# DS + AI Accelerators

## Lunch & Learn | August 3, 2023

Bryan Powell & Florencia Silveira
*Data Scientists - Data Science Professional Services at Domo, Inc.*

Davis Busteed
*Data Scientist - RXA*

# Agenda

- Intro to Domo's DS + AI Accelerators
  - GitHub Repository

- Data Profiler
  - Background

- Data Anonymizer
  - Background & Demo

- Model Insights
  - Background & Demo

# Domo's DS + AI Accelerators

- Accelerators
  - Tools to simplify, streamline, automate, & integrate different parts of the data management process.
    - Accelerators automate repetitive work to free Domo users' time to focus on insights and decision making.

  - Requires access to & use of Jupyter Workspaces

- Available at domo.developer.com and the Domo DS GitHub.
  - https://developer.domo.com/portal/pg9t2vdbix5un-data-science-accelerators

  - https://github.com/domoinc/domo-data-science-resources/tree/main/data_science_accelerators

# Accelerator: Data Profiler

- Creates a profile of the input dataset on a Domo dashboard (HUGE time saver!)

- Conducts an automated exploratory data analysis, giving a first glance at the distribution and relationships between variables.

- Several cards are created and can be used or discarded at the user's discretion.

- Demo of the tool available at

https://developer.domo.com/portal/pg9t2vdbix5un-data-science-accelerators#data-profiler
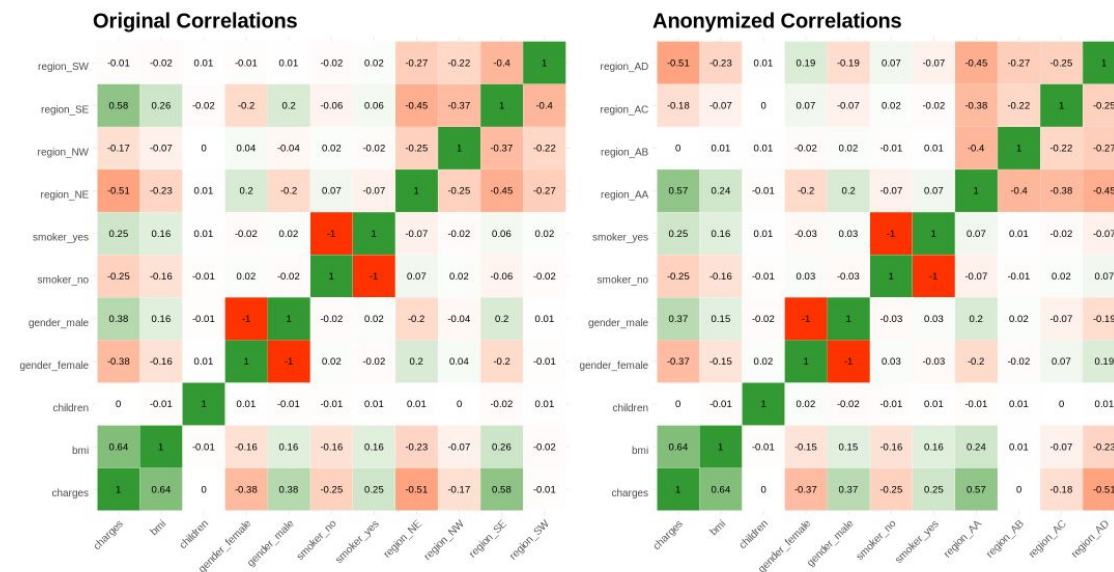
# Accelerator: Data Anonymizer

- Use case: You have sensitive data on an individual level that you would like to analyze on an aggregate level or use to train a model for prediction on future individuals.

- The output anonymized dataset maintains the aggregate characteristics of the original dataset without revealing information about individuals within the dataset.

