MARKET

ESPRIT Project Number EP 24456

SPSS

IT Innovation

Somerfield Stores

D5.4 Final Demonstration System Documentation

Paul Allen and Steve Taylor
IT Innovation
1 November 1999

Executive Summary

This document D5.4 “Final Demonstration System Documentation” describes the public demonstrator for the EC Project MARKET (EP 24456). The MARKET project started on September 1 1997 and ended on June 30 1999. The MARKET project includes the following partners: SPSS Inc (formerly ISL Ltd, UK), IT Innovation Centre (formerly PAC), UK and Somerfield Stores Ltd, UK.

Data mining is a maturing technology and application area. However, there are few public case studies demonstrating the power and benefit of the technique. In the MARKET project we set out to demonstrate how non-IT specialists using data mining techniques can derive business benefit from a large corporate data warehouse.

Within the supply chain department, Somerfield have investigated the bread ordering process to improve bread availability and reduce wastage. Based on the experience of Somerfield we have implemented a public demonstrator as a set of interactive Web HTML pages.

The demonstrator deals with a fictitious supermarket chain "TTN Supermarkets Ltd". The issues in this demonstrator are real; however, specific data, results and conclusions are fictitious and are independent of the specific conclusions reached by Somerfield.

The demonstrator has been designed to be viewed by a business user in approximately 15 minutes and describes the business problem, the data understanding, data preparation, modelling and deployment stages of the data mining process.

This report describes the demonstrator, showing the power and benefits of data mining. In particular, it shows how “TTN Supermarkets Ltd” achieve ROI through reducing wastage costs and increasing revenue due to increased sales and increased customer satisfaction.

1 Introduction

This document D5.4 “Final Demonstration System Documentation” describes the public demonstrator for the EC Project MARKET (EP 24456). The MARKET project started on September 1 1997 and ended on June 30 1999. The MARKET project includes the following partners: SPSS Inc (formerly ISL Ltd, UK), IT Innovation Centre (formerly PAC), UK and Somerfield Stores Ltd, UK.

2 Approach

Data mining is a maturing technology and application area. However, there are few public case studies demonstrating the power and benefit of the technique. In the MARKET project we set out to demonstrate how non-IT specialists using data mining techniques can derive business benefit from a large corporate data warehouse.

Within the supply chain department, Somerfield have investigated the bread ordering process to improve bread availability and reduce wastage. Based on the experience of Somerfield we have implemented a public demonstrator as a set of interactive Web HTML pages.

The demonstrator deals with a fictitious supermarket chain "TTN Supermarkets Ltd". The issues in this demonstrator are real; however, specific data, results and conclusions are fictitious and are independent of the specific conclusions reached by Somerfield.

The demonstrator has been designed to be viewed by a business user in approximately 15 minutes and describes the business problem, the data understanding, data preparation, modelling and deployment stages of the data mining process.

3 Demonstrator

3.1 Background

TTN Supermarkets Ltd is investigating their process for ordering bread. The current process is based on demand which is well understood under normal operation. However, the demand for bread is complex during promotional periods.

Bread is particularly significant because:

- Bread has a short shelf life. Over-ordering can lead to increased wastage and loss of revenue.
- Bread is a staple product and is contained in 75% of baskets. Under-ordering can lead to loss of sales, customer satisfaction and market share.

TTN Supermarkets Ltd require more advanced ordering methods for bread.

There are two main effects that cause the demand of bread to be complex during promotional periods: domino and steal. The concepts of domino and steal effects are already intuitively understood.

Steal is when sales of a particular product are affected due to a promotion on another similar product.

Domino is when customers switch to an alternative product if their first choice product is unavailable. This may cause the alternative product to sell out, shifting demand to a third product, and so on.

However these effects are not quantitatively understood and therefore not usable by the bread ordering process.

3.2 Introduction

This demonstrator shows the results of the data mining process used by TTN Supermarkets Ltd in improving its bread ordering process. TTN Supermarkets Ltd have four months of raw basket-level data and are using the Clementine tool from SPSS for the data mining. Clementine has a visual interface in which data transformations and analyses can be performed by building up graphical icons into 'streams'. This process is demonstrated in the following pages.

There are four phases to the demonstrator.

