
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

COM303

Monitor and predict health data using AWS AI services

Luca Bianchi (He/Him)

Chief Technology Officer
Neosperience

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Luca Bianchi

Chief Technology Officer at
Neosperience and Neosperience Health

AWS Serverless Hero, passionate about machine learning and AI

Loves the Opera and Puccini

github.com/aletheia

It.linkedin.com/in/lucabianchipavia

@bianchiluca

SCAN ME



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Safe harbor statement

- None of the presented methods have been validated by clinical study (their evaluation is underway)
- None of the suggested ideas in this talk, slides, or related material represent medical advice nor is it intended to substitute any medical advice
- The main focus of this talk is to present a solution to monitor health data and extract meaningful insights and/or outliers that need to be evaluated by a doctor prior any prescripted therapy
- The author and speaker decline any responsibility

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Health is broken

The age of connected devices

Time-series data collection and monitoring

Amazon Lookout for Equipment

A serverless architecture to handle health data

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Health is broken

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

People’s health is worsening

Metabolic diseases are spreading in the US and worldwide, posing a threat for people’s health and healthcare system sustainability https://pubmed.ncbi.nlm.nih.gov/19558267/

According to the CDC, metabolic diseases rose by double digits for every decade since the mid-1990s reaching a worrisome 35% increase between 2010 and 2019 https://www.cdc.gov/pcd/issues/2017/16_0287.htm



Photo by JESHOOTS.COM on Unsplash

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What causes most diseases?

- Metabolic diseases (MD) correlates with many health diseases such as diabetes
- Resolving MD improves people’s life quality and lowers healthcare
- Health issues should be prevented rather than treated
- Screening is costly and has been disrupted by the pandemics
- Screening could prevent Metabolic Syndrome and Chronic Obstructive Pulmonary Disease (COPD)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Metabolic syndrome

- Not an illness, but a condition spanning multiple indicators such as
  - heart rate variability
  - body weight and BMI
  - body response to glucose
  - blood pressure

- Often undetected for years
- Related to diabetes and hearth failures

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The age of connected devices

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Continuous glucose monitor (GCM)

Smart sensor, simple to use and remove

Developed for patients with diabetes, provides useful insights to healthy peoples about

- Fasting glucose
- Insulin response to food
- Glucose response to hydration



Photo courtesy of Flickr

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

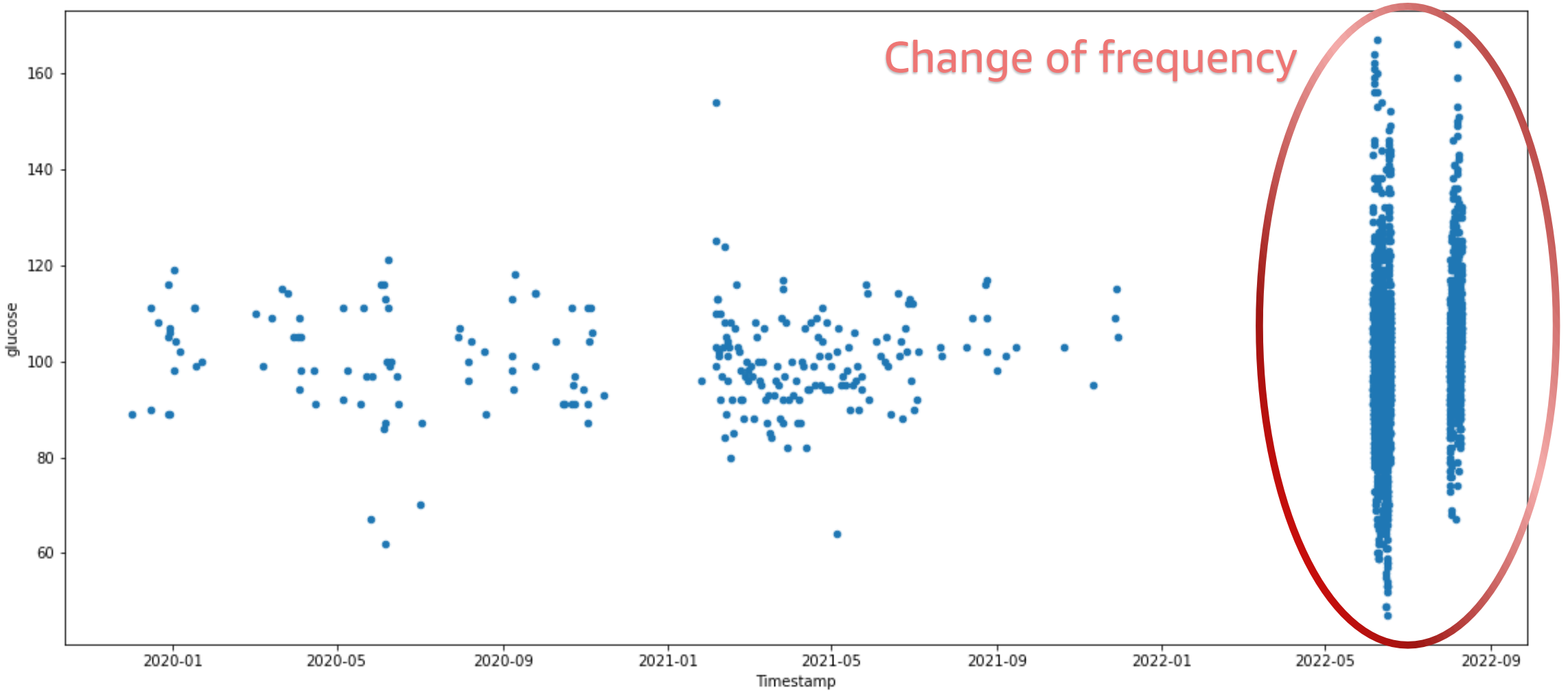
“Everyone should wear a GCM at least once in their lives to understand their body”

