



# AWS SF Learning Days

## AWS for Epidemiology – Low/No-Code Data Preparation and ML

Yuan Yao (he/him)

Senior Solutions Architect  
AWS Non-Profit Healthcare

# Workshop Details

How do I use advanced statistical models to support data-driven decision-making?

In this workshop, you will explore:

- How to clean and transform data
- Run a machine learning algorithm

All without writing any code

# Epidemiological Data Challenges

## Epidemiology

Study of distribution and determinants of health-related states among specified populations and the application of that study to the control of health problems.

## Data

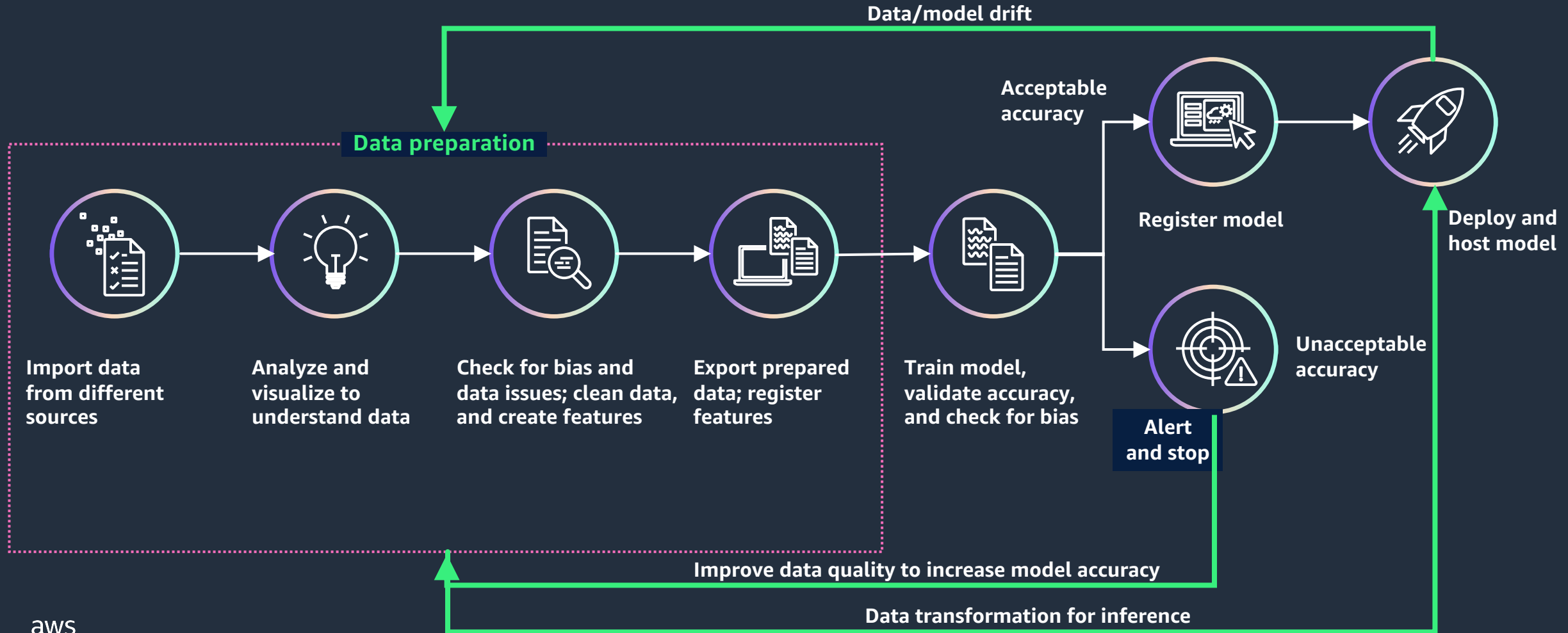
- Variety – data disparity
  - Quality
  - Format
- Volume – data magnitude
  - Scaling automation
- Velocity – data generation
  - Real-time

