8.3 Final Project Step 1

Amelia Farrell

October 25th 2021

Topic: Supermarket Shrink - How supermarkets can reduce waste by ordering the right amount at the right time.

The problem with shrinkage in supermarkets can be seen in every department. From transport to checkout, dairy to cereal. According to research by the FMI and The Retail Control Group, 64% of store shrink can be traced back to ineffective store operating practices (wheresmyshrink.com, 2012). The highest percentage of this comes from Ordering and Production Planning inefficiencies. Meaning that if we can simply order better and plan to stock the right items, we could reduce 64% of supermarket shrink! However, we know it is not as simple as that. We need to break this large task into smaller more manageable pieces in order to build ourselves a starting place to solving this wide spread problem. Meat and Produce the largest contributors to shrink from perishable goods. For the purposes of this assignment (and my own troubles with keeping purchasing fruit right before it goes bad), we will be focusing on the shrink steaming from the produce section of the standard American supermarket.

Research questions

1. How much does the average grocery store lose in fresh fruit/vegetables per year (in $ value and pounds)?
2. What are the causes (5 Whys) behind the high shrinkage/waste from produce in grocery stores?
3. What produce produces the most amount of waste (only waste occurring within the supermarket, does not include farming, transit, customer waste)?
4. How are the high waste items correlated with one another?
5. Is there any relationship to the % of waste (pounds of produce lost) and the time of year? Seasonality.
6. Are there any correlations between the amount of produce sold and other non-perishable goods? E.g. does cereal sales increase at the same rate as bananas?
7. Can we make any predictions on what and how much produce to stock based off other predictors?
8. How can we use the information we gathered from the above questions to make better stocking decisions and reduce shrinkage for produce?

Probem Statment: How can we decearse waste (shrink) in the produce department?

Approach (Plan to address this problem)

How do we plan to address this problem statement? By first understanding what specific produce is contributing to 80% of the produce shrink (Pareto chart). I.e. Can we make the greatest impact by focusing on just the top shrink items? Next we will want to understand why the shrink/loss is so high for these items (is this due to inconsistent buying patterns of customers? Or simply an extremely short shelf life of the good, making it more difficult to stock correctly). Next we will want to look for customer buying patterns to try to better stock these high shrink items and reduce waste. Each portion will require data scrapping from multiple sources and identifying trends using the skills we have learned in this class.

How does this proposed approach address the problem?

It addresses the problem by identifying the top contributors to shrink, uncovering patterns in sales for these items, and stocking the right amount at the right time in order to decrease the amount loss due to the short shelf life of produce.

3 datasets to address the issue.

1. Food Loss and Waste Database - The Food Loss and Waste Database is publish by The Food and Agriculture Organization (FAO). It was created to help anyone with the interest in learning more about food waste. This data base provides us with the option to pull food loss data for the specific country (we will be looking at the U.S.A only) and value chain (Retail for this analysis). The data then provides you with the source of the data (each line item may come from a different source) and 21 other variables (however, many are blank and will need to be removed) Link: https://www.fao.org/platform-food-loss-waste/flw-data/es/
2. Groceries dataset (Kaggle) - This data set was published by Kaggle and holds 38765 lines of orders from customers at grocery stores. The original purpose of this data was for Market Basket Analysis. It consists of 3 variables (Customer, Date, Product). This will help us identify trends in buying patterns for produce.There are no details on when/how was it collected. Link: https://www.kaggle.com/heeraldedhia/groceries-dataset
3. Shelf Life - This data set was published by fightbac.org and developed by the Food Marketing Institute along with Cornell University. It was originally created to help individuals select and store perishable goods from the grocery store. We will be using this document to scrape the average shelf life of the different product categories. Once we scrape the data from the PDF, we will have two variables, the product (fruit/vegetable) and it's average shelf life. Link: https://lee.ces.ncsu.edu/wp-content/uploads/2012/12/TheFoodKeeper.pdf?fwd=no&fbclid=IwAR2QE_yWd_E6kzD7Sp18AnLN36h7uLPpmM7CrsUZC91OQz_pHi_hT3jZvBU

Required Packages

We will be using R for our analysis. While we could spend hours had writing code for every transformation, plot, and statistical test in R, we will be using some additional packages to aid in this process. There may be package that we realize could be of great use during our analysis, at the very minimum we will need the below packages from R in order to make our analysis a bit easier.

- library(ggplot2) - ggplot2 will aid us in discovering and visualizing trends/findings in out data.
- library(GGally) - GGally is an extension of ggplot2 that helps reduce the complexity of combined objects.
- library(ppcor) - ggplot2 will help us identify relationships between variables through pairwise partial correlations.
- library(purrr) - purrr provides us with a complete kit for working with functions and vectors within our data sets.
- library(car) - The car package will be one of the final packages needed. This package will only be used if we find that we can make any liner predictions based off our data sets. This can be used to run an analysis-of-variance on liner models. Depending on what we find within the data sets chosen, this package may or may not be used for this analysis.

Plots and Table Needs

We have learned how to create many different visualizations in this course to help us better understand our data and uncover trends/relationships. I think a few of the important ones that we will use in this analysis will be the below;

- Histograms - for understanding the the distribution of the produce waste by type.
- Scatter plots - for visualizing relationships between all variables.
- Bar charts (Pedro chart) - for identifying top areas (specific produce fruit/vegetable) of focus/impact.
- Line charts - for identifying seasonality in sales data.
- Q-Q Plot - for assessing the performance of a model.

Questions for future steps

After going through what we have and need in order to begin this analysis, what are some of the things we don't know how to do that we need in order to answer our problem statement? In this course we have learned how to, look at the data, transform it, identify relationships, skewed distributions, and make linear predictions. But what about identifying seasonality? Time-series.In Research question 5, I want to look for possible seasonality in shoppers buying habits. We have not reviewed this so far in the course so I will need to do a bit of self guided learning to answer this question.

References

wheresmyshrink.com, 2012. Executive Summary. http://wheresmyshrink.com/executivesummary.html?fbclid=IwAR0w7KKjS-4Lr1wJ3JuJ2ZYbsZGZbc57Go4NuBinNwytYNG5911QUBtXXYE.

FAO, 2021. Food Loss and Waste Database. The Food and Agriculture Organization (FAO). https://www.fao.org/platform-food-loss-waste/flw-data/es/.

Dedhia H., 2020. Groceries dataset. Kaggle.com. https://www.kaggle.com/heeraldedhia/groceries-dataset

Food Marketing Institute & Cornell University, 2020. The Food Keeper. fightbac.org. https://lee.ces.ncsu.edu/wp-content/uploads/2012/12/TheFoodKeeper.pdf?fwd=no&fbclid=IwAR2QE__yWd__E6kzD7Sp18AnLN36h7uLPpmM7CrsUZC91OQz__pHi__hT3jZvBU