Dr. Peter Attia, MD

Physician – Longevity expert

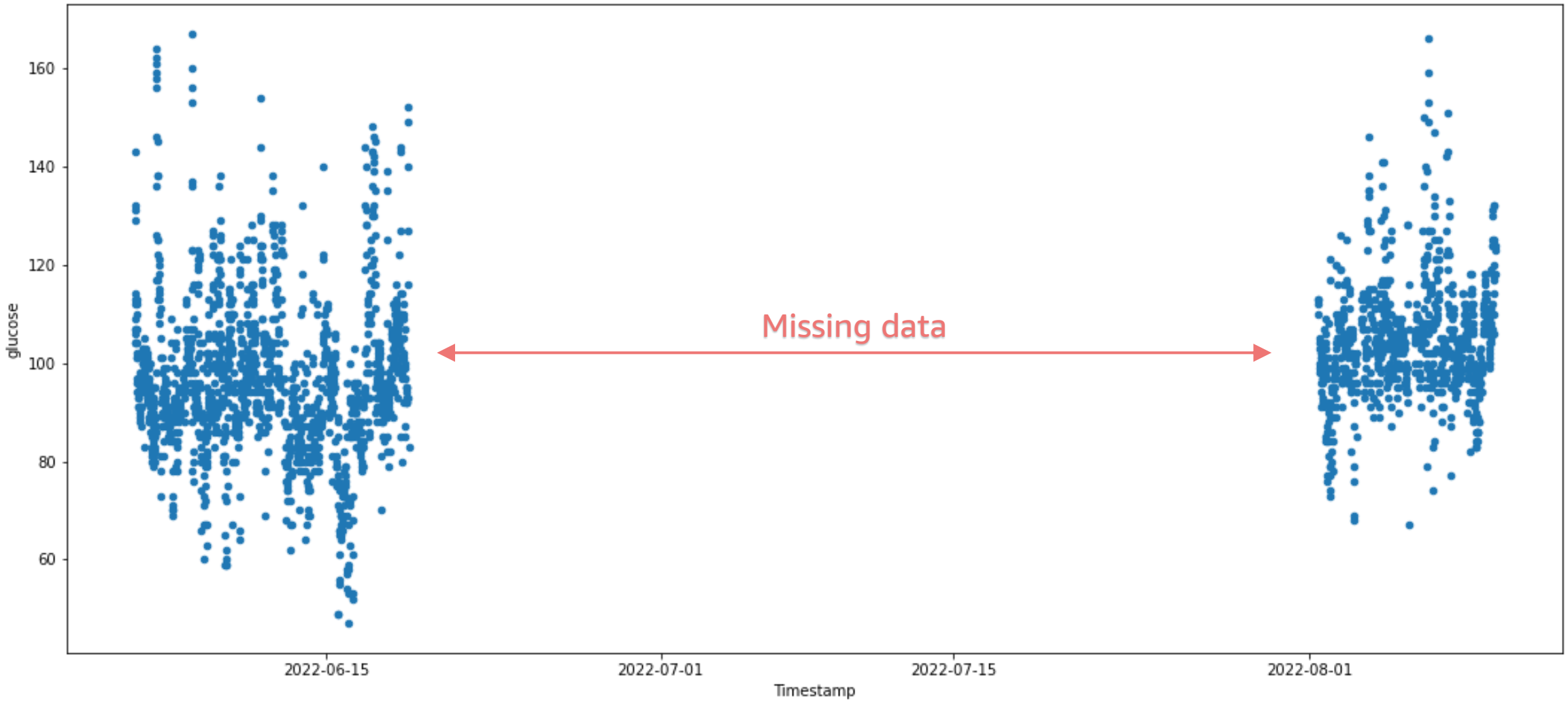
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Glucose data



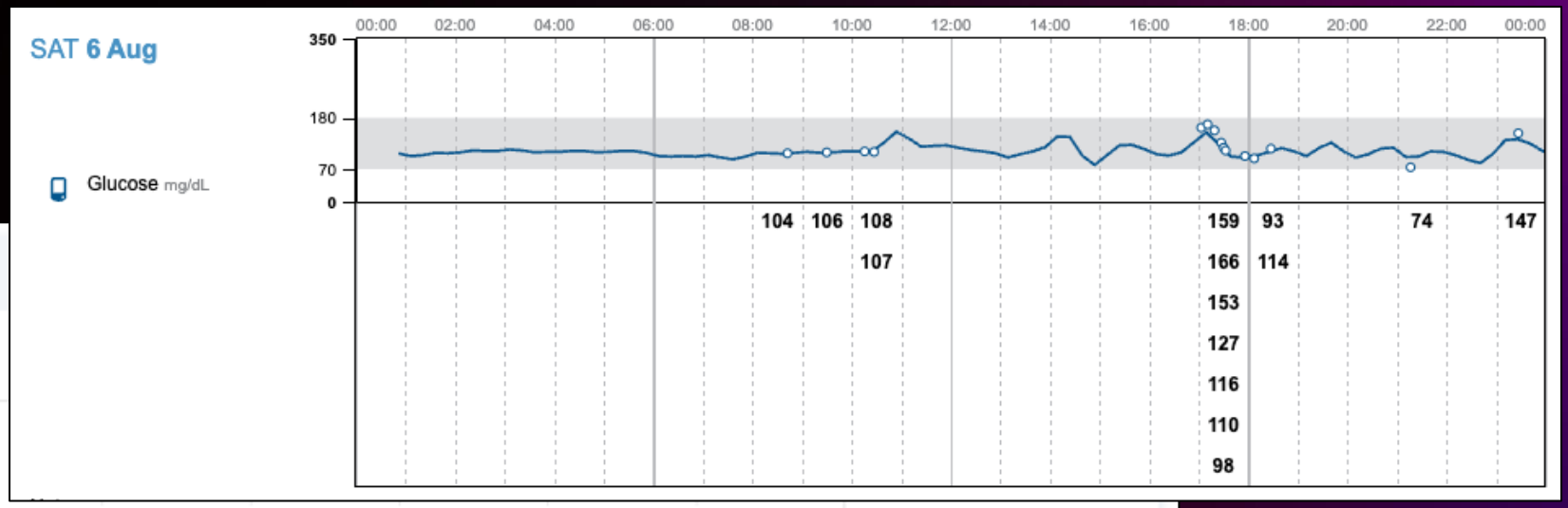
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

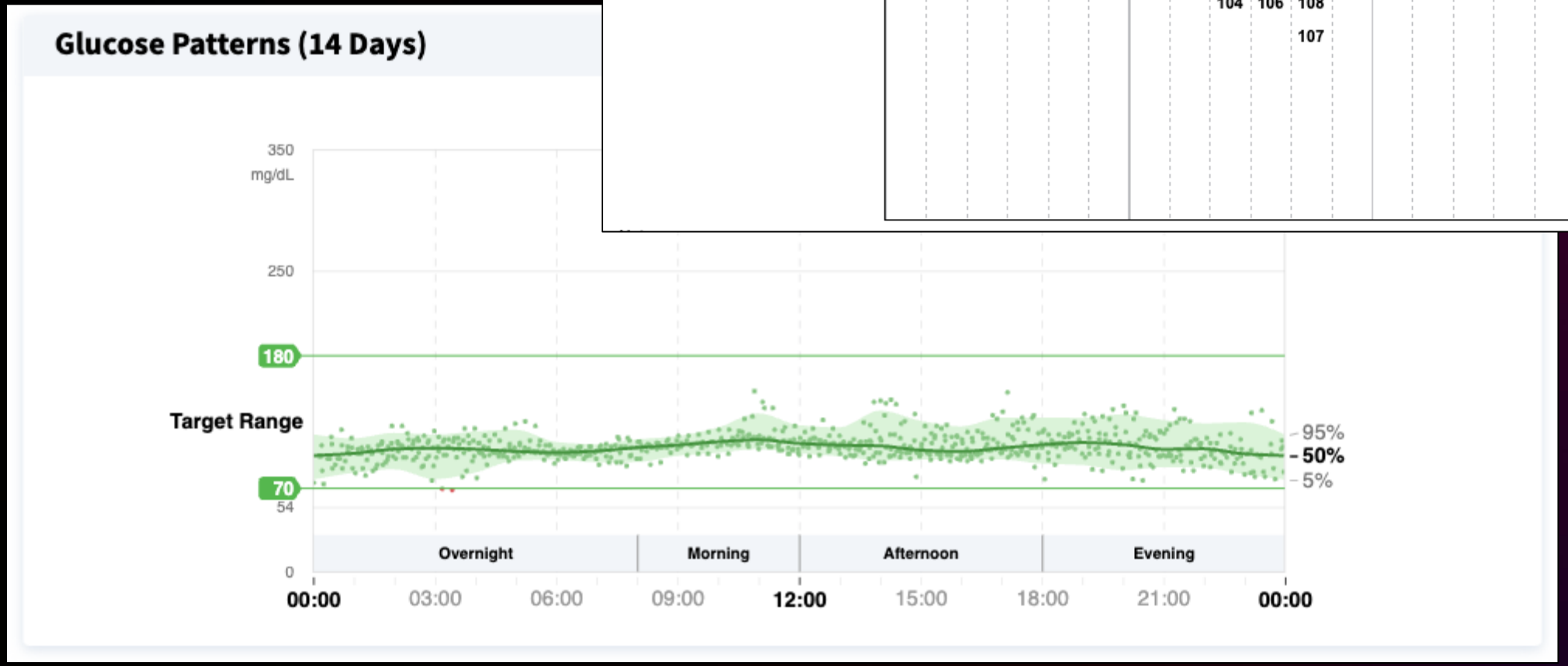
Glucose data – drill down



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Glucose patterns





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why monitor glucose?

