

# OUTPUT:

## Created the Final project folder in GCP:

console.cloud.google.com/welcome?authuser=1&project=cps585-finalproject-group7

Google Cloud cps585 finalProject group7 Search (/) for resources, docs, products, and more

Welcome

You're working in [cps585 finalProject group7](#)

Project number: 950687480788 Project ID: cps585-finalproject-group7

[Dashboard](#) [Recommendations](#)

Privacy Policy Terms of Service

CLOUD SHELL Terminal (cps585-finalproject-group7) Open Editor

```
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to cps585-finalproject-group7.
Use "gcloud config set project [PROJECT_ID]" to change to a different project.
mounikachaitia06@cloudshell:~ (cps585-finalproject-group7) $
```

## Created the bucket:

console.cloud.google.com/storage/browser?authuser=1&project=cps585-finalproject-group7&prefix&forceOnBucketsSortingFiltering=true&cloudshell=false

Google Cloud cps585 finalProject group7 Search (/) for resources, docs, products, and more

Cloud Storage Buckets CREATE REFRESH HELP ASSISTANT LEARN

Try The New Cloud Storage Monitoring Dashboard

Check out the new Cloud Storage monitoring dashboard and Bucket Observability pages! Powered by Cloud Operations, you can customize these dashboards for each project.

TRY NOW

View security recommendations

Improve security by applying security recommendations to your buckets. The security insights column in the table describes which buckets have excess permissions.

VIEW IN TABLE LEARN MORE

Filter Filter buckets

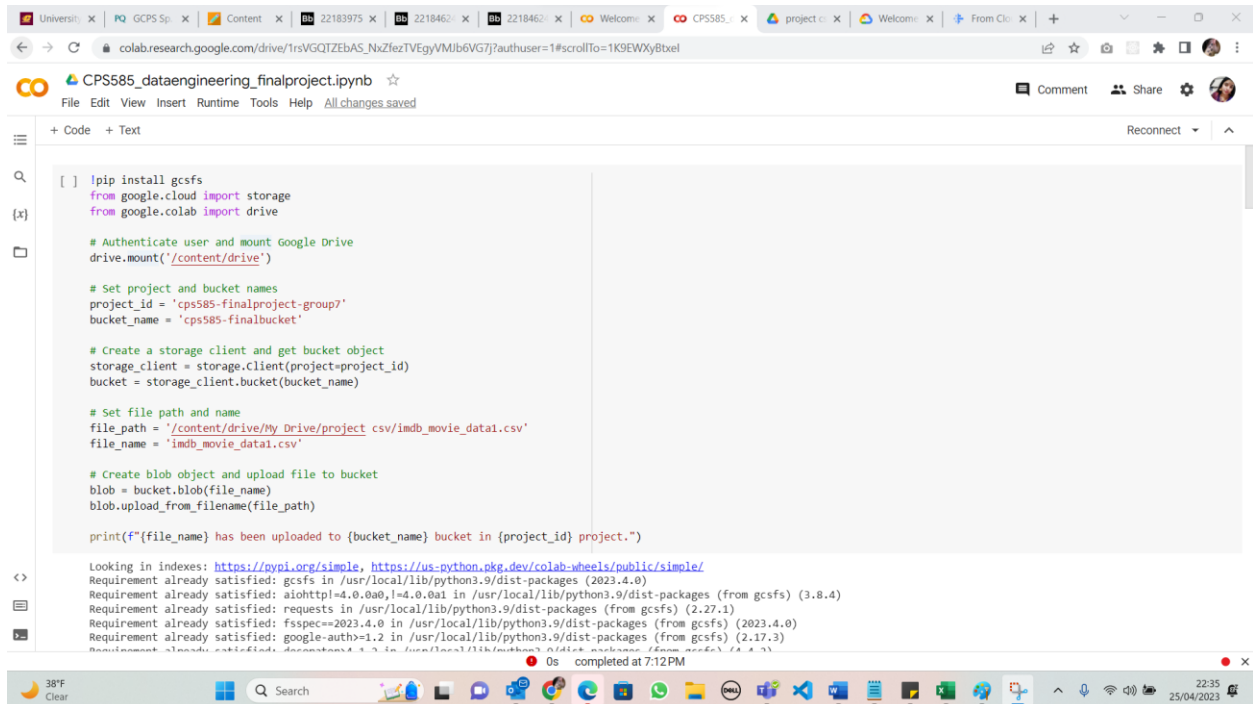
| <input type="checkbox"/> | Name  | Created                  | Location type | Location | Default storage class | Last modified            | Public access     |
|--------------------------|---|--------------------------|---------------|----------|-----------------------|--------------------------|-------------------|
| <input type="checkbox"/> | <a href="#">cps585-finalbucket</a>                    | Apr 17, 2023, 4:28:56 PM | Multi-region  | us       | Standard              | Apr 17, 2023, 4:28:56 PM | Not public        |
| <input type="checkbox"/> | <a href="#">dataprep-staging-77f5828d-7a1c-46a...</a> | Apr 17, 2023, 4:53:42 PM | Multi-region  | us       | Multi-regional        | Apr 17, 2023, 6:02:53 PM | Public to interne |

