Cardiovascular Disease Data Analysis

Product Engineering Project 2

by Nick, Alex, Amrita, Mario

Overview

- Red Shield, a hospital in Los Angeles, wants to find patterns in patients with cardiovascular disease

- Red Shield is looking a different factors to determine common features in patients with cardiovascular disease:
    - Physical features like Age, Height, Weight, and Gender
    - Habits like Smoking, Alcohol intake, and Physical activity

The Data

Patient

- attributes about a patient’s age, height, weight, and gender

Habits

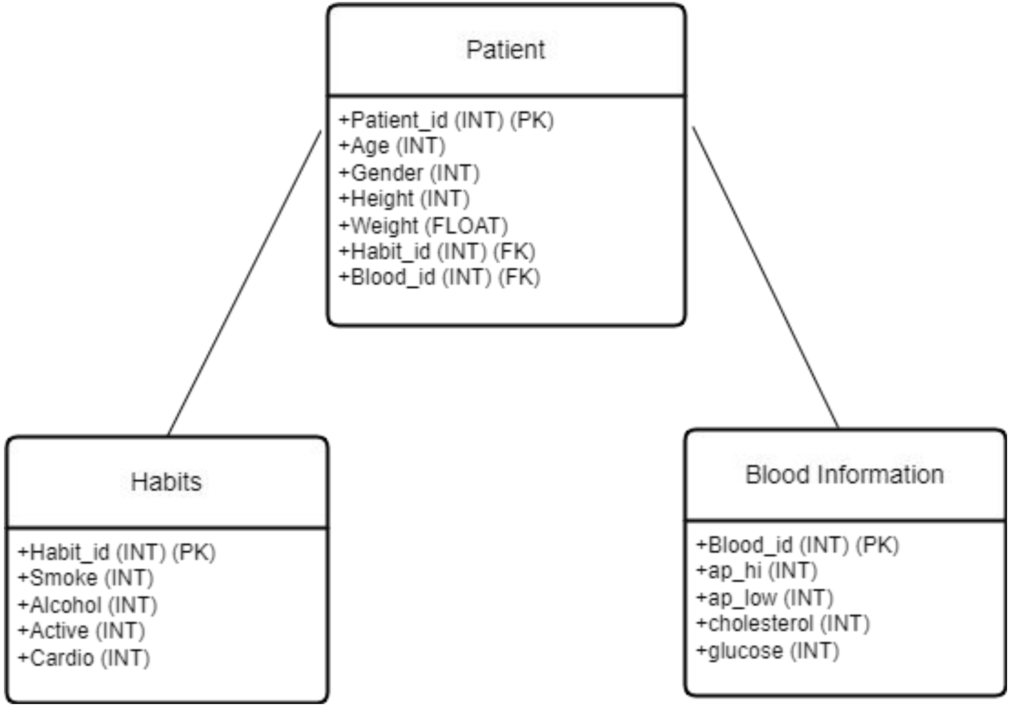
- Describes patients habits such as drinking and smoking

Blood Information

- information about blood such as blood pressure

variable	Data type
Age	int (days)
Height	int (cm)
Weight	float (kg)
Gender	1 = female, 2 = male
Systolic blood pressure	int
Diastolic blood pressure	int
Cholesterol	1: normal, 2: above normal, 3: far above normal
Glucose	1: normal, 2: above normal, 3: far above normal
Smoking	0 = no, 1= yes
Alcohol intake	0 = no, 1= yes
Physical activity	0 = no, 1= yes
Presence or absence of cardiovascular disease	0 = no, 1= yes

ER Diagram



Patient

+Patient_id (INT) (PK)
+Age (INT)
+Gender (INT)
+Height (INT)
+Weight (FLOAT)
+Habit_id (INT) (FK)
+Blood_id (INT) (FK)

Habits

+Habit_id (INT) (PK)
+Smoke (INT)
+Alcohol (INT)
+Active (INT)
+Cardio (INT)

Blood Information

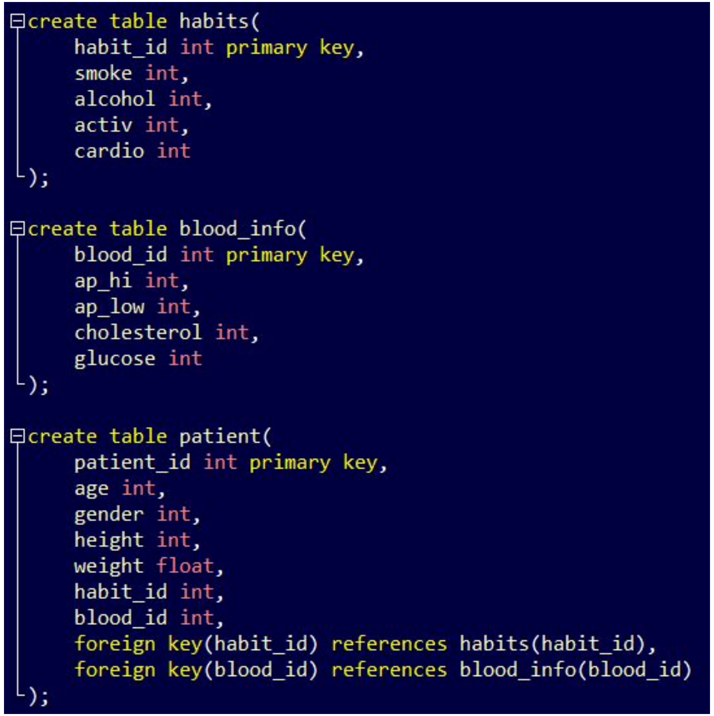
+Blood_id (INT) (PK)
+ap_hi (INT)
+ap_low (INT)
+cholesterol (INT)
+glucose (INT)

To normalize the database we split them into three tables:
Patient, Habits, and Blood info

- 1-1 relationship between Patient and Habits
- 1-1 relationship between Patient and Blood Information
- 2 foreign keys in Patient to make a reference between both of the tables

