSAULT 2ND
184786  2022 DECEMBER MAXIMUM EXPIRATION BRONX    MALE    31      ASSAULT 2ND
184787  2022 DECEMBER MAXIMUM EXPIRATION BRONX    MALE    53      ROBBERY 1ST

[184788 rows x 7 columns]
[ec2-user@ip-172-31-32-228 files]$ |

```

```

184787 2022 DECEMBER MAXIMUM EXPIRATION BRONX    MALE    53      ROBBERY 1ST

[184788 rows x 7 columns]
[ec2-user@ip-172-31-32-228 files]$ nano aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ python3 aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ nano aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ python3 aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ ls

```

```

ec2-user@ip-172-31-32-228:~ + 
GNU nano 5.8                                     aitfinal.py
import pandas as pd
import matplotlib.pyplot as plt
crime_data = pd.read_csv('crime_data_Last_decade.csv')
# Group the data by 'COUNTY_OF_COMMITTMENT' and count the number of cities with most crimes
county_counts = crime_data['County_of_Commitment'].value_counts()

# Sort the counties for a better-organized bar plot
sorted_counties = county_counts.head(10)

# Plotting the horizontal bar chart with the highest crimes county at first
plt.figure(figsize=(10, 6))
sorted_counties[-1].plot(kind='barh', color='skyblue') # Reverse the order
plt.title('Top 10 Counties with the Most Crimes')
plt.xlabel('Number of Crimes')
plt.ylabel('County')
plt.grid(axis='x', linestyle='--', alpha=0.7) # Use 'x' for grid on the horizontal axis
plt.savefig('topcrimecounties')

^G Help      ^O Write Out    ^W Where Is      ^K Cut          ^J Execute      ^C Location      M-U Undo      M-A Set Mark    M-] To Bracket
^X Exit      ^R Read File    ^W Replace      ^U Paste         ^J Justify      ^G Go To Line    M-E Redo      M-G Copy       ^Q Where Was
                                         Search

```

```

[ec2-user@ip-172-31-32-228:~] + -
Collecting importlib-resources>=3.2.0
  Downloading importlib_resources-6.1.1-py3-none-any.whl (33 kB)
Collecting fonttools>=4.22.0
  Downloading fonttools-4.45.1-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (4.6 MB)
| 4.6 MB 30.2 MB/s
Collecting pyparsing>=2.3.1
  Downloading pyparsing-3.1.1-py3-none-any.whl (103 kB)
| 103 kB 50.6 MB/s
Collecting kiwisolver>=1.3.1
  Downloading kiwisolver-1.4.5-cp39-cp39-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (1.6 MB)
| 1.6 MB 24.2 MB/s
Requirement already satisfied: python-dateutil>=2.7 in /home/ec2-user/.local/lib/python3.9/site-packages (from matplotlib) (2.8.2)
Collecting zipp>=3.1.0
  Downloading zipp-3.17.0-py3-none-any.whl (7.4 kB)
Requirement already satisfied: six>=1.5 in /usr/lib/python3.9/site-packages (from python-dateutil>=2.7->matplotlib) (1.15.0)
Installing collected packages: zipp, pyparsing, pillow, packaging, kiwisolver, importlib-resources, fonttools, cycler, contourpy, matplotlib
Successfully installed contourpy-1.2.0 cycler-0.12.1 fonttools-4.45.1 importlib-resources-6.1.1 kiwisolver-1.4.5 matplotlib-3.8.2 packaging-23.2 pillow-10.1.0 pyparsing-3.1.1 zipp-3.17.0
[ec2-user@ip-172-31-32-228 files]$ python3 aitfinal.py
   Year      Month     Release_Type County_of_Commitment   Sex  Age_at_Release Most_Serious_Crime
0    2013    JANUARY        PAROLE      ALBANY    MALE       34      BURGLARY 3RD
1    2013    JANUARY        PAROLE      ALBANY    MALE       35      SEXUAL_ABUSE 1ST
2    2013    JANUARY        PAROLE      ALBANY    MALE       36      BURGLARY 3RD
3    2013    JANUARY        PAROLE      ALBANY    MALE       42      RAPE 1ST
4    2013    JANUARY        PAROLE      ALBANY    MALE       48      BURGLARY 2ND
...    ...
184783  2022  DECEMBER  MAXIMUM_EXPIRATION  SULLIVAN    MALE       23      BURGLARY 3RD
184784  2022  DECEMBER  MAXIMUM_EXPIRATION  ULSTER     MALE       27      CONTRABAND 1ST
184785  2022  DECEMBER  MAXIMUM_EXPIRATION  ULSTER     MALE       35      ASSAULT 2ND
184786  2022  DECEMBER  MAXIMUM_EXPIRATION  BRONX     MALE       31      ASSAULT 2ND
184787  2022  DECEMBER  MAXIMUM_EXPIRATION  BRONX     MALE       53      ROBBERY 1ST
[184788 rows x 7 columns]
[ec2-user@ip-172-31-32-228 files]$ nano aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ python3 aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ nano aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ python3 aitfinal.py
[ec2-user@ip-172-31-32-228 files]$ ls
'Starbucks satisfactory survey.csv'  aitfinal.py  crime_data_Last_decade.csv  topcrimecounties.png
[ec2-user@ip-172-31-32-228 files]$ |