38°F Clear 22:39 25/04/2023

# OUTPUT:

We tried Two ways of uploading the CSV file into the Bucket:

1. Using Python
2. Manual uploading into the bucket.



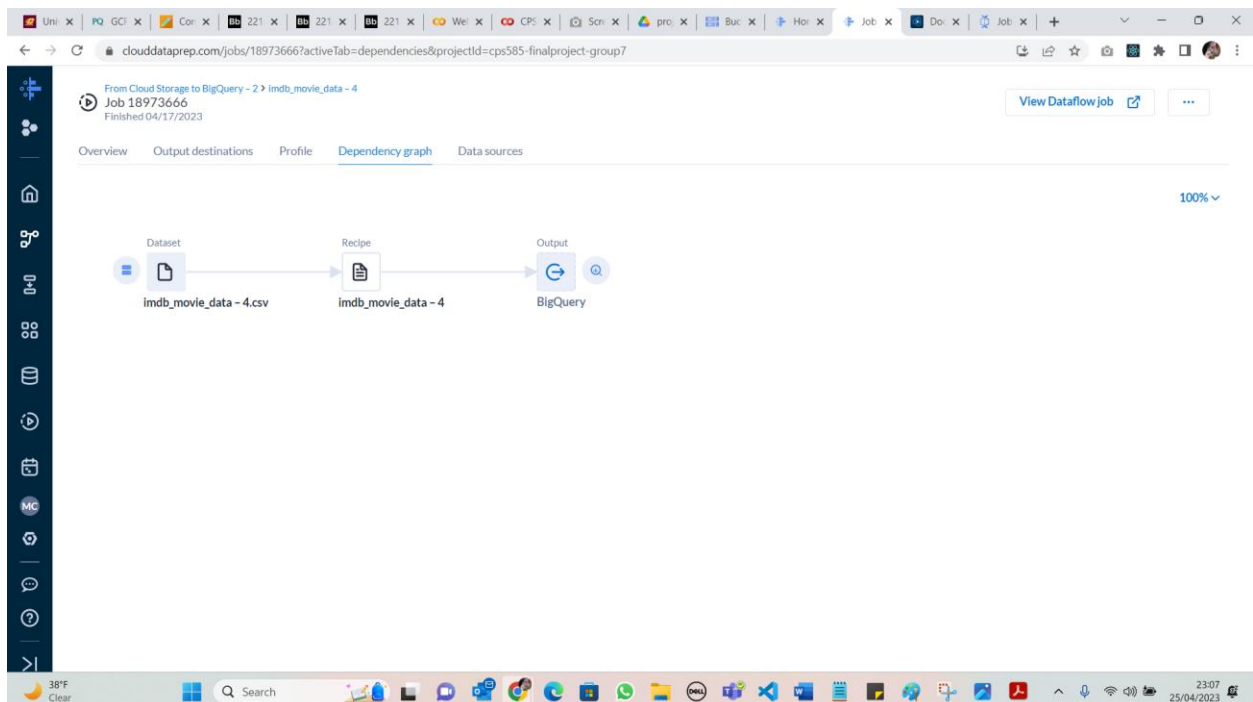
The screenshot shows a Google Colab notebook titled 'CPS585\_dataengineering\_finalproject.ipynb'. The code in the notebook is as follows:

```
[ ] | pip install gcsfs
    | from google.cloud import storage
    | from google.colab import drive
    |
    | # Authenticate user and mount Google Drive
    | drive.mount('/content/drive')
    |
    | # Set project and bucket names
    | project_id = 'cps585-finalproject-group7'
    | bucket_name = 'cps585-finalbucket'
    |
    | # Create a storage client and get bucket object
    | storage_client = storage.Client(project=project_id)
    | bucket = storage_client.bucket(bucket_name)
    |
    | # Set file path and name
    | file_path = '/content/drive/My Drive/project csv/imdb_movie_data1.csv'
    | file_name = 'imdb_movie_data1.csv'
    |
    | # Create blob object and upload file to bucket
    | blob = bucket.blob(file_name)
    | blob.upload_from_filename(file_path)
    |
    | print(f'{file_name} has been uploaded to {bucket_name} bucket in {project_id} project.')
```