## Actionable Insights

- ML learning curve
- Model validation
- Cost as a barrier

# Data prep takes 60%–80% of ML project time

THE MOST CRITICAL AND TIME-CONSUMING TASK



# AWS Glue DataBrew

SERVERLESS, NO CODE DATA PREPARATION FOR DATA ANALYSTS AND DATA SCIENTISTS



## Understand data quality

Understand patterns and detect anomalies using profiles



## Clean and normalize data

Over 250 built-in transformations



## Visually map data lineage

Understand steps that the data has been through



## Automate at scale

Save transformations and apply to new data as it comes in

# Challenges Analysts Face in Building ML



## **Analysts lack deep ML expertise, and learning curve is steep**

---

- Need to build understanding for ML concepts across data preparation, model development, and optimization
- Need expertise in choosing the right combination of feature engineering, type of model, and optimization technique
- Learning to write or decipher code is usually needed



## **Business needs explainability and validation from experts**

---

- Analysts prefer to partner with data scientists in order to learn and build trust in the process, but data scientists time is limited and typically devoted to a few key ML projects
- Analysts need to be able to explain ML model predictions to business executives



## **Available no-code ML tools tend to lack transparency and have upfront fees**

---

- Many no-code ML options lack code-level transparency making it difficult to inspect and productionalize models
- The UX for analysts and data scientists tends to be the same, requiring analysts to know the ML concepts and jargon
- Frequently, no-code ML tools come with licensing fees, so experimentation requires upfront investment