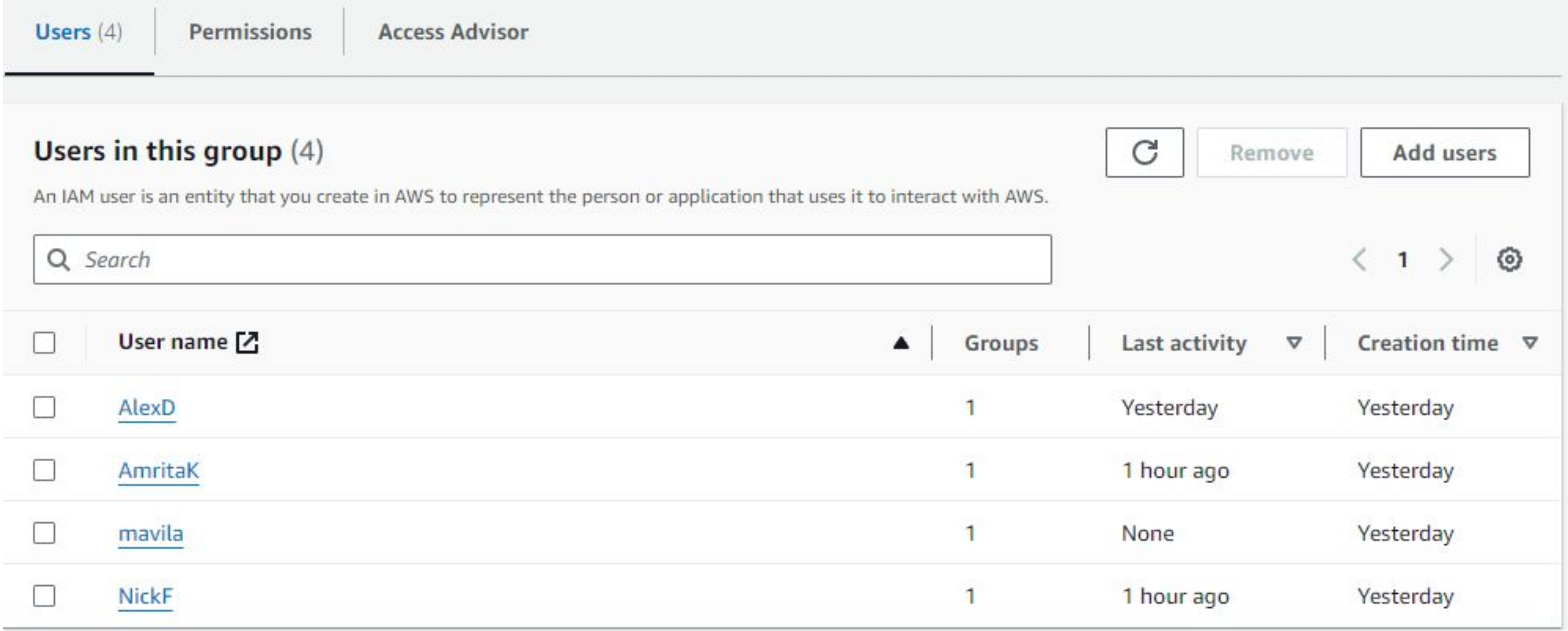
Database Setup

- Read data from S3 bucket
- Used pandas module to work with data
- Removed outliers
- Only able to insert about 3,000 entries due to time restrictions
- Created two foreign keys in patient to reference other tables.



```
create table habits(
    habit_id int primary key,
    smoke int,
    alcohol int,
    activ int,
    cardio int
);

create table blood_info(
    blood_id int primary key,
    ap_hi int,
    ap_low int,
    cholesterol int,
    glucose int
);

create table patient(
    patient_id int primary key,
    age int,
    gender int,
    height int,
    weight float,
    habit_id int,
    blood_id int,
    foreign key(habit_id) references habits(habit_id),
    foreign key(blood_id) references blood_info(blood_id)
);
```

AWS Setup - IAM accounts

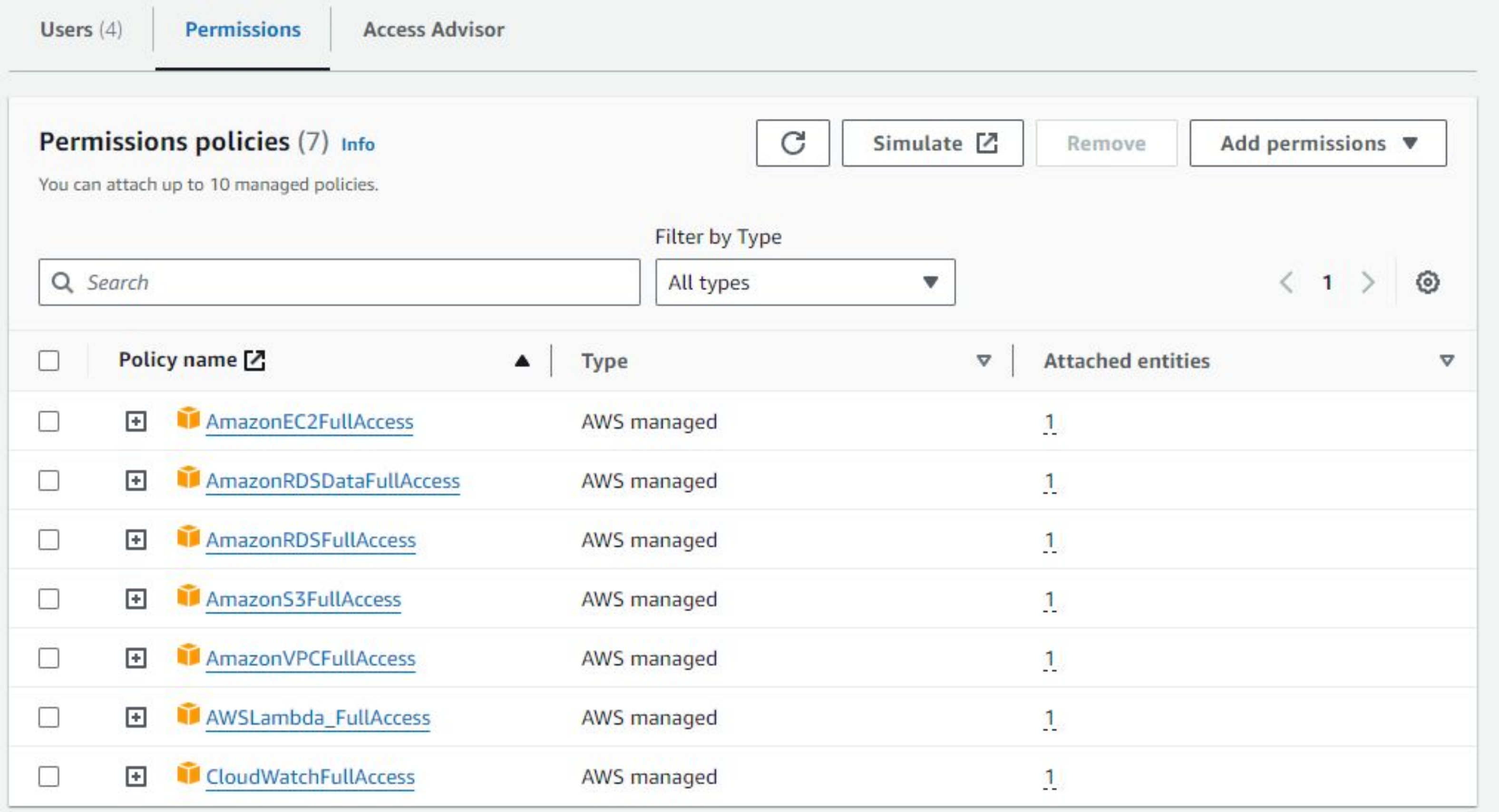


Users in this group (4)

An IAM user is an entity that you create in AWS to represent the person or application that uses it to interact with AWS.

□	User name ▲	Groups	Last activity ▽	Creation time ▽
□	AlexD	1	Yesterday	Yesterday
□	AmritaK	1	1 hour ago	Yesterday
□	mavila	1	None	Yesterday
□	NickF	1	1 hour ago	Yesterday