Below the code, the output shows the successful installation of gcsfs and the upload of the file 'imdb\_movie\_data1.csv' to the bucket 'cps585-finalbucket' in the project 'cps585-finalproject-group7'.

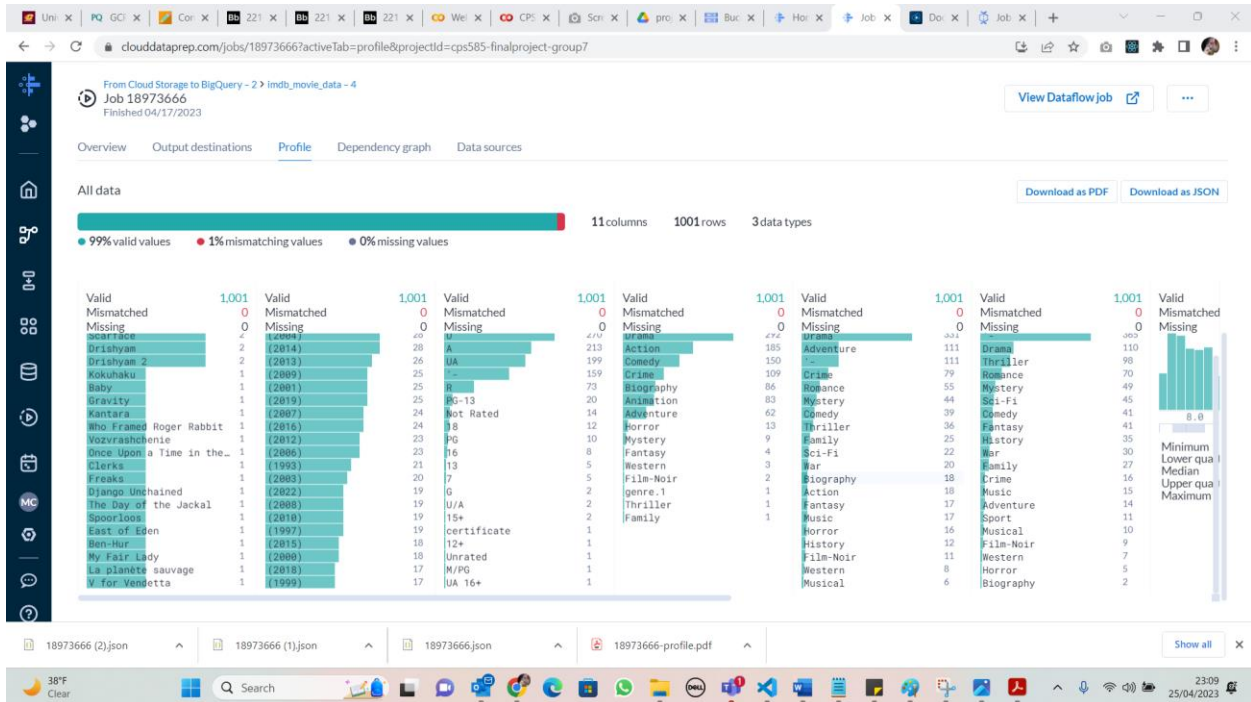
DATA PREP AND BIG QUERY PROCESS: loading the CSV file to DATA PREP and BIGQUERY

