

# Proactive Cost Management

Detecting Anomalies in Logs  
with Time Series Analysis

**CONF42**



Cloud AppMod Engineer



Love SRE, DevOps, Chaos  
Engineering, Observability,  
Reading, Writing and Teaching.



@yurnino



# How many of you?

Devops teams to their  
finance team after  
misconfiguring their  
cloud computing needs



# I would like to tell you a sad story ...

## Daily Activities

I was oncall putting out fires in production when the sound of nuts is scary.



## Where are the Logs?

I was investigating the root cause of the incident ... OMG!! There are no logs!



## This will not happen to me again:

I will activate all available logs admin activity, data access, and system event.



## What happened?

\$20K billing increased by 700%!!!  
... OMG the reason is Cloud Logging.



# The Hidden Costs of Anomaly-Only Detection & Response Systems

By D. Mark Durrett





# What is the cost of Inaction?



- **Downtime:** Lost revenue, customer churn, reputational damage.
- **Inefficient Resource Use:** Cloud bills exploding, wasted infrastructure.
- **Security Breaches:** Massive financial penalties, legal costs, irreparable harm.
- **Wasted Engineering Time:** Hours spent troubleshooting reactive problems.

# AGENDA

Topics will be covered

-  Cost Management Challenges
-  Logs can help you ... but ...
-  Machine Learning Techniques
-  Deeping on Time Series
-  Use Cases
-  Q & A

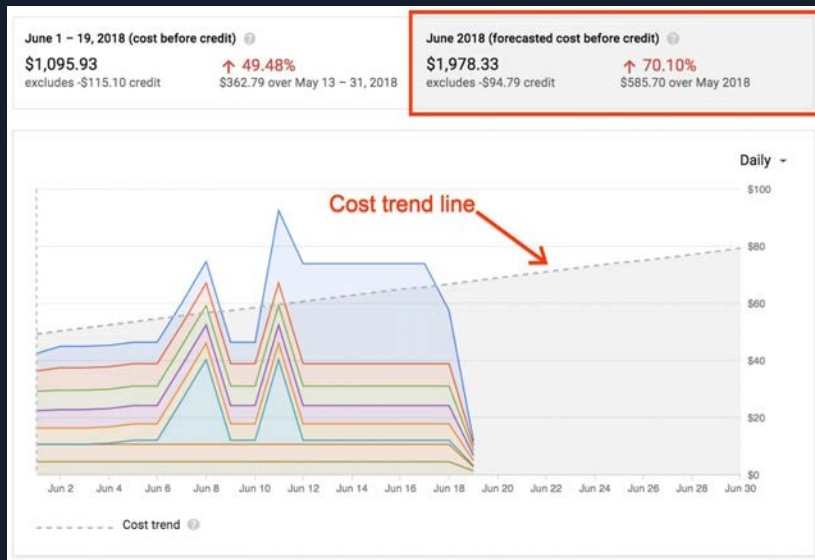
If the **problem** was the **logs** ... the **solution**  
should be in the **logs** also ...

Apply **proactive cloud cost** is useful here  
which involves continuously monitoring,  
analyzing, and optimizing spending on cloud  
resources ...



**Proactive Cost in Cloud** shifts the focus from "what did we spend?" to "how can we spend smarter and more efficiently from the outset? It's a critical component to maximize business value from cloud investments while keeping costs under control.

# Proactive Cost in Cloud



- Anticipating and Preventing Issues
- Continuous Optimization
- Predictive Analytics
- Establishing Budgets and Alerts
- Leveraging Cloud Provider Tools and Third-Party Solutions
- Visibility and Monitoring

