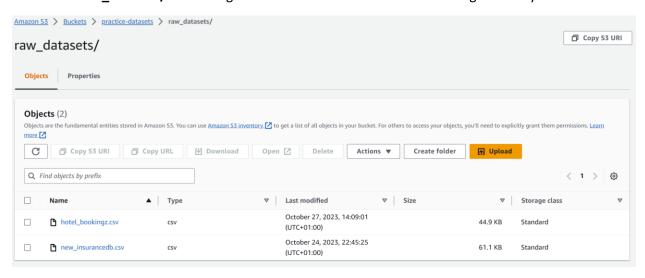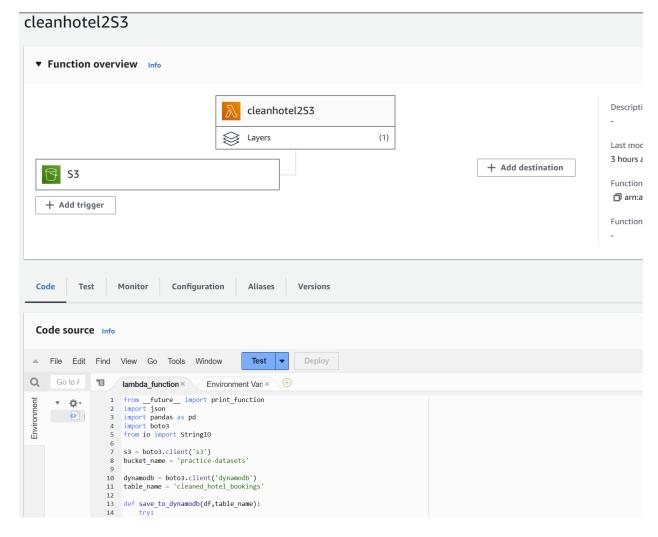# Python function to migrate all files (CSV) to S3 bucket

```python
import boto3
import os
import glob

s3 = boto3.resource('s3')
dir_path = './'

# Get list of specific files for migration

#
# Create S3 bucket via CLI
# aws s3 mb s3://[unique-bucket-name]
#
# Create dynamoDB table in CLI
# aws dynamodb create-table --table-name cleaned_hotel_bookings --attribute-definitions AttributeName=hotel,AttributeType=S AttributeName=arri

# NOTE: my key-schema is not unique, make yours unique to avoid your table rows from being overwritten by new entries in dynamoDB


csv_files = glob.glob(os.path.join(dir_path, '*.csv'))
#json_files = glob.glob(os.path.join(dir_path, '*.json'))
#all_files = csv_files + json_files

all_files = csv_files

# Iterate through list to upload objects to S3
for rawfile in all_files:
    if rawfile.endswith('.csv'):
        try:
            file = open(rawfile, 'rb')
            file_name = os.path.basename(rawfile)
            object = s3.Object('practice-datasets','raw_datasets/' + file_name)
            ret = object.put(Body=file,Metadata={'filename':file_name})
            print (file_name, " is uploaded")
        except Exception as e:
            print(f'Error uploading {file_name}: {str(e)}')
```

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   PORTS

PS F:\Data Engineering\Practices\Data Migration - GitHub> aws configure list
      Name                    Value             Type    Location
      ----                    -----             ----    --------
   profile                <not set>             None    None
access_key     ****************24H3 shared-credentials-file
secret_key     ****************qEqJ shared-credentials-file
    region            us-east-2      config-file    ~/.aws/config
PS F:\Data Engineering\Practices\Data Migration - GitHub> python migrate_to_s3.py
hotel_bookingz.csv  is uploaded
PS F:\Data Engineering\Practices\Data Migration - GitHub>
```

S3 folder **raw_datasets/** where migrated raw files are stored before cleaning on the fly

Amazon S3 > Buckets > practice-datasets > raw_datasets/

raw_datasets/

[ ] Copy S3 URI

Objects   Properties

Objects (2)

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 |

Q Find objects by prefix

⟨ 1 ⟩ ⚙

☐	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
☐	📄 hotel_bookingz.csv	csv	October 27, 2023, 14:09:01 (UTC+01:00)	44.9 KB	Standard
☐	📄 new_insurancedb.csv	csv	October 24, 2023, 22:45:25 (UTC+01:00)	61.1 KB	Standard

Lambda function **cleanhotel2S3,** showing its **S3 Trigger** and added **Python Layer**

cleanhotel2S3

▼ **Function overview**   Info

λ cleanhotel2S3

≋ Layers   (1)

S3

➕ **Add trigger**

➕ **Add destination**

Descripti
-

Last mod
3 hours a

Function
🗂 arn:a

Function
-

| **Code** | Test | Monitor | Configuration | Aliases | Versions |

Code source Info

File   Edit   Find   View   Go   Tools   Window   **Test** ▼   Deploy

🔍 Go to /   ▤   **lambda_function** ✕   Environment Vari ✕   ⊕

```
1  from __future__ import print_function
2  import json
3  import pandas as pd
4  import boto3
5  from io import StringIO
6
7  s3 = boto3.client('s3')
8  bucket_name = 'practice-datasets'
9
10 dynamodb = boto3.client('dynamodb')
11 table_name = 'cleaned_hotel_bookings'
12
13 def save_to_dynamodb(df,table_name):
14     try:
```

S3 folder **clean_datasets/** where clean CSV dataset and a JSON copy of same dataset are stored

Amazon S3 > Buckets > practice-datasets > clean_datasets/

clean_datasets/

🗂 Copy S3 URI

Objects | **Properties**

Objects (2)

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 ↗

🔄 | 🗂 Copy S3 URI | 🗂 Copy URL | ⬇ Download | Open ↗ | Delete | Actions ▼ | Create folder | 📤 Upload

🔍 Find objects by prefix

‹ 1 › ⚙

	Name	Type	Last modified	Size	Storage class
☐	📄 hotel_bookingz.csv	csv	October 27, 2023, 14:09:06 (UTC+01:00)	48.1 KB	Standard
☐	📄 hotel_bookingz.csv.json	json	October 27, 2023, 14:09:06 (UTC+01:00)	255.6 KB	Standard

Dynamo DB table where **clean** and **transformed** data are stored

customer_type	total_of_special_requests	total_guests	arrival_date_day_of_month	meal	days_in_waiting_list	count
Transient	2	2	1	BB	0	PRT
Transient	2	2	2	HB	0	PRT
Transient	0	2	3	BB	0	DEU
Transient-Party	1	2	4	BB	0	FRA
Transient	0	2	5	BB	0	ESP

Items returned (5)

Full source codes and steps are on my GitHub repository.