DEPENDENCY GRAPH :

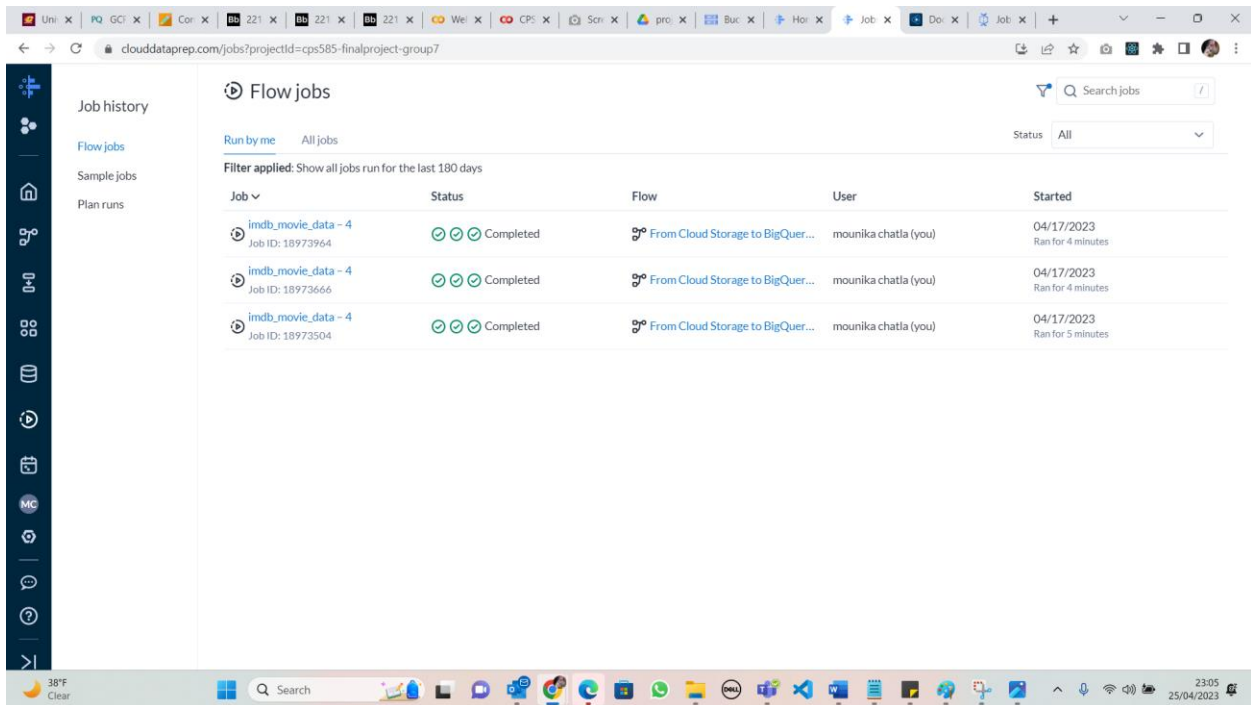


## OUTPUT:

1. After Running the Job , Here the output graph : Dataprep flow to clean and transform the IMDb movie data. Import the data from the Cloud Storage bucket into Dataprep, and use Data prep's features to remove duplicates, fill in missing values, and transform the data into the required format.



Flow of JOBS we runned:



### Report:

# OUTPUT:

clouddataprep.com/jobs/18973666?activeTab=overview&projectId=cps585-finalproject-group7

From Cloud Storage to BigQuery - 2 > imdb\_movie\_data - 4  
Job 18973666  
Finished 04/17/2023

View Dataflow job

Overview Output destinations Profile Dependency graph Data sources

**Schema validation**  
Completed 04/17/2023, started 04/17/2023 • Ran for 19 sec  
Datasets  
imdb\_movie\_data - 4.csv No schema changes found  
View all

**Transform with profile**  
Completed 04/17/2023, started 04/17/2023 • Ran for 4 min  
Environment Dataflow  
99% valid values 1% mismatching values 0% missing values  
View steps and dependencies View profile View dataflow job

