Grocery Association Rules

Allison Fischer
allisonfischer@lewisu.edu
DATA-51000001, Summer 2021
Data Mining and Analytics
Lewis University

I. Introduction

Retail sales, including grocers, are continuously seeking opportunities for increased customer satisfaction as well as increased sales. An important factor in increasing sales and satisfaction is understanding. A company has the need to understand customer desires and indeed predict customer wants in order to best serve the customer. The analysis of customer transactions is an important step in furthering understanding. I sought to provide a better understanding of customer behavior for the purpose of predicting future behaviors. This analysis involves grocery transaction history by product and product category [1]. With Orange 3 data visualization and analysis software, I have used association rule mining to investigate associated purchases [2].

In the future sections, I will describe the data, methodology, results, and conclusion. In section II, I provide the data used and descriptions. In section III, I present the analysis methodology. In Section IV, I describe the results and analysis is discussed. Finally, in section V the conclusions are discussed.

II. Data Description

The data examined includes products purchased over a one-month period. The data includes 1,361 transactions and 255 unique items. Data attributes can be seen in Table I. The data is formatted for association rule mining, in that the presence of the purchase is indicated along with the item. White bread was the most frequently purchased item. Celery, and Oats and Nuts Cereal were the least frequently purchased items (Fig. I)

\

TABLE I. Data attributes

Attribute	Type	Example Value	Description
PRODUCT	Nominal (string)	Frozen Chicken Thighs	Name of product
TRANSACTION ID	Nominal (string)	1	Transaction ID number
Purchased	Nominal (string)	1	Whether item was purchased

