

A top-down view of a desk with various financial trading items. A silver laptop is at the top. Below it are several papers: one with a candlestick chart and a green dot, another with a line chart and a green dot, and a third with a candlestick chart and a yellow dot. A black pen and a black marker are on the left. A pair of glasses is on the right. A magnifying glass is over a candlestick chart at the bottom. A white box with the text 'MARGINS' and 'Case Study - Marcelino Mayorga Quesada' is in the center.

MARGINS

Case Study - Marcelino Mayorga Quesada



AGENDA

1

About me

2

Case Study

Background, Requirements, Resources, Objectives

3

Data

Profiling, Pipeline, Measure development

4

Dashboard & Analysis

Widgets, observations and action items

5

Next Steps

Areas of improvements, and expanded pipeline

6

Q & A



MARCELINO MAYORGA QUESADA

- 15 Years of Experience in Software Development executing multiple technical and delivery management roles.
- Clients: Blackstone, Cambridge Associates, EA Sports, Citibank, CCI Global Channel Management, and BAC Credomatic.
- Passionate about Swimming , Artificial Intelligence , Music , and Video games .

Section 2

CASE STUDY

CONTEXT

Background

- Databricks deploys its products in multiple regions.
- Each region generates revenue and has dedicated cloud resource footprint.
- We measure the health of each region / product line by tracking the margins.
- $Margin = (Revenue - Cost) / (Cost)$.

Resources

- CSV File: Region, Product Line, Month, Cost, Revenue.
- Databricks Community Edition (Python, dbutils, SQL Functions, DataFrames and Visualization).

Requirements: Dashboard

- **Global** month-over-month margin trend.
- **Per region** month-over-month margin trend.
- **Per product line** month-over-month margin trend.
- **Per region, per product line** month-over-month margin trend.
- **Best & worst** performing regions & product lines.

Objective: Business Optimization

1. Understand regions and product lines **current health state**.
2. Improve regional and product line margins.

Section 3

DATA

DATA PROFILING

- **Rows:** 32
- **Columns:** 5
- **Date Range:** 8 Months (Jan - Aug 2023)
- **Regions:** us-west1 and us-east1
- **Product Lines:** sql and jobs
- **Complete Data**

region ▲	product_line ▲	month ▲	cost_m_usd ▲	revenue_m_usd ▲
us-west1	sql	2023-01-01	0.8	0.6
us-west1	jobs	2023-02-01	1.1	1.5
us-west1	jobs	2023-06-01	1.1	1.5
us-west1	jobs	2023-07-01	1.1	1.5
us-east1	sql	2023-01-01	1.5	1.4
us-east1	sql	2023-02-01	1.5	1.4

DATA PIPELINE



1 - Import Data

Forensic Data into Databricks Cluster

Results saved into SQL Table



2 - Dashboard for Actuals

Data Preprocessing

Metric Development (SQL Functions)

Visualization for dashboard

[Notebook File](#)

METRIC DEVELOPMENT

Margin and Month Over Month Margin (MoM)

pyspark.sql.functions.expr: used for margin calculation, allows SQL-like syntax over the data frame.

```
1 df = df.withColumn("margin", expr("(revenue_m_usd - cost_m_usd) / cost_m_usd"))
```

	region	product_line	month	cost_m_usd	revenue_m_usd	margin
1	us-west1	sql	2023-02-01	0.8	1	0.24999999999999994
2	us-west1	sql	2023-03-01	0.8	1.2	0.49999999999999999
3	us-west1	sql	2023-04-01	0.8	1.3	0.625
4	us-west1	sql	2023-05-01	0.8	1.3	0.625
5	us-west1	sql	2023-08-01	0.8	1.5	0.8749999999999999

pyspark.sql.window: used for **MoM** calculations, allows partitioning and ordering.

