



# Machine Learning Pipelines with Honey Science



Google Cloud

*honey*



# Armen Donigian

Head of Personalization & Discovery

Google Cloud

honey

# Who are We?

---

## Our Mission...

- Make The World More Fair

## What We Do?

- We're building a better way to shop, one that saves you time and money, even if there's less in it for us.

## Our Business Model...

- We earn a commission from stores when you find savings.
- Find out more at [www.joinhoney.com/about](https://www.joinhoney.com/about)

More than  
**\$1 billion**  
in savings found

---

Trusted by  
**10+ million**  
members worldwide

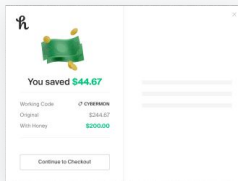
---

More than  
**\$8 million**  
rewarded to  
members

# Our user facing tools

## Savings Finder

Our original and most popular tool automatically finds and applies coupon codes on 40,000+ popular sites. One click is all it takes.



## Price History

See how much an item's price has gone up and down in the past. Based on past trends, you can choose to buy now or wait for another drop.



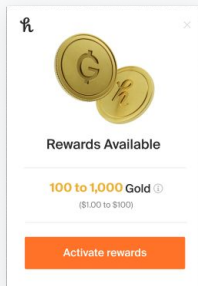
## Amazon Best Price

We compare every seller to show you the lowest price (even if Amazon doesn't). Plus, lots of our tools work together to make Amazon even better.

Original Total		New Total	
Price	\$399.99	Price	\$399.99
Shipping Fee	\$49.99	Shipping Fee	\$49.99
Amazon Prime	✓	Amazon Prime	✓
Free Delivery	✓	Free Delivery	✓
Free Returns	✓	Free Returns	✓
\$437.94		\$396.95	
Amazon Prime Shipping		Amazon Prime Shipping	

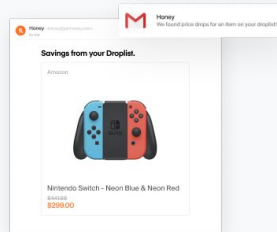
## Honey Gold

Earn rewards while you save with our exclusive Honey Gold program. Redeem your Gold for gift cards at your favorite stores.



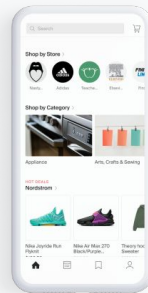
## Droplist

Who has time to track prices? Just add items to your Droplist and we'll watch them for you and notify you when there's a price drop.



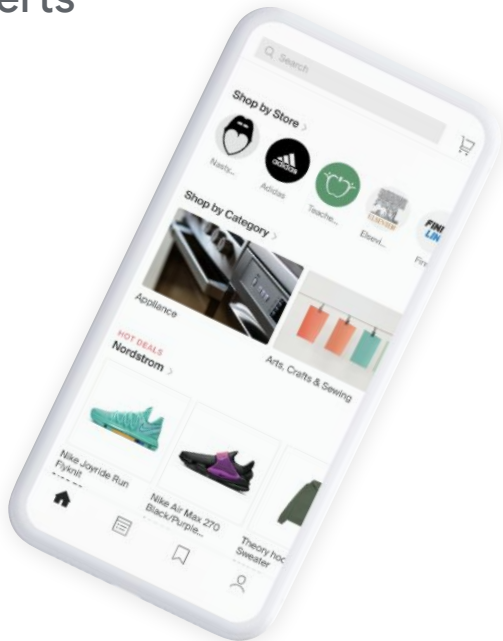
## Mobile App

Meet the world's most convenient way to buy. Shop your favorite stores, get amazing deals, discover new finds, and checkout - all in one app.



# Business Use Case

## Deal Alerts



### Goal

- Make joinhoney.com a destination for product/deal discovery, not just a waypoint to the extension
- Re-engage web & mobile users by generating actionable content by tapping into our rich history of product catalog data