# Visibility and Monitoring



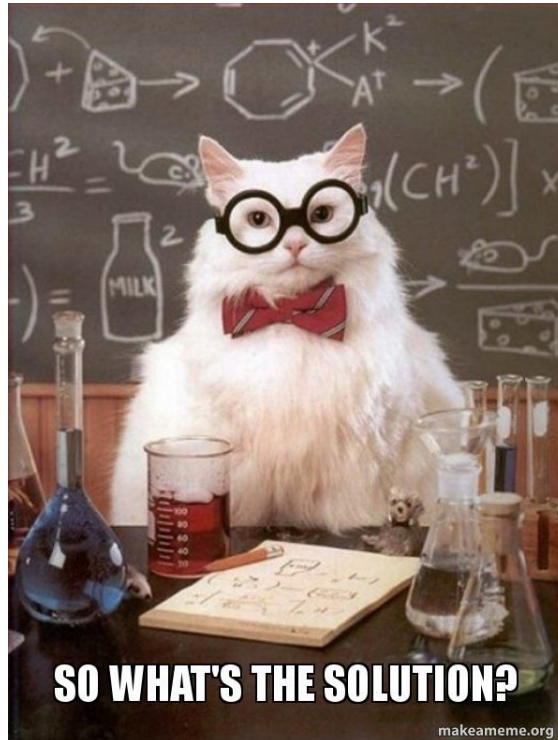
## Analyzing logs in **cloud** ...

represent several challenges compared to traditional environments, primarily due to the distributed, dynamic, and often ephemeral nature of cloud infrastructure.

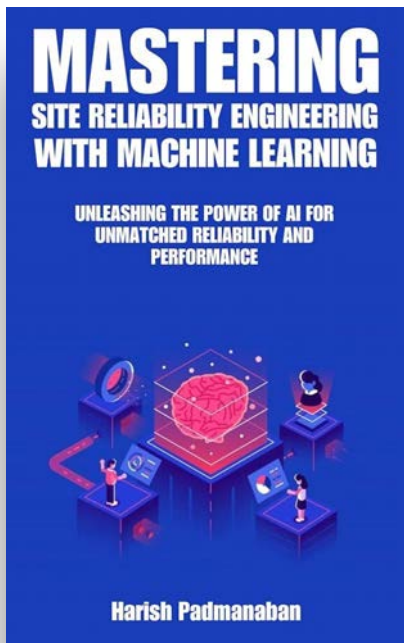


If the **problem** was the **logs** ... the **solution** should be in the **logs** also ...

# How many of you?

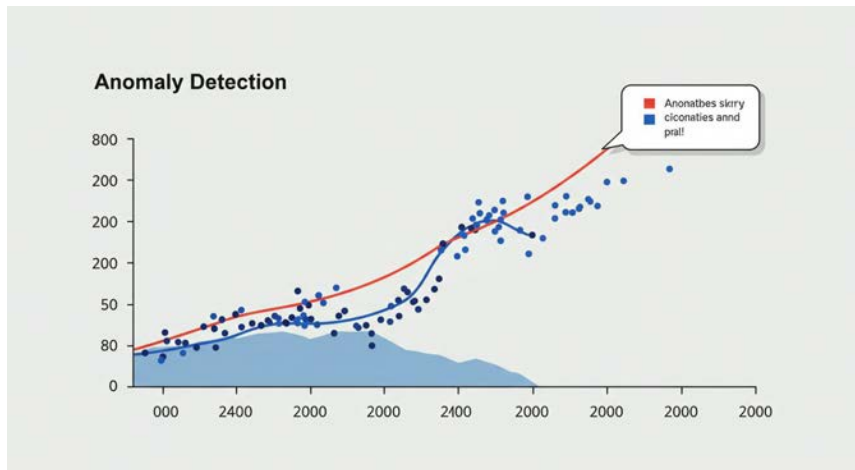


# The solution ...



**SRE Anomaly Detection** uses a combination of sophisticated **Machine Learning Techniques and statistical methodologies**. Statistical techniques that rely on departures from past data or pre-established criteria to find anomalies.

# What is Anomaly Detection?



Identifying patterns that significantly **deviate from expected behavior**.

Finding the "normal abnormal" – the subtle hints something's wrong.