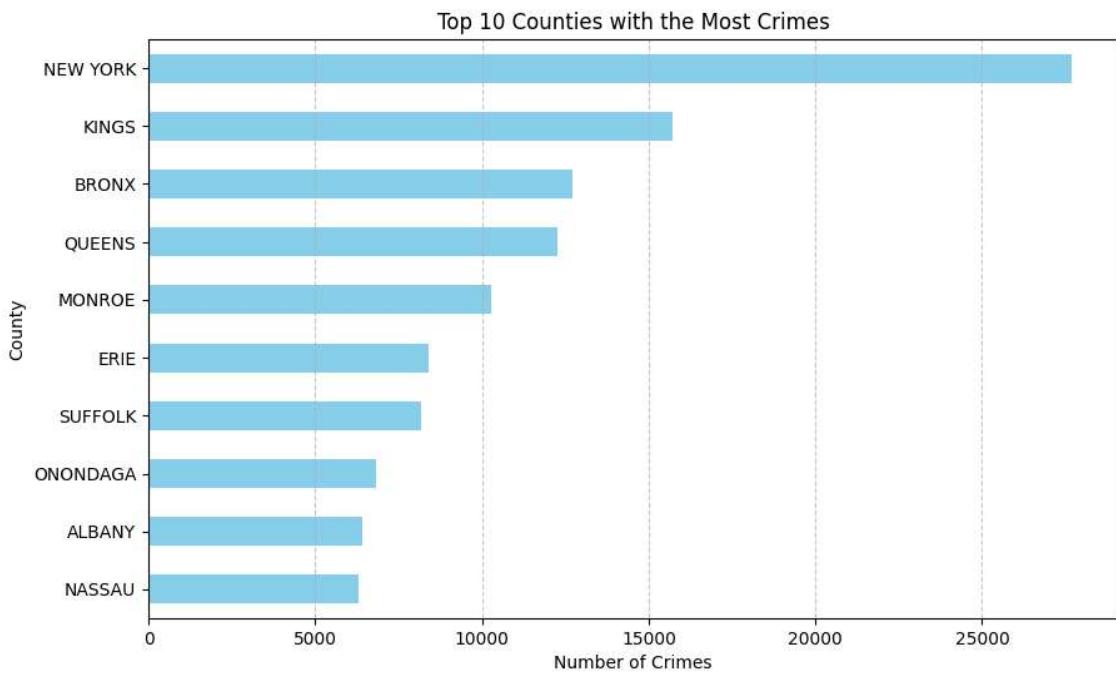
```

```

[ec2-user@ip-172-31-32-228 files]$ C:\Windows\System32\cmd.e
Microsoft Windows [Version 10.0.23580.1000]
(c) Microsoft Corporation. All rights reserved.