AWS Setup - IAM account Permissions



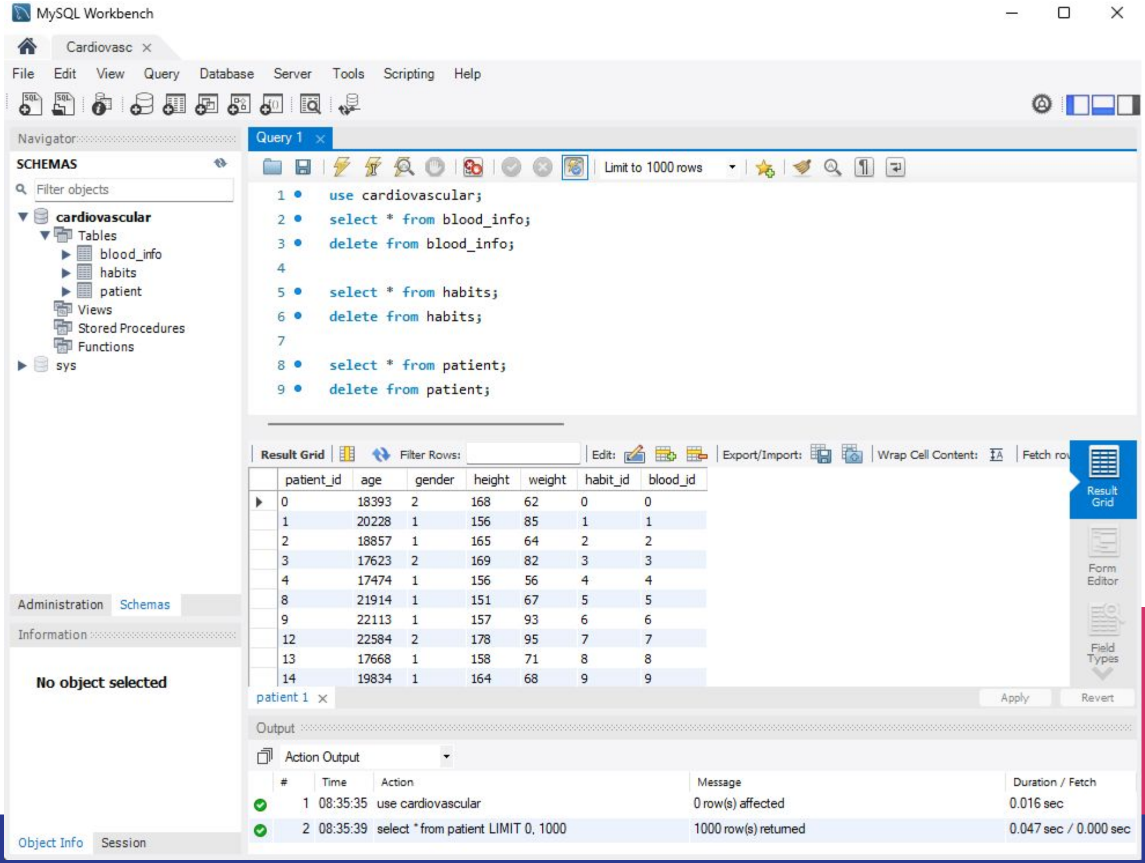
Permissions policies (7) Info

You can attach up to 10 managed policies.

Filter by Type: All types

□	Policy name	Type	Attached entities
□	AmazonEC2FullAccess	AWS managed	1
□	AmazonRDSDataFullAccess	AWS managed	1
□	AmazonRDSFullAccess	AWS managed	1
□	AmazonS3FullAccess	AWS managed	1
□	AmazonVPCFullAccess	AWS managed	1
□	AWSLambda_FullAccess	AWS managed	1
□	CloudWatchFullAccess	AWS managed	1

AWS Setup - Database

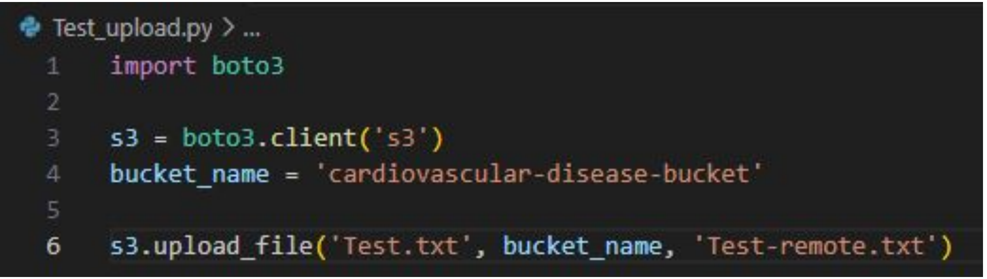


```sql
use cardiovascular;
select * from blood_info;
delete from blood_info;

select * from habits;
delete from habits;

select * from patient;
delete from patient;
```

patient_id	age	gender	height	weight	habit_id	blood_id
0	18393	2	168	62	0	0
1	20228	1	156	85	1	1
2	18857	1	165	64	2	2
3	17623	2	169	82	3	3
4	17474	1	156	56	4	4
8	21914	1	151	67	5	5
9	22113	1	157	93	6	6
12	22584	2	178	95	7	7
13	17668	1	158	71	8	8
14	19834	1	164	68	9	9

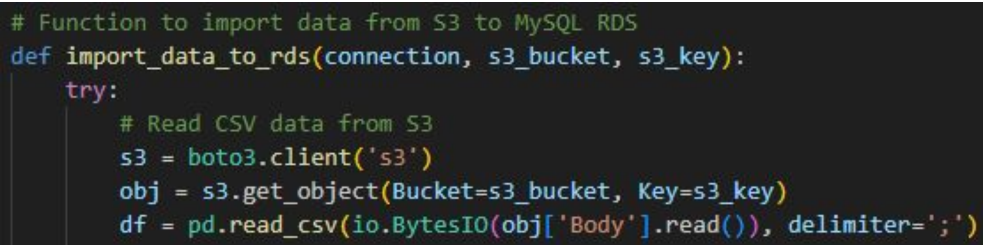
```
#   Time      Action                                Message                Duration / Fetch
1   08:35:35  use cardiovascular                    0 row(s) affected      0.016 sec
2   08:35:39  select * from patient LIMIT 0, 1000   1000 row(s) returned   0.047 sec / 0.000 sec
```

Lambda Function

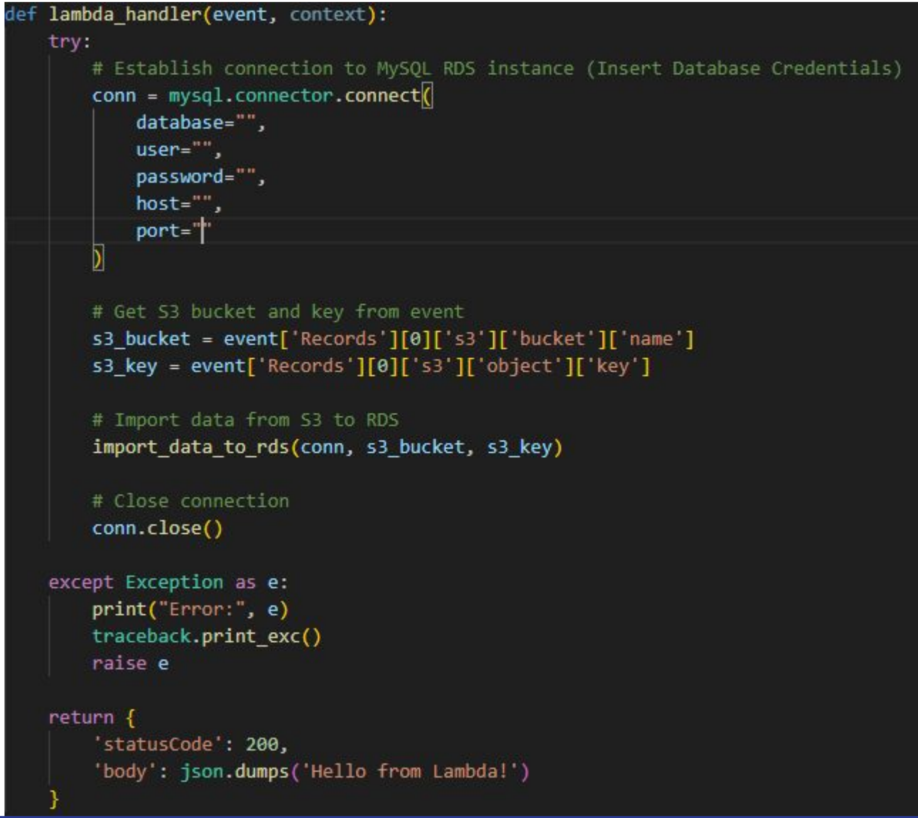


Test_upload.py > ...

```python
import boto3

s3 = boto3.client('s3')
bucket_name = 'cardiovascular-disease-bucket'

s3.upload_file('Test.txt', bucket_name, 'Test-remote.txt')
```

