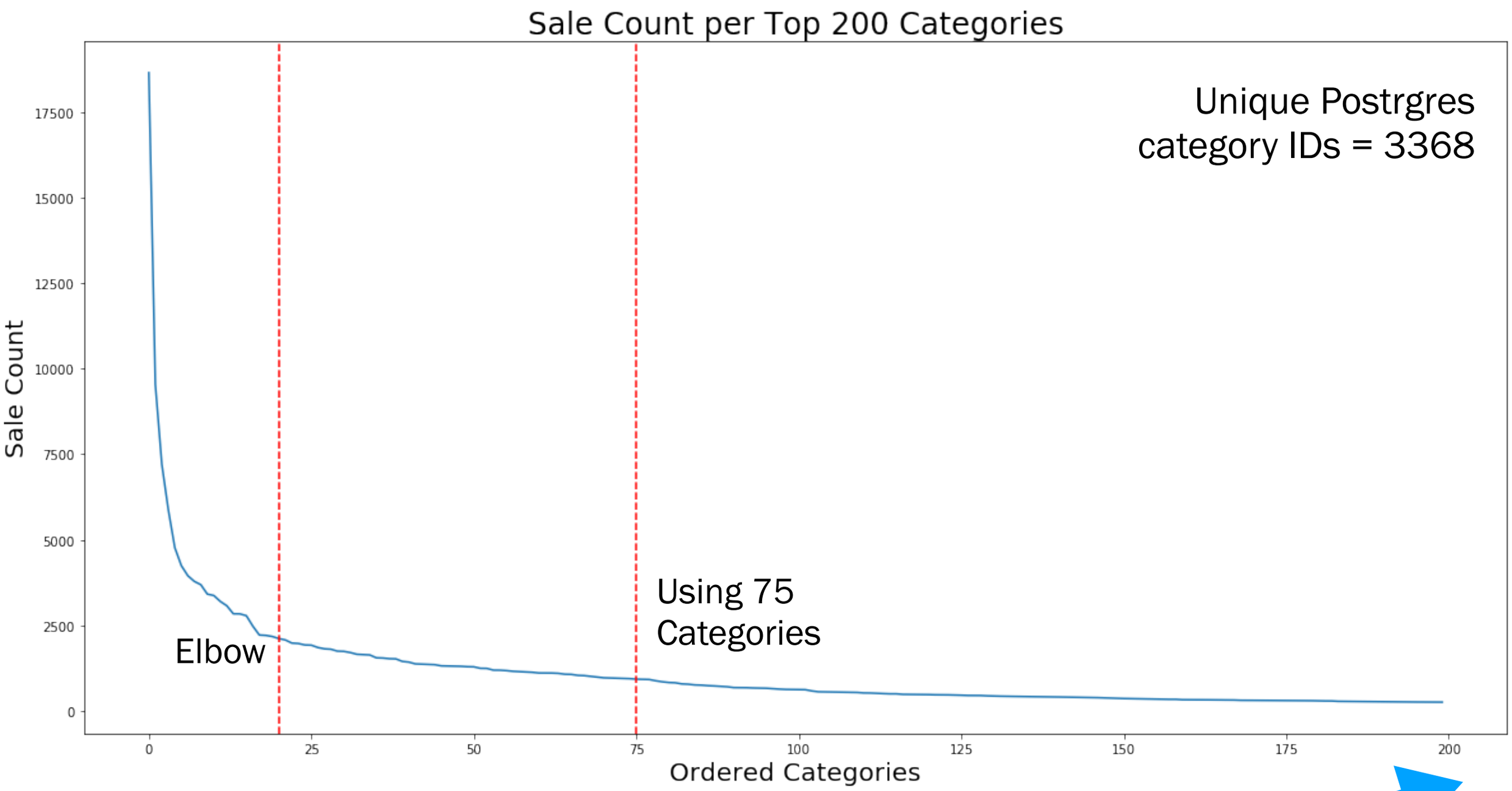
Engineering a Model

Demand Prediction - Machine Learning Team