- Data understanding
- Initial modelling
- Further exploration
- Final results

3.3 Data Understanding

An initial exploration is performed in order to better understand the data. Figure 1 shows a number of operations that have been carried out on the raw data in order to understand the data better.

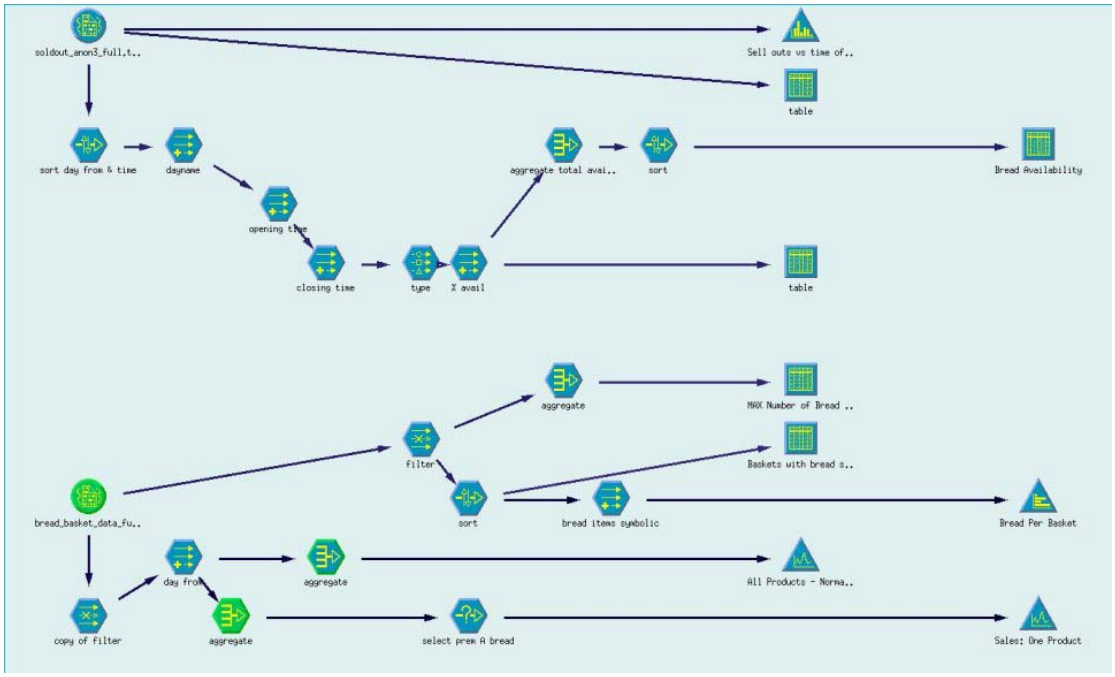


Figure 1 Clementine Stream for initial data exploration

Click on the icon "Bread Availability" to see the initial state of bread availability on the shelf at TTN Supermarkets Ltd

Click on the icon "Bread Per Basket to see the distribution of bread products amongst different baskets

Click on the icon "Sales: One Product" to see the sales of one bread product over time.

3.3.1 Bread Availability

Figure 2 shows the list of major bread products sold by TTN Supermarkets Ltd and the percentage of time that these products are available on the shelves. TTN Supermarkets Ltd does not record this information directly and has no conventional way of measuring it. Clementine has been used to analyse the sales data and detect patterns where no sales have been made but substantial sales are expected. These patterns have been used to deduce the availability shown here.