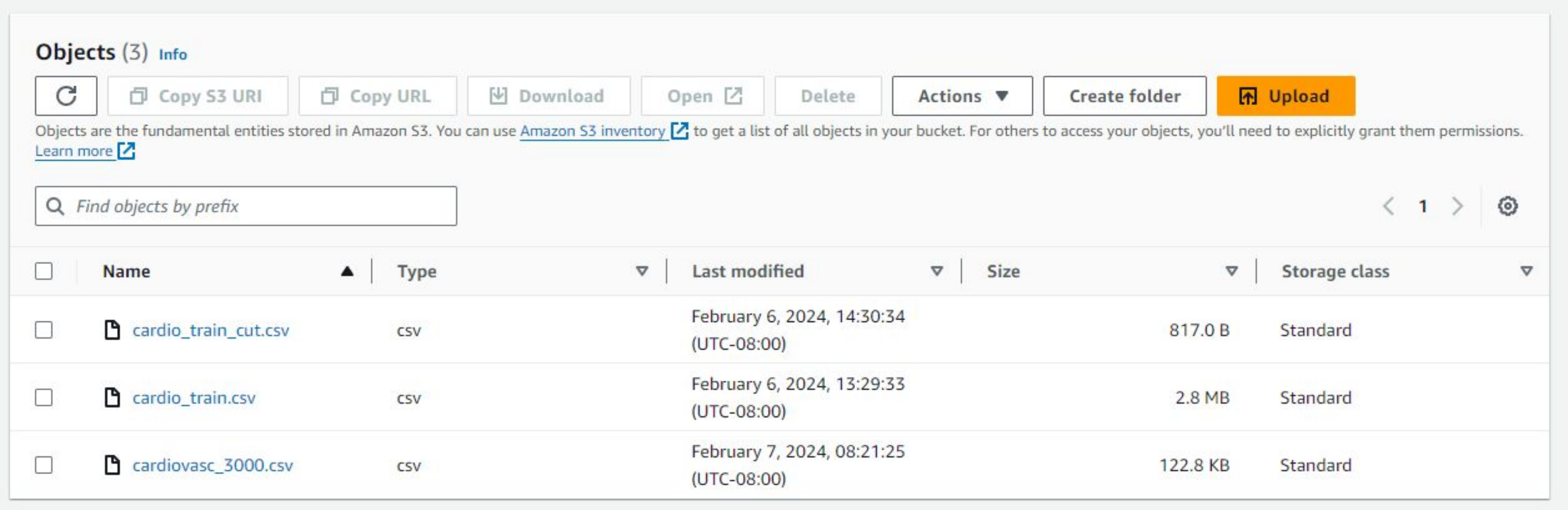


```python
# Function to import data from S3 to MySQL RDS
def import_data_to_rds(connection, s3_bucket, s3_key):
    try:
        # Read CSV data from S3
        s3 = boto3.client('s3')
        obj = s3.get_object(Bucket=s3_bucket, Key=s3_key)
        df = pd.read_csv(io.BytesIO(obj['Body'].read()), delimiter=';')
```



```python
def lambda_handler(event, context):
    try:
        # Establish connection to MySQL RDS instance (Insert Database Credentials)
        conn = mysql.connector.connect(
            database="",
            user="",
            password="",
            host="",
            port=""
        )

        # Get S3 bucket and key from event
        s3_bucket = event['Records'][0]['s3']['bucket']['name']
        s3_key = event['Records'][0]['s3']['object']['key']

        # Import data from S3 to RDS
        import_data_to_rds(conn, s3_bucket, s3_key)

        # Close connection
        conn.close()

    except Exception as e:
        print("Error:", e)
        traceback.print_exc()
        raise e

    return {
        'statusCode': 200,
        'body': json.dumps('Hello from Lambda!')
    }
```

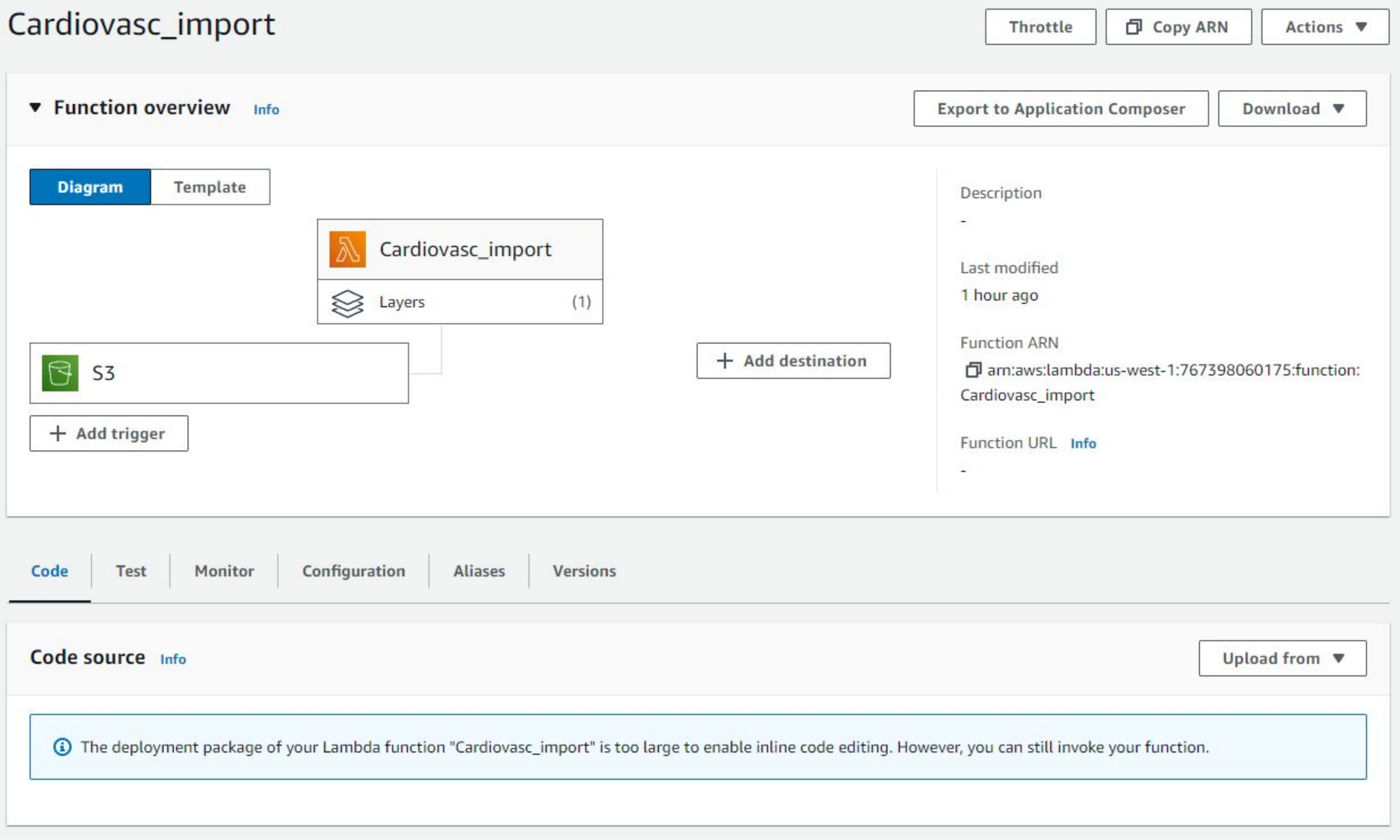
S3 Bucket



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

□	Name	Type	Last modified	Size	Storage class
□	cardio_train_cut.csv	csv	February 6, 2024, 14:30:34 (UTC-08:00)	817.0 B	Standard
□	cardio_train.csv	csv	February 6, 2024, 13:29:33 (UTC-08:00)	2.8 MB	Standard
□	cardiovasc_3000.csv	csv	February 7, 2024, 08:21:25 (UTC-08:00)	122.8 KB	Standard