pyspark.sql.functions.lag: operates over capture previous values rows within the defined window.

```
1 window_spec = Window.orderBy("month")
2 df = df.withColumn("margin_mom_change", col("margin") - lag("margin").over(window_spec))
```

	region	product_line	month	cost_m_usd	revenue_m_usd	margin	margin_mom_change
1	us-west1	sql	2023-01-01	0.8	0.6	-0.25000000000000006	null
2	us-west1	jobs	2023-01-01	1.1	1.1	0	0.25000000000000006
3	us-east1	sql	2023-01-01	1.5	1.4	-0.066666666666666672	-0.066666666666666672
4	us-east1	jobs	2023-01-01	2.1	3	0.4285714285714285	0.4952380952380952
5	us-west1	sql	2023-02-01	0.8	1	0.24999999999999994	-0.17857142857142855

Section 4

DASHBOARD & ANALYSIS

[Link](#)

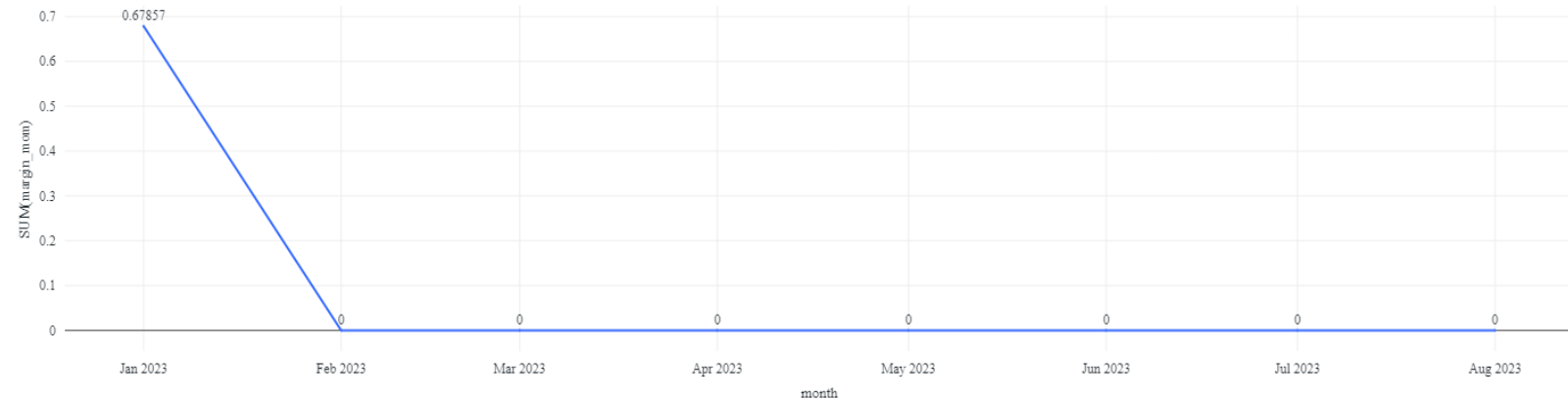
1 - GLOBAL MONTH-OVER-MONTH MARGIN TREND

Observations

- January **Peaked** with 0.25 margin MoM.
- Flat **zero** margin from February through August.

Actions

- Understand what drives **January's healthy** margin.
- Understand **revenue and cost** that lead to zero margin.
- Analyze differences in avg vs median for outliers. (eg: median shows August's uptrend)