- Glucose response is coupled to patterns in glucose behavior
- Unpaired glucose responses to fasting or food intake are NOT medical conditions
- It is related to diabetes development
- GCM devices provides understanding of body response to glucose and meaningful insights
- Collected data could present gaps or changes in frequency

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blood pressure monitor (BPM)

Non invasive, simple to use, cheap

Smart sensor to track

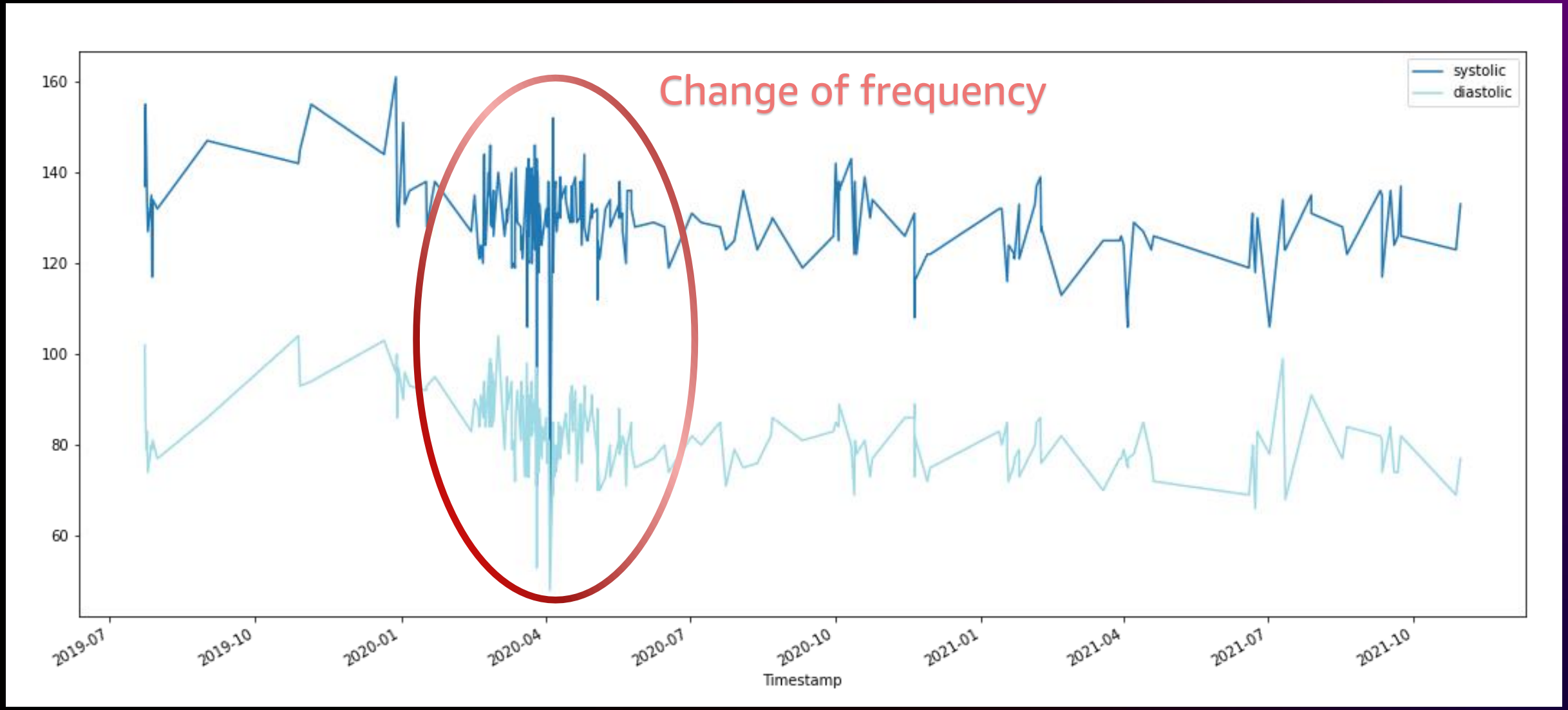
- **Systolic** blood pressure
- **Diastolic** blood pressure
- Heart rate

Allows detection of hypertension or bad heart rate variability



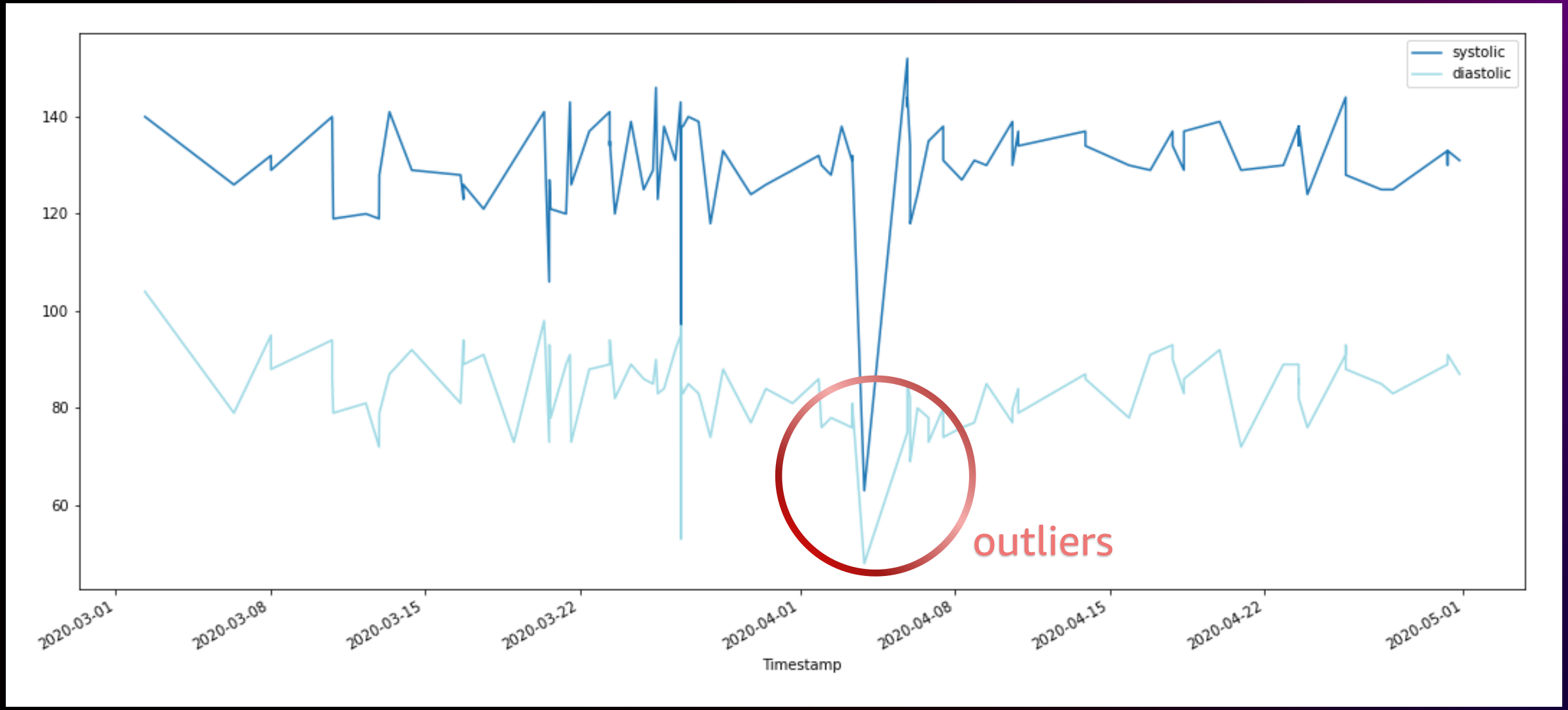
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blood pressure data