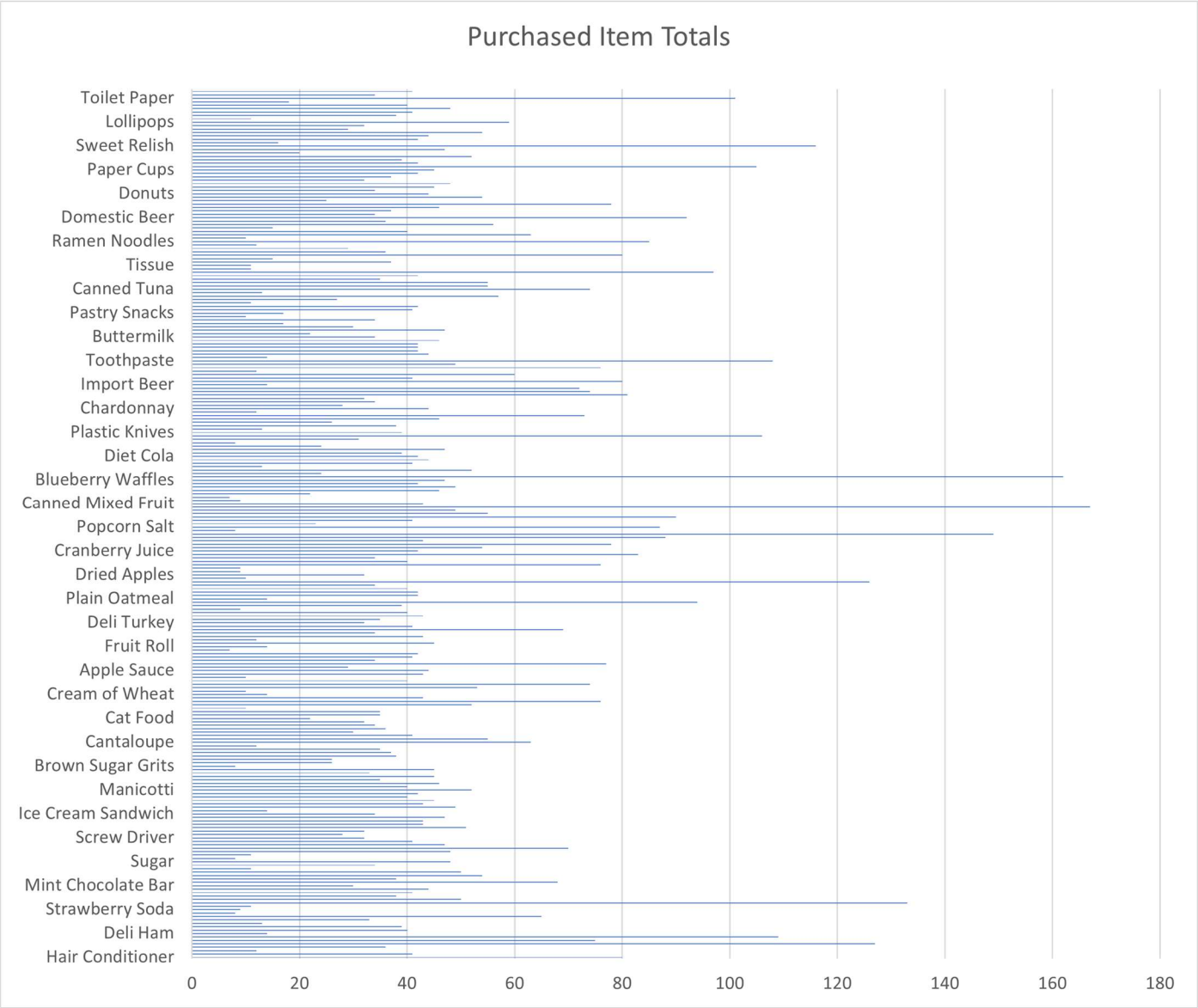

Fig I. Totals of Unique Items Purchased

III. Methodology

First, the data was checked for suitability of content and format. Next, a bar graph was created to visualize the totals of each purchased item (Fig I.). Association rules were run at low support and low confidence to ensure the data had associations. Next, support and confidence were adjusted to seek meaningful and interesting associations. Scatterplots were created to visualize results.

IV. Results and Discussion

The highest support for this dataset was 0.035 for toothpaste purchased when Cola was purchased (Table II).

TABLE II. TOP 5 ASSOCIATION RULES BY SUPPORT

Support	Confidence	Coverage	Strength	Lift	Leverage	Antecedent	Consequent
0.035	0.443	0.078	1.019	5.588	0.028	Cola	Toothpaste
0.035	0.435	0.079	0.981	5.588	0.028	Toothpaste	Cola
0.035	0.500	0.069	1.585	4.567	0.027	Pepperoni Pizza - Frozen	2% Milk
0.035	0.315	0.109	0.631	4.567	0.027	2% Milk	Pepperoni Pizza - Frozen
0.035	0.534	0.065	1.841	4.487	0.027	Bologna	White Bread

The highest confidence for this dataset was 0.496 for eggs purchased when potato chips were purchased (Table III).

TABLE III. TOP 5 ASSOCIATION RULES BY CONFIDENCE

Confidence	Support	Coverage	Strength	Lift	Leverage	Antecedent	Consequent
0.496	0.048	0.098	1.256	4.044	0.037	Potato Chips	Eggs
0.494	0.031	0.062	1.753	4.513	0.024	Ramen Noodles	2% Milk
0.489	0.034	0.069	1.723	4.111	0.026	Pepperoni Pizza - Frozen	White Bread
0.489	0.032	0.066	1.478	1.478	0.026	Tomatoes	Potato Chips
0.489	0.032	0.066	1.656	1.656	0.025	Tomatoes	2% Milk

The highest lift was 5.540 for sweet relish purchased when toothpaste was purchased (Table IV).

TABLE IV. TOP 5 ASSOCIATION RULES BY LIFT

Lift	Support	Confidence	Coverage	Strength	Leverage	Antecedent	Consequent
5.540	0.037	0.472	0.079	1.074	0.031	Toothpaste	Sweet Relish
5.540	0.037	0.440	0.085	0.931	0.031	Sweet Relish	Toothpaste
5.517	0.032	0.677	0.048	2.569	0.026	White Bread, Toothpaste	Eggs
5.517	0.032	0.263	0.123	0.389	0.026	Eggs	White Bread, Toothpaste
5.433	0.037	0.669	0.055	2.227	0.030	Sugar Cookies	Eggs

Using the scatterplot of Confidence vs. Lift (Fig II), the interesting association of toothpaste and sweet relish can be visualized. The association of toothpaste and sweet relish has confidence of 0.472, support 0.037 and lift 5.54.

