ANITA B.ORG 2023
GRACE HOPPER CELEBRATION

THE WAY FORWARD

Lead with Data

Using Goldman Sachs‘ Open-Sourced Legend Platform to Unlock Sustainability Insights

ANITA B.ORG 2023 GRACE HOPPER CELEBRATION

THE WAY FORWARD

Wazila Shikari

Vice President, Goldman Sachs

Leading Data Strategy for Liquidity Management

"Share the excitement and transformative potential of data and data modeling with fellow women engineers!"

Beeke-Marie Nelke

Vice President, Goldman Sachs

Leading Global Product Team for Legend

"Show the power of Legend and make more women excited about data (it's really cool!)"

Garima Swaroop

Vice President, Goldman Sachs

Leading Product Strategy for Liquidity Management

“Want to bring more women engineers into the world of data, design, and product. And it starts here!”

Manjula Nagineni

Senior Solutions Architect, AWS

Leading adoption of AWS GenAI & Data Analytics technologies in Financial Services Industry

“Strong believer in the significance of achieving equal representation of Women in the Tech World"

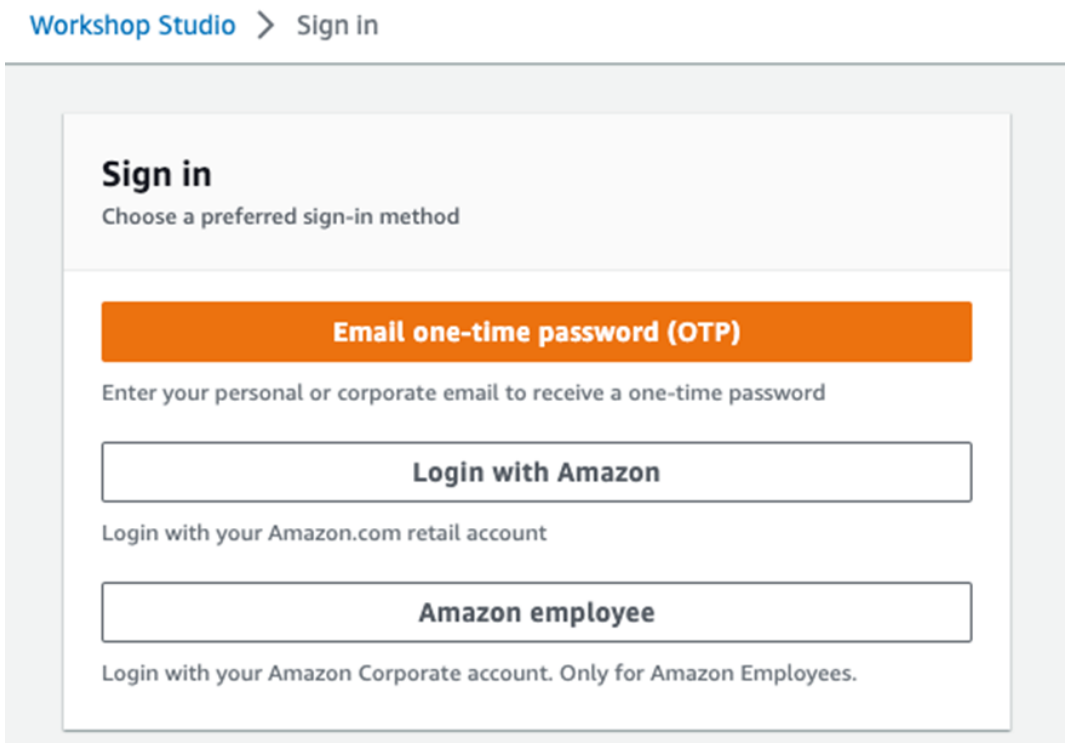
ANITA B.ORG 2023
GRACE HOPPER CELEBRATION
THE WAY FORWARD

AGENDA

1. Legend Set Up
2. Introduction to Legend
3. Lab Session Goals
4. Data Sets
5. Data Exploration
6. Data Modeling Basics
7. Data Queries
8. Data Visualization

LEGEND // Let's get you set up

1. Get access to your Workshop Studio AWS Account
2. Access the console of your AWS account
3. Set up Legend Workbench
4. Launch Legend



Go here → https://github.com/finos/legend/wiki/GHC-Cheat-Sheet

LEGEND

// One place to access ALL the data!

Data Engineer Workbench

Build Data Models, Connect to all Data Stores, Publish APIs in seconds

Self-service Data Queries

Build data queries, drag & drop style, no SQL needed!

Built-in Data Quality

Data Validations, Cardinality, Associations & more

Intuitive Interface

Build data models, diagrams & APIs with a few button clicks

Simple Integrations

Build model artifacts, generate other coding languages with a button click (e.g. JSON, Java, Python)

Streamlined Analytics

Use Legend data models in Business Intelligence & Developer tools

Lab Session Goals

Explore Data

Explore data models and data to capture values of interest.