E:\AKHIL\AIT580>scp -i "E:\AKHIL\AIT580\researchpaper.pem" ec2-user@34.203.239.142:files/topcrimecounties.png "E:\AKHIL\AIT580"
100% 35KB 1.2MB/s 00:00
E:\AKHIL\AIT580>

```



us-east-1.console.aws.amazon.com/databrew/home?region=us-east-1#create-dataset

DATASETS

Dataset name
crimedata
The dataset name must contain 1-255 characters. Valid characters are alphanumeric (A-Z, a-z, 0-9), hyphen (-), period (.), and space.

Connect to new dataset Info

File upload
Enter your source from S3Info
For you to select a folder, all files in the folder need to share the same file type. If there are different schemas, they will be merged.
S3://aitfinalproject/crime_data_Last_decade.csv

Amazon S3
Format is: s3://bucket/prefix
crime_data_Last_decade.csv is selected

S3 Buckets > aitfinalproject
Select the entire folder

Define dynamic dataset parameters

- What can I do with parameters?
 - Example of RegEx that can be added
 - Select files in parent folder only
 - Select files ending with .csv only
 - Select files ending with .csv in parent folder only

Choose filtered files Info

Specify number of files to include
Latest 10

Specify last updated date range
Past 24 hours

Gmail Imported from Goo... Imported from Goo... Imported from Goo... GMU SQL US

Services Search [Alt+S] Create custom parameter

DATASETS

PROJECTS RECIPES DQ RULES JOBS WHAT'S NEW

Additional configurations

Selected file type Format of the selected file CSV delimiter

CSV Comma (,)

JSON

PARQUET

EXCEL

ORC

Preview data

Column header values

Treat first row as header The first row in your dataset will be treated as column header values

Add default header Default headers will be added with values Column_1, Column_2 ...

Tags - optional

Metadata that you can define and assign to AWS resources. Each tag is a simple label consisting of a customer-defined key (name) and an optional value. Using tags can make it easier for you to manage, search for, and filter resources by purpose, owner, environment, or other criteria.

Cancel **Create dataset**

Gmail Imported from Goo... Imported from Goo... Imported from Goo... GMU SQL US

Services Search [Alt+S] N. Virginia vocabs/user2868445=aarekat@gmu.edu @ 13

Created dataset "crimedata".

DataBrew > Datasets

Datasets (4) Info

View details Create project with this dataset Run data profile Actions Connect new dataset

Find datasets

Dataset name	Data type	Data profile	Source	Location	Create date
crimedata	CSV	-	S3	s3://aitfinalproject/crime_data_Last_decade.csv	in a few seconds November 29, 2023, 9:41:35 pm
imdb-dataset	CSV	-	S3	s3://my-ait-s3-bucket/Glu-folder/Top 100 IMDB Rated Series.csv	23 days ago November 6, 2023, 11:37:48 pm
USH	CSV	-	S3	s3://my-ait-s3-bucket/Glu-folder/	a month ago November 3, 2023, 11:44:39 pm