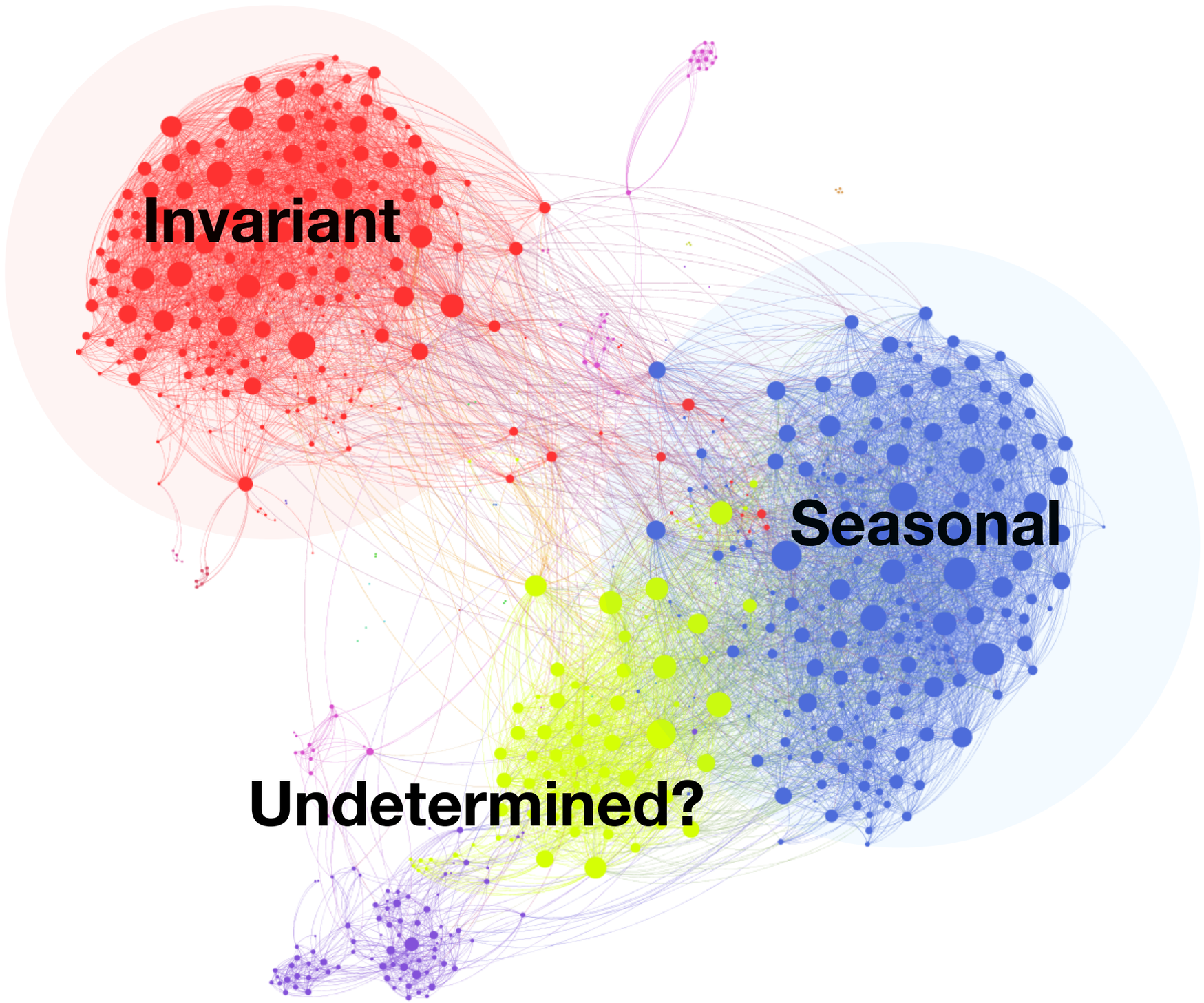
Chris Chen
Tony Reina
Kyle Shannon
Suman Gunnala
Anil Luthra