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blood pressure data – drill down



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Heart rate data



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Apple Health data

- Many smart sensors sync data to Apple Health app
- XML data with 30+ features
- Big data export (~ 1.5GB)
- Redundant data (almost 4x)
- Pre-processing to filter and convert into CSV is required

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Using AI to monitor health patterns

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data preparation

- Convert data to CSV
- Remove outliers and measurement errors (i.e., glucose values below 50mg/dL)
- Repack time series data into [‘Timestamp’, ‘value’] tuples
- Upload data to Amazon Simple Storage Service (Amazon S3)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Time-series analysis - issues

- Sparse datasets
  - Lots of **null** values
  - Sampling differences
- Unbalanced datasets
- Consider the whole patient, no only a metric or a set of anomalies

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Lookout for Equipment

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Lookout for Equipment

- Analyzes the data from industrial equipment (e.g., pressure, flow rate, RPMs)
- Auto model training
- Handles data from up to 300 sensors in one model
- Performs anomaly detection
- Async batch inference using Amazon S3
- No prior data science knowledge required
- Multi-variate data inference
- Why not use Amazon Lookout for Metrics?

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Lookout for ~~Equipment~~ Health

- Map “sensors” to metrics and “assets” to people
- Datasets from commercial devices
- Train, fine tune, and deploy a model for batch inference
- Build a digital twin capable of detecting health anomalies
- Build a serverless architecture for data ingestion and anomalies notification

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Parse Apple Health data to pandas

```
def load_and_convert_apple_health(patient):
    path = os.path.join(raw_data_path, patient)
    apple_data_file = os.path.join(path, 'apple_health/latest/export.xml')
    tree = ET.parse(apple_data_file)
    root = tree.getroot()
    record_list = [x.attrib for x in root.iter('Record')]
    record_data = pd.DataFrame(record_list)

    for col in ['creationDate', 'startDate', 'endDate']:
        record_data[col] = pd.to_datetime(record_data[col])
        record_data['value'] =
                pd.to_numeric(record_data['value'], errors='coerce')
    record_data['value'] = record_data['value'].fillna(1.0)

    record_data['type’] =
        record_data['type'].str.replace('HKQuantityTypeIdentifier', '')
    record_data['type’] =
        record_data['type'].str.replace('HKCategoryTypeIdentifier', '')
    return record_data
```

Convert data to number, remove null values, then sanitize metric names

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Parse GCM data to pandas

```
def load_abbott_data(patient):
    abbott_path = os.path.join(raw_data_path, patient, 'abbott_freestyle')
    patient_path = os.path.join(prepared_data_path,patient)
    for file in os.listdir(abbott_path):
        # […] read files

    df['Timestamp'] = pd.to_datetime(df['Timestamp'], format='%d-%m-%Y %H:%M')
    df['value'] = df['value_scan'].combine_first(df['value_historical'])
    df = df[['Timestamp', 'value']].sort_values('Timestamp')
    df = df.drop_duplicates().dropna()
    df = df.set_index('Timestamp')
    df['glucose_daily'] = df['value']
    df = df.drop(columns=['value'])
return df
```

Combine manually triggered measures with automatic sampling into one single column

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dataset schema discovery

Amazon Lookout for Equipment

Projects

- reinvent2022_partials
  - Dataset
  - Models
  - Inference schedules New

Data ingestion in progress
Depending on the size of your dataset, ingestion can take up to 20 minutes. Navigating away from this page will not interrupt the ingestion process.

Amazon Lookout for Equipment > Projects > reinvent2022_partials

reinvent2022_partials Info

How it works



Step 1. Create project

A *project* is a collection of resources associated with a single industrial asset that you want to monitor.

Created



Step 2. Add dataset

Each project contains a *dataset*—a collection of historical data that you ingest into Lookout for Equipment.

Ingestion in progress



Step 3. Train model

You train a custom model on your asset's historical data to establish a baseline for normal behavior.

Get started



Step 4. Schedule inference

You schedule continuous inference on new measurement data to monitor your asset in real-time.

Schedule inference

Project details Info

Project name	Status	Creation date
reinvent2022_partials	Ingestion in progress	October 13, 2022 at 11:59 (UTC+2:00)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dataset schema creation

```
dataset_schema = {
    "Components": [

        {"ComponentName": "glucose_daily",
            "Columns":     [  {"Name": "Timestamp", "Type": "DATETIME”          },
                              {"Name": "glucose_daily", "Type": "DOUBLE”        }]},

        {"ComponentName": "glucose_fasting",
            "Columns":     [  {"Name": "Timestamp", "Type": "DATETIME"},
                              {"Name": "glucose_fasting", "Type": "DOUBLE"}     ]},

        {"ComponentName": "heart_rate",
            "Columns": [      {"Name": "Timestamp", "Type": "DATETIME"},
                              {"Name": "heart_rate", "Type": "DOUBLE"}          ]},

        {"ComponentName": "blood-pressure",
            "Columns": [      {"Name": "Timestamp", "Type": "DATETIME"},
                              {"Name": "systolic", "Type": "DOUBLE"},
                              {"Name": "diastolic", "Type": "DOUBLE"}           ]}
    ]
}
```

