

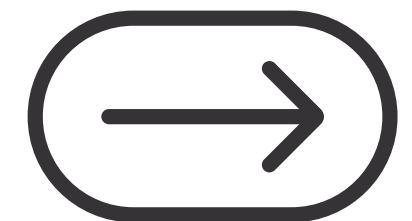
---

# PREDICTING PEDIATRIC APPENDICITIS

*Using AWS Cloud Based Machine Learning*

*DATS 6450 Cloud Computing - Group3*  
*December 1, 2025*

**Presented By: Abirham Getie, Haeyeon Jeong, Yonathan Shimelis**





01

**Project  
Summary**

02

**AWS  
Architecture  
+ Team Roles**

03

**Demo  
Workflow**

04

**Results &  
Takeaways**

# Project Summary

## Goal

- Build an end-to-end AWS ML pipeline to predict pediatric appendicitis.

## Main AWS Services

- **S3**: Shared central storage for datasets, model results, and visuals.
- **EC2 with Shared AMI** (Jupyter): Common environment for preprocessing and model training
- **IAM Policies**: Enable cross-account read/write to S3 for all team members.

## Pipeline

- Data Load → Preprocess → Model → Evaluate → Visualize

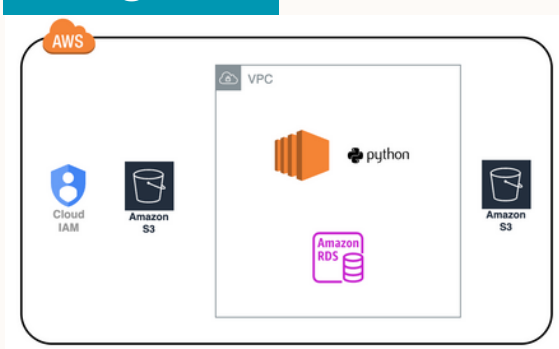
## Data Source

- **Dataset: Regensburg Pediatric Appendicitis (UCI ID 938)**
  - 782 rows | 53 features | Target = Diagnosis (Appendicitis / No)
  - Variables: WBC, CRP, Appendix Diameter, Ultrasound findings