Fig II. Confidence vs. Lift.

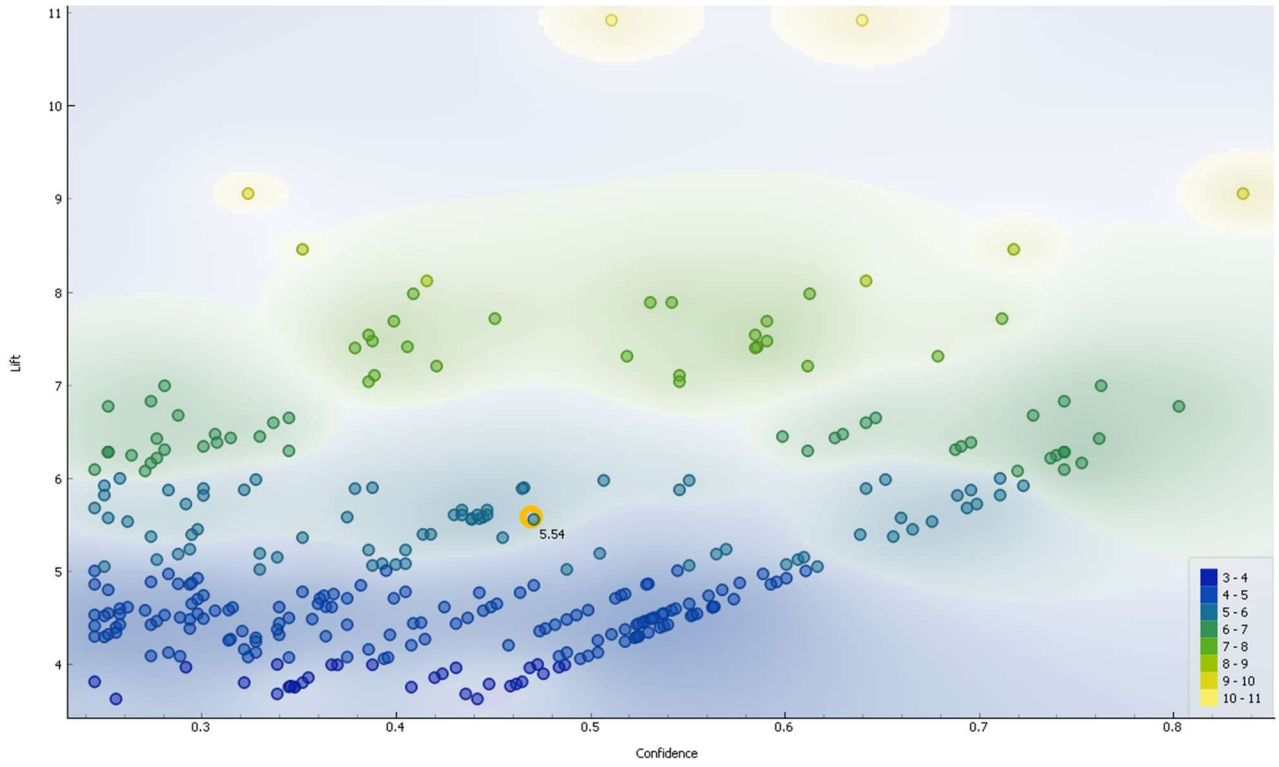

Fig III. Confidence vs. Lift.

V. Conclusions

The most interesting association rule found was purchase of sweet relish when toothpaste was purchased. This association occurred 47.2% of the time. Another notable association was the purchase of 2% milk when Ramen Noodles were purchased. This occurred with a 0.494 confidence. This information can be used for future product placement, as well as targeted couponing and product replacement suggestions.

[1] G. T.G, "Grocery Store Data," *Kaggle*, 24-Jan-2021. [Online]. Available: https://www.kaggle.com/gustavotg/grocery-store-data. [Accessed: 14-Jun-2021].

[2] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, 1993.