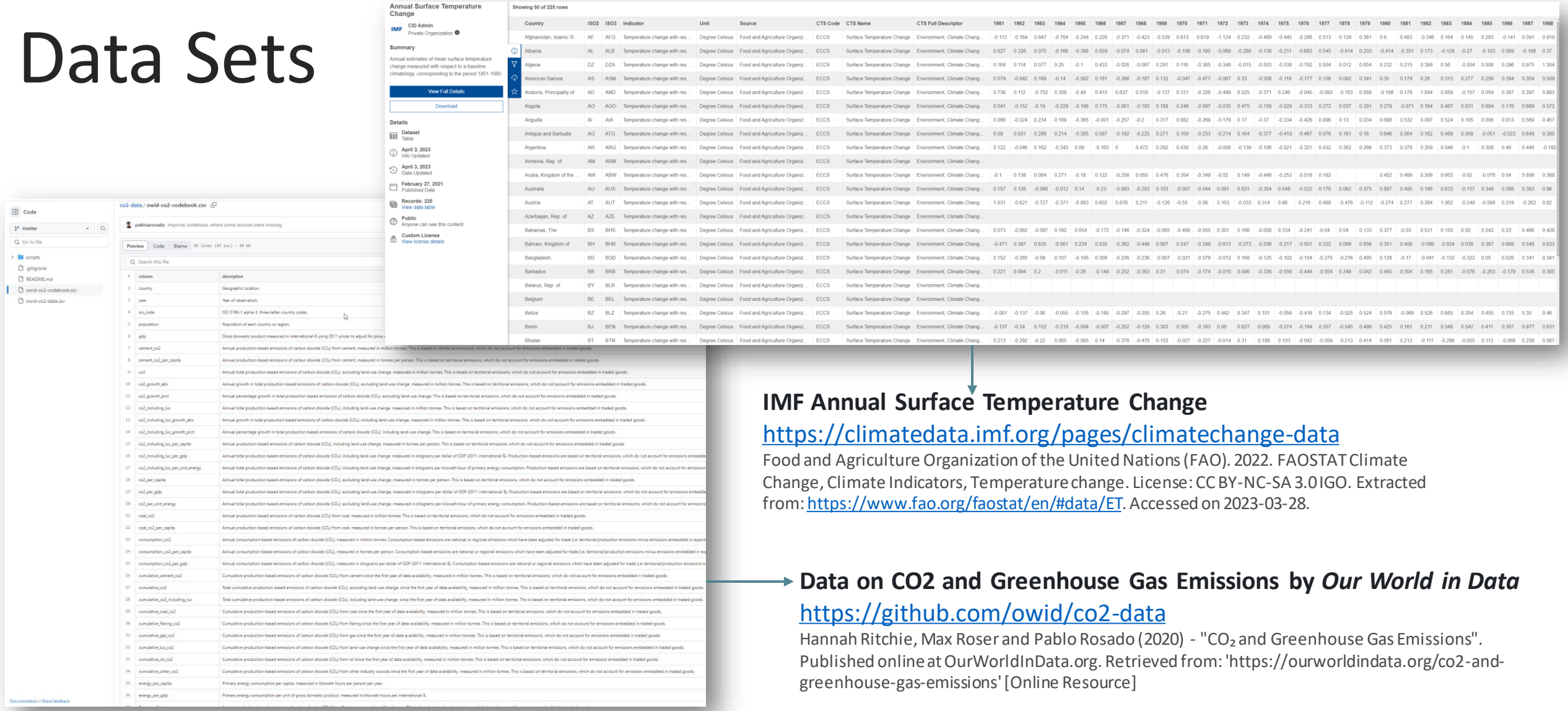
Build Data Relationships

Formalize data relationships using joins to unlock meaningful insights.

Generate Data Insights

Execute queries to find insights, trends, and interesting patterns in data.

Data Sets



IMF Annual Surface Temperature Change

https://climatedata.imf.org/pages/climatechange-data

Food and Agriculture Organization of the United Nations (FAO). 2022. FAOSTAT Climate Change, Climate Indicators, Temperature change. License: CC BY-NC-SA 3.0 IGO. Extracted from: https://www.fao.org/faostat/en/#data/ET. Accessed on 2023-03-28.

Data on CO2 and Greenhouse Gas Emissions by *Our World in Data*

https://github.com/owid/co2-data

Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "CO_2 and Greenhouse Gas Emissions". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/co2-and-greenhouse-gas-emissions' [Online Resource]

Explore Data

Let's take a quick look at our emissions dataset!

Let's build data relationships!

// Expand data models

// Create consistency

// Unlock insights

Go here for detailed instruction:
https://github.com/finos/legend/wiki/GHC-Cheat-Sheet#data-modelling-guide

Follow along:

1. Define the schema for surface temperature
2. Build joins between the datasets

Try it yourself:

3. Create the new surface temperature class
4. Map the data model to datastore
5. Visualize the expanded model in a diagram

Generate Data Insights

Part 1:

Get the number of distinct countries that have production-based carbon dioxide emission statistics

Part 2:

Get the list of distinct countries that have production-based carbon dioxide emission statistics

***Hint**: There are countries with empty values*

Generate Data Insights

Query 2:

Get the top 10 countries with highest carbon dioxide emissions in 2021

***Hint**: There are countries with empty values*

Generate Data Insights

Query 3:

Get top 10 countries by average emissions across all years available in the data set.

***Hint**: There are countries with empty values*

Generate Data Insights

Query 4:

List all countries and their surface temperature change in 2020

***Hint**: There are countries with empty values*

Generate Data Insights

Query 5:

Get top 10 countries with largest surface temperature change between 2016 and 2021

***Hint**: There are countries with empty values*

Generate Data Insights

Query 6:

Query and sort countries by average (1)surface temperature & then by average (2)carbon dioxide across all years since 2001

***Hint**: There are countries with empty values*

1. Take data relationships + insights all the way to the front end

2. Use the right data viz to showcase the story

3. Highlight the insights but plan for extensibility

ANITA B.ORG 2023
GRACE HOPPER CELEBRATION

THANK YOU