# Machine Learning Techniques

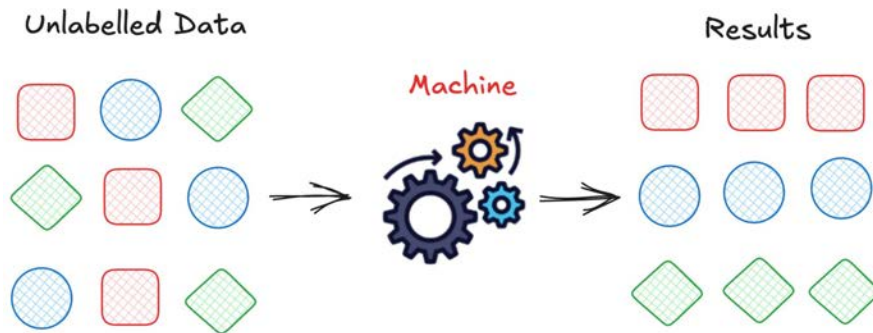
For Anomaly Detection



# Unsupervised Learning Algorithms

Since unsupervised learning does not require labelled data, it is especially well-suited for anomaly detection applications.

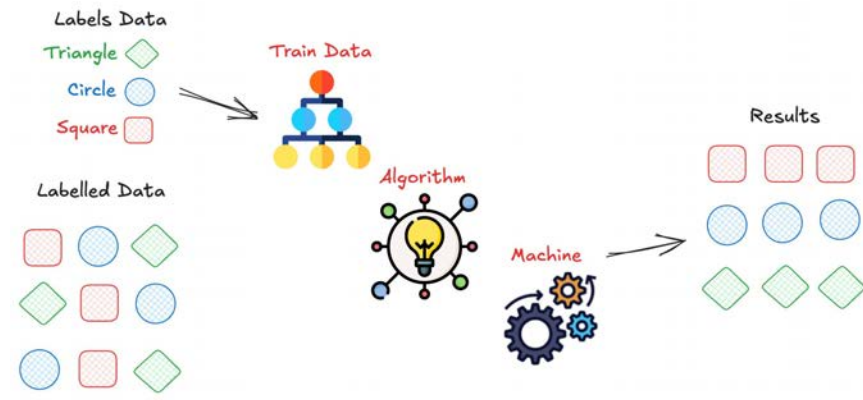
- Clustering Algorithms
- Autoencoders



# Supervised Learning Algorithms

They can be used when historical data with labelled anomalies is available, albeit they are less frequent because they require labelled anomaly data.

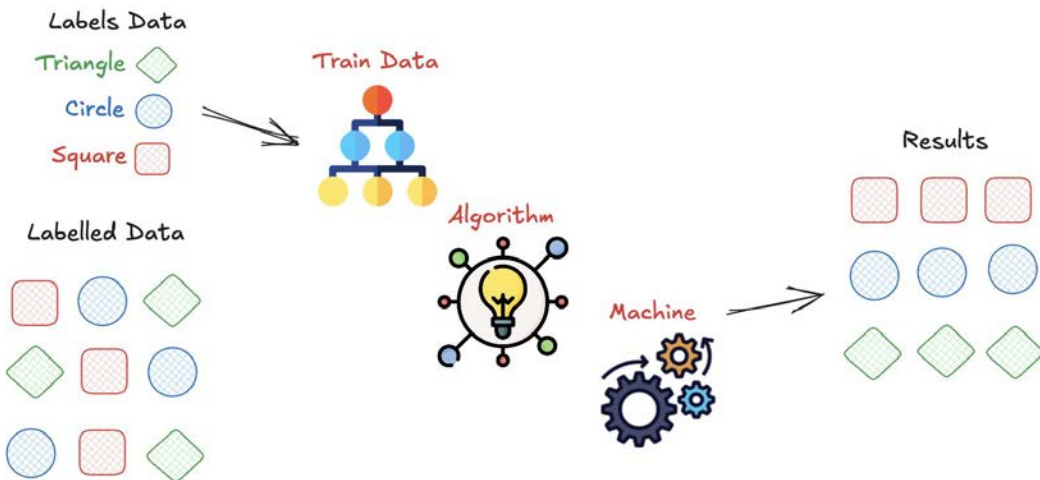
- Classification Algorithms



# Semi Supervised Learning Algorithms

With this method, which combines elements of supervised and unsupervised learning, anomalies.