# Data Anonymizer

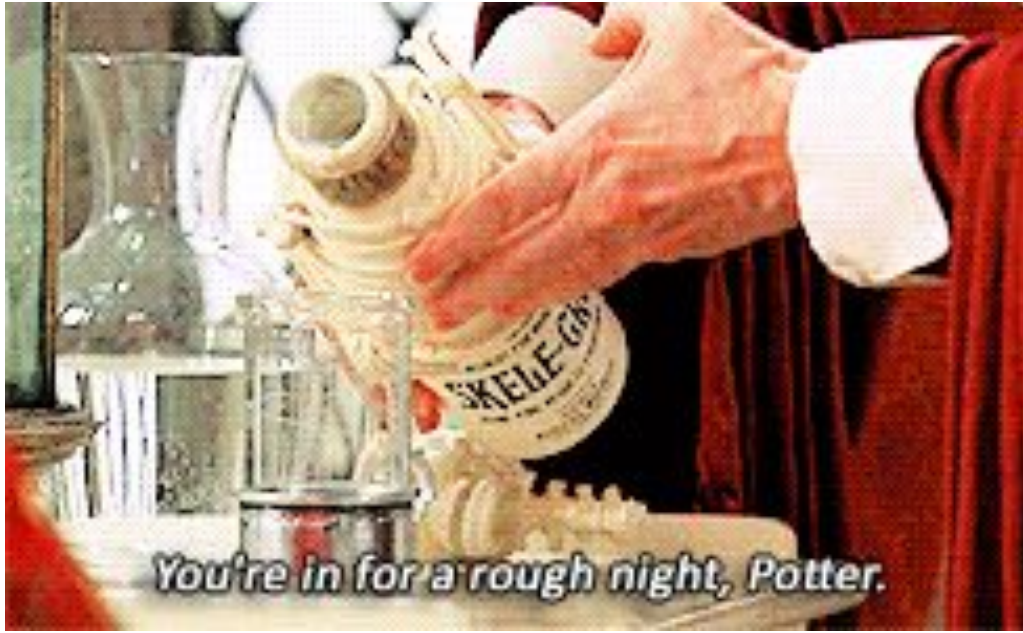

You're in for a rough night, Potter.

Well, what do you expect? Pumpkin juice?

- Demo Use Case
  - You, a Muggle data scientist, have been tasked with creating a model that predicts whether a Hogwarts student is at risk for diabetes.

  - You have training data available from past students, but to protect their privacy you must anonymize this data before training your model.

