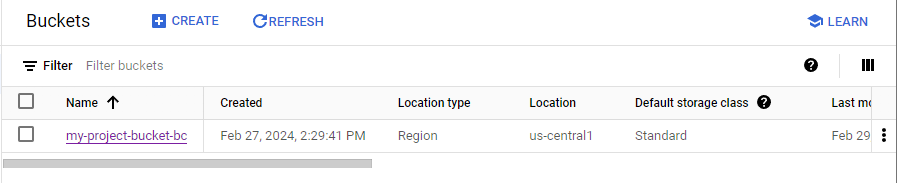
**Project ID:** cis-4130-semester-project

**Bucket Name:** my-project-bucket-bc

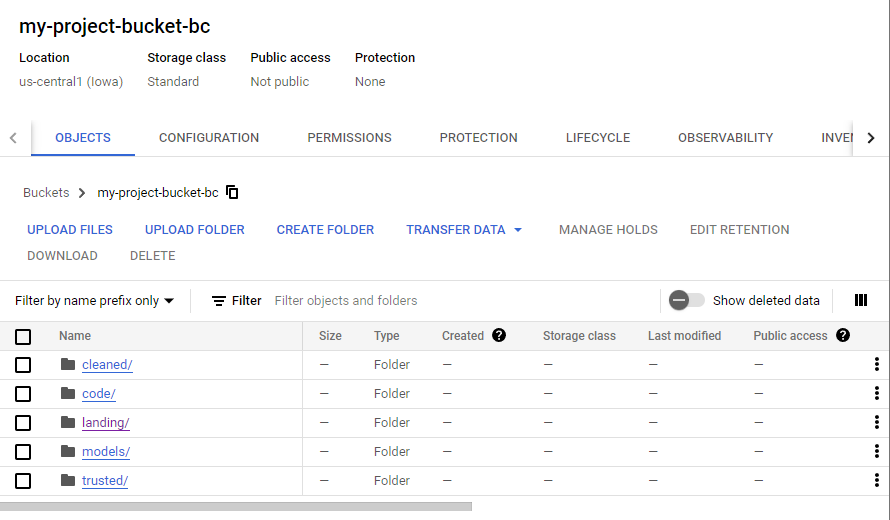
**File in Bucket:** my-project-bucket-bc/landing/nypd\_mv\_collisions

Step 1: Navigate to Cloud Storage >> Buckets

Step 2: Create a new bucket with the name: my-project-bucket-bc



Step 3: Create folders: cleaned, code, landing, models, trusted



Step 4: Go to the Dataset Link: [https://console.cloud.google.com/marketplace/product/city-of-new-york/nypd-mv-collisions](https://console.cloud.google.com/marketplace/product/city-of-new-york/nypd-mv-collisions?authuser=1&project=cis-4130-semester-project)

Step 5: Select View Dataset and It should open up BigQuery (after selected Project)

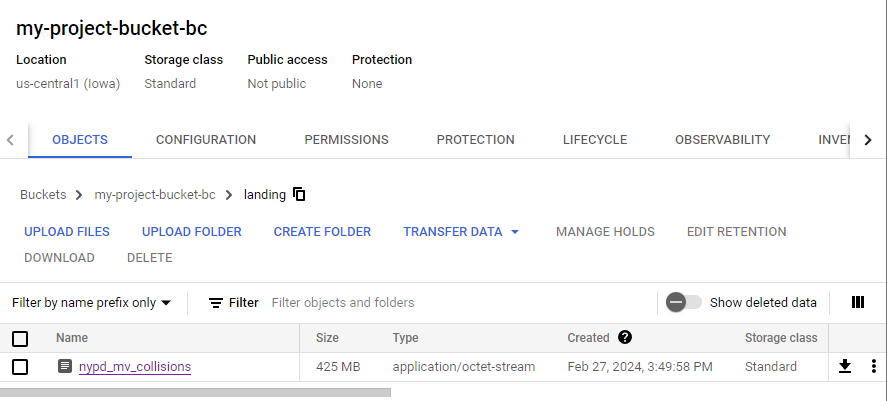
Step 6: Open the dataset with the dropdown error and select on the Table

Step 7: Select the Export icon in the top ribbon 

Step 8: Select Export to GCS

Step 9: Browse Buckets and select: my-project-bucket-bc

Step 10: Select landing/ folder save file name “nypd\_mv\_collisions”

Step 11: Make Sure the table is exported to Buckets; Navigate to Cloud Storage >> Buckets 

Step 12: Activate the Cloud Shell Editor on the top 

Step 13: Install Google Cloud Storage in Cloud Shell Editor

| bellachen1212@cloudshell:~ (cis-4130-semester-project)$ pip3 install google-cloud-storage |
| --- |

Step 14: Authenticate with Google Account

| bellachen1212@cloudshell:~ (cis-4130-semester-project)$ gcloud auth application-default login |
| --- |

| Ran into error:  ERROR: (gcloud.auth.application-default.login) PERMISSION\_DENIED: Service Usage API has not been used in project cis-4130-semester-project before or it is disabled. Enable it by visiting https://console.developers.google.com/apis/api/serviceusage.googleapis.com/overview?project=cis-4130-semester-project then retry. If you enabled this API recently, wait a few minutes for the action to propagate to our systems and retry.  - '@type': type.googleapis.com/google.rpc.Help  links:  - description: Google developers console API activation  url: https://console.developers.google.com/apis/api/serviceusage.googleapis.com/overview?project=cis-4130-semester-project  - '@type': type.googleapis.com/google.rpc.ErrorInfo  domain: googleapis.com  metadata:  consumer: projects/cis-4130-semester-project  service: serviceusage.googleapis.com  reason: SERVICE\_DISABLED   * Went to the website provided and enabled the Services with API * Re-run the code to authenticate |
| --- |

Step 15: Create and Avtivate Python Virtual Enviroment and Install Libraries

| **Create a Python Virtual Environment:** python3 -m venv pythondev  **Activate the Python VE:** cd pythondev; source bin/activate  **Install python libraries:** pip3 install jupyter pandas numpy |
| --- |

Step 16: Activate the Python Enviroment

| (pythondev) bellachen1212@cloudshell:~/pythondev (cis-4130-semester-project)$ python |
| --- |

Step 17: Import storage from google.cloud

| from google.cloud import storage |
| --- |

| Ran into error:  Traceback (most recent call last):  File "<stdin>", line 1, in <module>  ModuleNotFoundError: No module named 'google'   * Exited Python and Installed Google: * (pythondev) bellachen1212@cloudshell:~/pythondev (cis-4130-semester-project)$ pip install google |
| --- |

Step 18: List all Buckets

| # Create a client object that points to GCS  >>> storage\_client = storage.Client()  >>> buckets = storage\_client.list\_buckets()  >>> print(buckets)  <google.api\_core.page\_iterator.HTTPIterator object at 0x7b9114948bb0> |
| --- |

Step 19: Load file from GCS as DataFrame in Pandas (Import Pandas and io)

| >>> bucket = storage\_client.get\_bucket("my-project-bucket-bc")  >>> blob = bucket.blob("landing/"+"nypd\_mv\_collisions")  >>> import pandas as pd  >>> from io import StringIO  >>> df = pd.read\_csv(StringIO(blob.download\_as\_text())) |
| --- |

Step 20: Get info on DataFrame

| >>> df.info() |  |
| --- | --- |