

# DATA PORTFOLIO

# About Me



Hey, I'm John, an Australian data professional with 3 years experience in **data analytics** and **engineering** and a background in data science.

Data has always been my area of passion. There are always hidden insights to be uncovered and depending on the way you slice it, it can tell polarising stories. I'm here to help tell these stories.

I am immensely curious and always looking to learn new skills, knowledge and interests. I love being hands on and building things, whether it be physical or digital.

Beyond work, I'm a big fan of **badminton**, **tennis** and **board games** and I thoroughly enjoy expressing myself creatively through singing, dancing and the occasional pottery.



# Data Skills

## -Engineering

I leveraged **Google BigQuery** & **Dataform**, **SQL** and **Apache Airflow** to build resilient and robust data pipelines for data migration.

## -Analytics

I used **Alteryx** and **SQL** to build ETL pipelines to empower internal employees and **Python** to conduct statistical analysis.

## -Visualisation

I brought disconnected data points together in **Tableau** and **ThoughtSpot** to tell compelling narratives and drive data-backed decisions

# Portfolio Projects

Below are three projects reflecting the full analytics-engineering lifecycle from building reliable data infrastructure to applying rigorous experimentation and translating findings into business action.

# Project:

## Building New Ingestion Pipelines for Third-Party Aggregator Data



### What's the problem

Third-party aggregator data for verifying home loan customer information was arriving manually as **Excel** files, and updated manually in our **Alteryx** workflows with no automated process, making it prone to human errors.

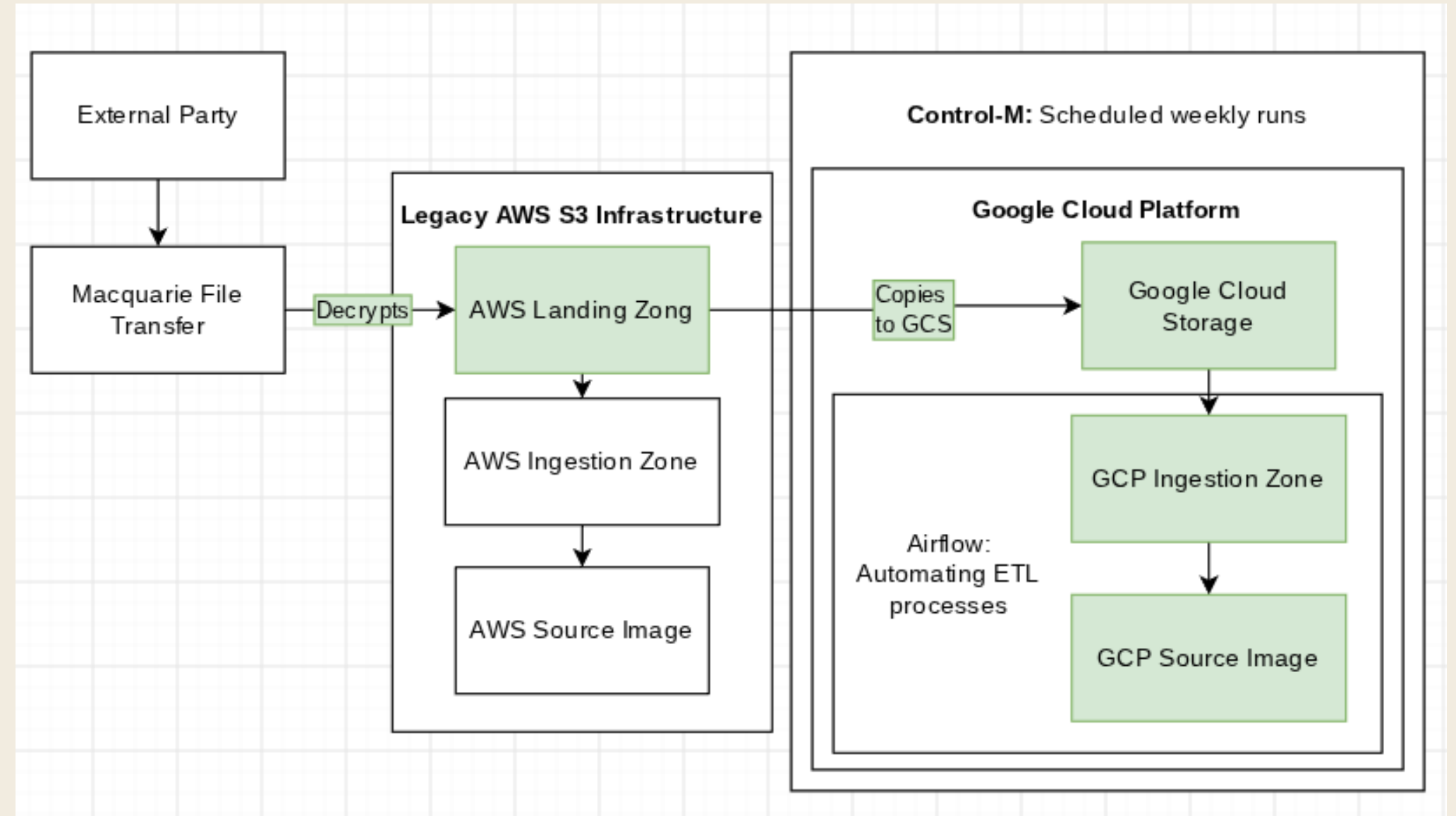
# Project:

## Building New Ingestion Pipelines for Third-Party Aggregator Data



### What did I do

Built and scheduled the data pipeline to decrypt and ingest the file into our **GCP** landing zone and modified the **Alteryx** workflow to transform and load a format-corrected file into **BigQuery** whilst working with external and internal stakeholders regarding frequency and content.



# Project:

## Building New Ingestion Pipelines for Third-Party Aggregator Data



### The Result

Replaced a manual, human-error prone process with scheduled, consistent and resilient pipeline with format checks and incident raising capabilities built into the automated pipeline.

### The Impact

Removed a manual bottleneck and made third-party data reliably available for downstream use in loan processing and BDM trigger initiative.

# Project:

## Trigger-Based Intervention - A/B Testing (Frequentist vs. Bayesian)



### What's the problem

A trigger system flagging at-risk home loan applications lifted broker re-engagement from 30% to 80%, but did the operational win actually drive more completed loans?

### What did I do

Tested the causal impact using both frequentist and Bayesian A/B testing approaches in Python, comparing loan volume and value for intervention vs. non-intervention cohorts.

# Project:

## Trigger-Based Intervention - A/B Testing (Frequentist vs. Bayesian)



### The Result

Using standard Frequentist and Bayesian with an uninformative prior distribution, yielded a one-sided p-value of more than 0.05 and a probability of greater than 70%. The results were inconclusive at both the overall experiment level and at more granular customer segments as sample size was too small to leverage either framework.

### The Impact

Leadership continued the rollout based on the operational uplift and BDM feedback. While the experiment demonstrated initial success, we need more volume and time to elucidate the exact impact of trigger-based intervention.

# Project:

## Home Loan Variable Rate Review Segment Level Discount Analysis



### What's the problem?

The Data Science team built a variable rate model that provisioned home loan interest rates to customer segments. We needed a tool to visualise and further inform us of our pricing strategy and sensitivity of each customer segments.

### What did I do?

Built a segment-level analysis dashboard in **ThoughtSpot** visualising key metrics such as *discount cost*, *external refinance rate*, *margin retention* and *settlement volume* across successive experiment cohorts.