**Publish**  
Completed 04/17/2023, started 04/17/2023 • Ran for 13 sec  
Activity  
imdb\_movie\_data - 4\_1.csv Completed • 13 sec

**Job summary**  
Job ID 18973666  
Job status Completed  
Flow From Cloud Storage to BigQu...  
Output imdb\_movie\_data - 4

**Execution summary**  
Job type Manual  
User mounika chatla  
Start time April 17th 2023, 5:16 pm  
Finish time April 17th 2023, 5:20 pm  
Last update April 17th 2023, 5:20 pm  
Duration 4 minutes  
vCPU usage 0.025 vCPU hours (00:01:30)  
Environment Dataflow

**Optimization summary**

18973666 (2).json 18973666 (1).json 18973666.json 18973666-profile.pdf Show all

console.cloud.google.com/dataflow/jobs?authuser=1&project=cps585-finalproject-group7&supportedpurview=project

Google Cloud cps585 finalProject group7 Search (/) for resources, docs, products, and more

Dataflow Overview Jobs Pipelines Workbench Snapshots SQL Workspace

Star the Apache Beam GitHub repository!  
Hello, fellow Dataflow enthusiasts! If you're a regular user of Apache Beam and would like to share your love for it, please consider starring the GitHub repository.  
OPEN REPOSITORY

