Introduction

Project Overview

In the realm of modern commerce, the digitization of transactions has become abundant, providing businesses with invaluable insights into consumer behavior. Leveraging this data, businesses can refine their sales strategies to stay competitive in the market. For our project, we chose to analyze sales the realm of a Scottish bakery, utilizing association analysis techniques such as Support, Confidence, Lift metrics, and FP-Growth algorithm to uncover patterns among products in a dataset belonging to a bakery in Edinburgh, Scotland, spanning from December 2003 to January 2011."

Dataset Overview

The dataset our group chose contains transaction details of customers who ordered various items online from the bakery. Acquired via Kaggle.com, this dataset serves as the foundation for our analysis, and gives us information such as the number of products purchased, which products were purchased, and at what days and times they were purchased. With over 20,000 rows and 9,000 transactions, the dataset offers a variety of customer purchasing preferences and behaviors, without being overly-complex in how descriptive the specific sales were when they are recorded.

Motivation Overview

In the fast-paced bakery industry, understanding product associations is paramount for optimizing sales strategies, since pastry-related foods can expire so rapidly, with some items within our dataset needing to sell the day that its made. This make concepts such as shrinkage, the term used within the grocery industry representing how much percent of your product is expiring out, rather than selling out, extra important within a bakery-setting, and thus exciting to choose to focus on. Through discerning which products tend to sell together, businesses can restructure production schedules, selling techniques, minimize their shrink, and enhance customer satisfaction. For instance, the urgency to sell perishable items like donuts and cakes requires thoughtful planning of production and sales, which can be driven by datamining sales datasets such as our own.

Basic Objectives

In our project, we will be addressing two fundamental questions through data mining techniques:

1. Which products exhibit strong associations in sales?
2. Which products should be grouped in combinations and sold together to maximize revenue and minimize waste?

Challenges/Technical Approach Overview

Navigating through our chosen dataset presents several challenges which we faced, including a lack of detailed product categorization and inconsistencies in recording similar items. To mitigate these challenges, we employed data preprocessing techniques using Python's Pandas library. We were required to use preprocessing techniques in extracting attributes such as date, time, and day of the week into our program, and created visuals that help us break down the results of our dataset analysis easier, with attributes such as date, time, and day labeled in the charts.

Our approach involves using a variety of methods to dig into our data to find relevant information. The data is preprocessed, and then we subsequently employ association analysis techniques such as Support, Confidence, Lift metrics, and an FP-Growth algorithm to uncover associations and patterns among items sold from our Scottish bakery.

In clarifying the connections between bakery products and understanding how consumers behave when choosing to purchase them, we can optimize the sales strategies and minimize waste in the landscape of a bakery. Moreover, this same approach could be expanded to other industries of modern commerce as well with few adjustments needed, showing that our approach of Bakery Sales Analysis can be used for other commerce industries with similar sales techniques, even if those industries and businesses are not necessarily bakery themselves.