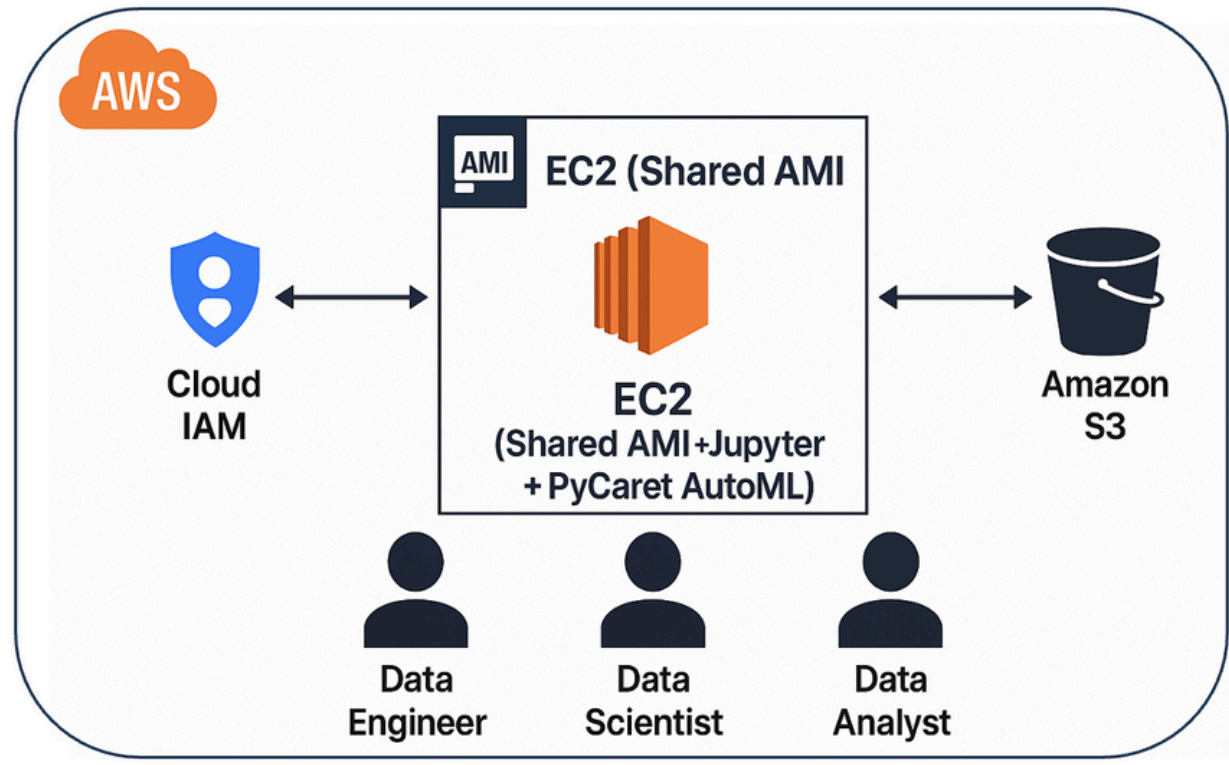
# AWS Architecture + Team Roles

## Services Used

### Original



### Improved



**Used:** S3, EC2 (**Shared AMI** + Jupyter + **PyCaret AutoML**), IAM Policy  
**Removed:** RDS

## Role Summary

Role	Main Tasks
Data Engineer	<ul style="list-style-type: none"><li>Creates shared AMI</li><li>Sets IAM policy</li><li>Uploads raw dataset to S3</li></ul>