Jobs CREATE JOB FROM TEMPLATE ENABLE SORTING REFRESH LEARN

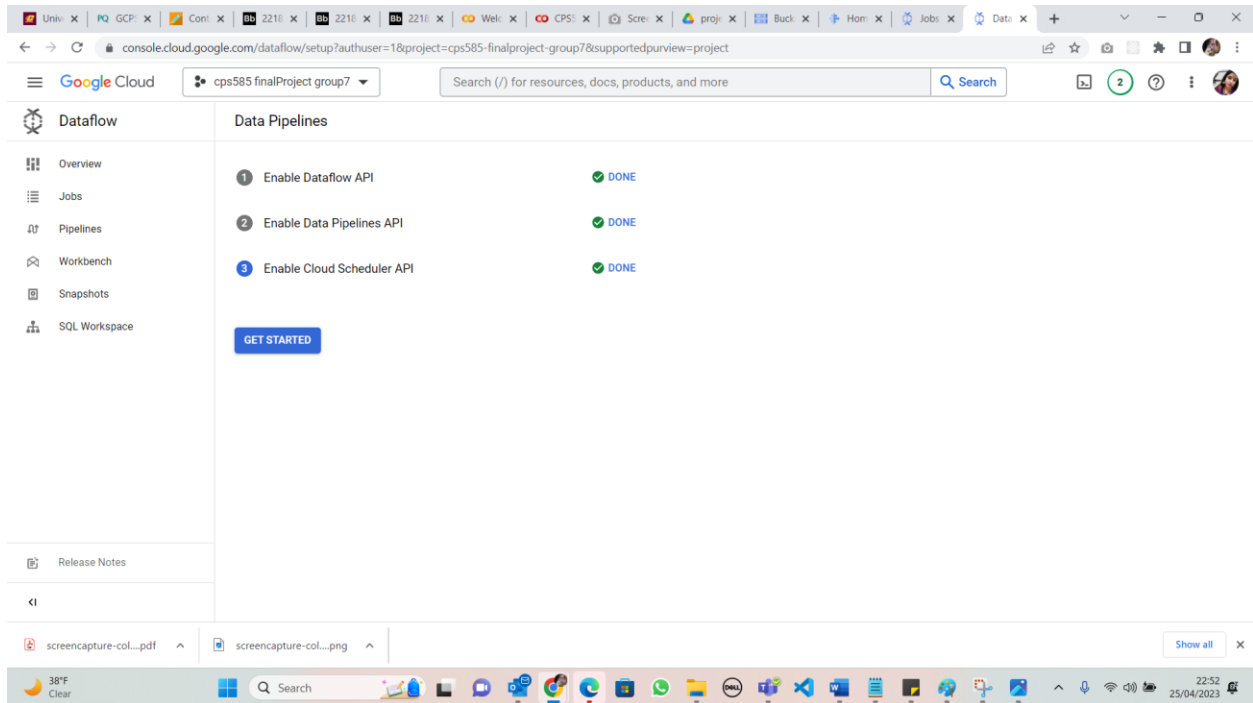
Running Filter Filter jobs

| Name  | Type  | End time                | Elapsed time | Start time               | Status    | SDK version | ID                                       | Region      | Insights |
|---|-------|-------------------------|--------------|--------------------------|-----------|-------------|--|-------------|----------|
| cloud-dataprep-from-cloud-storage-to-b-18973964-by-mounikachattai | Batch | Apr 17, 2023, 5:39:13PM | 3 min 2 sec  | Apr 17, 2023, 5:36:11 PM | Succeeded | 2.35.0      | 2023-04-17_14_36_08-2694037585455793478  | us-central1 |          |
| cloud-dataprep-from-cloud-storage-to-b-18973666-by-mounikachattai | Batch | Apr 17, 2023, 5:20:10PM | 3 min 14 sec | Apr 17, 2023, 5:16:56 PM | Succeeded | 2.35.0      | 2023-04-17_14_16_54-12971840881693165183 | us-central1 |          |
| cloud-dataprep-from-cloud-storage-to-b-18973504-by-mounikachattai | Batch | Apr 17, 2023, 5:11:36PM | 4 min 5 sec  | Apr 17, 2023, 5:07:31 PM | Succeeded | 2.35.0      | 2023-04-17_14_07_28-14782884761085079758 | us-central1 |          |

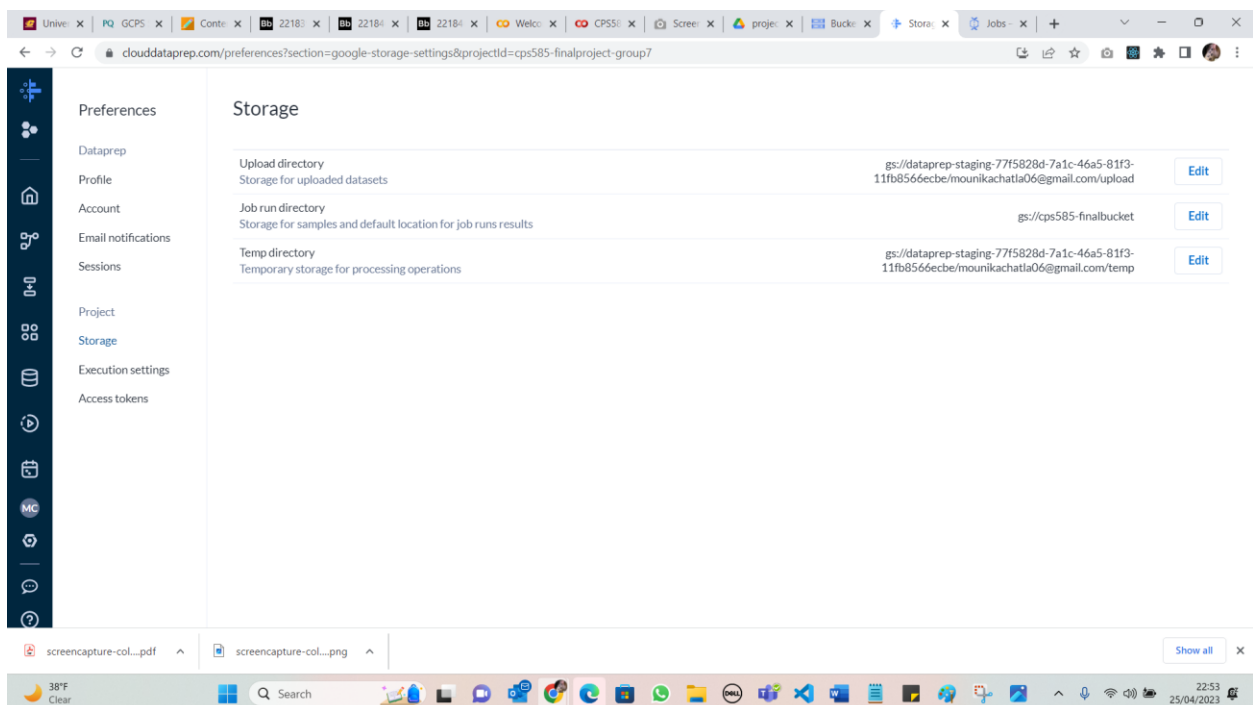
Release Notes

screenshot-col.png screenshot-col.png Show all

# OUTPUT:



We set the storage path in the DATA Prep: Here we assigned the path which stores the output file generated after the transformation of data directed to the GCP bucket.