# Project:

## Home Loan Variable Rate Review Segment Level Discount Analysis

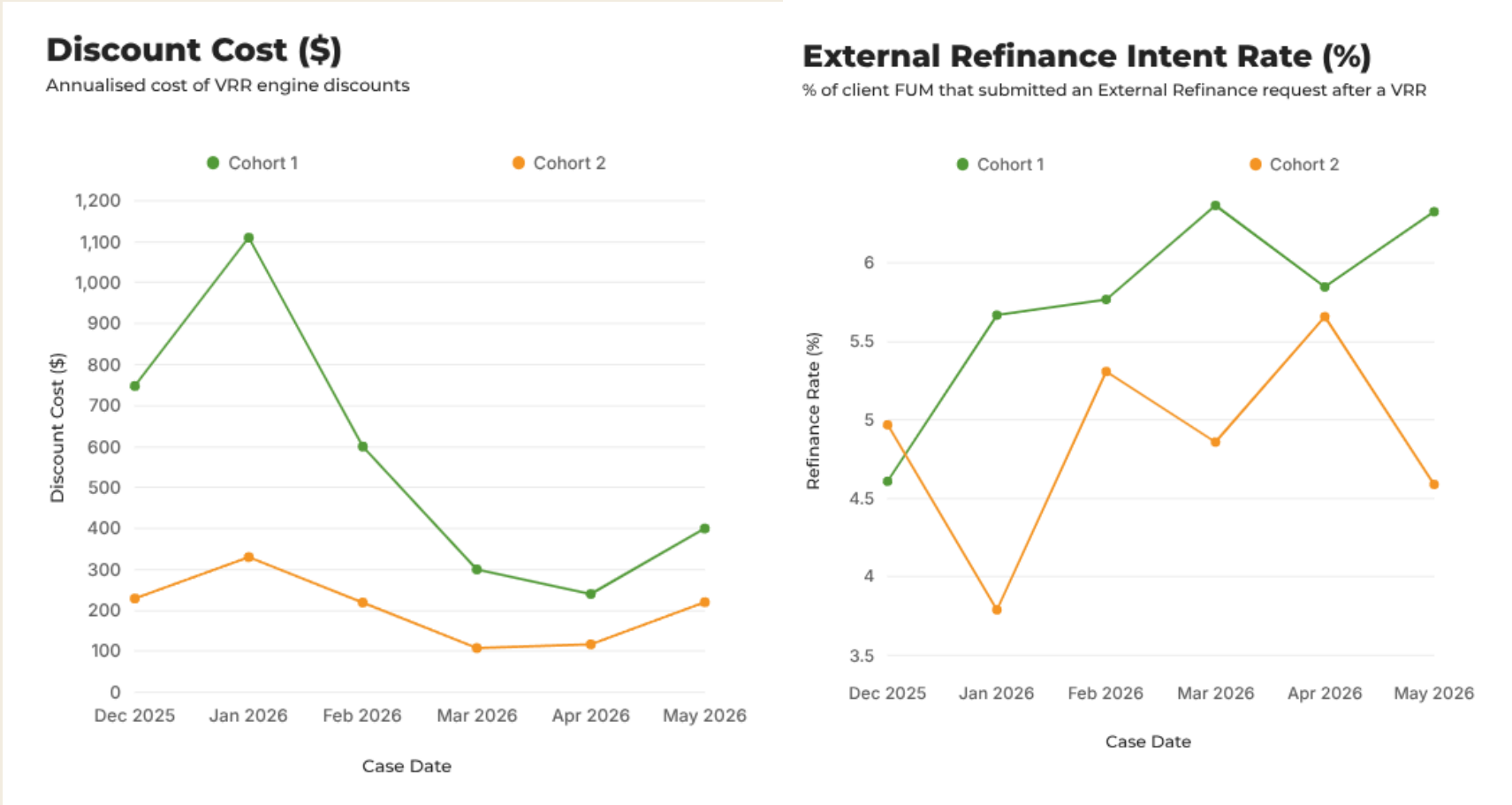


### The Result

Identified cohorts with disproportionately higher refinance rates and discount sizes relative to settlement volume, these cohorts were given to marketing and data science, leading to reduction in average discount given and influenced feature engineering for future VRR models

### The Impact

Protected lending margin without hurting conversion by cutting discounts where they weren't changing customer behaviour, and helped improve future iterations of the model.



# Contact Me

**email:**      [johnwei0327@gmail.com](mailto:johnwei0327@gmail.com)  
**phone:**    +49 178 9091534  
**linkedin:** [linkedin.com/in/john-wei-01](https://www.linkedin.com/in/john-wei-01)