Seboca AI

Grocery Data Analysis

OVERVIEW

The aim of this analysis is to perform data exploration on a Groceries dataset[1] and provide actionable insights based on the results. The aforementioned dataset has one month worth of recorded transactions containing "baskets" of various sizes. Insights and suggestions are provided at each step of the data exploration process.
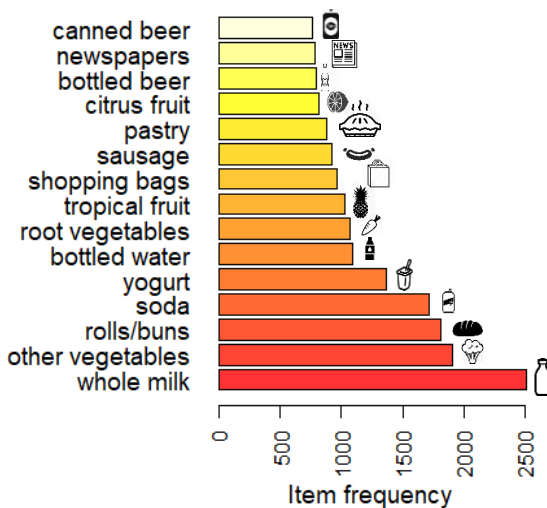
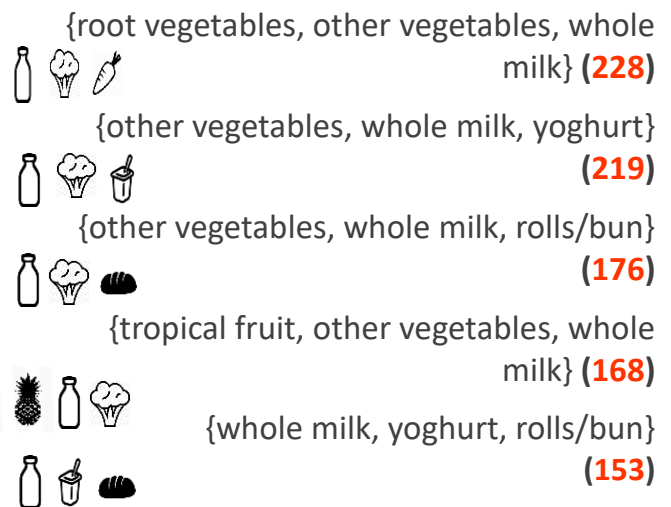**DATA CONTENTS**

Transactions	Grocery items	Item categories (sub-categories)
9835	**169**	**10 (55)**

FREQUENT ITEMS

TOP 5 FREQUENT BASKETS (NUMBER OF OCCURRENCES)

{root vegetables, other vegetables, whole milk} **(228)**

{other vegetables, whole milk, yoghurt} **(219)**

{other vegetables, whole milk, rolls/bun} **(176)**

{tropical fruit, other vegetables, whole milk} **(168)**

{whole milk, yoghurt, rolls/bun} **(153)**

FINDINGS

- *Whole milk* tops the list by appearing in just over 2500 grocery baskets
- Items from the *fruits and vegetables, fresh produce* and *drinks* (alcoholic and non-alcoholic) item categories frequent the list

- Most common basket size had **3** grocery items
- Majority of baskets will at least contain *milk, yoghurt* or a *fruit and vegetable.*

INSIGHTS

The grocery store should consider discounting some of these items individually or through a bundle deal to attract potential customers

The grocery store layout should be such that the *fruit and vegetables, drinks* and *fresh produce* sections should be located relatively close to each other for optimise shopping experience.

Grocery Data Analysis

Seboca AI

Executive Summary

HIGHLY PROBABLE ITEM COMBINATIONS

Findings

- It is more likely to have **sugar** than **butter** in a basket that already contains **rice** and **whole milk**

- Combinations including **vegetables** will most likely include a fruit as well as a dairy produce

- **Whole milk** can be bought with **non-food** items such as **hygiene articles**

Suggestions

- Dairy produce section (more particularly whole milk) should be centrally located within the grocery store

Probability of finding combination
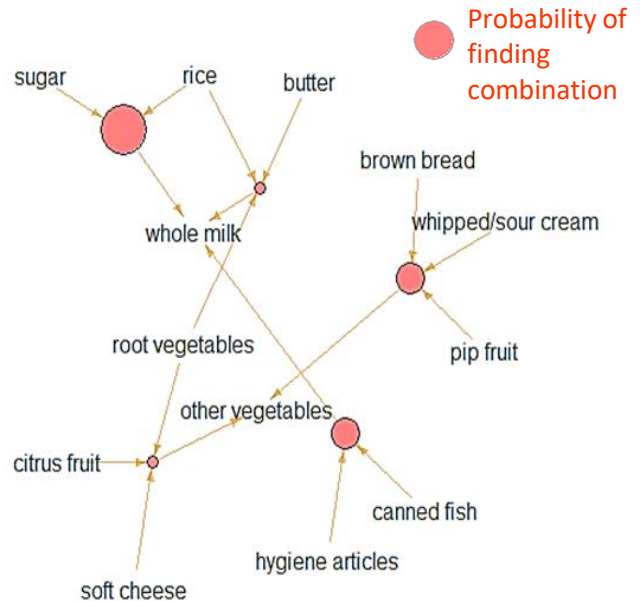
COMBINATIONS OF FREQUENT ITEMS

Findings

- A basket with;
  - **Whole milk**
    - has a high probability of including a meat product
    - Has small but significant chance of including items used for baking or already baked in store
  - **Vegetables (other and root)** will contain
    - A dairy produce and another fruit or vegetable
    - Items, such as rice and oil, used for home-cooked meals
  - **Yogurt**
    - Will contain a dairy produce

Suggestions

- The **meat** section needs to be close to the **fresh/dairy produce** section
- The aisle that houses the family meal ingredients (such as rice and cooking oil) should be in close proximity to that of the **fruits and vegetables** section

Probability of finding combination

3 rules: {oil, rice, +4 items}
4 rules: {sliced cheese, margarine, +7 items}
47 rules: {whole milk, grapes, +33 items}
88 rules: {sugar, pastry, +42 items}
164 rules: {hamburger meat, chicken, +55 items}

{root vegetables}
{yogurt}
{other vegetables}
{whole milk}

CONCLUSION

The report provides numerous insights, which would assist in designing an optimal layout of the store, based on the transaction data of customers.