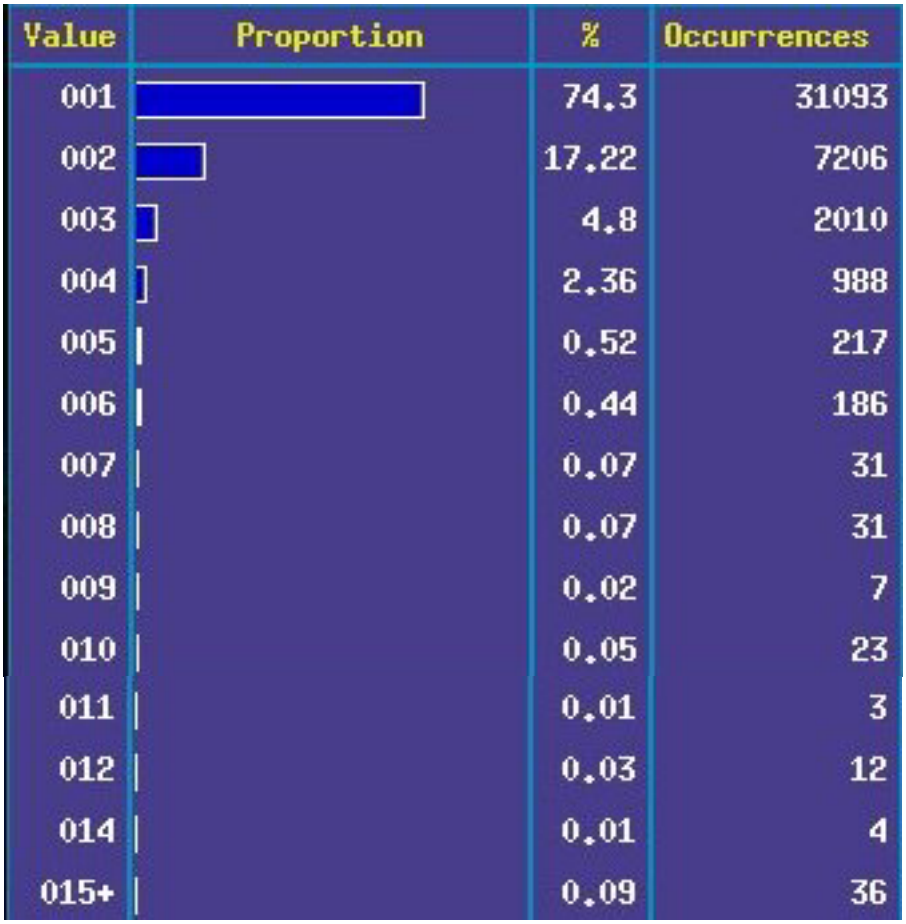
Product Description	% avail
SUPERMARKET BRAND BASIC WHITE BREAD	84.285
SUPERMARKET BRAND WHITE BREAD THICK	81.843
SUPERMARKET BRAND WHITE BREAD MEDIUM	80.952
PREMIUM A WHITE BREAD MEDIUM	78.082
PREMIUM B PREM WHITE BREAD THICK	76.943
PREMIUM A WHITE BREAD THICK	76.264
PREMIUM B PREM WHITE BREAD MEDIUM	74.791
PREMIUM A FARMHOUSE GOLD	72.987
PREMIUM A SQ CUT THICK	71.125
PREMIUM A SQ CUT BREAD PREMIUM	63.844
PREMIUM B HALF LOAF WHITE	63.805
SA GRAIN WHITE BREAD MEDIUM	58.824
SUPERMARKET BRAND BB PREM WHITE BREAD THICK	58.756
SUPERMARKET BRAND BB PREM WHITE BREAD MEDIUM	58.741
PREMIUM A PREM WHITE BREAD MED	55.539
SUPERMARKET BRAND BB TRAD BREAD THICK	53.702
SUPERMARKET BRAND BB TRAD BREAD MEDIUM	52.2
PREMIUM B FARMHOUSE	52.098

Figure 2 Bread availability at TTN Supermarkets Ltd

TTN Supermarkets Ltd are not happy with these availability figures and aim to improve them, as we shall see.

3.3.2 Bread Distribution In Baskets

Figure 3 shows how many bread products people buy. Only baskets that contain at least one bread product are shown.



Value	Proportion	%	Occurrences
001		74.3	31093
002		17.22	7206
003		4.8	2010
004		2.36	988
005		0.52	217
006		0.44	186
007		0.07	31
008		0.07	31
009		0.02	7
010		0.05	23
011		0.01	3
012		0.03	12
014		0.01	4
015+		0.09	36

Figure 3 Bread distribution in baskets

It can be seen that when people buy bread, nearly three-quarters buy only one bread product, and 99% buy five or fewer bread products. There is a small, yet non-negligible figure of 36 occurrences of 15+ bread products in a single basket. This 'wholesale purchase pattern' was investigated by TTN Supermarkets Ltd and was found to be a contributing factor to unpredictable demand and sell-outs. This was an early result from the data mining process.

