**MARKET BASKET ANALYSIS**

**SIMULATED COLES DATA**

**Table of Contents**

[Executive Summary 1](#_Toc32067897)

[Introduction 1](#_Toc32067898)

[Description of the data set 2](#_Toc32067899)

[Original data 2](#_Toc32067900)

# **Executive Summary**

This report provides an analysis of the Simulated Coles data which consists of 58,100 observations and 53 variables. The aim of the analysis is to explore questions about which products customers are buying and what kind of customers are shopping at Coles. The answers to these questions can be used to enhance customer experience and/or to maximize profits which can be achieved by improving store layouts, offering discounts or marketing different products together, and determining the best place to shelf the specific products.

After cleaning and preprocessing the data, Clustering analysis and Market Basket analysis have been applied to the data. Clustering analysis results revealed that there 4 groups of customers based on income, age and spending value. These 4 groups are named **Elderly High Spenders, High Spending Youth, Young & Broke and Rich & Wise Youth. These groups are named based on their characteristics of the customers in the groups. Market Basket Analysis results uncovered that there is a high correlation between products like bread, milk & banana and nappies & baby food. Some less obvious product combinations were fish, vegetables & household cleaners and coffee, bread, vegetables, frozen meals & household cleaners.**

**This report also offers a detailed explanation of data cleaning and analytical processes, data mining techniques, analysis outcome, and suggestions for future analysis.**

# **Introduction**

Coles is one of the leading supermarkets in Australia, generally dealing with the provision of everyday goods. The emergence of competitors in the market has meant that the competition has become much fiercer and It is crucial for Coles to take strategic actions to beat the competition. One way to do that is to use the data Coles has been collecting from its customers and analyze data to gain a competitive edge. This report aims to use unsupervised knowledge discovery to find hidden patterns in customers’ transactions and to form customer segments. We’ll be seeking a better understanding of two aspects:

* **Customer Segments**: - What are the different segments or groups and their demographic as well as socio-economic features.
* **Products Patterns**: - Popularity of the products, what kind of products are being bought together, the likelihood of two or more products being purchased together.

By better understanding these two aspects and finding answers to the questions related to these two aspects, Coles would be able to form sophisticated strategies that would help Coles to lead the competition. A better understanding of the customer segments would help Coles create more effective and successful market campaigns and design eye-catchy catalogs of the products whereas better knowledge of product patterns would allow coles to improve in-store layouts and product placements.