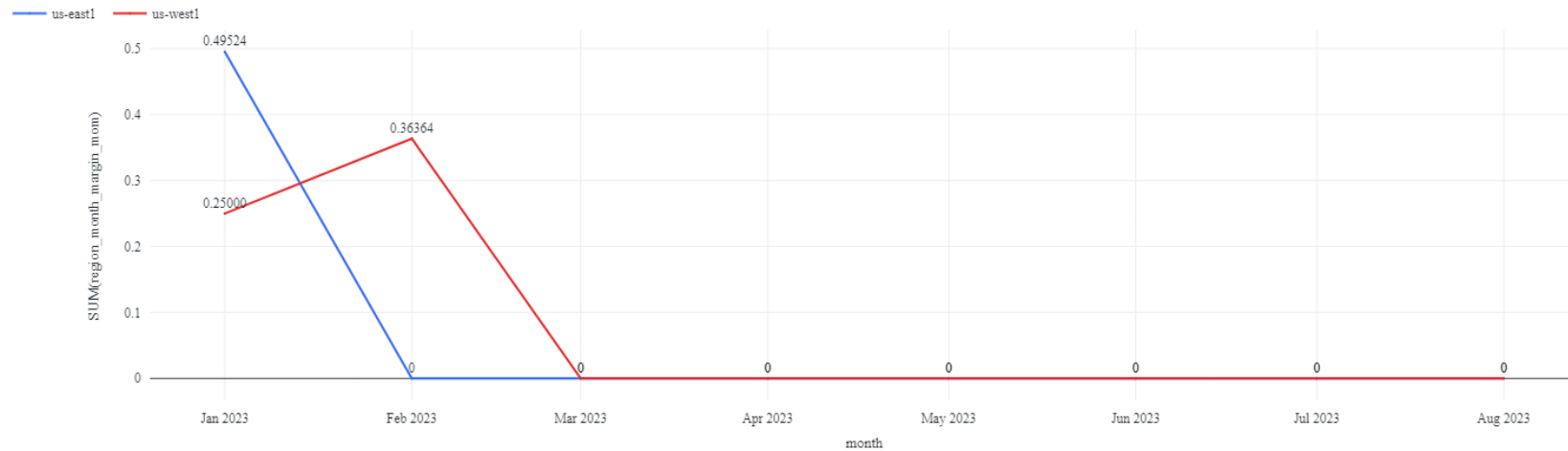
2 - PER REGION MONTH-OVER-MONTH MARGIN TREND

Observations

- **us-east1 peaked positive margin**(blue) in January dropped to zero by February and remained flat.
- **us-west1 dropped to zero**(red) progressively over 2 months and remained flat.

Actions

- Understand what drives both **regions' healthier margins** in the first quarter.
- Understand **zero margins from March-August**.



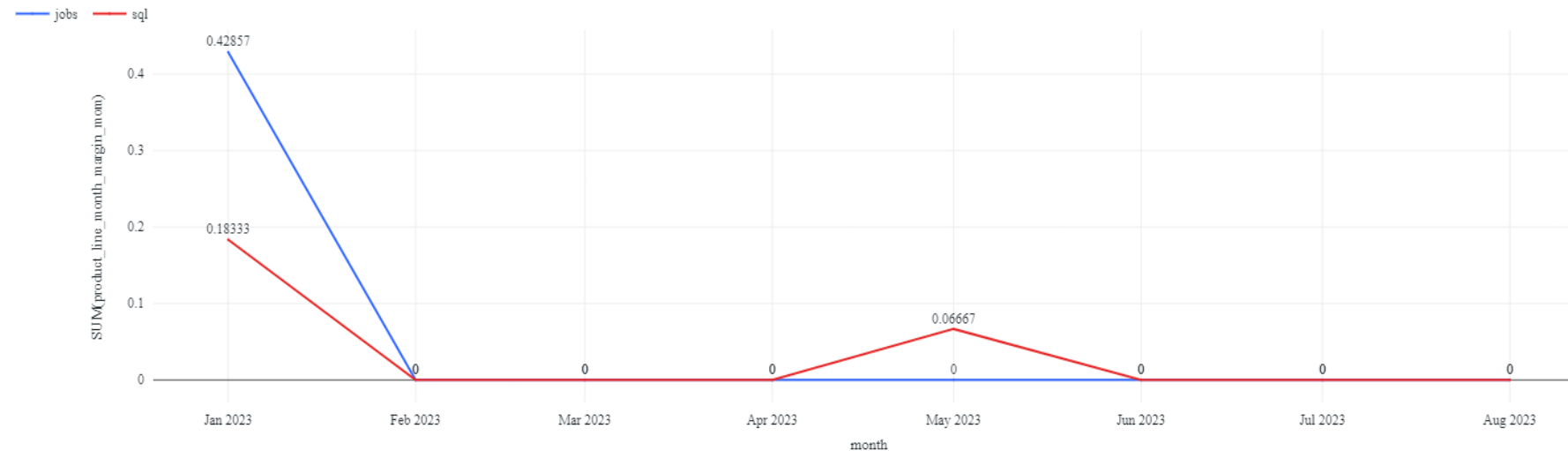
3 - PER PRODUCT LINE MONTH-OVER-MONTH MARGIN TREND

Observations

- **Positive Margins** in January for both product lines
- **'Jobs'**(blue) product line is **slightly healthier** than 'SQL'(red).
- Both product lines **dropped to zero** in February.