3.3.3 Product Sales Over Time

Figure 4 shows the sales of one bread product over a period of about 4 months. We can see that the graph is very unpredictable and that there does not appear to be a consistent pattern.

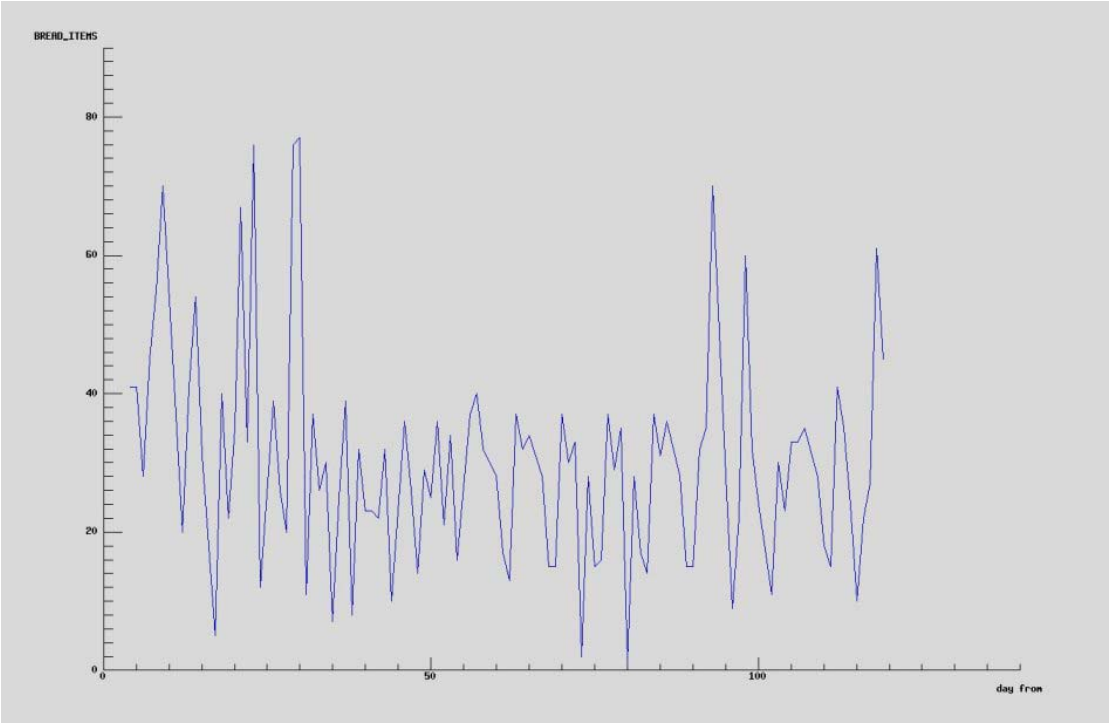


Figure 4 Product sales over time

The business specialist notes that this graph contains some 'abnormal' periods such as:

- Easter, where demand is unusually high;
- periods where the product in question was on promotion;
- days where one customer purchased very large quantities of the product; and
- days where the product sold out at some point.

We hypothesize that if we remove these 'abnormal' periods we may be able to better predict demand.

3.4 Initial modelling

Following the data exploration, an initial model is produced which is valid during non-promotional periods. Figure 5 shows the Clementine stream that filters out the days on which 'abnormal' sales occurred, such as:

- Easter, where demand is unusually high;
- periods where the product in question was on promotion;
- days where one customer purchased very large quantities of the product; and
- days where the product sold out at some point.