Lambda Function Deployment



Cardiovasc_import

Function overview Info

Description
-

Last modified
1 hour ago

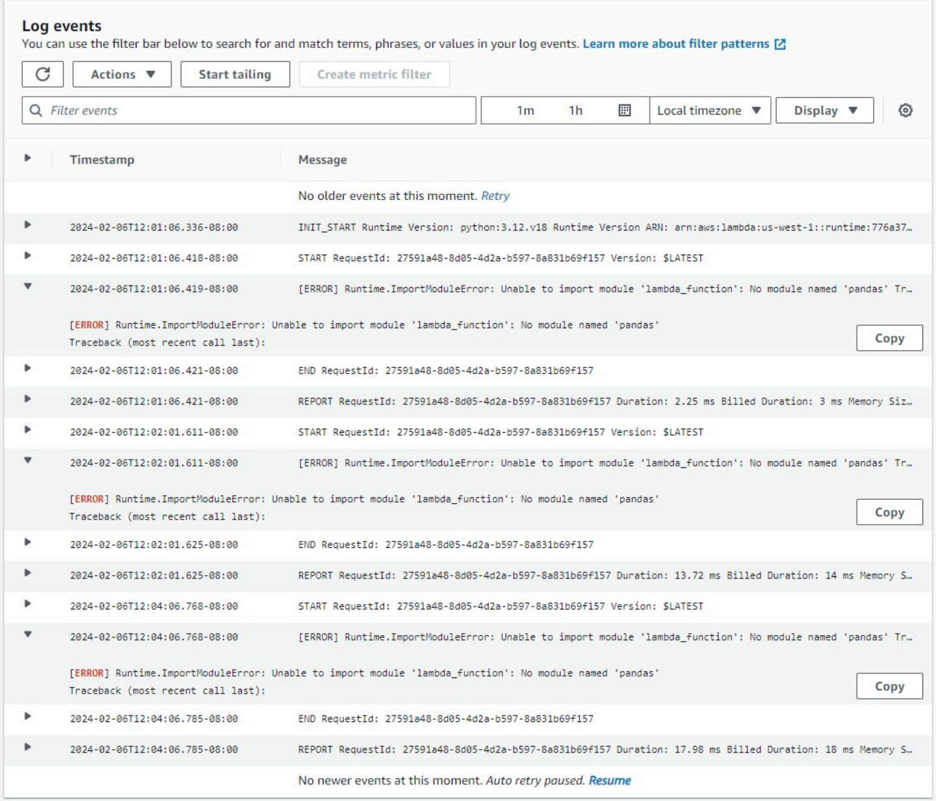
Function ARN
arn:aws:lambda:us-west-1:767398060175:function:Cardiovasc_import

Function URL Info
-

Code source Info

The deployment package of your Lambda function "Cardiovasc_import" is too large to enable inline code editing. However, you can still invoke your function.

CloudWatch Logs - Failed Module Imports



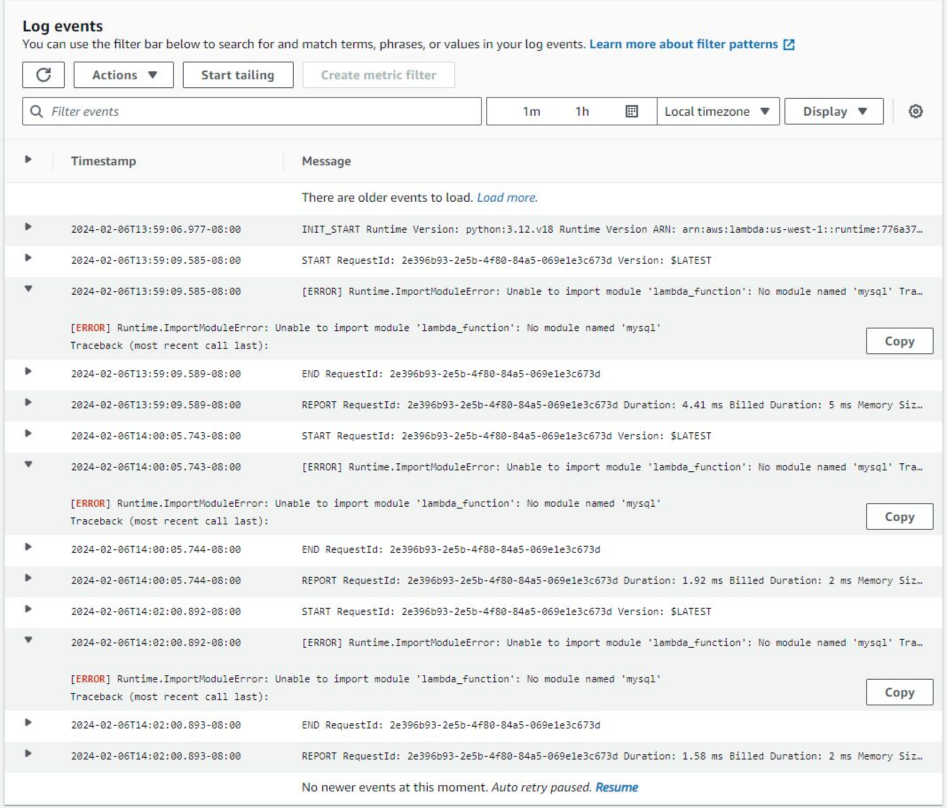
Log events

You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns

No older events at this moment. *Retry*

```
2024-02-06T12:01:06.336-08:00    INIT_START Runtime Version: python:3.12.v18 Runtime Version ARN: arn:aws:lambda:us-west-1::runtime:776a37…
2024-02-06T12:01:06.418-08:00    START RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157 Version: $LATEST
2024-02-06T12:01:06.419-08:00    [ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'pandas' Tr…

[ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'pandas'
Traceback (most recent call last):

2024-02-06T12:01:06.421-08:00    END RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157
2024-02-06T12:01:06.421-08:00    REPORT RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157 Duration: 2.25 ms Billed Duration: 3 ms Memory Siz…
2024-02-06T12:02:01.611-08:00    START RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157 Version: $LATEST
2024-02-06T12:02:01.611-08:00    [ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'pandas' Tr…

[ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'pandas'
Traceback (most recent call last):

2024-02-06T12:02:01.625-08:00    END RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157
2024-02-06T12:02:01.625-08:00    REPORT RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157 Duration: 13.72 ms Billed Duration: 14 ms Memory S…
2024-02-06T12:04:06.768-08:00    START RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157 Version: $LATEST
2024-02-06T12:04:06.768-08:00    [ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'pandas' Tr…

[ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'pandas'
Traceback (most recent call last):

2024-02-06T12:04:06.785-08:00    END RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157
2024-02-06T12:04:06.785-08:00    REPORT RequestId: 27591a48-8d05-4d2a-b597-8a831b69f157 Duration: 17.98 ms Billed Duration: 18 ms Memory S…
```