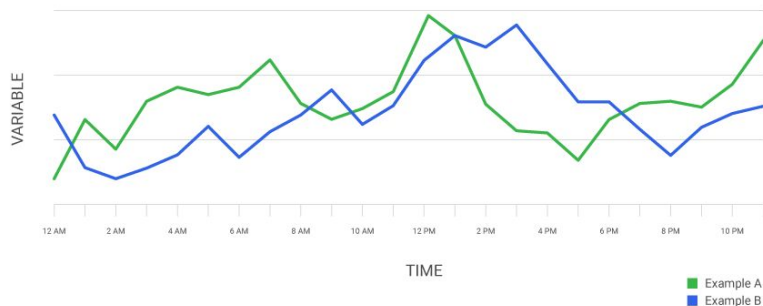
- Isolation Forests



# Time Series

Time-series analysis techniques are essential for identifying abnormalities over time since many SRE measures have a temporal component.

- Seasonal Decomposition

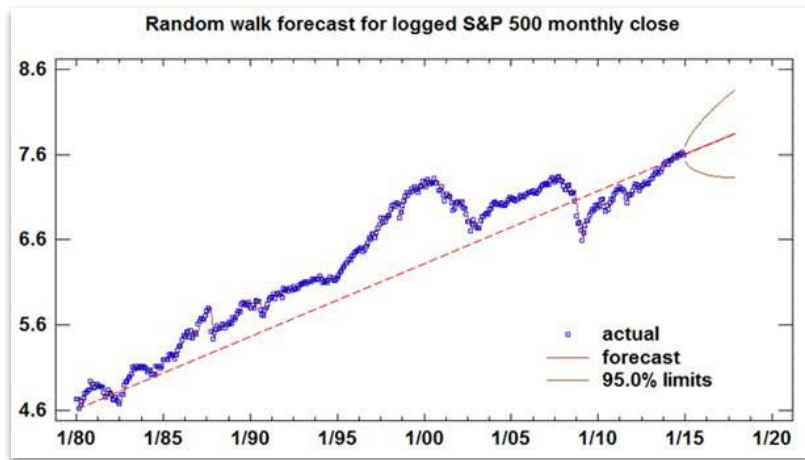


# How Google manage

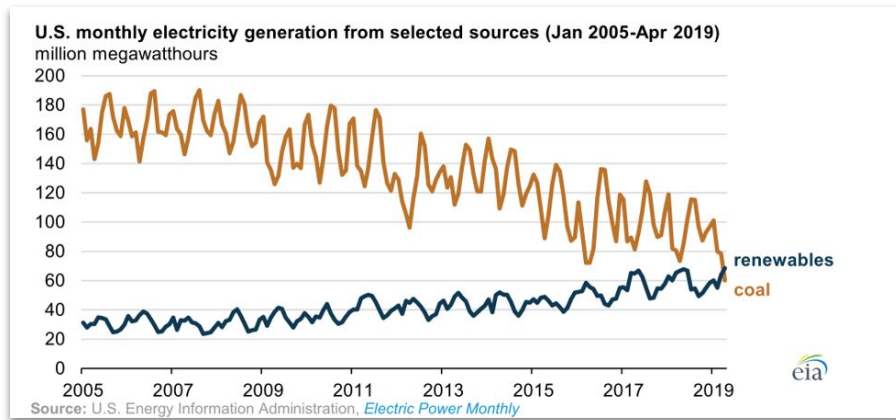
## Time Series

# Most time-series are non-stationary

Financial time series as “random walk with drift”



Energy production influences by wind & solar supply



# Dragon Kings or Black Swans?



Dragon King Theory

Theory developed by Didier Sornette, Physicist at ETH Zurich. Extreme events can be predicted  
- we just look at wrong data or not at dynamics



Dragon King Theory

Theory developed by Didier Sornette, Physicist at ETH Zurich. Extreme events can be predicted  
- we just look at wrong data or not at dynamics