# OUTPUT:

Univer x

PQ GCPS x

Conte x

2218 x

2218 x

2218 x

Welco x

CPSS x

Scre x

proj x

datap x

Stora x

Jobs x

+ x

console.cloud.google.com/storage/browser/dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe?tab=objects?forceOnBucketsSortingFiltering=true&authuser=1&project=cp...

Search (/) for resources, docs, products, and more

Search

2

?

?

?

Cloud Storage

Buckets

Monitoring

Settings

Marketplace

Release Notes

⌵

Bucket details

REFRESH

HELP ASSISTANT

LEARN

dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe

Location

Storage class

Public access

Protection

us (multiple regions in United States)

Multi-regional

Public to internet

None

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

OBSERVABILITY

NEW

INVENTORY REPORTS

NEW

Buckets > dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

TRANSFER DATA

MANAGE HOLDS

DOWNLOAD

DELETE

Filter by name prefix only

Filter

Filter objects and folders

Show deleted data

| <input type="checkbox"/> | Name                       | Size | Type   | Created | Storage class | Last modified | Public access | Version history | Encryption | Retention expiration |
|--------------------------|----------------------------|------|--------|---------|---------------|---------------|---------------|-----------------|------------|----------------------|
| <input type="checkbox"/> | mounikachatia06@gmail.com/ | -    | Folder | -       | -             | -             | -             | -               | -          | -                    |

38°F Clear

Search

22:56 25/04/2023

console.cloud.google.com/storage/browser?authuser=1&project=cps585-finalproject-group7&prefix=forceOnBucketsSortingFiltering=true&cloudshell=false

Search (/) for resources, docs, products, and more

Search

?

?

?

Cloud Storage

Buckets

Monitoring

Settings

Marketplace

Release Notes

⌵

Buckets

CREATE

REFRESH

HELP ASSISTANT

LEARN

Try The New Cloud Storage Monitoring Dashboard

Check out the new Cloud Storage monitoring dashboard and Bucket Observability pages! Powered by Cloud Operations, you can customize these dashboards for each project.

TRY NOW

View security recommendations

Improve security by applying security recommendations to your buckets. The security insights column in the table describes which buckets have excess permissions.

VIEW IN TABLE

LEARN MORE

Filter

Filter buckets

?

?

| <input type="checkbox"/> | Name  | Created                  | Location type | Location | Default storage class | Last modified            | Public access      |
|--------------------------|---|--------------------------|---------------|----------|-----------------------|--------------------------|--------------------|
| <input type="checkbox"/> | cps585-finalbucket                                    | Apr 17, 2023, 4:28:56 PM | Multi-region  | us       | Standard              | Apr 17, 2023, 4:28:56 PM | Not public         |
| <input type="checkbox"/> | dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe | Apr 17, 2023, 4:53:42 PM | Multi-region  | us       | Multi-regional        | Apr 17, 2023, 6:02:53 PM | Public to internet |

38°F Clear

Search

22:39 25/04/2023

# OUTPUT:

Univer x GCPS x Cont x 2218 x 2218 x 2218 x Welc x CPSS x Scre x proje x datap x Stora x Jobs x +

console.cloud.google.com/storage/browser/dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe/mounikachatta06@gmail.com?authuser=1&project=cps585-finalproject-gr...

