

Copying the CSV data to the S3 Bucket

```
In [ ]: # View our data
import pandas as pd

data = pd.read_csv("csv/new_dataset.csv")
data.head()
```

```
Out[ ]:
```

	track_id	artists	popularity	duration_ms	explicit	danceabili
0	5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	73	230666	False	0.6
1	4qPNDBW1i3p13qLCt0Ki3A	Ben Woodward	55	149610	False	0.4
2	1iJBSr7s7jYXzM8EGcbK5b	Ingrid Michaelson;ZAYN	57	210826	False	0.4
3	6lfxq3CG4xtTiEg7opyCyx	Kina Grannis	71	201933	False	0.2
4	5vjLSffimiIP26QG5WcN2K	Chord Overstreet	82	198853	False	0.6

Check Pre-Requisites from an earlier notebook

```
In [ ]: %store -r setup_s3_bucket_passed
```

```
In [ ]: try:
        setup_s3_bucket_passed
except NameError:
    print("+++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are")
    print("+++++")
```

```
In [ ]: print(setup_s3_bucket_passed)
```

True

```
In [ ]: import boto3
import sagemaker
import pandas as pd

sess = sagemaker.Session()
bucket = sess.default_bucket()
role = sagemaker.get_execution_role()
region = boto3.Session().region_name
account_id = boto3.client("sts").get_caller_identity().get("Account")

sm = boto3.Session().client(service_name="sagemaker", region_name=region)
```

Set S3 Source Location (Local CSV File)

```
In [ ]: data_path = "/root/AAI-540/Module2/csv"
        print(data_path)
```

/root/AAI-540/Module2/csv

```
In [ ]: %store data_path
```

Stored 'data_path' (str)

Set S3 Destination Location (Our Private S3 Bucket)

```
In [ ]: s3_private_path_csv = "s3://{}/module2_data/csv".format(bucket)
        print(s3_private_path_csv)
```

s3://sagemaker-us-east-1-004608622582/module2_data/csv

```
In [ ]: %store s3_private_path_csv
```

Stored 's3_private_path_csv' (str)

Copy Data From the Local File to Private S3 Bucket

```
In [ ]: !aws s3 cp --recursive $data_path/ $s3_private_path_csv/ --exclude "*" --include "n
```

upload: csv/new_dataset.csv to s3://sagemaker-us-east-1-004608622582/module2_data/csv/new_dataset.csv

Check to see if the files are copied over successfully

```
In [ ]: !aws s3 ls $s3_private_path_csv/
```

2024-05-18 03:30:50 14676903 new_dataset.csv

```
In [ ]: from IPython.core.display import display, HTML
```

```
display(
    HTML(
        '<b>Review <a target="blank" href="https://s3.console.aws.amazon.com/s3/buc
        region, account_id, region
    )
)
```

/tmp/ipykernel_1780/1772462402.py:1: DeprecationWarning: Importing display from IPython.core.display is deprecated since IPython 7.14, please import from IPython display

```
from IPython.core.display import display, HTML
```

Review [S3 Bucket](#)

Store and Close Notebook

```
In [ ]: %store
```

Stored variables and their in-db values:

data_path	-> '/root/AAI-540/Module2/csv'
ingest_create_athena_db_mod2_passed	-> True
ingest_create_athena_table_csv_passed	-> True
s3_private_path_csv	-> 's3://sagemaker-us-east-1-00460
8622582/module2_dat	
setup_dependencies_mod2_passed	-> True
setup_s3_bucket_passed	-> True

```
In [ ]: %%html

<p><b>Shutting down your kernel for this notebook to release resources.</b></p>
<button class="sm-command-button" data-commandlinker-command="kernelmenu:shutdown">

<script>
try {
  els = document.getElementsByClassName("sm-command-button");
  els[0].click();
}
catch(err) {
  // NoOp
}
</script>
```

Shutting down your kernel for this notebook to release resources.

```
In [ ]:
```