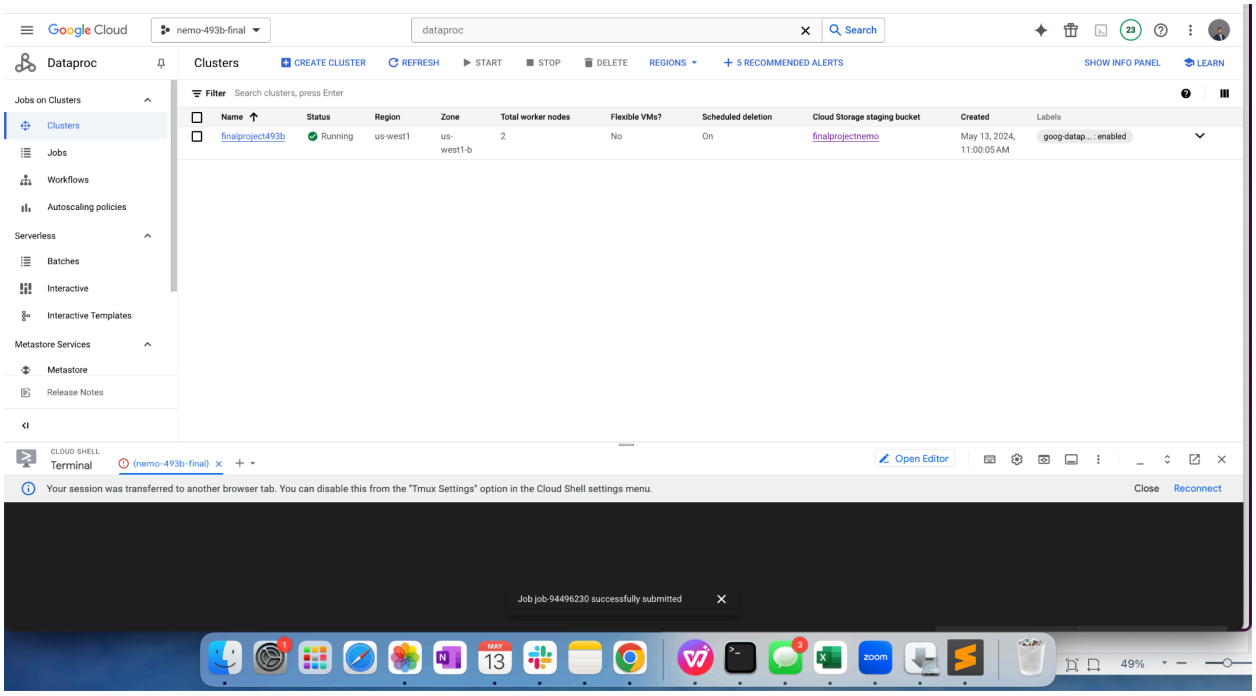
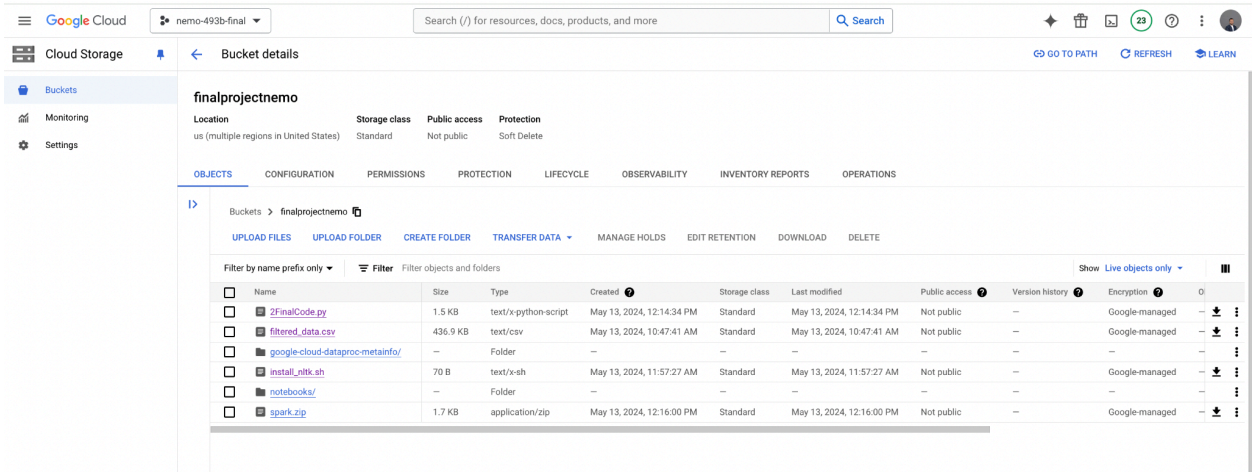
Google Cloud Steps and Screenshots of Steps

1. Create a cluster inthe Google Cloud Platform



2. Create a bucket for storing files



3. Run spark job

4. We also used Google Colab to run our code by connecting Google Colab with Google Cloud Platform after authentication:

```
[ ]  1 ! gcloud auth login
```

```
Go to the following link in your browser, and complete the sign-in prompts:

    https://accounts.google.com/o/oauth2/auth?response_type=code&client_id=32555940559.apps.googleusercontent.com&redirect_uri=https%3A%2F%2Fsdk.cloud.google.com%2Fau

Once finished, enter the verification code provided in your browser:

Command killed by keyboard interrupt

^C
```

3. Colab, uncomment and run:

```
1 from google.colab import auth
2 auth.authenticate_user()
```

Create a Cloud Storage bucket

Create a storage bucket to store intermediate artifacts such as datasets.

```
[ ]  1 BUCKET_URI = f"gs://finalprojectnemo-{PROJECT_ID}-unique"  # @param {type:"s
```

BUCKET_URI: f"gs://finalprojectnemo-{PROJECT_ID}-unique"

Only if your bucket doesn't already exist: Run the following cell to create your Cloud Storage bucket.

```
[ ]  1 ! gsutil mb -l {REGION} -p {PROJECT_ID} {BUCKET_URI}
```

```
11
```

6. Then, we got our results for sentiment analysis

```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.4)
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
Sentiment Percentages:
sentiment
positive    51.329442
neutral     26.588846
negative    22.081712
Name: count, dtype: float64

Confusion Matrix:
sentiment  negative  neutral  positive
target
0               524      323       397
4               157      497      1186
```