Wheaton College Department of Mathematics and Computer Science

Forecasting FMCG Demand Using Generalized Additive Models

Carson Slater

Abstract

AwanTunai provides supply chain business solutions with embedded financing to suppliers and merchants. Here we try to forecast demand for the most popular stock-keeping units (SKU's) given fourteen months of time-series data. We determine the 50 most-popular SKU's by volume of sales. We then use additive regression models to estimate the volume of sales for a future time period of eight weeks and evaluate the results.

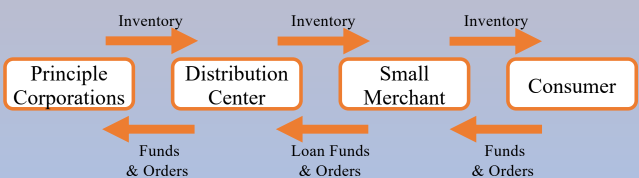
Background

Introduction

Consumer staples in Indonesia flow from principle corporations, to distributors, to merchants, and finally to the consumer. What is atypical about Indonesia is over 50% of FMCG sales come from small, single store merchants, as opposed to large retail outlets (1). Around 80% of the Indonesian population purchases goods from these smaller merchants.

Problem Statement

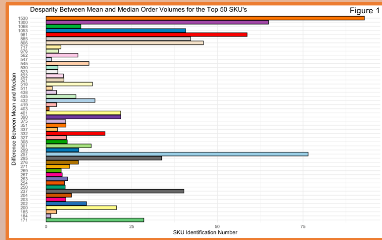
Many of these small merchants have very tight profit margins, making it difficult to restock goods to grow their customer base. AwanTunai offers cloud-based solutions for small merchants to restock their inventories with options to finance their orders. This enables them to keep up with demand, attract new customers, and pay for the goods when they turn a larger profit. AwanTunai's services provide a financial buffer for the "mom and pop" shops during market shocks. We hope to forecast the quantity ordered for the most popular SKU's, to optimize efficiency and mitigate sunk costs for distribution centers and merchants.

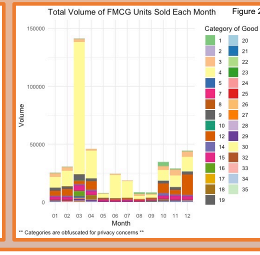


(Left) A smaller merchant taking delivery of an order they placed. (Right) A small merchant store.

The Data

These raw data contain order information for 62,048 orders from a particular distribution center (DC) that uses AwanTunai solutions. This DC serviced 2,138 unique merchants over the fourteen months our data was observed. In cleaning these data we found that there were 178 duplicate entries, which were removed for analysis.





Preprocessing Steps

Selecting "Relevant" SKU's

The goal was to select goods to forecast that were purchased often, and were purchased in higher volumes. The methods we used are known to struggle with low daily volumes. In practice it is also more useful to forecast goods that 'move faster' with larger volumes being distributed more often. Defining the most "relevant" SKU's from the data involved picking the fifty SKU's with the highest total volume of units sold during the sample period with a difference between the mean and median daily order quantities below 100. Figure 1 shows the chosen SKU's and the difference between mean and median daily volumes. This measure was taken to ensure our working data was not highly skewed, and was more representative of the most frequently purchased items.

Observations From Exploratory Data Analysis

We found many Indonesian FMCG's experience seasonal trends. particularly with the Ramadan fast (See Figure 2). During the last few days of March there is a large demand shock with an uptick in purchases and on the last days of April, the same phenomena occurs. There is also an increase in order quantities during the latter months of the year.

Methods

The Models

We explored a variety of time series models, including seasonal naïve models, exponential smoothing models and ARIMA models in R from the `forecast` (2) package, we observed that, for the FMCG data, Facebook's `Prophet` (3) API produced models that best considered seasonal trends. Some of the problems with forecasts from ARIMA and exponential smoothing was how for stabilized data, they would eventually forecast a singular mean value for the entire forecast, which neglects seasonal change points and holidays. From our exploratory data analysis, we know that they FMCG data experiences high volumes in certain seasons. `Prophet`'s generalized additive model (GAM),

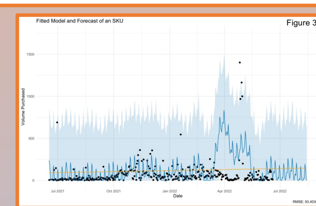
$$\max\{0, y(t_i)\} = \beta_0 + g(t_i) + s(t_i) + h(t_i) + \epsilon_i,$$

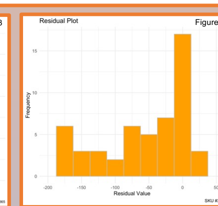
is flexible, robust to seasonality, but because of its additive nature can miss non-linear interactions between predictors. In the model,

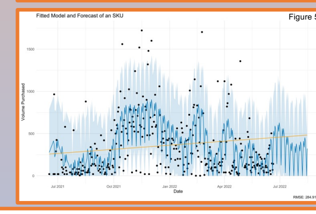
- $g(t_i)$ refers to trend (changes over a long period of time)
- $s(t_i)$ refers to seasonality (periodic or short term changes)
- $h(t_i)$ refers to effects of holidays in the time series.

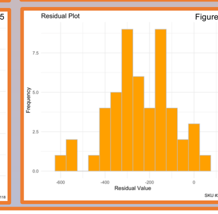
Results

We produced forecasts for eight weeks and found the models often overfit the data. Figure 3 shows the data and forecast of an SKU with low daily volumes, and Figure 4 shows the eight weeks of residuals are negative. Additionally, Figure 5 shows an SKU with high volumes of daily purchases, and the residuals also take on negative values. `Prophet`'s GAM performed best when trained on at least a full year of data, for SKU's that were purchased often, and through non-automated purchasing. In practice, it may be better to use the distribution when making inventory decisions to prevent shortages, such as the 75% upper confidence threshold estimate.









(Top) An SKU with lower daily volumes. (Bottom) An SKU with higher daily volumes.

Discussion

Future research might include additional exploratory data analysis to pick up on hidden trends in the data. Additionally, further research might experiment with data transformations to assess the suitability of the algorithm. These models show, especially from the plots, promising results to be able to adapt to shocks and seasonal trends. Lastly, although we can forecast univariate distributions, subsequent research should explore more power algorithms, such as Convolutional Neural Networks for multivariate probabilistic forecasting.

References

1. USDA Foreign Agricultural Service. (July 5, 2022). Retail sales value of grocery retailers in Indonesia in 2021, by type (in billion U.S. dollars) [Graph]. In *Statista*. Retrieved September 22, 2022, from https://www.statista.com/statistics/1228382/indonesia-grocery-retailers-sales-value-by-type/.
2. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2022). _forecast: Forecasting functions for time series and linear models_. R package version 8.17.0, <URL:https://pkg.robjhyndman.com/forecast/>.
3. Sean Taylor and Ben Letham (2021). prophet: Automatic Forecasting Procedure. R package version 1.0. https://CRAN.R-project.org/package=prophet.
4. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Acknowledgements

Special thanks to The Wheaton College Alumni Association, Wheaton College, and AwanTunai.

Scan to go to the Github Repository:

Adviser: Dr. Ying Li