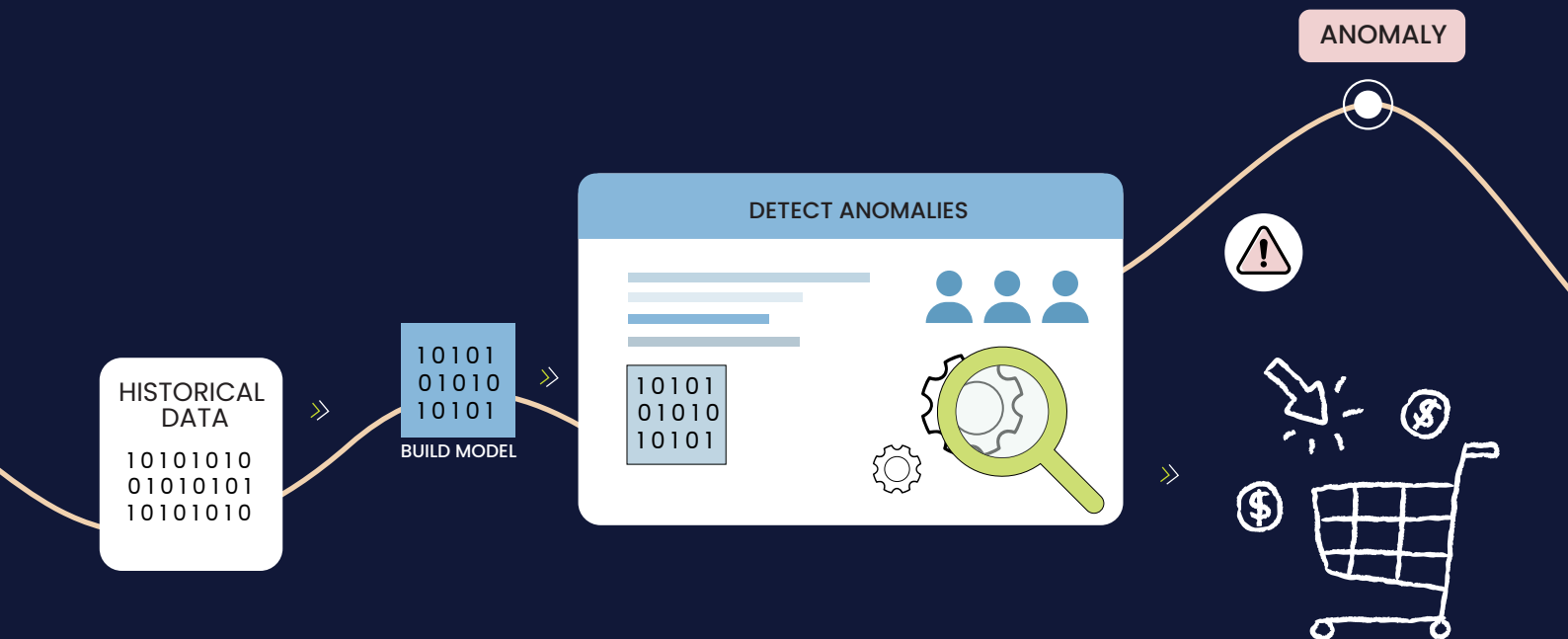


Case Study

Finding **Conversion Anomalies** at a Large E-Commerce Firm



About the Client

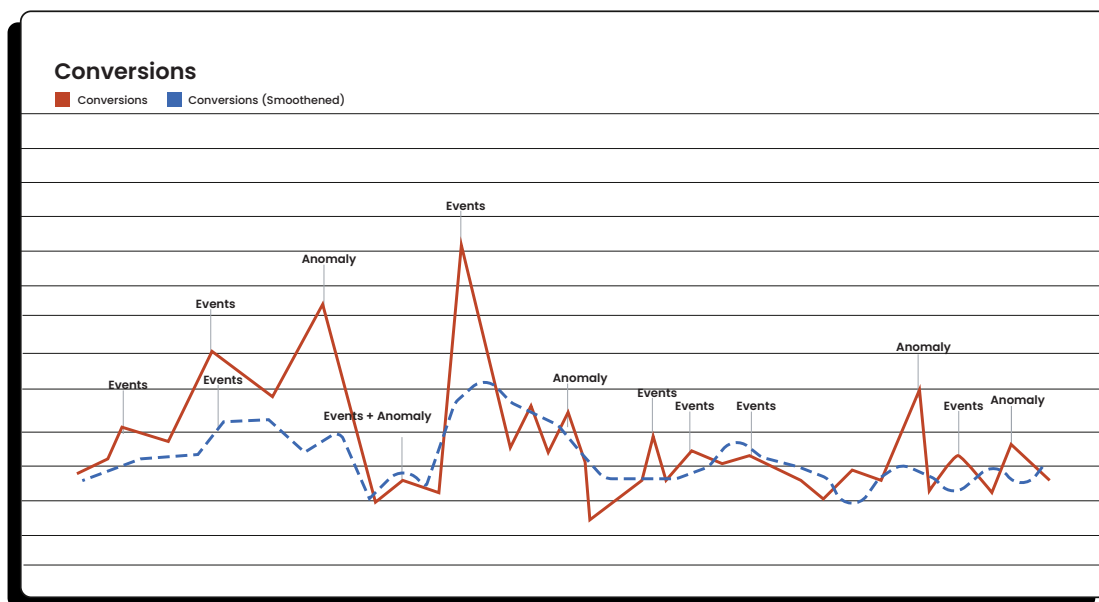
The client is a multi-billion dollar e-commerce company that sells across multiple categories such as consumer electronics, fashion, home essentials, lifestyle, and more. It is also considered the market leader in the apparel segment.

They have a 200+ member analytics team and a 1000+ engineering team to handle their data infrastructure, networking, data platforms and data science team.



Challenge

With competition intensifying, there is increased pressure on the client to understand their north star metric: conversion rates. This is a ratio of the number of customers visiting the client's website or app, and the total number of conversions (or confirmed orders). Conversions are critical to each business function, and they're tracked very closely across the organization. To drive conversions, the client organizes a number of sales events throughout the year, where they provide attractive discounts to customers.



While a sudden surge and a subsequent drop in conversion rates is expected during and after sale events, there had been many recent incidents where the conversion rates had gone outside of the thresholds set by the team. As a result, they were facing the following challenges:

Granularity and scale of analysis:

In a typical scenario – the conversion rates over a period of time are plotted on a graph, and when there is a sudden drop in conversions the relevant team performs a root cause analysis (RCA) which is a time and resource intensive activity, often taking weeks to complete. This is because of the large volumes of raw data that needs to be processed to generate the relevant features.

Most often, this analysis is only at the top / company-wide level and sometimes goes into details of specific business units (BUs) like furniture, electronics, and more. However, the product categories or sub-categories within these BUs are often ignored since this is a cumbersome process, taking 2 to 3 weeks for each investigation. As frequency of conversion rates dropping below the threshold has increased in recent times, the number of RCA requests has significantly increased, with additional pressure on the team to improve unit economics, which is the cost to revenue ratio per unit sold.

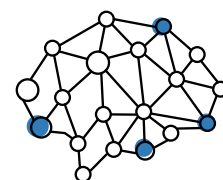
Conversion rate optimization:

With periodic drops in the conversion rates, the team would like to understand the underlying cause behind these sudden changes – whether it's because of internal issues, competition, stockouts, or some other factors. This would help them proactively avoid such situations and respond better during future events.