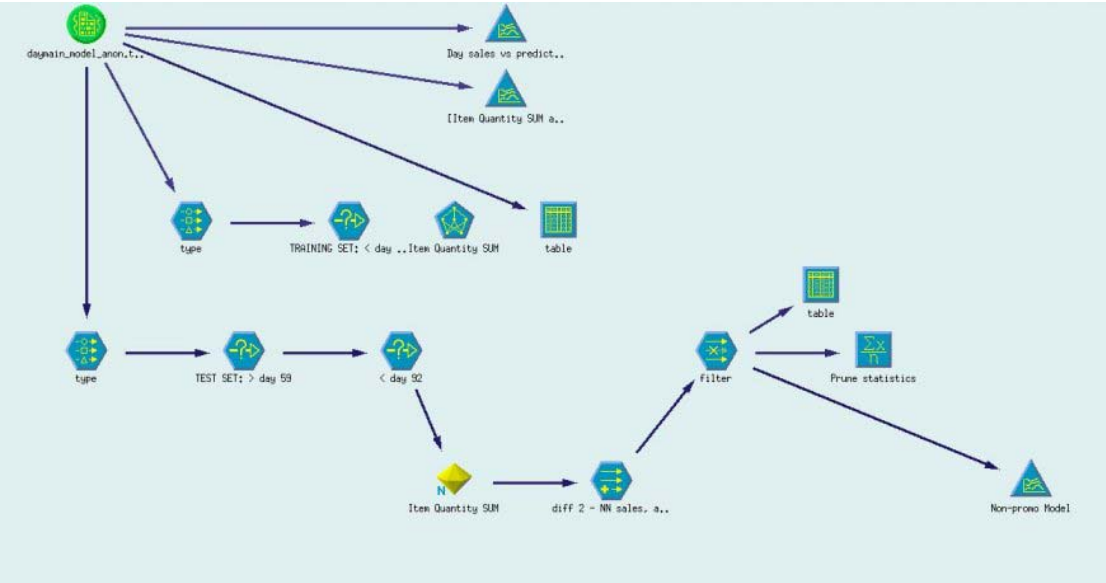


Figure 5 Clementine Stream for non-promotion neural network modelling

The sales demand is then modelled using a neural network. Click on the icon "Non-promo Model" to display the final result.

3.4.1 Modelling Bread During Non-promotional Periods

Figure 6 shows the 'cleaned' sales data along with the output of the neural network model.

- The actual sales are shown in blue
- The neural network output is shown in red
- The error between the neural network and the actual sales is shown in green

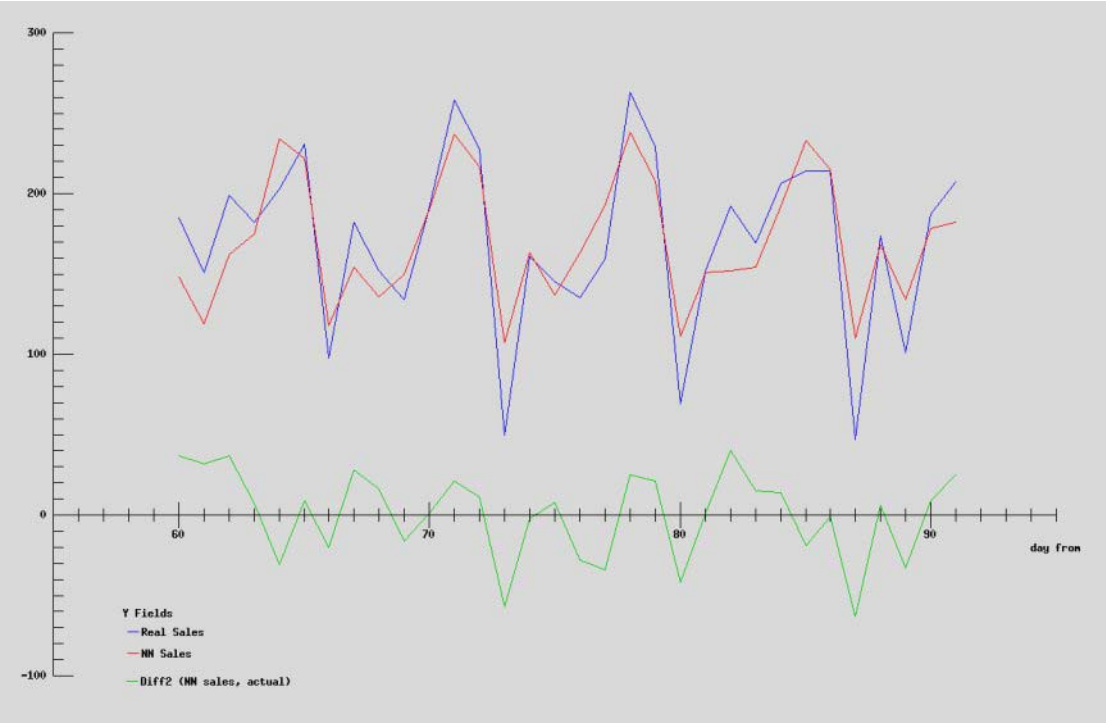


Figure 6 Neural network output

We have now verified that we can model the sales data during non-promotional periods. This is a good first step, but we also need to model the promotional periods. We will now focus in on the domino and steal effects.