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dataset schema creation

```
lookoutequipment = boto3.client('lookoutequipment’)

region = 'us-east-1'
config = Config(
    region_name = region
)

data_schema = {
    'InlineDataSchema': json.dumps(dataset_schema, config=config),
}

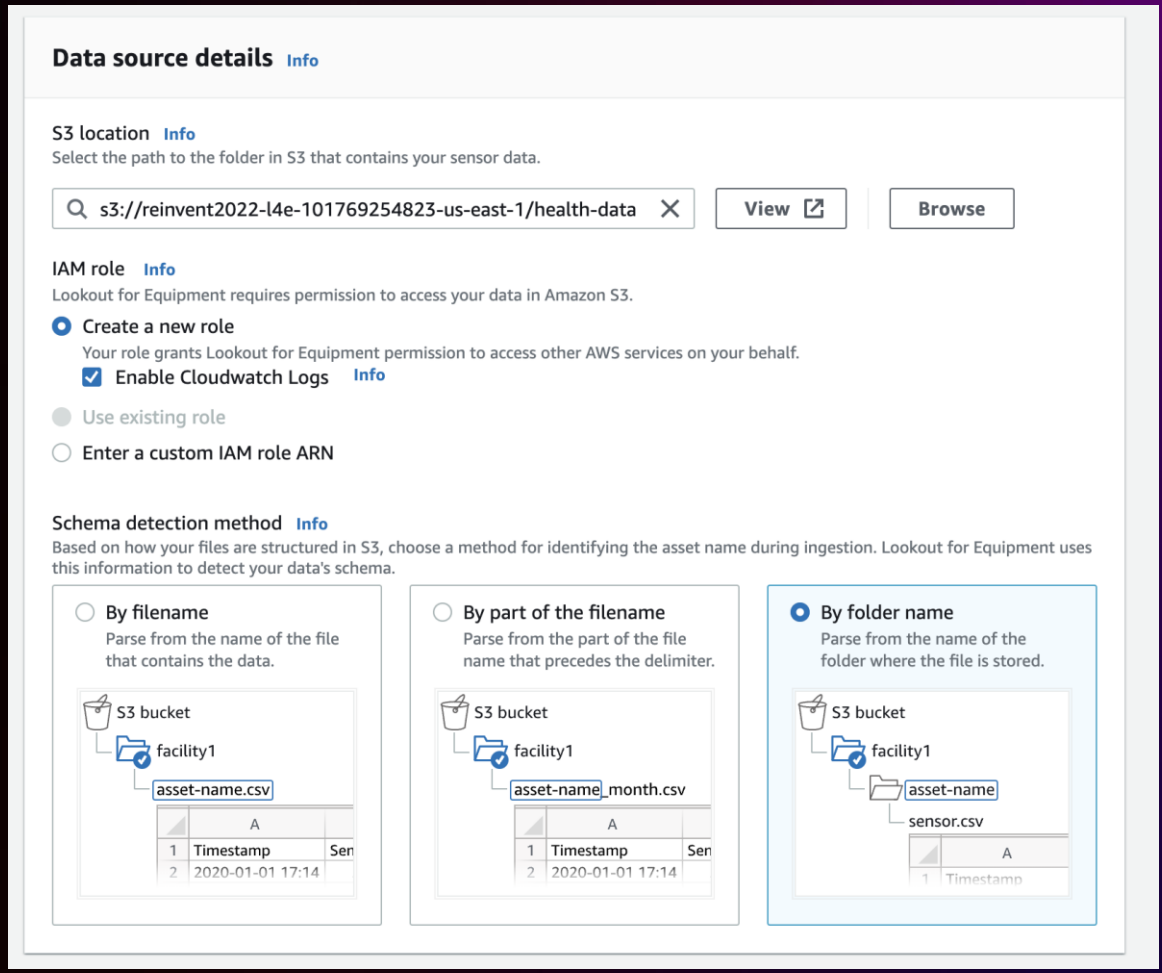
create_dataset_response = lookoutequipment.create_dataset(
                                DatasetName= ‘reinvent2022_partials',
                                DatasetSchema=data_schema
)
```

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data ingestion

NO NEED TO SPECIFY DATA STRUCTURE

- Select Amazon S3 bucket containing data
- Enable Amazon CloudWatch Logs on ingested data (improves ingestion failures explainability)
- Select a schema detection method, remember asset → patient and metric → sensor association



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data ingestion

Amazon Lookout for Equipment

Projects

- reinvent2022_partials
  - Dataset
  - Models
  - Inference schedules New

Data ingestion in progress
Depending on the size of your dataset, ingestion can take up to 20 minutes. Navigating away from this page will not interrupt the ingestion process.

Amazon Lookout for Equipment > Projects > reinvent2022_partials

reinvent2022_partials Info

How it works



Step 1. Create project

A *project* is a collection of resources associated with a single industrial asset that you want to monitor.

Created



Step 2. Add dataset

Each project contains a *dataset*—a collection of historical data that you ingest into Lookout for Equipment.

Ingestion in progress



Step 3. Train model

You train a custom model on your asset's historical data to establish a baseline for normal behavior.

Get started



Step 4. Schedule inference

You schedule continuous inference on new measurement data to monitor your asset in real-time.

Schedule inference

Project details Info

Project name	Status	Creation date
reinvent2022_partials	Ingestion in progress	October 13, 2022 at 11:59 (UTC+2:00)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data ingestion

```
import time

data_ingestion_job_id = response['JobId']
data_ingestion_status = response['Status']

print(f'=====Data Ingestion job is started. Job ID: {data_ingestion_job_id}=====\n')

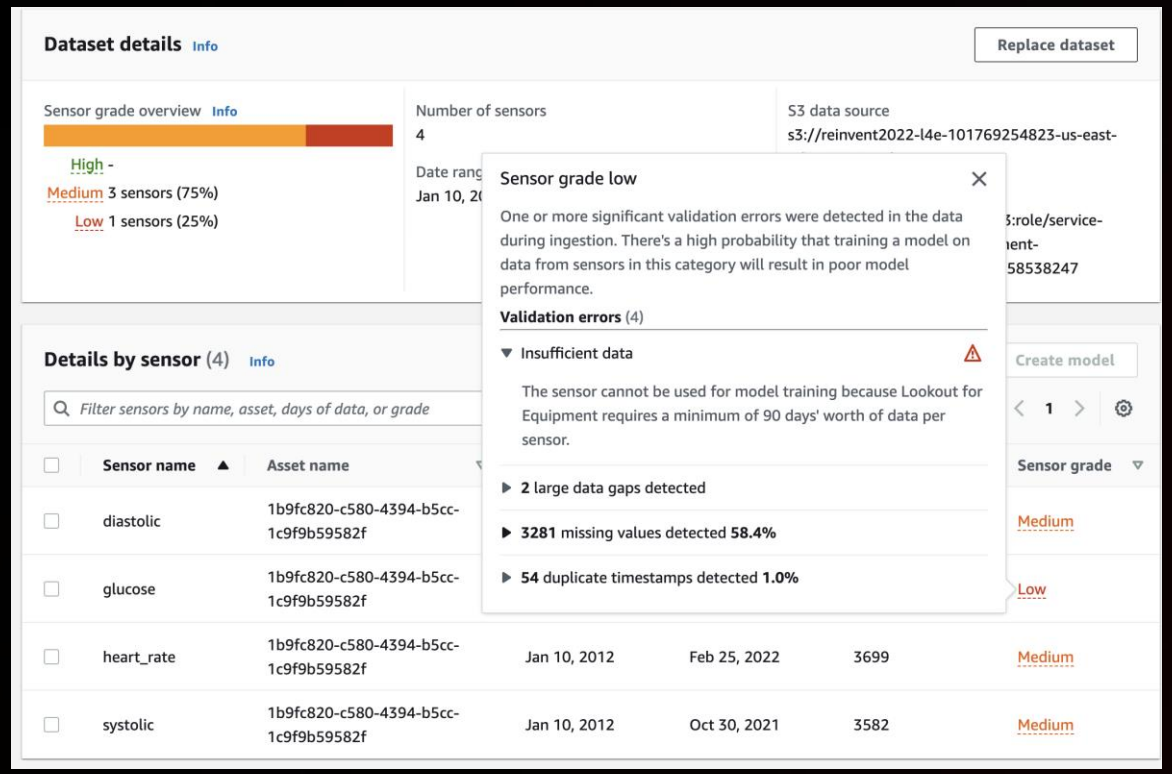
# Wait until completes
print("=====Polling Data Ingestion Status=====\n")
print("Data Ingestion Status: " + data_ingestion_status)
while data_ingestion_status == 'IN_PROGRESS':
    time.sleep(30)
    describe_data_ingestion_job_response = l4e.describe_data_ingestion_job(JobId=data_ingestion_job_id)
    data_ingestion_status = describe_data_ingestion_job_response['Status']
    print("Data Ingestion Status: " + data_ingestion_status)
    print("\n=====End of Polling Data Ingestion Status=====")
```

