

# Data Analysis & ML Workflow Experiences

EPIC Retreat – April 17-18, 2023

Barbara Stanley – Engineering Director for BigQuery Experience

## A little about my work...

- *Role:* BigQuery Experience Pillar
- *Priority:* Data Analytics & ML workflows
- *Goal:* a simplified, seamless experience



Barbara Stanley

# Pressures facing our customers today



## Market Pressures

---

- Cost cuts
- Global regulations
- Security: Data breaches



## Technical Pressures

---

- Complex data stacks
- Data size growth
- Multi-structured data sources



## People Pressures

---

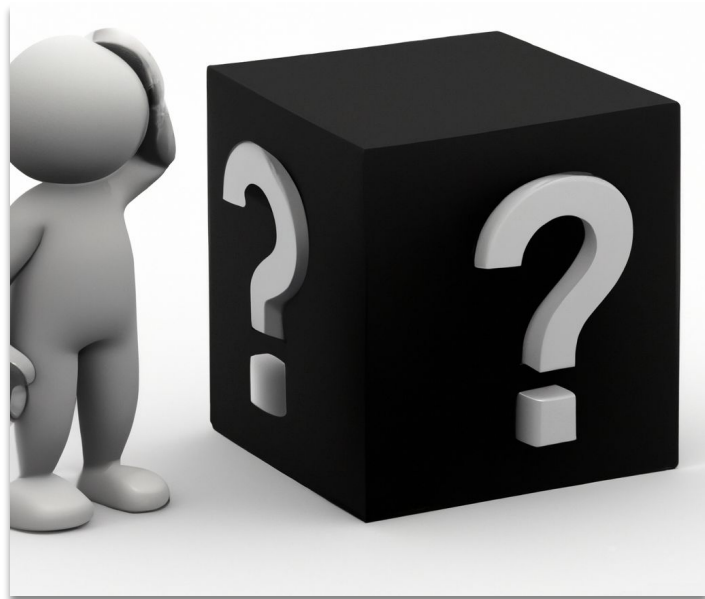
- How to operationalize ML
- Experienced talent hard to find
- Roles blurring

# Fear of black box solutions

- Self-limited to easy-to-understand models
- Stakeholders insist on understanding the models
- Resort to simple & safe linear regressions

## What if...

- We had assistive technology to explain models
- If improved models were *amazingly* better, would people care if they didn't 100% understand them?

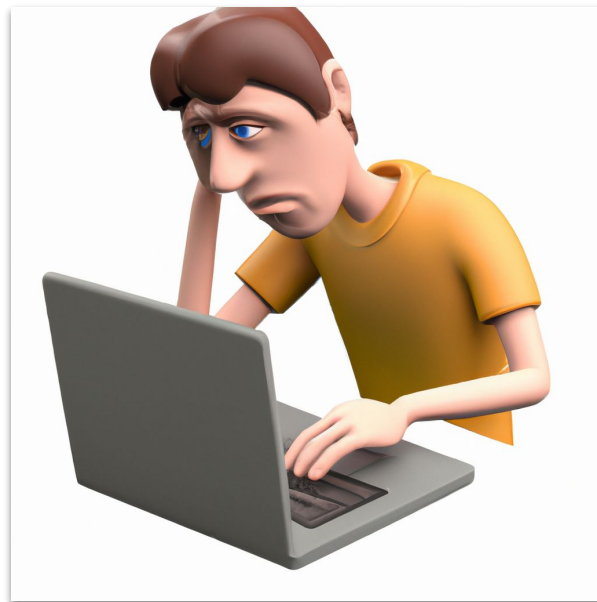


# Toilsome data prep

- ~80% of ML work can be data prep
- Independent tools exist but often not used
- Have the tools not yet replaced human expertise?

