GRUPO BIMBO INVENTORY DEMAND

BY JUAN SERRANO, GIOVANNY MORENO AND NICOLAS GUEVARA

GATHERING DATA

The dataset for this competition was primarily sourced from Kaggle. Additionally we utilized web scraping techniques to augment our data with external variables. Specifically we collected biweekly inflation rates and consumer confidence indices from an API. These additional variables were gathered for the period from March 31st to June 1st 2016.



EXPLORATION DATA

We perform an inspection in the features in the dataset, counting the total of missing values, unique values, and cardinality. Then we proceed to merge the different dataframes in our disposal in one unique big dataset

PREPROCESSING

We change the columns type to tthe one needed, and fill the missing values if its required

FEATURE ENGINEERING

HOW DID WE DECIDE:

Upon exploring the dataset w realized that were several columns that were of the type "object" something that we needed to avoid since a regression ML model was used, also, we also evaluated wich columns were relevant to pass to the model so we could get an accurate prediction.



MODEL BUILDING

ataset in test and training les that were fitted in the GB model and landomForest model

TRAINING MODEL

showed a 28% margin of error

MODEL EVALUATION

MICH MODEL 10 025

Since we need a numerical answer from the data it was decided that a Regression Machine Learning Model. At first we evaluated between Random Forest and XGBoost. After meassuring the performance of each one of then, we decided to opt for the Random Forest model



BUSINESS QUESTIONS

The study aimed to address several business questions elated to inventory nanagement and demand

What are our best and wors

- customers?

 What is the relationship her
- sales growth rates (1 and 2) and adjusted demand?
- How are our sales going against the returns, and our adjusted demand
 What are our most used routes?

REFERENCE

https://www.kaggle.com/competitions grupo-bimbo-inventory-demand/overv ew