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data ingestion – feedback

POOR DATA QUALITY LOWERS MODEL PERFORMANCE



- Dataset is checked for the minimum amount of data (90 days)
- Large gaps between data clusters worsen analysis (because introduce a no-value bias)
- Missing data points are forward filled with estimated values based on available data. Aim for < 10% missing values
- Duplicated data is omitted, no need to remove duplicates
- Considering health data, target for medium sensor grade

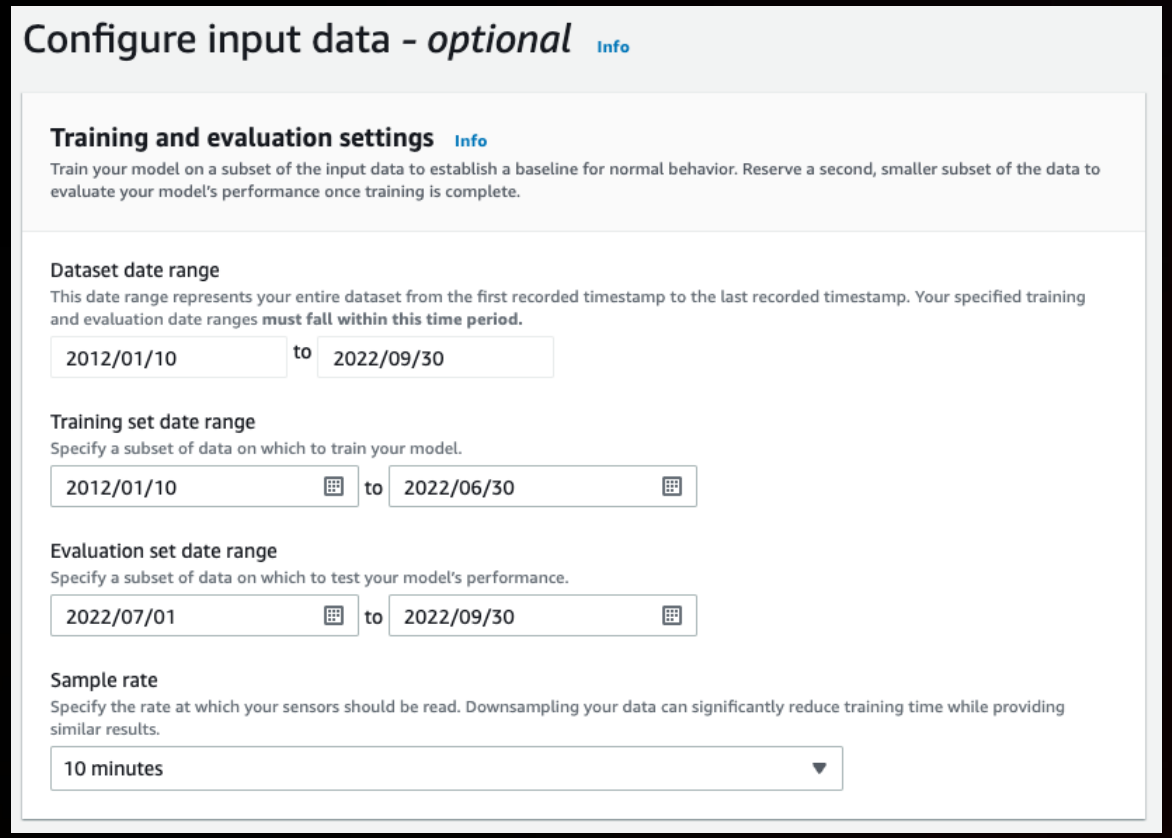
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Training a model

- Unsupervised anomaly detection model
- Hyperparameters auto optimization
- Amazon Lookout for Equipment encodes the dataset, different datasets lead to different models
- Data labels force Amazon Lookout for Equipment to use a neural network as base model
- Train a model for each patient

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Training parameters

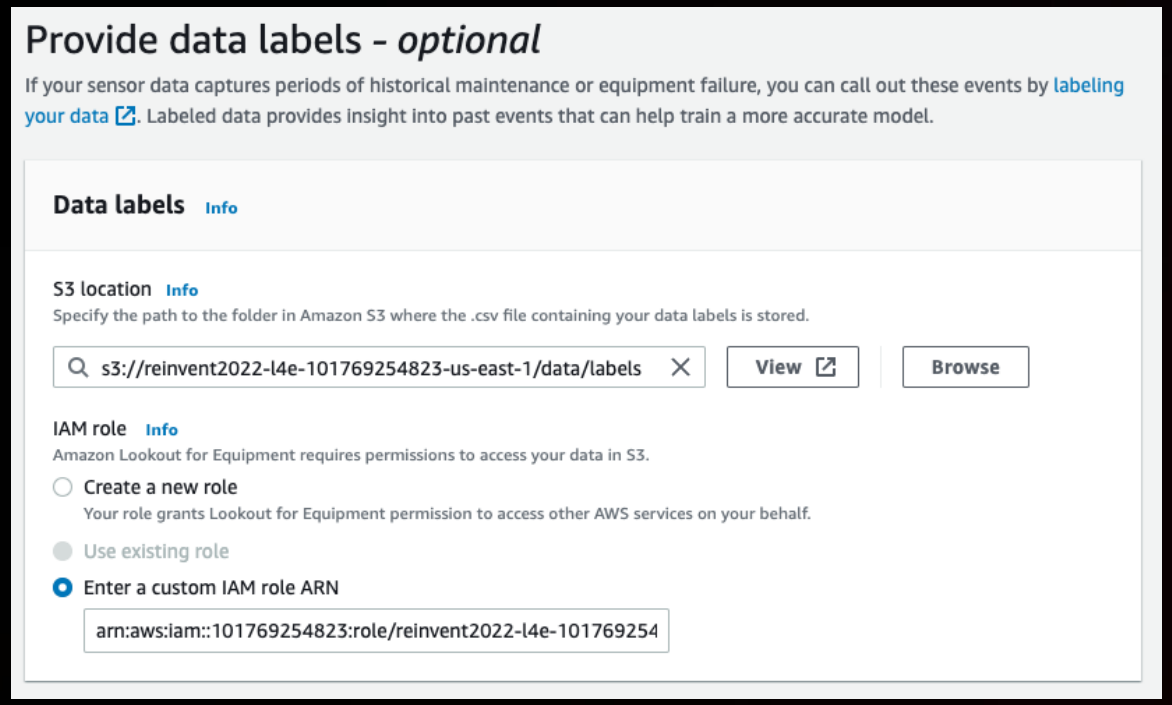


- Dataset can be used entirely for training, but model won't be evaluated to confirm behavior
- Selecting date ranges for training and evaluation provides better performances
- Data sampling should be set to the highest data frequency
- Missing data are estimated during training to fill gaps