Data Scientist	<ul style="list-style-type: none"><li>Read the data from S3 and preprocesses data</li><li>Trains ML models and saves results to S3</li></ul>
Data Analyst	<ul style="list-style-type: none"><li>Reads model results from S3</li><li>Visualizes performance</li><li>Uploads visuals back to S3</li></ul>

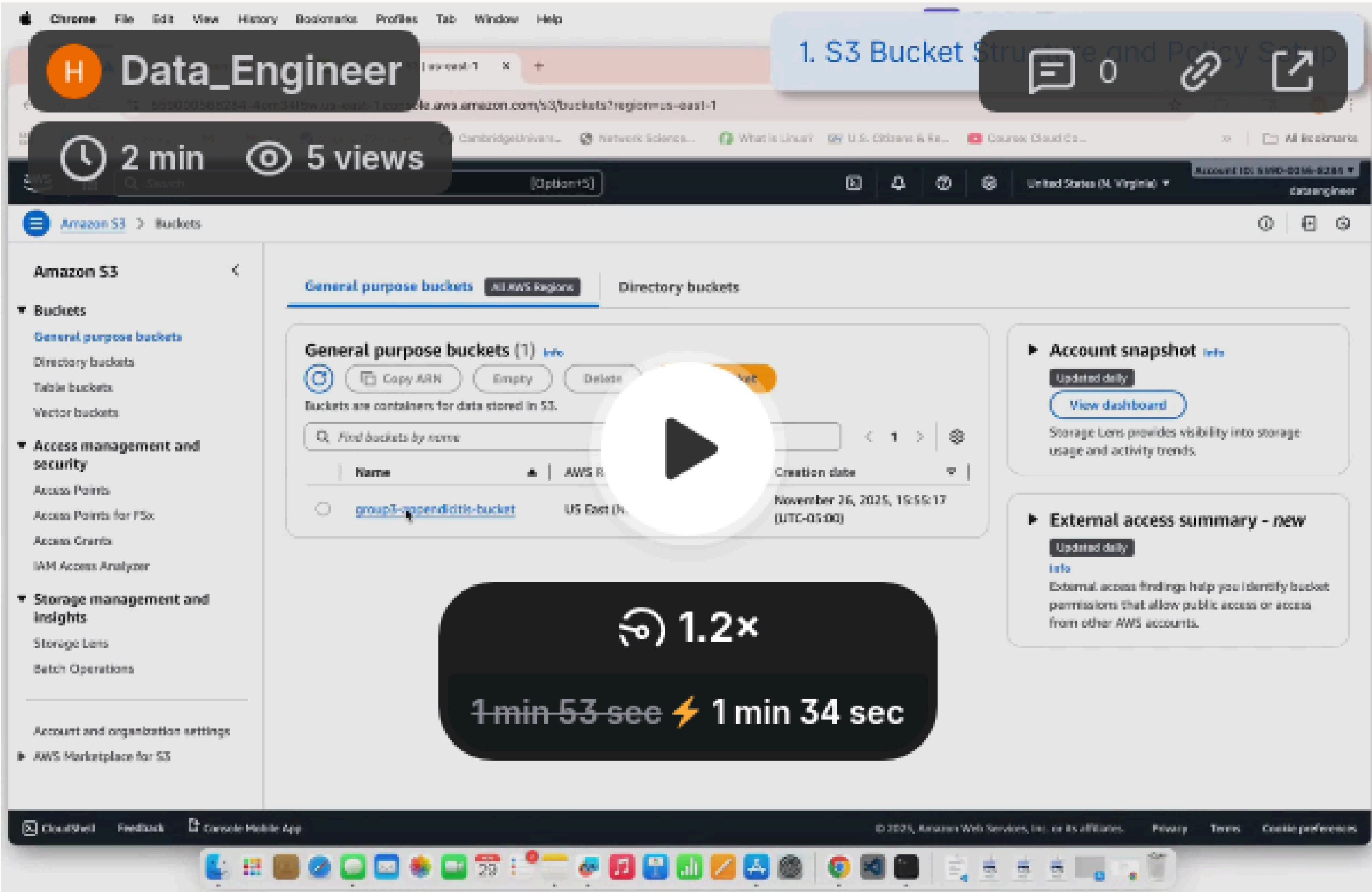
<https://www.loom.com/share/7fae3d4e12274e4289ddfe67eee854a3>