Model Constraints



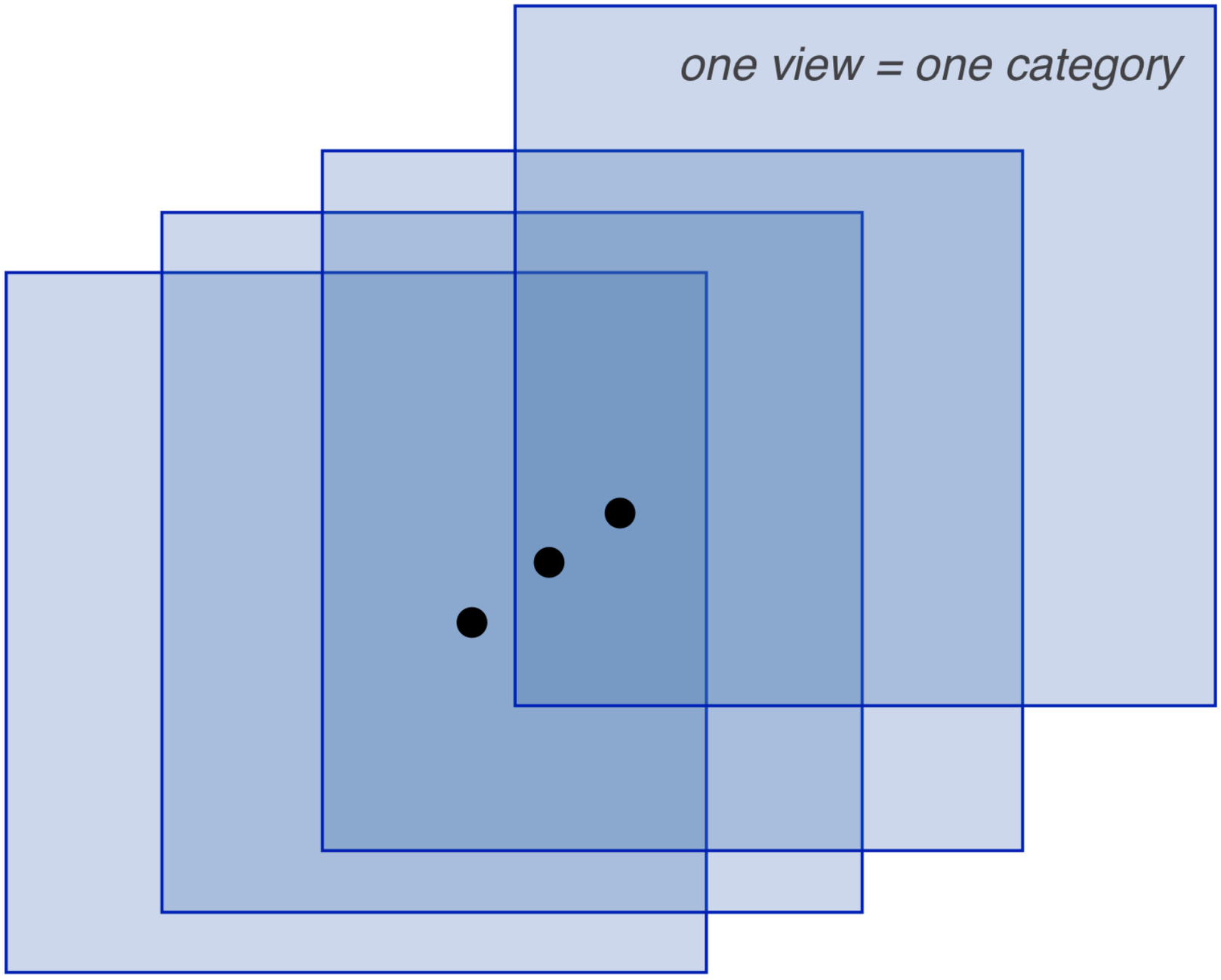
Remaining 3168 is the long tail

Cluster Top 75 Categories



How we would like Data?