Assetization of data:

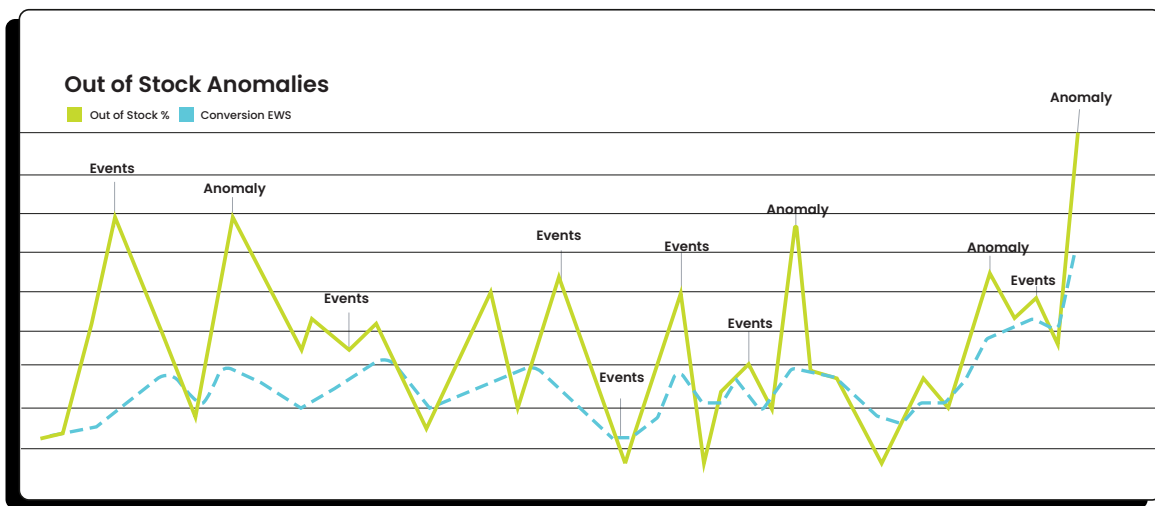
The client's analytics team is also looking to create assets or reusable components that can be used for future analytics initiatives to:

- a. Increase the amount of automation; and
- b. Improve the productivity of the team



Scribble's Solution: Anomaly Detection







The client's team wanted Scribble Data's Enrich Intelligence Platform to help them define and flag anomalies without having to use ad hoc processes. They expected the system to help them identify anomalies in their north star metric, and also compare it to other parameters such as pricing, product availability, delivery time, and more. This would help them identify and analyze the drivers of conversion rates.



For example, they wanted to understand how “Out of stock” contributes to conversion rates, and if so, they wanted the ability to identify which categories needed attention, and proactively take action. However, for this to happen, the team would need to understand if products being “out of stock” are indeed the problem. Additionally, they wanted to identify if there were other factors such as serviceability (or the lack thereof) in the customer’s zip code which were leading to lower conversion rates, and use this knowledge to take remedial action, such as expanding the logistics network.

Identification of causal factors

The data received for ingestion was over hundreds of terabytes in size. However, with Scribble Data’s feature engineering capabilities, the client’s team was able to zero in on 15 drivers of conversion rates, including:

