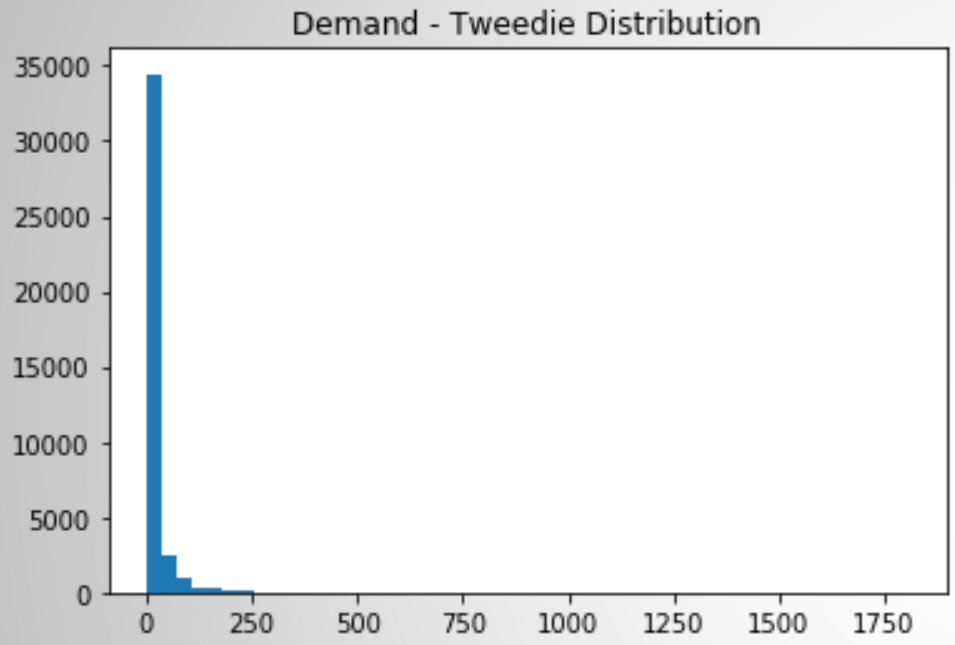
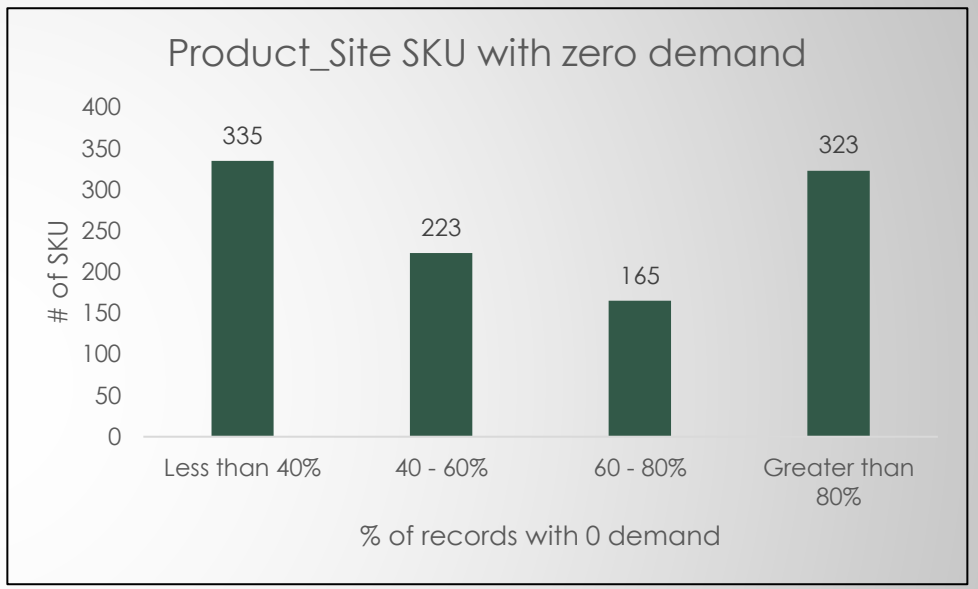
USAID - Intelligent Forecasting Competition Results Summary

Aishwarya Sivasurya

Exploratory Data Analysis





- Demand follows a tweedie distribution.
- Traditional and Machine Learning Algorithms – Global Models - on SKUs with less than 80% of records with zero demand
- Naïve Forecasting Algorithm(Traditional) – Local Model - on SKUs with greater than 80% of records with zero demand

Feature Engineering

- Product Causals
- Site Causals
- Trend
- Seasonality
- Categorical features
  - Product code
  - Site code
  - District

- Average Monthly consumption lags (3 - 5)
- Opening Stock Lags (3 - 5)
- Simple Moving Average of Demand Lags (6 - 8)
- Demand Lags (3 - 8)
- Calendar Features
  - Year
  - Month
  - Quarter

Algorithms – Local and Global Models

- Naïve Forecasting (Local Model)
- Croston's Method – Intermittent Forecasting (Local Model)
- Catboost Regressor with Tweedie loss function (Global Model)
- LightGBM Regressor with Tweedie loss function (Global Model)
- Random Forest (Global Model)
- XG Boost (Global Model)

Best Model Selection

- Model with the lowest MASE on test data

Algorithms	Less than 40%	40 - 60%	60 - 80%	40 - 80%	Greater than 80%	Total
Naïve Forecasting	62	49	23		323	**457**
RandomForest Regressor	68	2	2	86		**158**
Catboost Regressor	58	21	15	35		**129**
Croston's Method	75			49		**124**
LightGBM Regressor	35	26	16	28		**105**
XGBoost Regressor	37			36		**73**
Total	**335**	**98**	**56**	**234**	**323**	**1046**

Records with zero demand	Average MASE
Less than 40%	0.61
40 - 80%	0.91
Greater than 80%	0.58