US-Hurricanes.csv	CSV	-	S3	s3://my-ait-s3-bucket/	a month ago November 3, 2023, 11:32:49 pm

Gmail Imported from Goo... Imported from Goo... Imported from Goo... GMU SQL US

Services Search [Alt+S] N. Virginia vocabs/user2868445=aarekati@gmu.edu @ 1378-3207-5

Created dataset "crimedata".

DataBrew > Projects > Create project

Create project Info

Project details

Project name The project name must contain 1-255 characters. Valid characters are alphanumeric (A-Z, a-z, 0-9), hyphen (-), period (.), and space.

Recipe details Info
Data cleaning steps in DataBrew are stored as a recipe. A recipe is connected to a project by default. An existing recipe with no associated project could also be applied to a project.

Attached recipe Recipe name The recipe name must contain 1-255 characters. Valid characters are alphanumeric (A-Z, a-z, 0-9), hyphen (-), period (.), and space.

Import steps from recipe Import recipe steps from an existing recipe into your project. The existing recipe that you chose will not be edited.

Gmail Imported from Goo... Imported from Goo... Imported from Goo... GMU SQL US

Services Search [Alt+S] N. Virginia vocabs/user2868445=aarekati@gmu.edu @ 1378-3207-5469

Created dataset "crimedata".

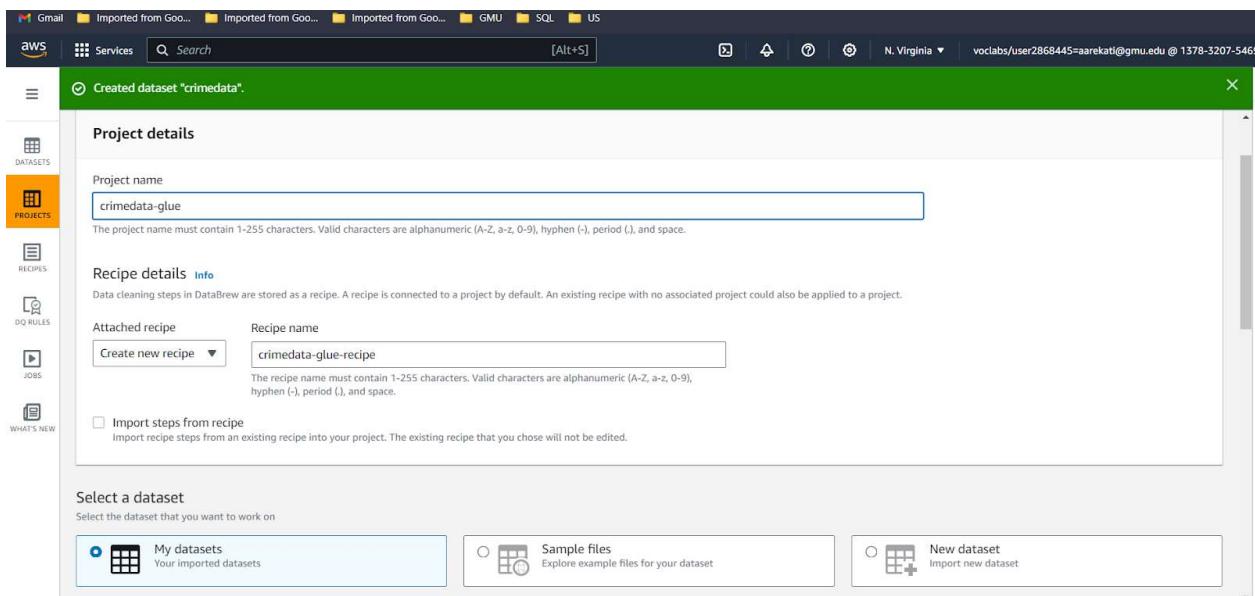
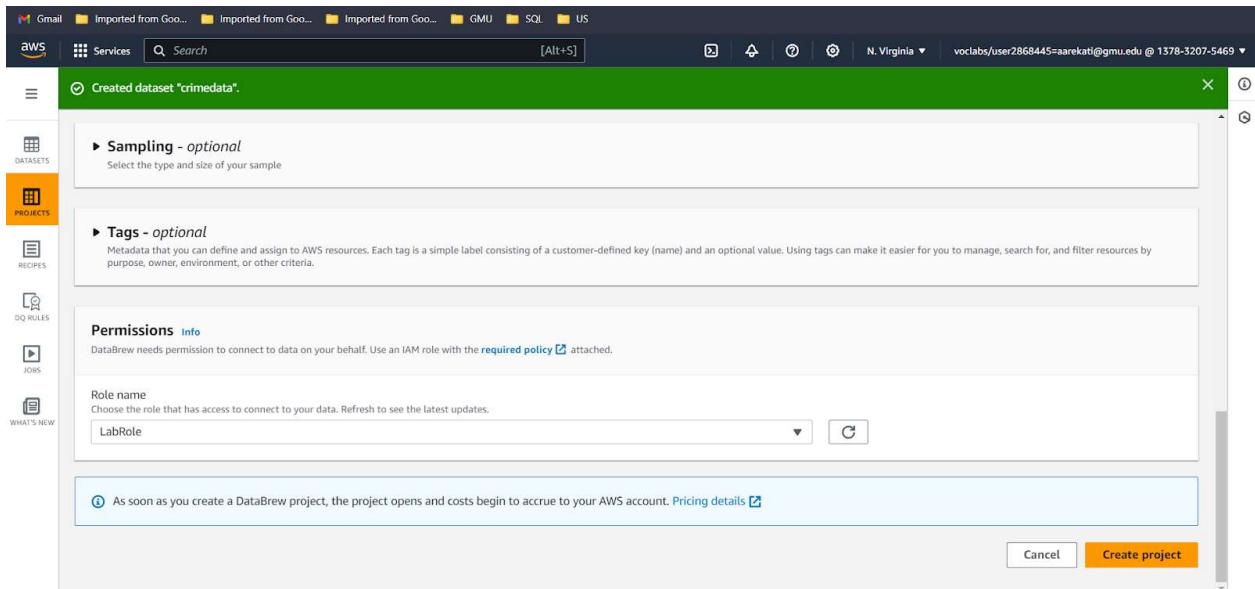
Select a dataset

Select the dataset that you want to work on

My datasets Your imported datasets Sample files Explore example files for your dataset New dataset Import new dataset

Dataset name	Data type	Source	Create date
crimedata	csv	S3	a minute ago November 29, 2023, 9:41:35 pm
imdb-dataset	csv	S3	23 days ago November 6, 2023, 11:37:48 pm
USH	csv	S3	a month ago November 3, 2023, 11:44:39 pm
US-Hurricanes.csv	csv	S3	a month ago November 3, 2023, 11:52:49 pm

Sampling - optional
Select the type and size of your sample



Created project "crimedata-glue".

crimedata-glue

Dataset: crimedata Sample: First n sample (0 rows)

Create job LINEAGE ACTIONS

UNDO REDO FILTER SORT COLUMN FORMAT CLEAN EXTRACT MISSING INVALID DUPLICATES OUTLIERS SPLIT MERGE CREATE FUNCTIONS CONDITIONS NEST-UNNEST PIVOT GROUP JOIN UNION TEXT SCALE MAPPING ENCODE MORE RECIPE

Viewing 0 rows SAMPLE GRID SCHEMA PROFILE

Your session will be ready soon!
Provisioning compute

Your session will take about a minute to be ready. Once ready there will be no additional load time.

0% Zoom 100%

Recipe (0)

crimedata-glue-recipe Working version More

Build your recipe

Start applying transformation steps to your data. All your data preparation steps will be tracked in the recipe