No newer events at this moment. *Auto retry paused.* *Resume*



Log events

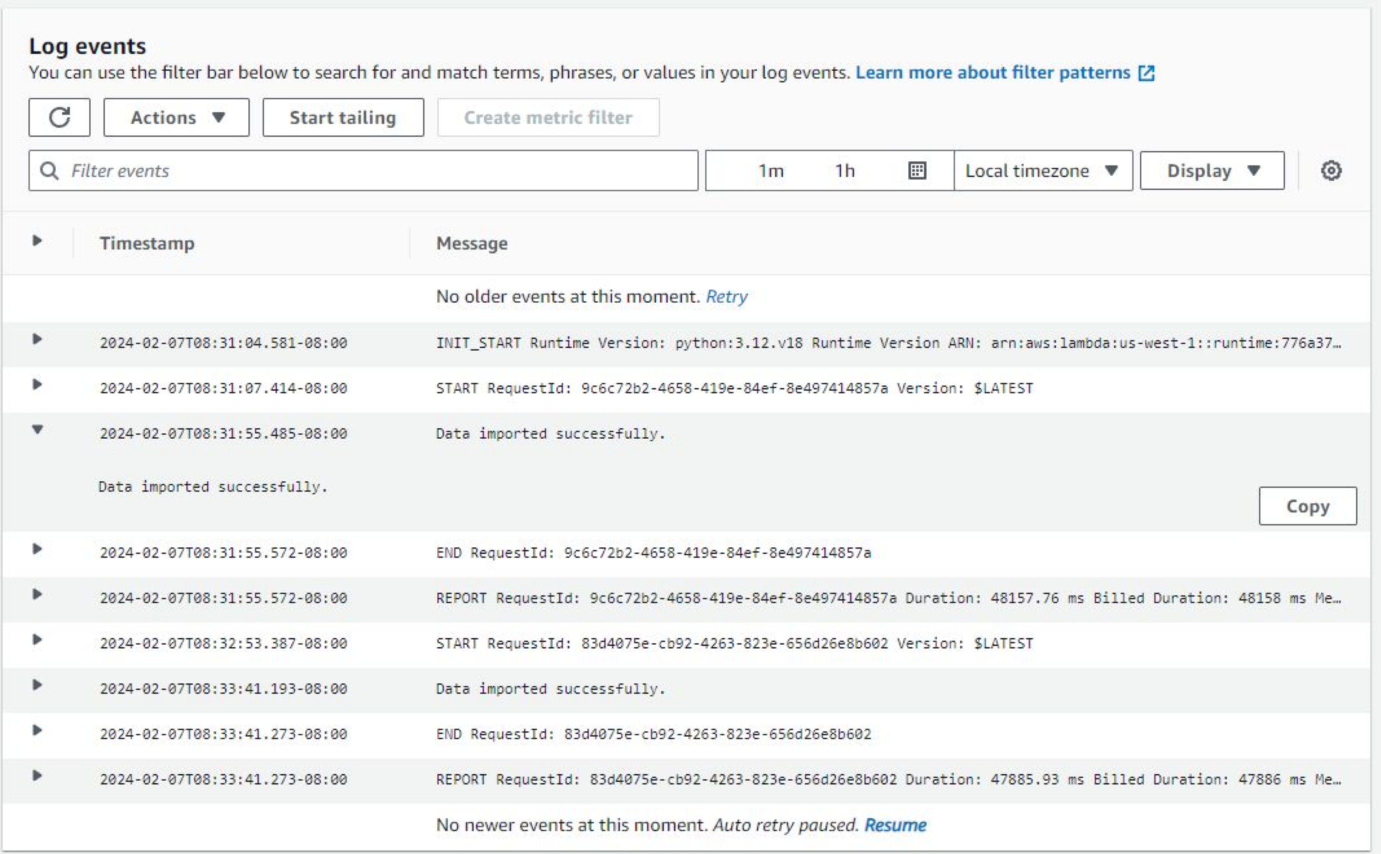
You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns

There are older events to load. *Load more.*

```
2024-02-06T13:59:06.977-08:00    INIT_START Runtime Version: python:3.12.v18 Runtime Version ARN: arn:aws:lambda:us-west-1::runtime:776a37…
2024-02-06T13:59:09.585-08:00    START RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d Version: $LATEST
2024-02-06T13:59:09.585-08:00    [ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'mysql' Tra…

[ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'mysql'
Traceback (most recent call last):

2024-02-06T13:59:09.589-08:00    END RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d
2024-02-06T13:59:09.589-08:00    REPORT RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d Duration: 4.41 ms Billed Duration: 5 ms Memory Siz…
2024-02-06T14:00:05.743-08:00    START RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d Version: $LATEST
2024-02-06T14:00:05.743-08:00    [ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'mysql' Tra…

[ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'mysql'
Traceback (most recent call last):

2024-02-06T14:00:05.744-08:00    END RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d
2024-02-06T14:00:05.744-08:00    REPORT RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d Duration: 1.92 ms Billed Duration: 2 ms Memory Siz…
2024-02-06T14:02:00.892-08:00    START RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d Version: $LATEST
2024-02-06T14:02:00.892-08:00    [ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'mysql' Tra…

[ERROR] Runtime.ImportModuleError: Unable to import module 'lambda_function': No module named 'mysql'
Traceback (most recent call last):

2024-02-06T14:02:00.893-08:00    END RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d
2024-02-06T14:02:00.893-08:00    REPORT RequestId: 2e396b93-2e5b-4f80-84a5-069e1e3c673d Duration: 1.58 ms Billed Duration: 2 ms Memory Siz…
```

No newer events at this moment. *Auto retry paused.* *Resume*

CloudWatch Logs - Success



Log events

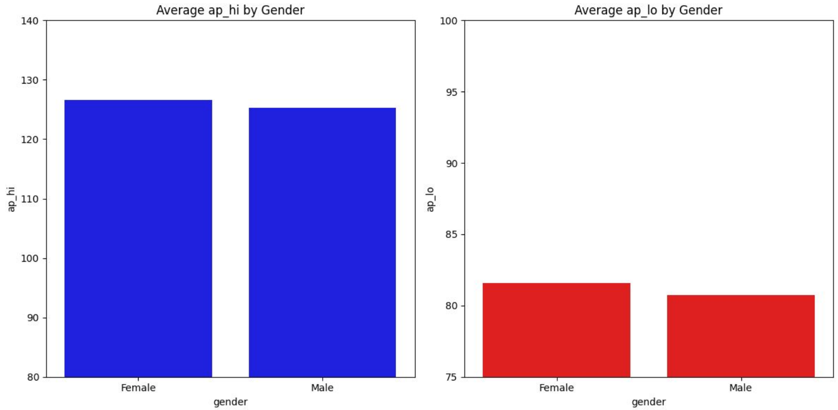
You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns