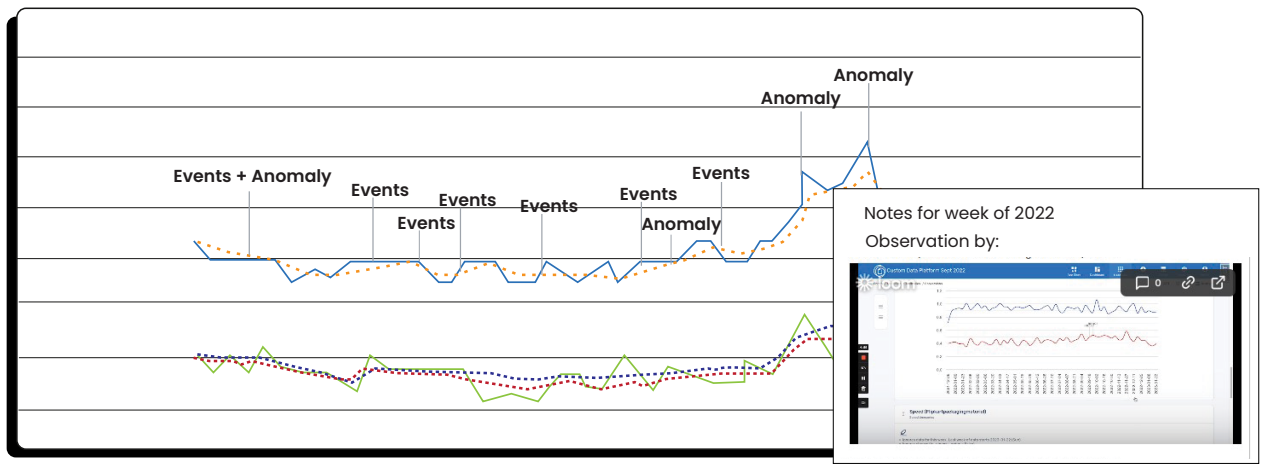
 Availability of stock	 Product discovery: Order confirmation or conversions from different tools for product discovery such as
 Product assortment	<ul style="list-style-type: none">○ Product search○ Category pages○ Product recommendations; and more
 Customer demographics	
 Competitor pricing	
 Delivery time	

With the Enrich Platform, the client’s team had a central place to see all of the causal factors along with the north star metric.

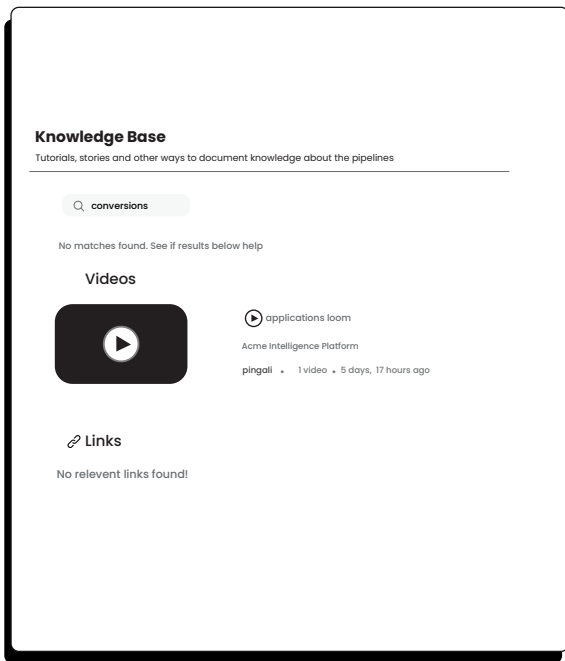
Additional Benefits of the Enrich Platform

Annotation and Knowledge Base Capabilities

The ability to add annotations is extremely helpful for the client’s analytics team as they can make sense of the visualized data, and further be able to share relevant notifications to the relevant business teams. This also allows them to add more context to certain anomalies caused due to factors such as seasonality and annual sales, where a surge in conversion rates is usually observed.



They can build a central knowledge base with explanations for each dataset or report from the voice of the analyst themselves. Thus, eliminating tribal knowledge and allowing everyone on the team to be on the same page.



Ease of integration and collaboration

With Enrich, clients can set user controls, download output for analysis on third-party tools, and can share reports among team members for collaborative investigation.

A Central Marketplace for all Processed Data

To ensure the trustworthiness of the datasets used for analysis and to facilitate further collaboration, the Enrich platform also provides the client with a centralized repository or marketplace for all the transformed datasets. This also allows the team to view:

- All the datasets generated to drive the data product
- The corresponding source
- Other entities such as data products or reports that might be accessing the dataset

Results



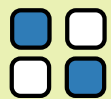
6x faster time to value with an operationalized anomaly detection product delivered within 4 weeks, instead of 6 months.



Time to create a root cause analysis report reduced from 2 weeks to 5 minutes, with the depth of RCAs increasing 4x



2x the number of tables, & sources analyzed



10x product categories analyzed