# Accelerator: Model Insights

- Creates an entire dashboard that evaluates and monitors the performance of input data science models.
- The tool details training performance, live performance, model drift monitoring, model feature details, and individual model predictions.
- Compatible with Regression, Binary Classification and Multi-Classification Models.
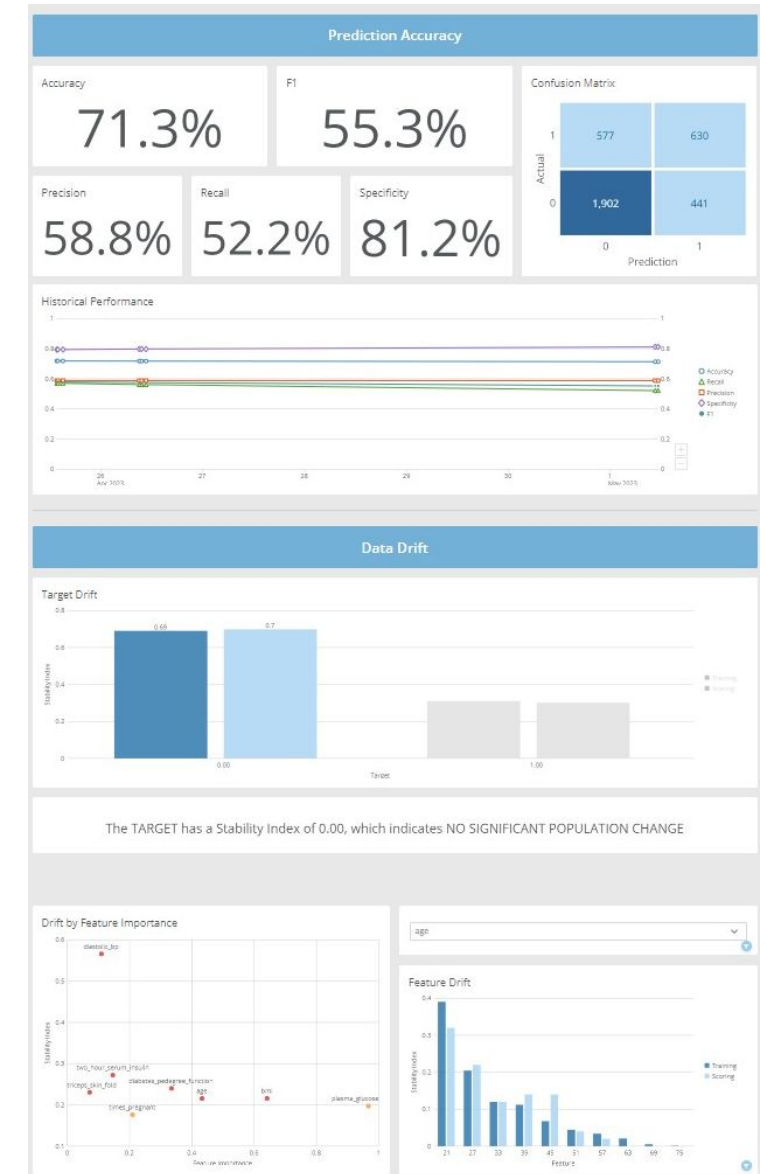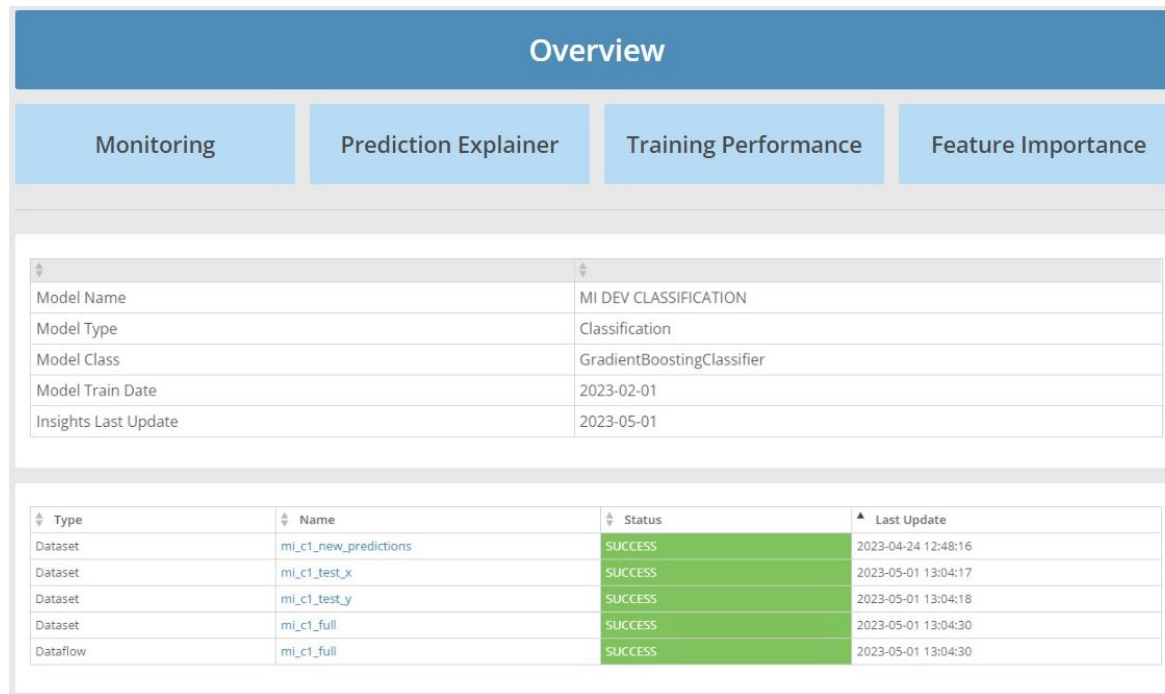
# Accelerator: Model Insights

- Inputs:
  - Pickled (saved) model
  - Training data
  - Live data
    - Model predictions on live data

- Outputs:
  - ~30 cards organized neatly into 5 dashboards detailing:
    - Model pipeline health
    - Model drift and performance on live data
    - Training performance
    - Model feature details
    - Individual prediction details

# Model Insights Demo



- 10 years ago, Divination Professor Sybill Trelawney foresaw that the Sorting Hat would be cursed, and unable to place students in their respective houses.