The screenshot shows the Databricks Data Preview interface. At the top, a green banner indicates "Created project *crimedata-glue*". Below it, the dataset name "crimedata-glue" is displayed along with a note "Dataset: crimedata" and "Sample: First n sample (500 rows)". On the right, there are buttons for "Create job", "LINEAGE", "ACTIONS", and a "RECIPE" section which is currently empty. The main area shows a table with 7 columns and 500 rows. The columns are labeled "ABC Month", "ABC Release_Type", "ABC County_of_Commitment", and others. Each column has a summary row at the top showing metrics like Total (500), Distinct (1), Unique (0), and a distribution chart. The table body contains several rows of data, mostly showing "JANUARY" for ABC Month and "PAROLE" for ABC Release_Type. To the right of the table, there's a "SAMPLE" section with a grid view, a "GRID" button, and tabs for "SCHEMA" and "PROFILE". A sidebar on the left shows navigation through "JETS", "RES", "JLES", and "NEW" projects. On the far right, a "Recipe (0)" panel is open, titled "crimedata-glue-recipe Working version". It includes a "More" button and a "Build your recipe" section with a "Start applying transformation steps to your data. All your data preparation steps will be tracked in the recipe" message.

The screenshot shows the Databricks Data Preview interface. At the top, a green header bar displays the message "Created project *crimedata-glue*." Below the header, the dataset name "crimedata-glue" is shown, along with a "Dataset: crimedata" link and a "Sample: First n sample (500 rows)" link. On the right side of the header, there are "Create job", "LINEAGE", and "ACTIONS" buttons.

The main interface features a toolbar with various data manipulation and visualization tools: UNDO, REDO, FILTER, SORT, COLUMN, FORMAT, CLEAN, EXTRACT, MISSING, INVALID, DUPLICATES, OUTLIERS, SPLIT, MERGE, CREATE, FUNCTIONS, CONDITIONS, NEST-UNNEST, PIVOT, GROUP, JOIN, UNION, TEXT, SCALE, MAPPING, ENCODE, and MORE. A "RECIPE" button is also present.

The data preview area displays a table with 7 columns and 500 rows. The columns are labeled: SOURCE, # Year, ABC Month, ABC Release_Type, and ABC County_of_Comp. The first column, SOURCE, contains a single row with the value "crimedata". The other four columns each show a distribution of values across 500 rows. For example, the ABC Month column shows a distribution where JANUARY is the most frequent month. The ABC Release_Type column shows a distribution where PAROLE is the most frequent type. The ABC County_of_Comp column shows a distribution where NEW YORK is the most frequent county.

On the right side of the interface, there are two panels: "Column details" and "Value distribution". The "Column details" panel provides a summary of valid, invalid, and missing values for each column. The "Value distribution" panel shows a stacked bar chart of the values for the ABC County_of_Comp column, with the largest segment being NEW YORK at 37%.