# Time Series Models

$$y'_t = \underbrace{c}_{\text{intercept}} + \underbrace{\phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p}}_{\text{lagged values}} + \underbrace{\theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}}_{\text{lagged errors}} + \varepsilon_t$$

differentiated time series

## Statistical Methods

ARIMA, Exponential Smoothing, etc  
Very popular and mature (>50 years of research)

```
%%bigquery
SELECT *
FROM ML.EVALUATE(MODEL retail.arima)
```

	store	non_seasonal_p	non_seasonal_d	non_seasonal_q	has_drift	log_likelihood	AIC	variance	seasonal_periods
0	1	0	1	5	True	-1653.911896	3321.823792	2.949013e+09	[YEARLY]
1	2	0	1	2	False	-1681.388460	3368.776921	4.558555e+09	[YEARLY]
2	3	0	1	2	False	-1510.527920	3027.055840	3.576431e+08	[YEARLY]
3	4	1	1	1	True	-1692.968739	3393.937477	5.298193e+09	[YEARLY]
4	5	2	1	3	True	-1453.827626	2921.655253	1.471647e+08	[YEARLY]
5	6	0	1	2	False	-1671.258646	3348.517291	3.914811e+09	[YEARLY]
6	7	0	1	4	True	-1577.753490	3167.506979	9.551661e+08	[YEARLY]
7	8	0	1	2	True	-1585.820089	3179.640178	1.065770e+09	[YEARLY]
8	9	0	1	1	True	-1541.173574	3088.347149	5.514750e+08	[YEARLY]
9	10	0	1	1	True	-1718.989964	3443.979928	7.836701e+09	[YEARLY]

## ARIMA (p, d, q)

**p:** The number of lag observations included in the model, also called the lag order.

**d:** The number of times that the raw observations are differenced, also called the degree of differencing.

**q:** The size of the moving average window, also called the order of moving average.



# Time Series Models available in Google CCloud

## Time-series forecasting on Google Cloud

### BigQuery ML

[BigQuery ML](#) enables users to create and execute machine learning models in BigQuery by using standard SQL queries. It supports a model type called [ARIMA\\_PLUS](#) to perform time-series forecasting and anomaly detection tasks.

With ARIMA\_PLUS modeling in BigQuery ML, you can make forecasts on millions of time series within a single SQL query, without leaving your data warehouse.

ARIMA\_PLUS is essentially a time-series modeling pipeline, which includes the following functionalities:

- Infer the data frequency of the time series
- Handle missing data, irregular time intervals, and duplicated timestamps
- Detect spike and dip outliers and abrupt level changes, and adjust them
- Handle holiday effects, seasonality, and trends

Tens of millions of time series can be forecast at once with a single query. Different modeling pipelines run in

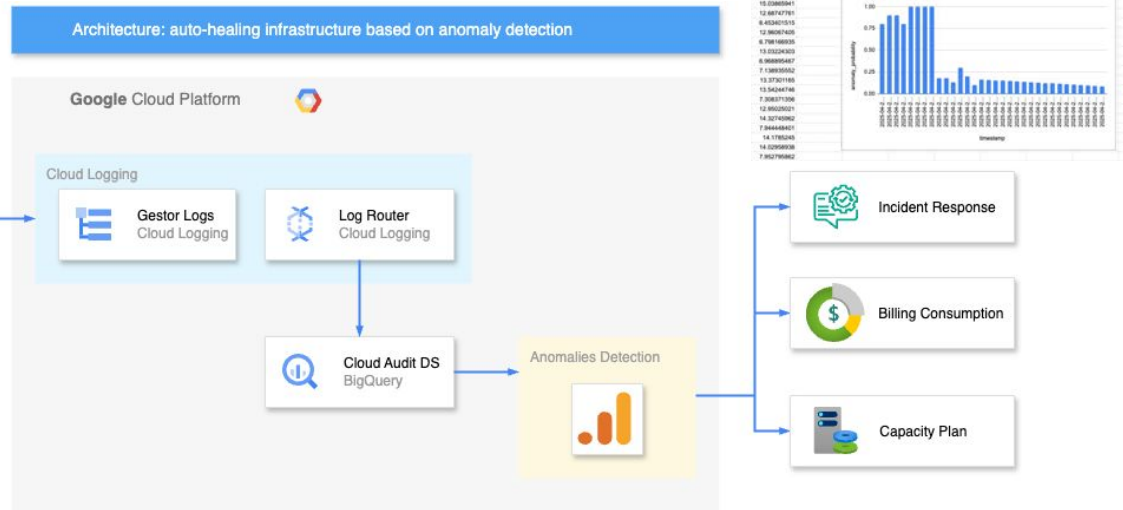
You can get started with BigQuery ARIMA\_PLUS with the following tutorials:

- [Single time-series forecasting](#) of Google Analytics web traffic with BigQuery ML ARIMA\_PLUS
- [Multiple time-series forecasting](#) of NYC bike trips with BigQuery ML ARIMA\_PLUS
- [Scalable forecasting](#) with millions of time series in BigQuery
- [Demand forecasting](#) of retail sales with BigQuery ML ARIMA\_PLUS

The modeling pipeline for the [ARIMA\\_PLUS](#) time series models performs the following functions:

- Infer the data frequency of the time series.
- Handle irregular time intervals.
- Handle duplicated timestamps by taking the mean value.
- Interpolate missing data using local linear interpolation.
- Detect and clean spike and dip outliers.
- Detect and adjust abrupt step (level) changes.
- Detect and adjust holiday effect.
- Detect multiple seasonal patterns within a single time series by using [Seasonal and Trend decomposition using Loess \(STL\)](#), and extrapolate seasonality by using [double exponential smoothing \(ETS\)](#).
- Detect and model the trend using the ARIMA model and the [auto.ARIMA](#) algorithm for automatic hyperparameter tuning. In auto.ARIMA, dozens of candidate models are trained and evaluated in parallel. The model with the lowest [Akaike information criterion \(AIC\)](#) is selected as the best model.

## Time Series



# Use Cases

Real-World Examples of Anomaly Detection

# Time Series in ...



## Retail & eCommerce

### Use Cases:

- Sales/Demand forecasting.
- Churn rate prediction.

### Typical Challenges:

- Forecasting new products.
- Complex hierarchy of products.



## Financial Services

### Use Cases:

- Asset Management.
- Product Sales Forecasting.

### Typical Challenges:

- Noisy data, state not observable.
- Many are 'Partially observable Markov decision processes'.



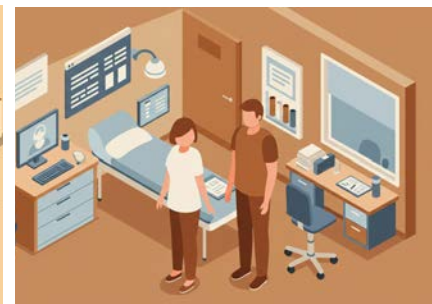
## Manufacturing

### Use Cases:

- Predictive Maintenance, Yield Opti.
- Adaptive controls.

### Typical Challenges:

- Poor data quality, very large data.
- Different sensor types and generations.



## Healthcare

### Use Cases:

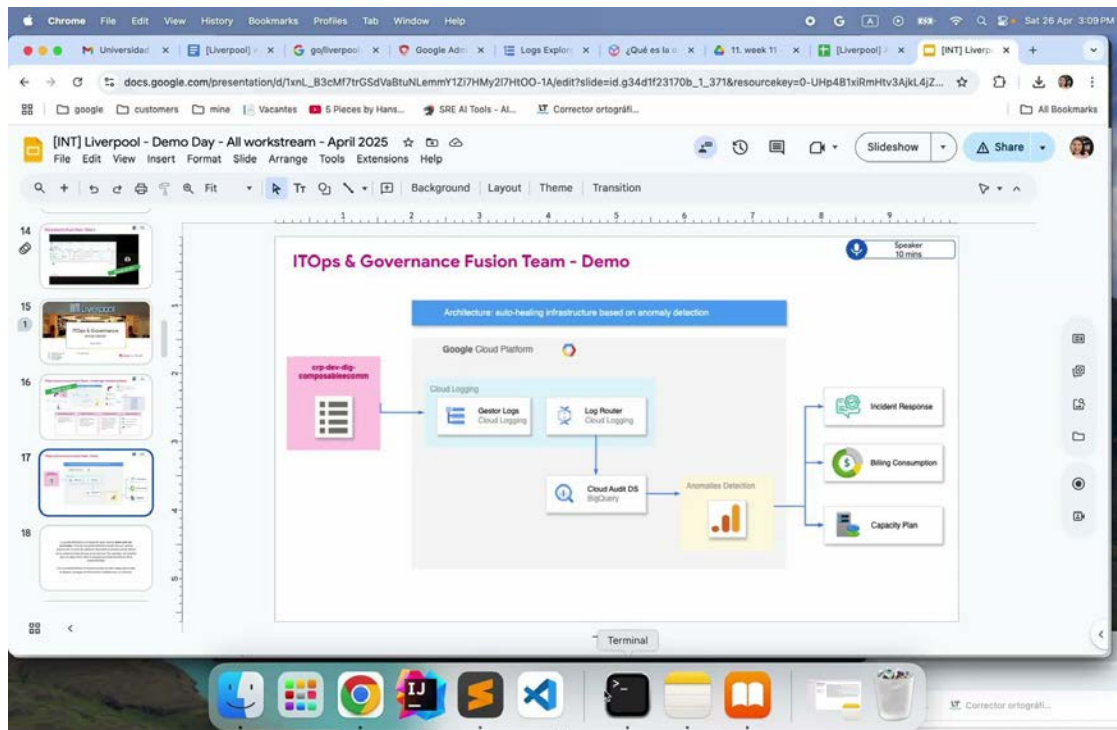
- Bed/emergency occupancy
- Demand for drugs for a pharm

