# Corporación Favorita Grocery Sales Forecasting

## A casestudy in modelling grocery sales in Ecuadorian supermarkets

David Scroggs

4-12/2022

# Introduction

- Competition is to forecast item sales in 54 different grocery stores.
  - 4,100 sale items
  - 4.5 years, 125 million rows of training data
  - Forecast period of 2 weeks past final training data

# Data

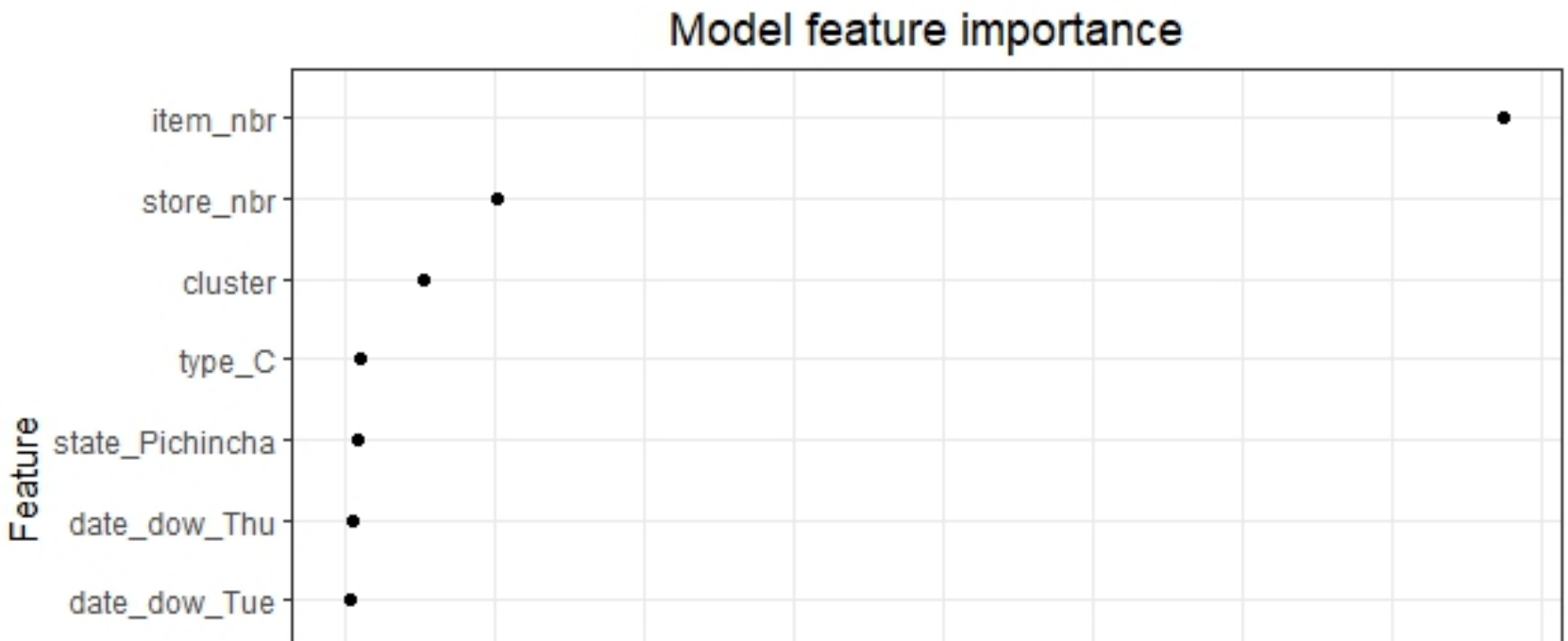
# Evaluation metric

# Models

# Results

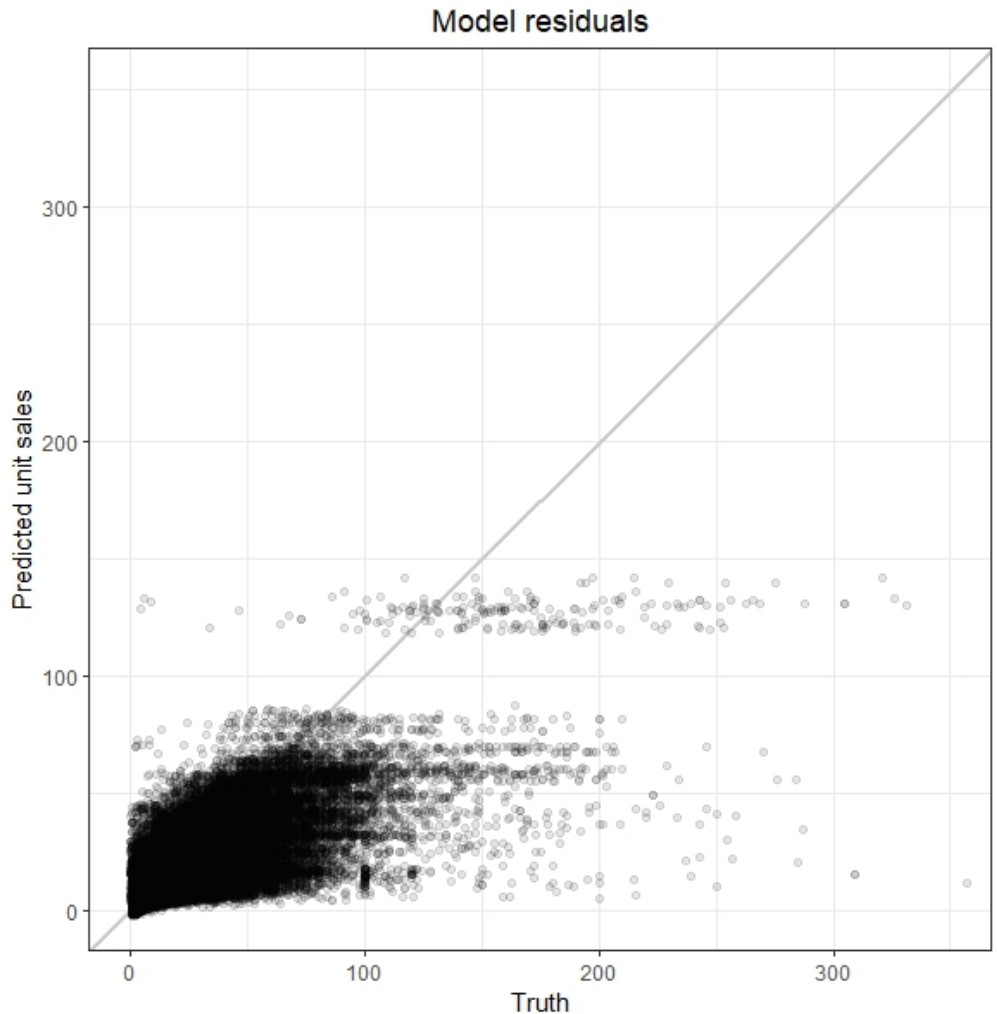
## Final model - Tuned XGBoost

- Features modelled - item number, store number, temporal information (weekday, month, year), store type, store cluster, location (state)
- hyper-parameters tuned: tree depth, min data points in a node, randomly sampled predictors
- Model evaluation: nwrmsle = 0.696



# Residuals

- Model tends to under-estimate the result
- Stratification of residuals in higher unit sales
  - Band of results where predicted sales $> 120$ from 2 stores (44, 49)



# Residual outliers