Actions

- Understand **zero margins from February-August**.
- Review **April-June potential anomaly/trend**.



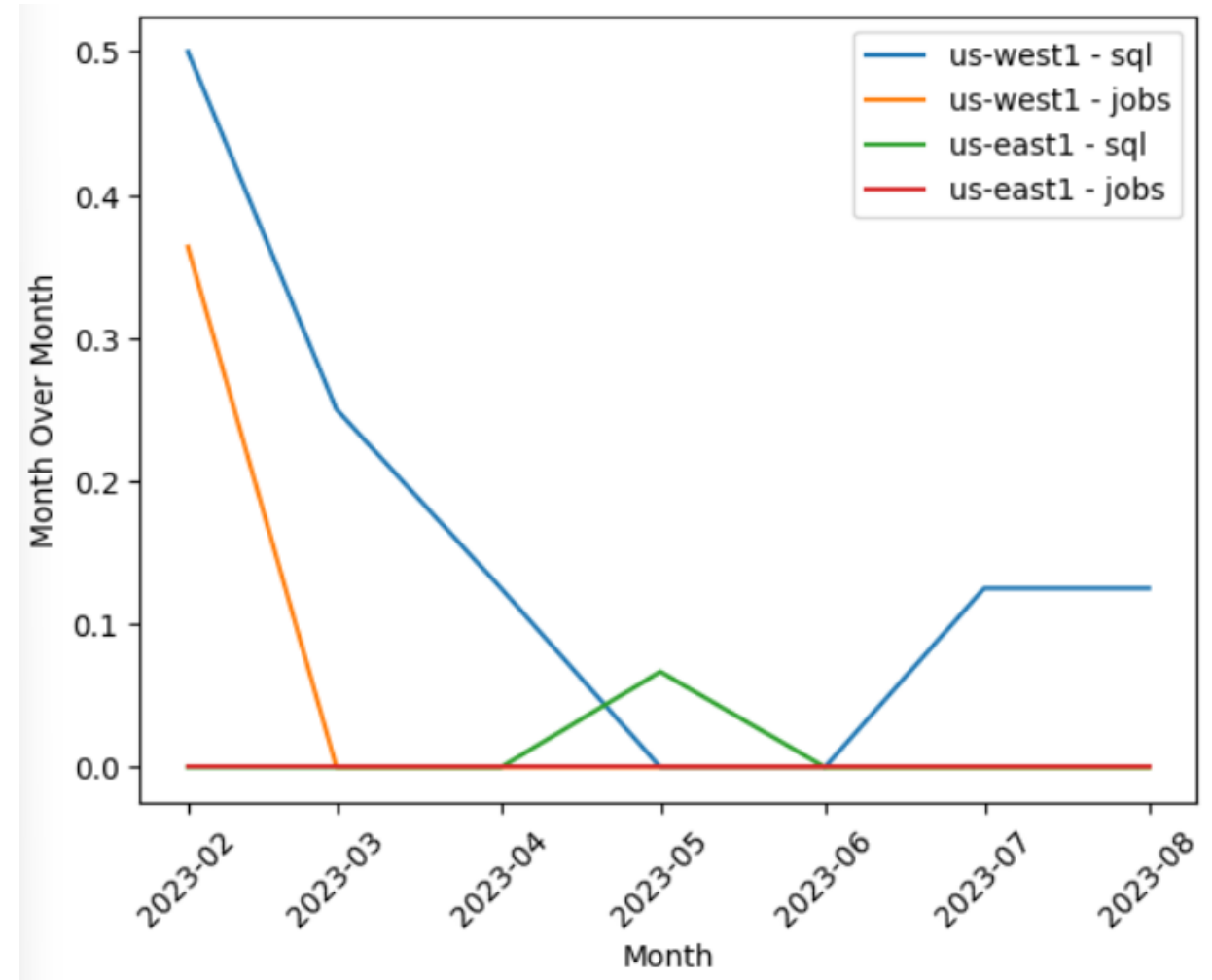
4 - PER REGION, PER PRODUCT LINE MONTH-OVER-MONTH MARGIN TREND

Observations

- **us-west1:** Both product lines had the **healthiest margin** trend.
 - sql(blue): Leads the **uptick** in the global MoM margin.
 - Jobs(yellow): Drops to **zero margin** in March, remains flat.
- **us-east1:** close to zero margin throughout the year.
 - sql(green): Shows **positive margins in Q2** (aligns with Product Line MoM chart).
 - jobs(red): **remained zero** margins