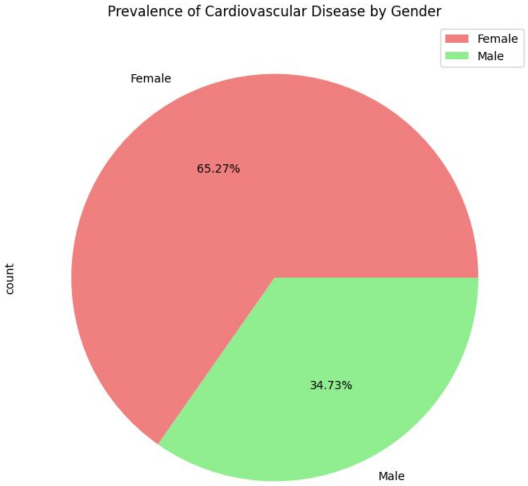
```
   Timestamp                        Message

                                    No older events at this moment. Retry
▶  2024-02-07T08:31:04.581-08:00    INIT_START Runtime Version: python:3.12.v18 Runtime Version ARN: arn:aws:lambda:us-west-1::runtime:776a37…
▶  2024-02-07T08:31:07.414-08:00    START RequestId: 9c6c72b2-4658-419e-84ef-8e497414857a Version: $LATEST
▼  2024-02-07T08:31:55.485-08:00    Data imported successfully.

   Data imported successfully.
                                                                                                             Copy
▶  2024-02-07T08:31:55.572-08:00    END RequestId: 9c6c72b2-4658-419e-84ef-8e497414857a
▶  2024-02-07T08:31:55.572-08:00    REPORT RequestId: 9c6c72b2-4658-419e-84ef-8e497414857a Duration: 48157.76 ms Billed Duration: 48158 ms Me…
▶  2024-02-07T08:32:53.387-08:00    START RequestId: 83d4075e-cb92-4263-823e-656d26e8b602 Version: $LATEST
▶  2024-02-07T08:33:41.193-08:00    Data imported successfully.
▶  2024-02-07T08:33:41.273-08:00    END RequestId: 83d4075e-cb92-4263-823e-656d26e8b602
▶  2024-02-07T08:33:41.273-08:00    REPORT RequestId: 83d4075e-cb92-4263-823e-656d26e8b602 Duration: 47885.93 ms Billed Duration: 47886 ms Me…
                                    No newer events at this moment. Auto retry paused. Resume
```

Statistics graphs

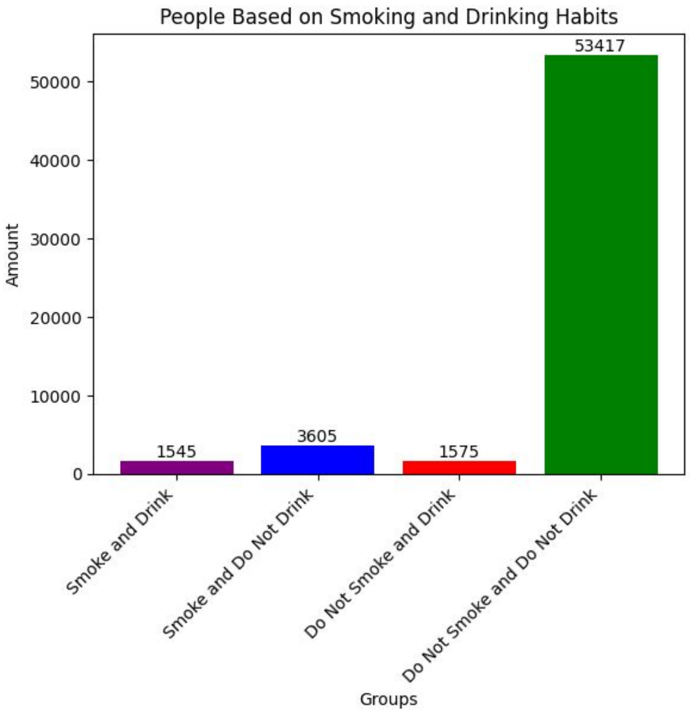
Had to clean the data when I noticed that there was a few records that made little sense such as:

- Negative blood pressure, height, and weight
- Blood pressure up to 14,000 (normally up to ~190)
- People who where over 10 feet tall
- People who were over 700 pounds





Smoking and Drinking Habits



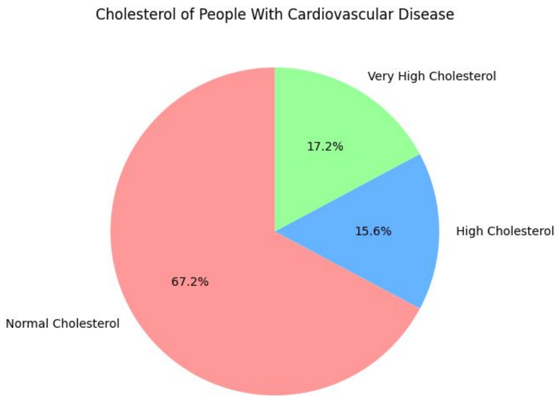
- People who do NOT drink nor smoke are much more common than all other categories

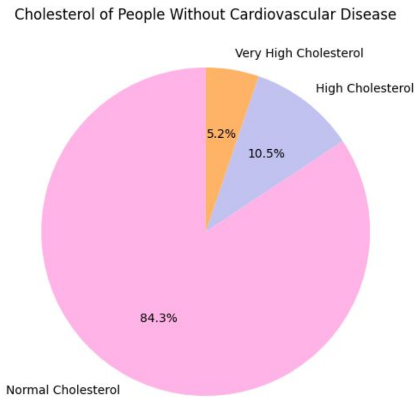
- This matches with other data found such as those in the next slide

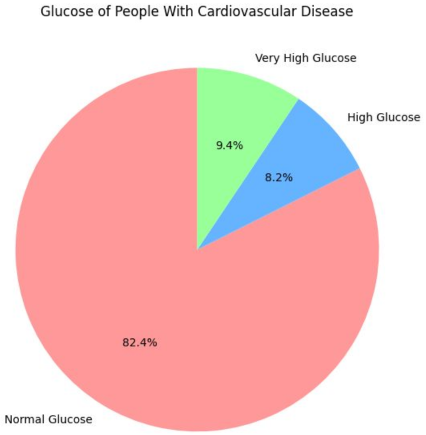
More Graphs

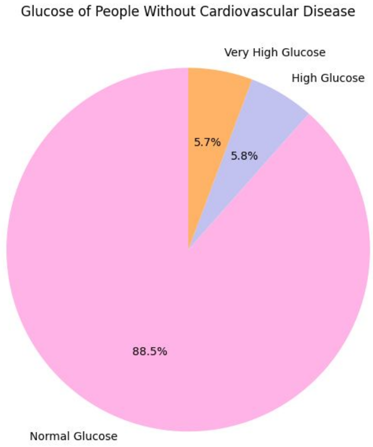
- Suggestions:
  - Added bmi, change age to years, store gender as female or male not 1 and 2

- Cholesterol seems to have a more important role than glucose when looking at people who have cardiovascular disease