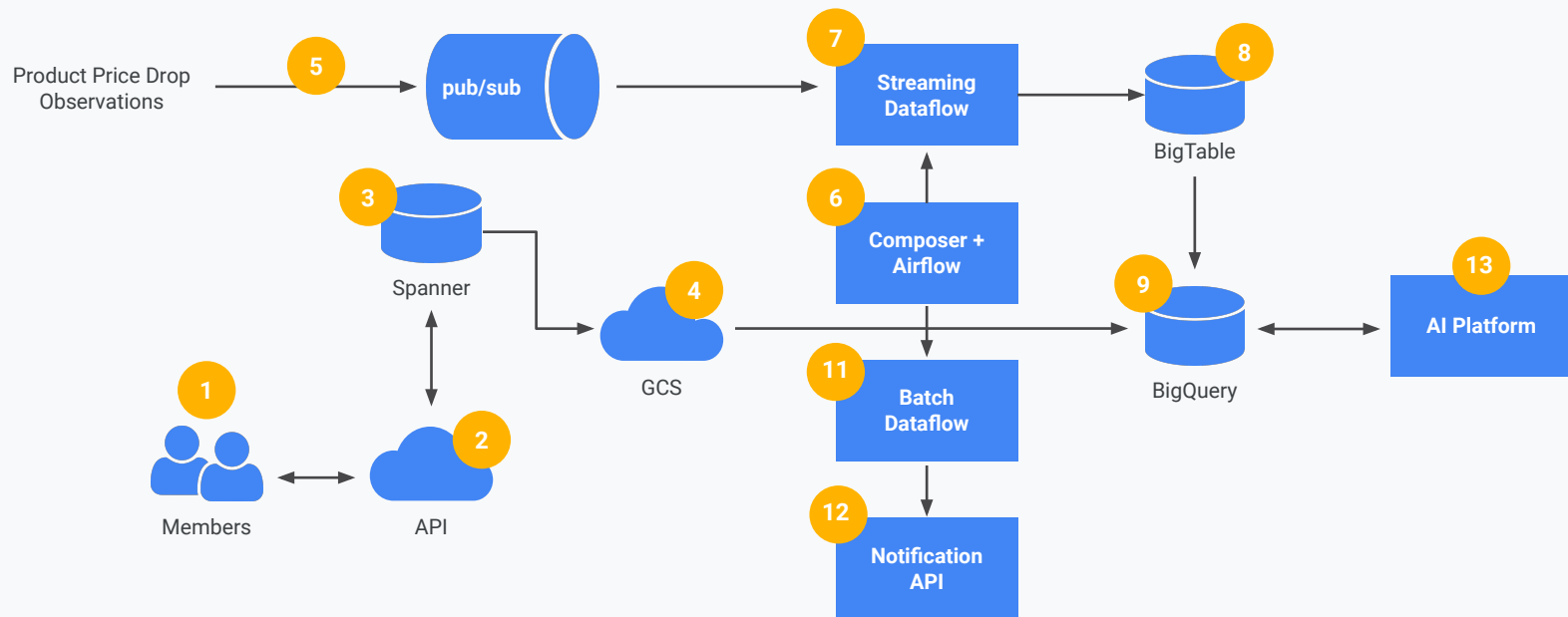
### What

- Feature that allows users to follow combinations of product brands/stores/categories/keywords.
- Notify users whenever there is a deal for any product that matches the specified criteria.

### Why

- Tens of thousands of product coupons monthly & product price drops daily
- Tons of product signals not discoverable by our audience in a timely manner!

# High Level Architecture



# Challenges

## Biggest Challenge to Machine Learning Adoption

---

- Finding a dataset with good “labels” for the use case you’re solving for.

## Not Possible to Reuse Data as we Reuse Code

---

- Software engineers find leverage in re-using libraries/packages.
- What can data professionals do?

## Product Feasibility given Quality of Data

---

- What can the data we have help us do?
- Given our data, how difficult is this problem to solve?

## Democratizing Innovation Across our Organization

---

- Does the organizational structure of the teams enable innovation bottom up?
- How can we increase the data literacy of individual contributors to expedite learning & experimentation?

## Data Discovery

---

- What data do we have?
- Which fields are still being maintained?
- What transformations have been applied to this data previously?
- What are the known issues/gotchas with this dataset?

## X Sounds Great but we’re busy doing Y...

---

- How do we enable teams across the organization to prioritize important but not urgent work?

# Which GCP Products & Why

## AI Platform

- Model Training & Exploratory work (via [AI Notebooks](#))
- Pre-built CMLE Airflow operator (model re-training & inference)
- Model Inference (Streaming & Batch)
- Integrated w/ other GCP products (GCS, Dataflow, BigQuery etc)
- Leverage State of the Art assets such as embeddings, models, pipelines (via [AI Hub](#))



# Which GCP Products & Why

Cloud Composer + Airflow

- Produce daily snapshots of operational single source of truth tables
- Perform aggregate queries & compute roll ups
- Generate feature stores in (BigQuery) to be used downstream by machine learning algorithms.
  - Perform batch prediction.
- Perform data sanity checks to ensure single source of truth is preserved on an ongoing basis
- Orchestrate batch & streaming Dataflow jobs
- Enables orchestration of data pipelines which with completion sensors to most GCP products

Time for Q&A