# Demo Workflow

01

## Data Engineer

- Check 3 bucket structure + Policy
- Launch EC2 instance from the shared AMI
- Open Jupyter Notebook and run the code
- Confirm collected dataset uploaded to S3



# Demo Workflow

02

## Data Scientist

- Use Jupyter Notebook to load data from S3
- Train and evaluate ML models on EC2
- Used AutoML to identify best performing model
- Save model results to S3

<https://www.loom.com/share/6609e95ea75b40918c3afa6e61185093>

The screenshot shows the AWS Management Console interface for the 'Data Scientist\_Appendicitis Prediction...' project. The video player overlay indicates a duration of 2 min 38 sec and 1x playback speed. The instance 'group3\_4' is highlighted as 'Running'.

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...
group3-project	i-0a9941c47d11n1a16	Stopped	t3.micro	-	View alarms +	us-east-1f	-	-
olafom2	i-02c19e0c5447d974	Stopped	t3.micro	-	View alarms +	us-east-1f	-	-
group3_4	i-08ae785b70126023b	Running	t3.micro	-	View alarms +	us-east-1f	ec2-3-235-238-65.com...	3-235-238-65

**i-08ae785b70126023b (group3\_4)**

**Instance summary**

Instance ID: i-08ae785b70126023b

IPv4 address: -

Hostname type: -

Public IPv4 address: 3-235-238-65

Instance state: Running

Private IP address: 172.31.74.123

Public DNS: ec2-3-235-238-65.compute-1.amazonaws.com | open address



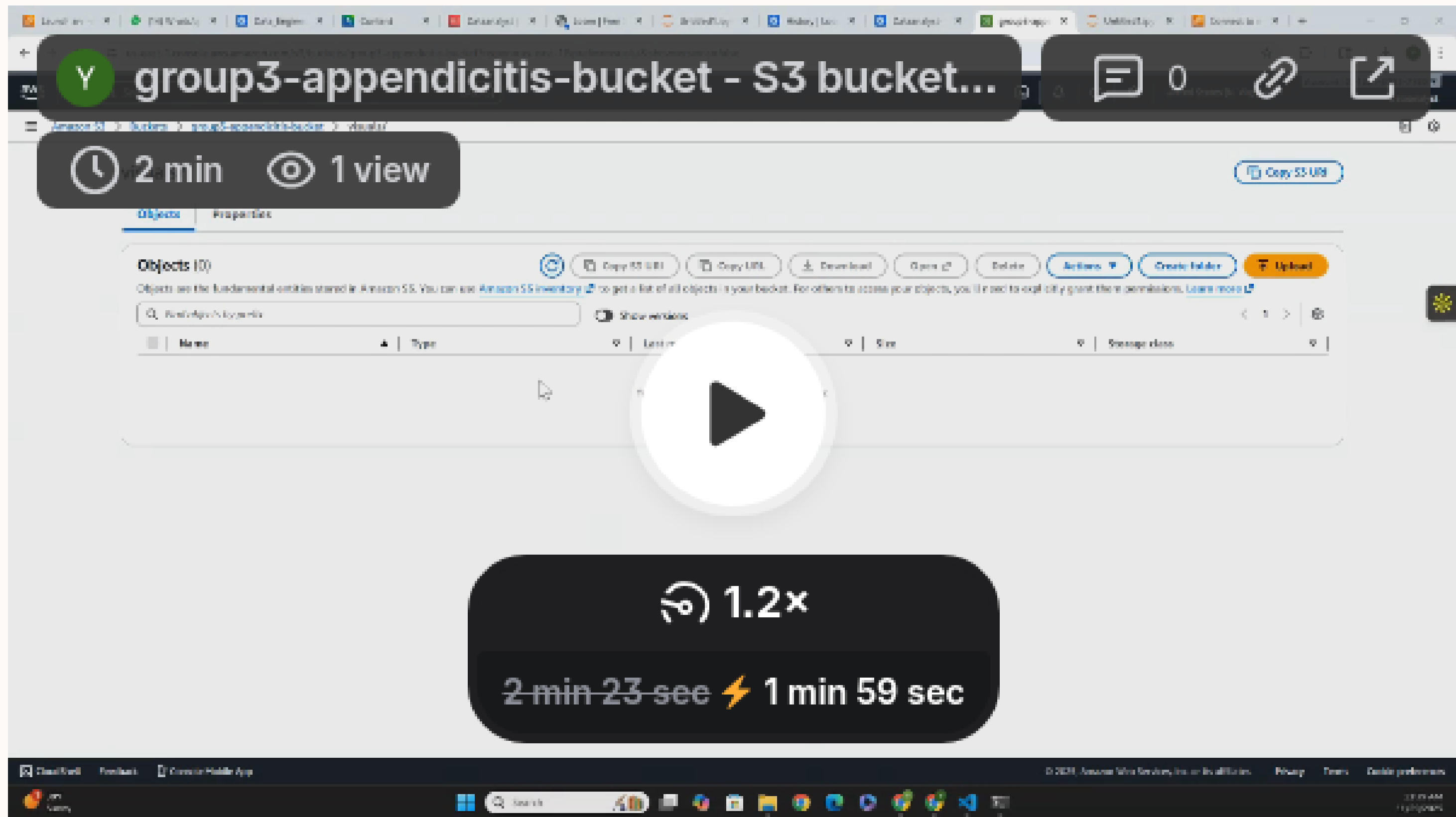
# Demo Workflow

<https://www.loom.com/share/eccb860892134ec197c660359b8c1a60>

03

## Data Analyst

- Access model results csv from S3
- Create performance charts and comparison plots
- Upload visuals to S3



# Results & Takeaways

## Folders in our S3 bucket include:

- data/ → cleaned datasets
- models/ → stored models
- results/ → model performance CSV
- visuals/ → plots and charts

## Outcomes:

- Fully shared AWS workflow across 3 accounts
  - End-to-end automation of data cleaning and upload
  - Collaboration through S3 instead of local files
- 
- Key features for model prediction include appendix diameter, white blood cell count, and C-reactive protein measurements
  - Used AutoML to evaluate a broad range of classification models and the search identified HistogramGradientBoostingClassifier as the best-performing model with a CV F1 Macro of 0.95
- 
- Demonstrated how S3 and EC2 enable team-based machine learning.
  - Achieved secure, reproducible workflow without extra services like RDS.
  - Future step: deploy model for real-time prediction using AWS Lambda.

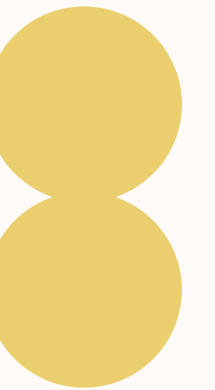
### Outputs

### Modeling takeaways

### Conclusion



# Thank You



Group 3: Abirham Getie, Haeyeon Jeong, Yonathan Shimelis