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Training parameters

DATA LABELING



- Labels are meaningful events (such as a fitness session, a party, or a trip) which can led to anomalies in data
- Use labels to switch from traditional machine learning model to a neural network

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Train a model

```
from datetime import datetime

create_model_request = {
    'ModelName': 'reinvent-model-01',
    'DatasetName': wd,
    'DataPreProcessingConfiguration':{
        'TargetSamplingRate': 'PT10M'
    },
    'TrainingDataStartTime': datetime(2012, 1, 10, 0, 0, 0),
    'TrainingDataEndTime': datetime(2022, 6, 30, 0, 0, 0),
    'EvaluationDataStartTime': datetime(2022, 7, 1, 0, 0, 0),
    'EvaluationDataEndTime': datetime(2022, 9, 30, 0, 0, 0)

}
create_model_response = l4e.create_model(**create_model_request)

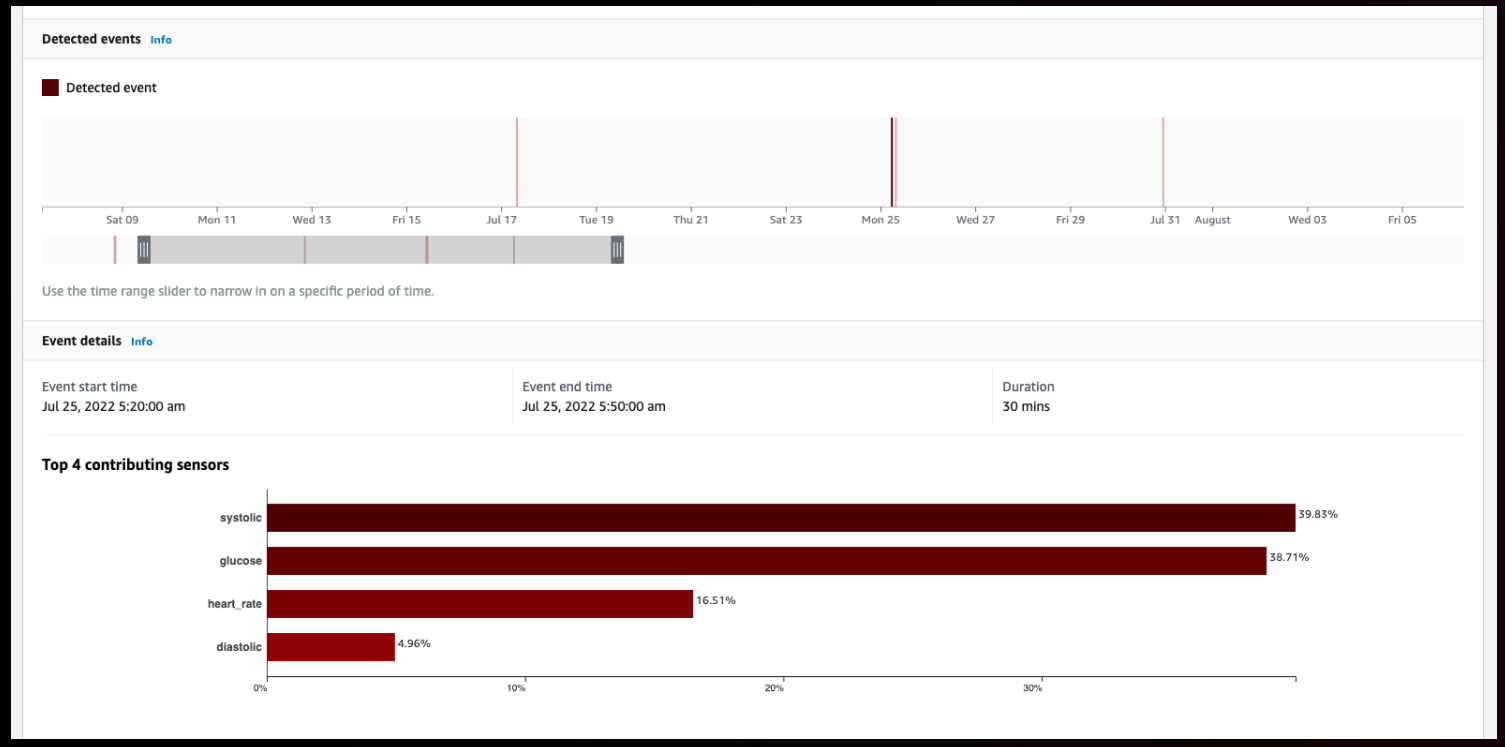
# Wait until complete
```

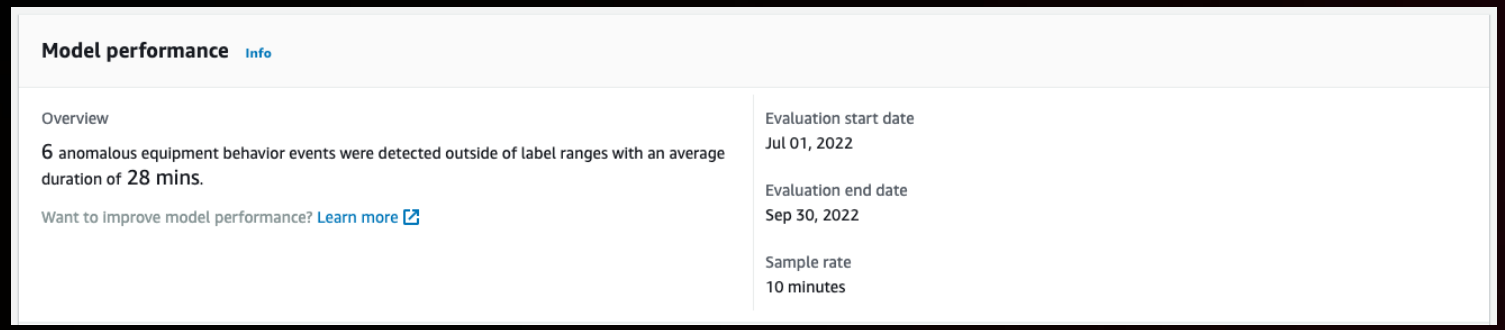
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Training results