## What if...

- No upfront configuration were required
- The data prep tools automated 90+% of the manual effort



# Regulatory requirements

- How can I be sure that I am meeting all regulatory requirements?
- Fear of hefty fines!
- *Much* effort required to track and adhere to regulations

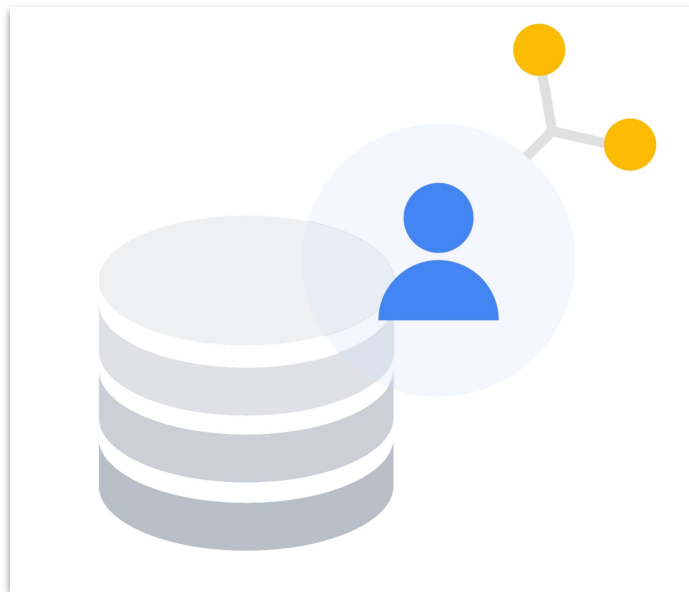
## What if there were...

- Service providers had cryptographically signed attestations trusted by regulators
- A general regulatory enforcer abstraction



# JOINing Unstructured Data with Structured Data

- Unstructured data: log files, 5G, IoT, clickstreams, business events, transactions
- To join with structured data: account data, CRM data, etc



## What if...

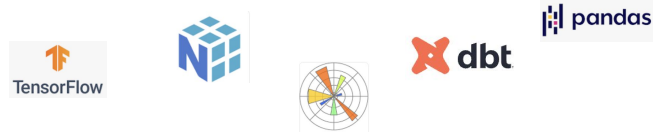
- Users didn't have to convert unstructured data to structured data to do JOINS

# How to get started?

- So many tools to choose from
- No time to investigate them all
- Jupyter notebook + Pandas + spreadsheets seems like a “safe” path
- Or back to spreadsheets

## What if there were...

- “Zero to hero” onboarding for new users
- Expert guidance offered throughout DA/ML workflows





# How to improve my model?

- Model looks pretty good; can it be better?
- Just use canned models; they are “safe”
- Others don't trust canned models

## What if there were...

- Assistive technologies to help users improve their models

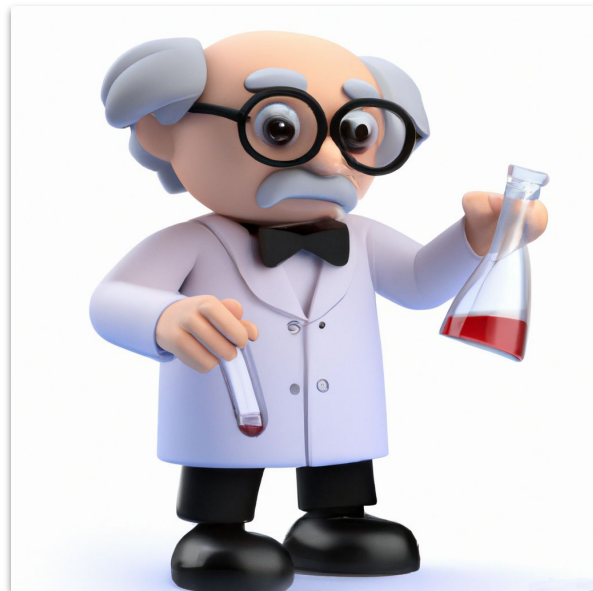


# How to operationalize ML?

- Many users never advance beyond experimentation or analysis
- Fear of negatively impacting the business
- Lack of self-confidence
- Unfamiliar best practices

## What if there were...

- Guidance for operationalizing ML models



# Summary

## Goals:

- Simplify users' data analytics->ML workflows end-to-end
- Happy, successful customers

## To connect:

- [Barbara Stanley @LinkedIn](#)
- barbara.stanley@gmail.com