3.5 Further exploration

Now that we have a non-promotional model we need to investigate domino and steal effects.

3.5.1 Domino stream

Figure 7 show the Clementine stream that analyses the sales data where sell-outs occurred. An 'association rules' algorithm has been used to derive the relationships between different products.

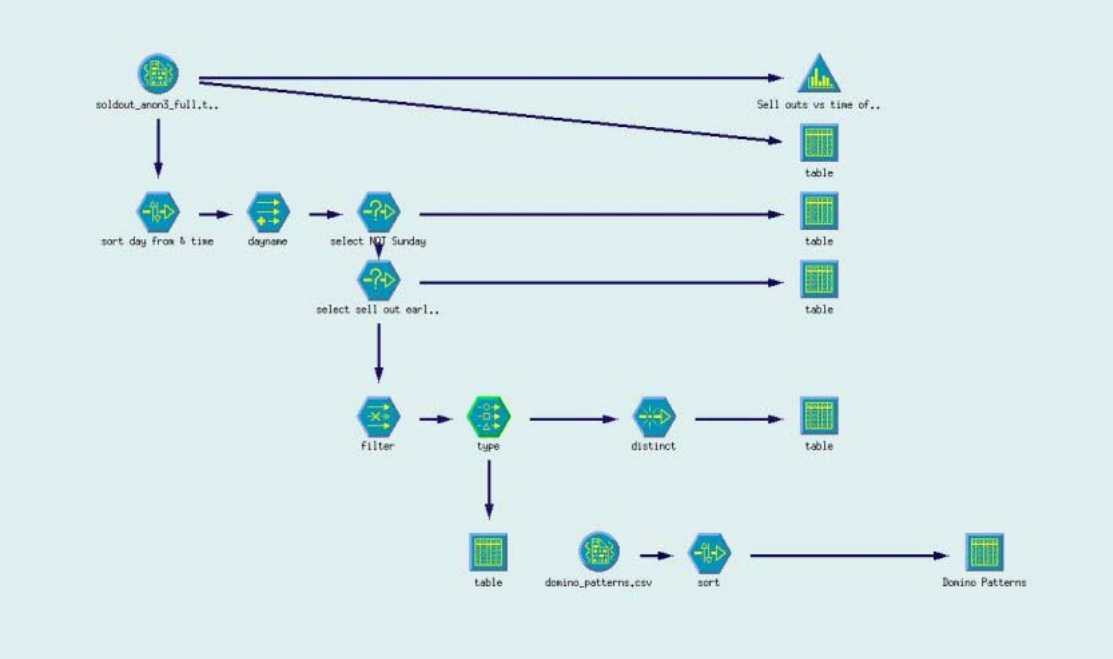


Figure 7 Clementine Stream for domino pattern analysis

Click on the icon "Domino Patterns" to display the most frequent 'domino chains' that occur in the period in question.

3.5.2 Domino Patterns

Figure 8 shows the most common 'domino chains' that have occurred during the period, listed in reverse order of frequency. It can be seen for example that the most common product to sell out first is "Prem A Sq Cut Bread Premium", which is usually followed by a sell-out of "Prem A Sq Cut Thick".