ANOMALY EVENTS DETECTED IN HISTORICAL DATA





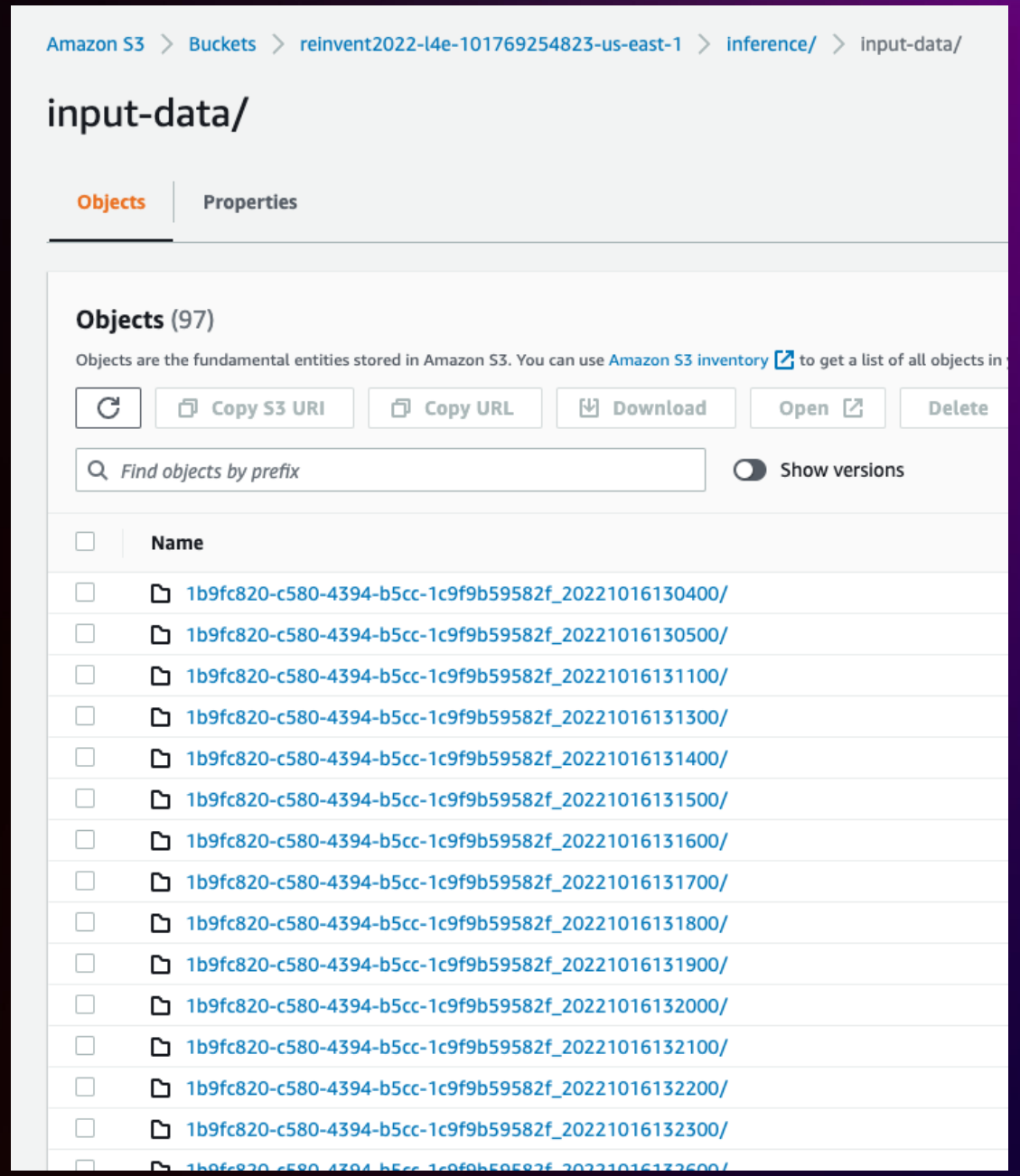
- Events are identified with top contributions to the specified anomaly
- A few detected events means irrelevant changes in health status
- Labels are helpful to detect the cause of the anomaly (in this case a fitness session)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Inference

- Inference is performed using a batch behavior
- Data must be collected from a sensor and loaded to Amazon S3
- Data structure must resemble original training data structure in the form of <asset_id>_<timestamp>
- Anomalies are saved in the destination bucket as json lines as

```
{
    "timestamp": "2022-10-16T15:20:00.000000",
    "prediction": 0,
    "prediction_reason":
    "NO_ANOMALY_DETECTED",
    "anomaly_score": 0.34672
}
```

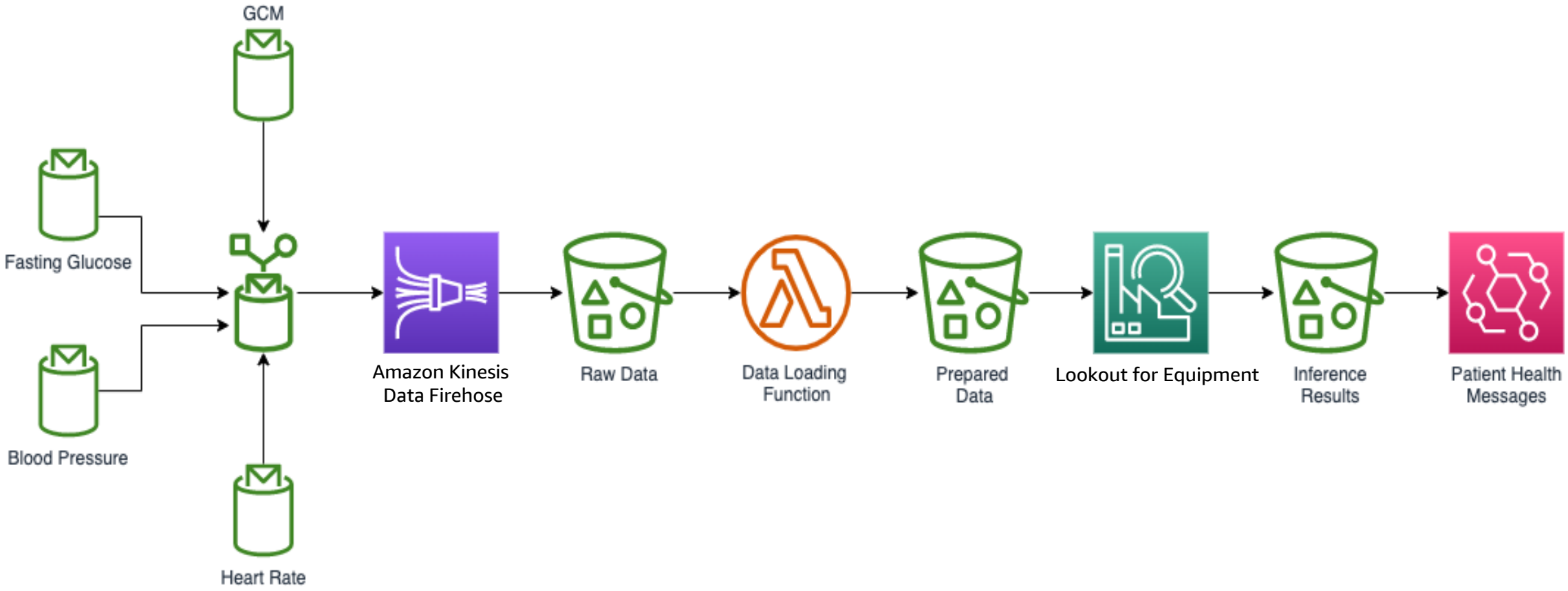


aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Putting it all together

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

System architecture



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Wrap up

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Wrap up

- A lots of data from personal devices
- Time-series is the most common type of data
- Multi-variate analysis to process data (i.e., glucose, blood pressure)
- Insights from real-life behaviors are tracked
- Detect conditions far before health issues arise
- Extend to patients' cohorts monitoring
- Fill data into Amazon Forecast to predict single metric behavior, then send predictions back to Amazon Lookout for Equipment to detect anomalies

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Luca Bianchi
@bianchiluca
about.bianchiluca.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.