Google Cloud cps585 finalProject group7 Search (/) for resources, docs, products, and more

Cloud Storage

Buckets

Monitoring

Settings

Bucket details

REFRESH HELP ASSISTANT LEARN

dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe

Location us (multiple regions in United States) Storage class Multi-regional Public access Public to internet Protection None

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY NEW INVENTORY REPORTS NEW

Buckets > dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe > mounikachatta06@gmail.com

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA MANAGE HOLDS DOWNLOAD DELETE

Filter by name prefix only Filter objects and folders Show deleted data

| Name    | Size | Type   | Created | Storage class | Last modified | Public access | Version history | Encryption | Retention expiration |
|---------|------|--------|---------|---------------|---------------|---------------|-----------------|------------|----------------------|
| jobrun/ |      | Folder |         |               |               |               |                 |            |                      |
| temp/   |      | Folder |         |               |               |               |                 |            |                      |
| upload/ |      | Folder |         |               |               |               |                 |            |                      |

screencapture-col...pdf screencapture-col...png Show all

38°F Clear

Univer x GCPS x Cont x 2218 x 2218 x 2218 x Welc x CPSS x Scre x proje x datap x Stora x Jobs x +

console.cloud.google.com/storage/browser/dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe/mounikachatta06@gmail.com/jobrun?pageState=(\"StorageObjectListTable\"...

Google Cloud cps585 finalProject group7 Search (/) for resources, docs, products, and more

Cloud Storage

Buckets

Monitoring

Settings

Bucket details

REFRESH HELP ASSISTANT LEARN

dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe

Location us (multiple regions in United States) Storage class Multi-regional Public access Public to internet Protection None

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY NEW INVENTORY REPORTS NEW

Buckets > dataprep-staging-77f5828d-7a1c-46a5-81f3-11fb8566ecbe > mounikachatta06@gmail.com > jobrun

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA MANAGE HOLDS DOWNLOAD DELETE

Filter by name prefix only Filter objects and folders Show deleted data

| Name                        | Size | Type   | Created | Storage class | Last modified | Public access | Version history | Encryption | Retention expiration |
|-----------------------------|------|--------|---------|---------------|---------------|---------------|-----------------|------------|----------------------|
| New_Folder/                 |      | Folder |         |               |               |               |                 |            |                      |
| imdb_movie_data - 4.csv/    |      | Folder |         |               |               |               |                 |            |                      |
| imdb_movie_data - 4.1.csv/  |      | Folder |         |               |               |               |                 |            |                      |
| imdb_movie_data - 4.2.csv/  |      | Folder |         |               |               |               |                 |            |                      |
| imdb_movie_data_4_18973504/ |      | Folder |         |               |               |               |                 |            |                      |
| imdb_movie_data_4_18973666/ |      | Folder |         |               |               |               |                 |            |                      |
| imdb_movie_data_4_18973964/ |      | Folder |         |               |               |               |                 |            |                      |

screencapture-col...pdf screencapture-col...png Show all

38°F Clear

# OUTPUT:

job-report-18973666.pdf - Adobe Acrobat Reader (64-bit)

File Edit View Sign Window Help

Home Tools job-report-1897366... x

1 / 3 76.2%

### Profile report - imdb\_movie\_data - 4

From Cloud Storage to BigQuery - 2 - Job ID: 18973666

All Data 11 columns 1,001 rows 3 data types

99% valid values 1% mismatching values 0% missing values

#### Output destinations

| Name                      | Location  | Status |
|---------------------------|---|--------|
| imdb_movie_data - 4_1.csv | gs://dataprep-staging-77f5c2bd-7a1c-46a5-81f3-11fa8966ecbe/mounlachattarak@gmail.com/jobrun/imdb_movie_data - 4_1.csv | -      |

#### Data sources

| Name                    | Location                                    | Schema validation       |
|-------------------------|---|-------------------------|
| imdb_movie_data - 4.csv | gs://tps585-finalbucket/imdb_movie_data.csv | No schema changes found |

#### Execution summary

Job Status: Completed

Job Type: Manual

Environment: Dataflow

Start time: Mon, Apr 17, 2023 5:16 PM -04:00

End time: Mon, Apr 17, 2023 5:20 PM -04:00

Duration: 4 min

Search tools

Export PDF

Create PDF

Comment

Fill & Sign

More Tools

Report | job 18973666 | From Cloud Storage to BigQuery - 2 | imdb\_movie\_data - 4 1 of 3

38°F Clear

Search

23:05 25/04/2023