cols = x Features + response var *y* + [MM/YYYY] DateTime



Features

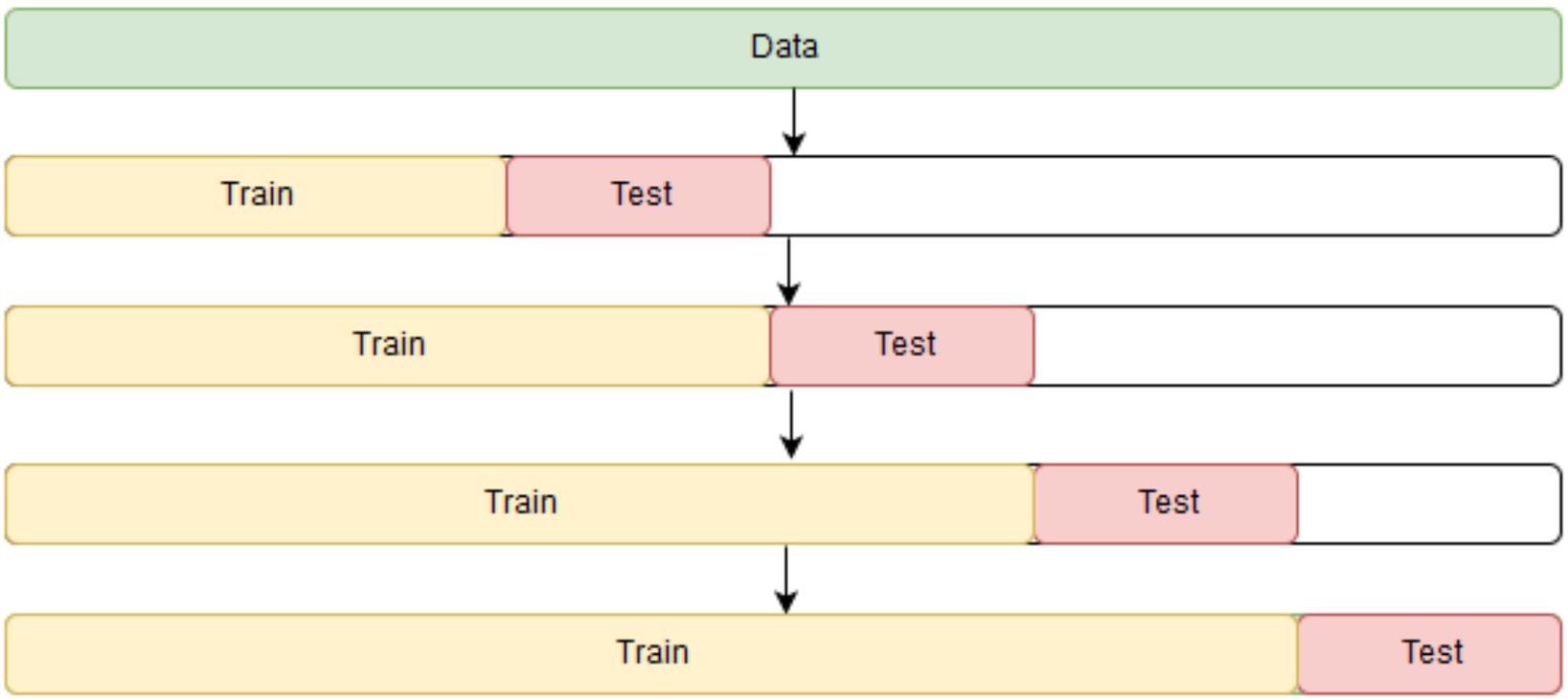
inventory_sold_ratio
dollar_sold_ratio
volume_moved
product_rating_average
product_rating_delta
total_sales
contains_sold_out_product
large_inventory_drop
is_pos_sentiment
is_neg_sentiment
is_neutral_sentiment
count_of_nodeIDs
is_in_campaign

Testing Training Regimens

The Problem. Latent Temporal Components.

The Solution. **Walk-Forward Validation**.

Does not assume independence or identical distributions.



Therefore… `sklearn.model_selection`.TimeSeriesSplit

Collaborations

- EDA team
  - Cluster Analysis of Top 75 Categories
  - Correlation heat map matrix
- Schema & Query Team
  - Given accepted/answerable queries, ensure mediated schema will return from wrappers/ source DBs in the correct format.

Thank You