# Model Insights Demo



What exactly is the function of a rubber duck?
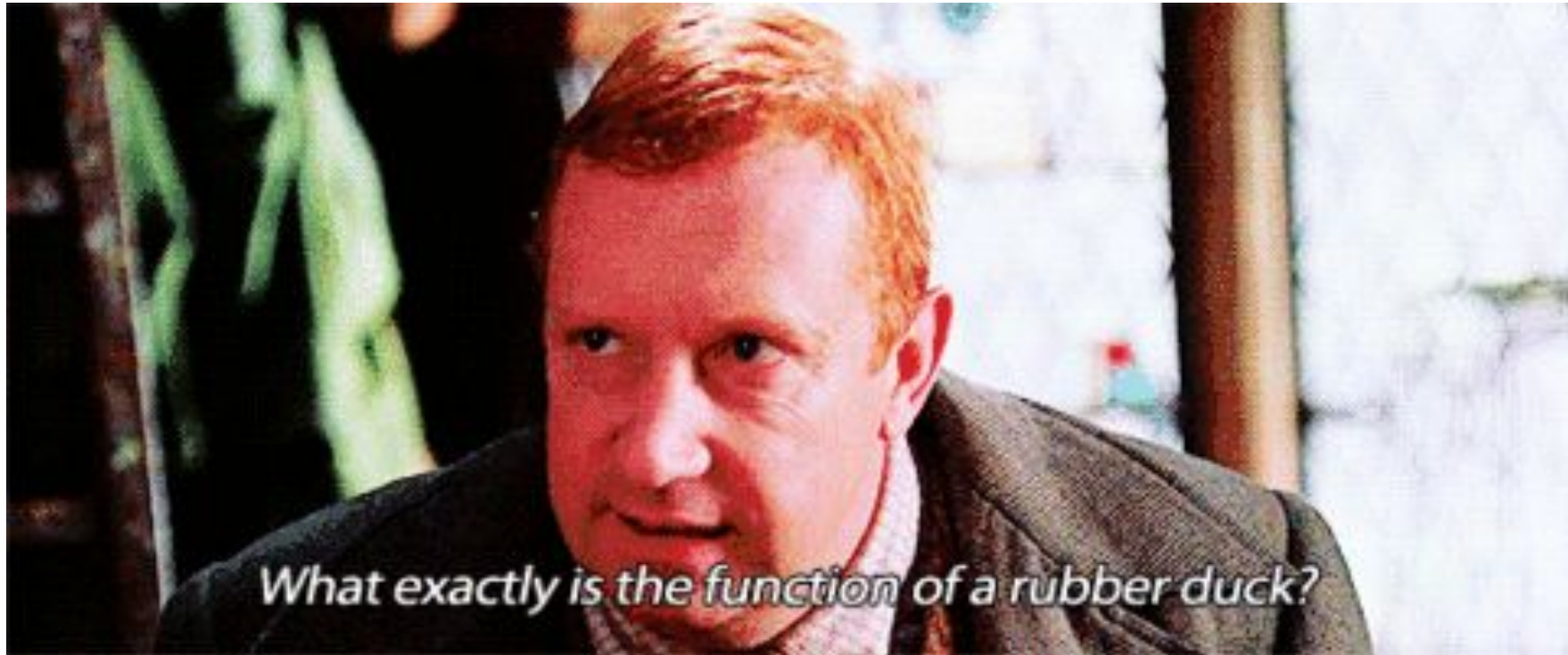
- In preparation, Arthur Weasley (Muggle Studies expert and data science dabbler) suggested that incoming students take a subject aptitude exam prior to being sorted into their Hogwarts House.

# Model Insights Demo

- Thanks to Mr. Weasley's fascination with all things Muggle, a training set with 10 years of incoming students' aptitude exams and the Sorting Hat's house decisions was compiled before the foretold curse fell.

- A model was built with this data and was used to sort students into their Hogwarts House last year.

- You, a Muggle data scientist, have been tasked with monitoring the performance of this model and deciding whether the model should be retrained.