### Typical Challenges:

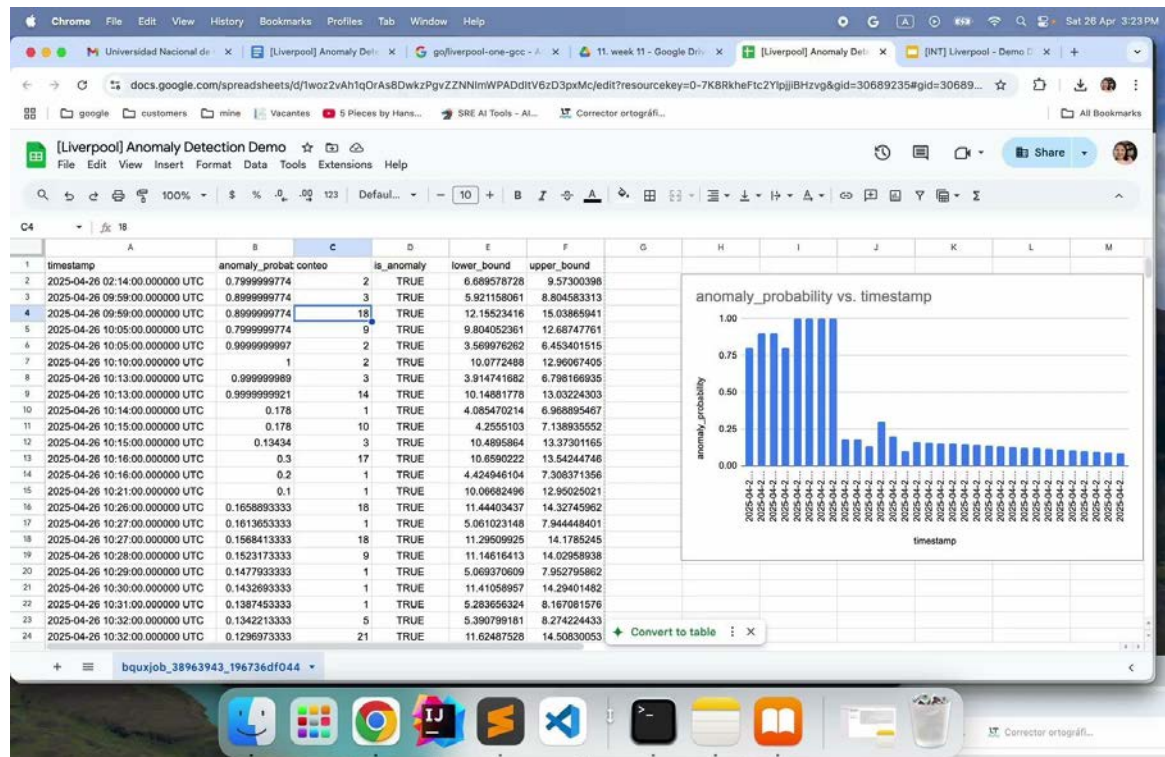
- Disparate data sources
- Data privacy PII



# Happy ending for the Sad story



# Time Series





# Thank you so much!



@yurnino