# Amazon SageMaker Canvas

**Build ML models and generate accurate predictions — no code required**



**Quickly access and prepare data for Machine Learning**



**Built-in AutoML to build models and generate accurate predictions**

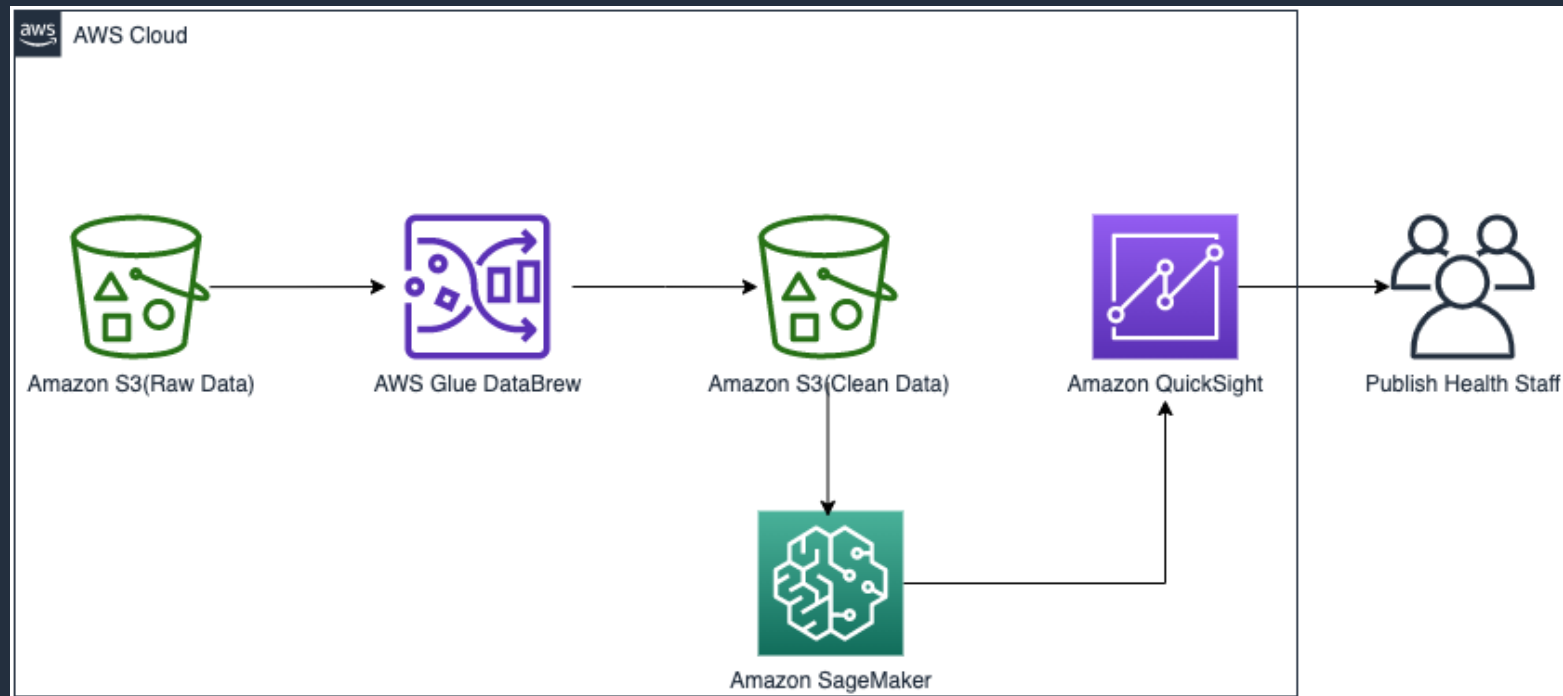


**Share ML models and collaborate with data science teams**



**Usage-based pricing to avoid licensing fees and reduce TCO**

# Workshop Architecture



# Getting started with this workshop



You have access to an AWS account with any optional pre-provisioned infrastructure and IAM policies needed to complete this workshop.



The AWS account is only available for the duration of this workshop. **You will lose access to the account once the workshop is complete.**



Any optional pre-provisioned infrastructure is deployed to a specific AWS Region. Make sure that you are working in this Region; other Regions are blocked.



Review the terms and conditions of the event. **Do not upload any personal or confidential information to the account.**

# [https://s12d.com/UPDATE\\_ME](https://s12d.com/UPDATE_ME)

**QR  
Code  
Here**

# Please Provide Your Feedback

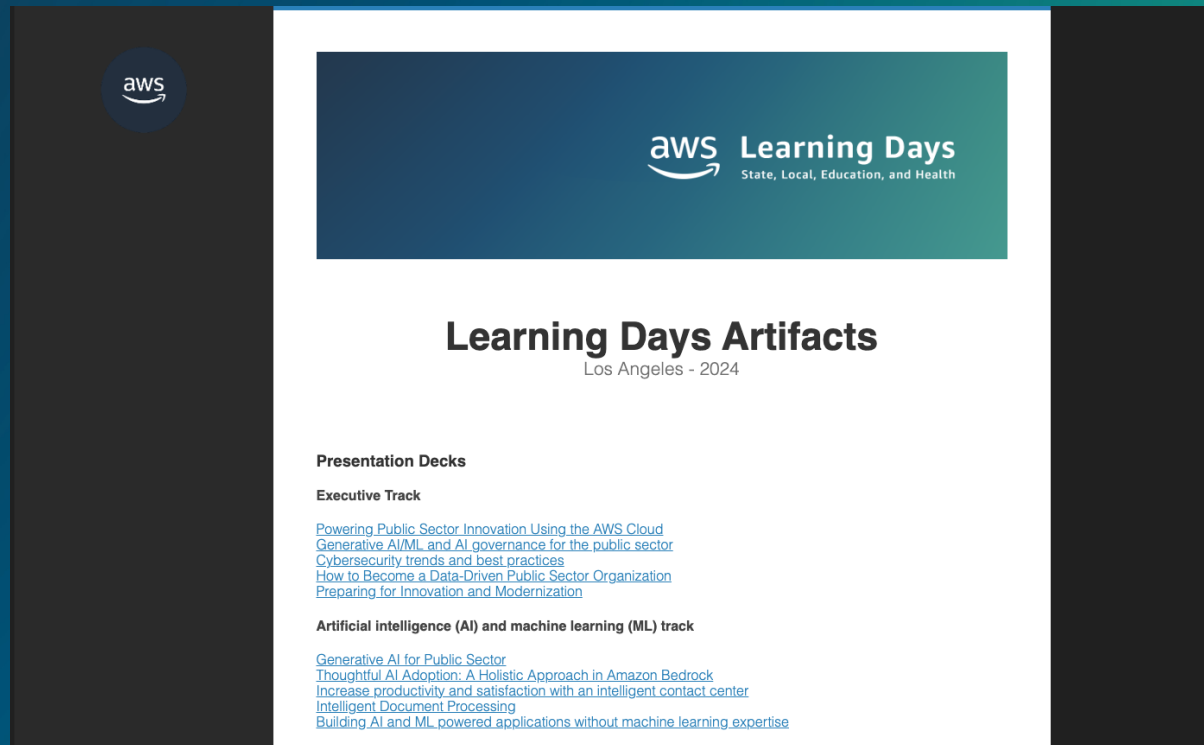


Step 1: Select **Healthcare Track**

Step 2: Select **AWS for Epidemiology – Low/No-Code Data Preparation and ML**

# Learning Day Content

<https://sanfrancisco2024.awslearningday.com/>





# Thank you!

Yuan Yao

[yno@amazon.com](mailto:yno@amazon.com)