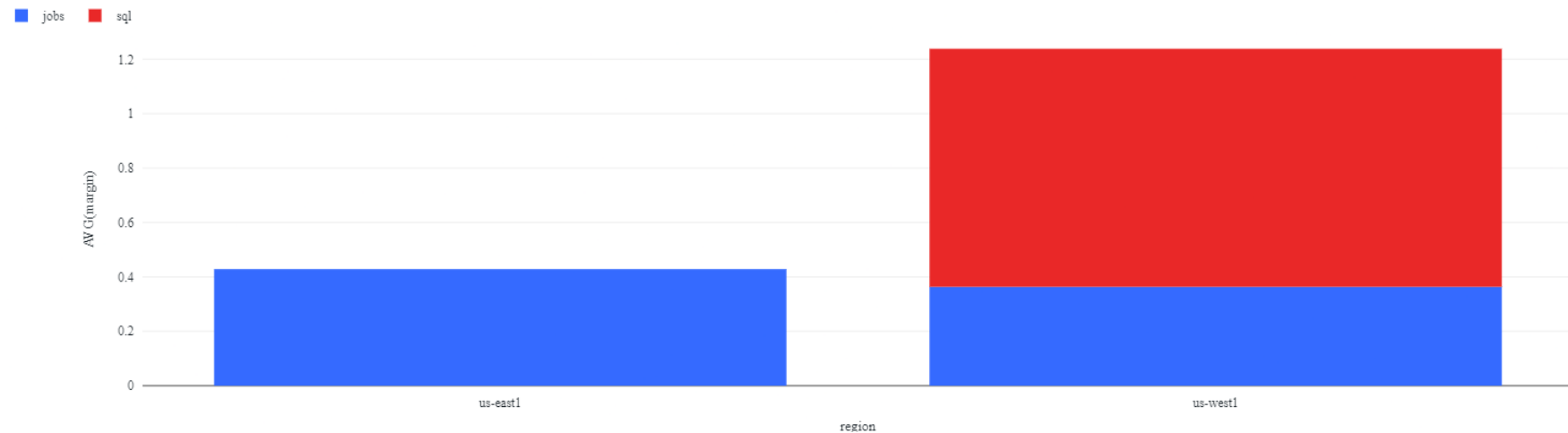
Actions

- Identify **trends, seasonality, special events & anomalies**



5 - BEST & WORST PERFORMING REGIONS & PRODUCT LINES

- **Best Region / Product Line:**
sql(red) in us-west1
- **Worst Region / Product Line:**
sql(red) in us-east1 (zero)
- *Based on margin



Section 5

NEXT STEPS

AREAS OF IMPROVEMENTS

- Set an **iterative automated data pipeline** for analysis and continuous forecasting.
- **Incorporate DatabricksUnit(DBU) and prices**, energy consumption, utilization, and distribution for more insights.
- Revisit revenue and cost **calculation**.
- Identify trends, seasonality, special events & anomalies. Leverage tools such as **Pycaret, TSFresh, or Darts**.

EXPANDED DATA PIPELINE



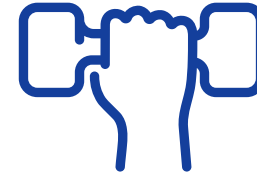
1 - Import Data

Forensic Data and define
Metrics



2 - Dashboard for Actuals

Data Preprocessing
Widget Development
[Notebook File](#)



3 - Model Training

Data Preprocessing
Model Selection, Training and
Fine-Tuning
Validation
[Notebook File](#)



4 - Forecast Margins

Run Forecasts and compare with
actuals
Consume Insights and handling
for Adaptability

Section 6

Q & A