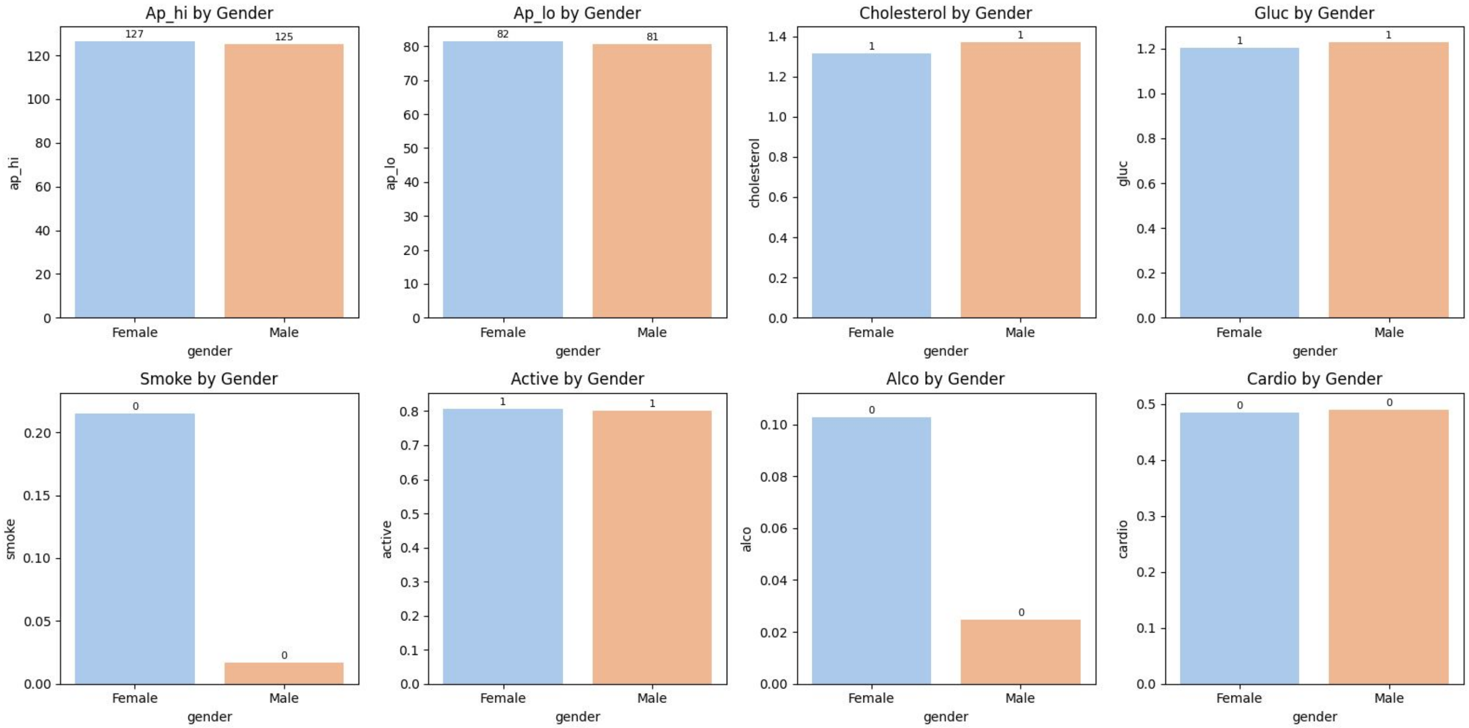

Cholesterol of People With Cardiovascular Disease


Cholesterol of People Without Cardiovascular Disease


Glucose of People With Cardiovascular Disease


Glucose of People Without Cardiovascular Disease

Everything by gender



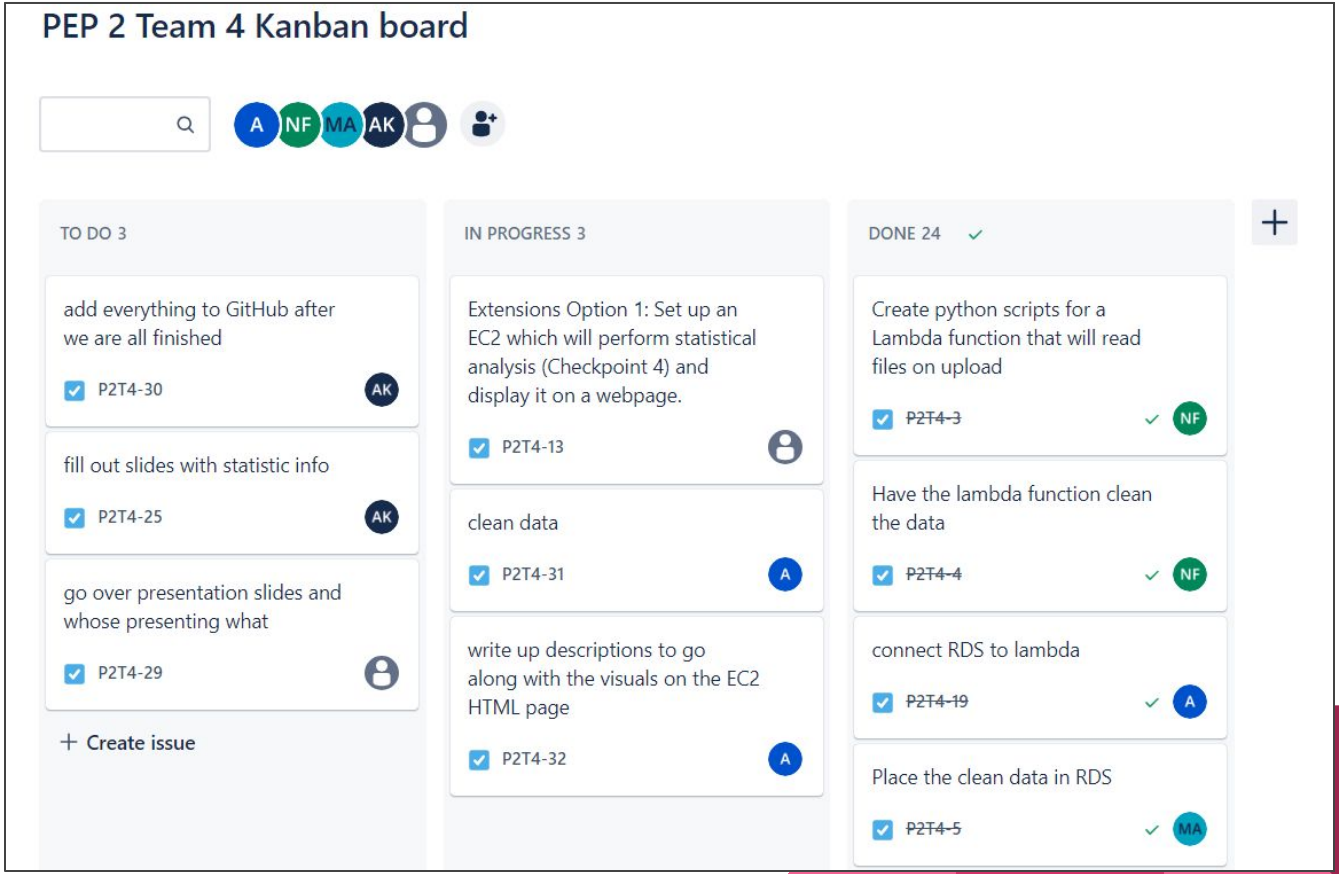
EC2 Website

54.183.30.121

Challenges we faced

- Connecting to AWS was an issue in the beginning
- Importing modules on AWS lambda
- Dealing with outliers in the data

Kanban Board



PEP 2 Team 4 Kanban board

TO DO 3

add everything to GitHub after we are all finished

☑ P2T4-30 AK

fill out slides with statistic info

☑ P2T4-25 AK

go over presentation slides and whose presenting what

☑ P2T4-29

+ Create issue

IN PROGRESS 3

Extensions Option 1: Set up an EC2 which will perform statistical analysis (Checkpoint 4) and display it on a webpage.

☑ P2T4-13

clean data

☑ P2T4-31 A

write up descriptions to go along with the visuals on the EC2 HTML page

☑ P2T4-32 A

DONE 24 ✓

Create python scripts for a Lambda function that will read files on upload

☑ ~~P2T4-3~~ ✓ NF

Have the lambda function clean the data

☑ ~~P2T4-4~~ ✓ NF

connect RDS to lambda

☑ ~~P2T4-19~~ ✓ A

Place the clean data in RDS

☑ ~~P2T4-5~~ ✓ MA

Any Questions?