FIRST SELL OUT	SECOND SELL OUT	THIRD SELL OUT
PREM A SQ CUT BREAD PREMIUM	PREM A SQ CUT THICK	PREM A WHITE BREAD THICK
PREM A SQ CUT BREAD PREMIUM	PREM A SQ CUT THICK	SPMKT WHITE BREAD MEDIUM
PREM A SQ CUT BREAD PREMIUM	PREM A SQ CUT THICK	PREM A WHITE BREAD MEDIUM
PREM A SQ CUT BREAD PREMIUM	PREM A WHITE BREAD MEDIUM	SPMKT WHITE BREAD MEDIUM
PREM B PREM WHITE BREAD MEDIUM	PREM A SQ CUT BREAD PREMIUM	PREM A SQ CUT THICK
PREM A SQ CUT BREAD PREMIUM	PREM A WHITE BREAD MEDIUM	PREM A WHITE BREAD THICK
PREM B PREM WHITE BREAD MEDIUM	PREM A SQ CUT BREAD PREMIUM	PREM A WHITE BREAD THICK
PREM B PREM WHITE BREAD MEDIUM	PREM A SQ CUT THICK	PREM A WHITE BREAD THICK
PREM A SQ CUT BREAD PREMIUM	PREM A WHITE BREAD THICK	SPMKT WHITE BREAD MEDIUM
PREM A SQ CUT THICK	PREM A WHITE BREAD THICK	SPMKT WHITE BREAD MEDIUM
PREM A SQ CUT THICK	PREM A WHITE BREAD MEDIUM	SPMKT WHITE BREAD MEDIUM
PREM A SQ CUT BREAD PREMIUM	PREM B PREM WHITE BREAD THICK	SPMKT WHITE BREAD MEDIUM
PREM B PREM WHITE BREAD THICK	PREM A WHITE BREAD MEDIUM	SPMKT WHITE BREAD MEDIUM

Figure 8 Domino Patterns at TTN Supermarkets Ltd

Often, several "Prem A" products are on promotion together. We notice that "Spmkt White Bread Medium" is often the final domino in the chain and that this product has never been on promotion. Therefore, when "Prem A" products are on promotion, orders for "Spmkt White Bread Medium" should be increased to ensure that it does not sell out.

3.5.3 Domino Example

Figure 9 shows an example day on which the domino effect took place.

- At 13:00 "Premium A Sq Cut Bread Premium" (red) sold out.
- This sell-out caused an increase in the sales of "Premium A Sq Cut Thick" (yellow) from 14:00.
- At 16:00 "Premium A Sq Cut Thick" sold out.
- This sell-out caused an increase in the sales of "Premium A White Bread Thick" (pale blue) from 17:00.

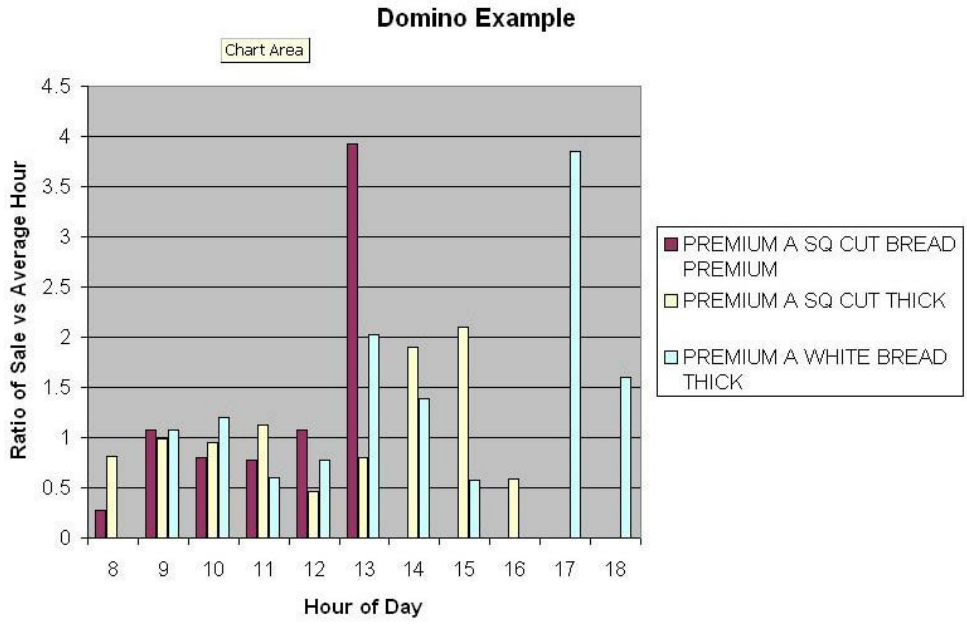


Figure 9 An example day where the domino effect can be observed.

3.5.4 The Steal Effect

Steals are subsequently investigated using further Clementine streams. Figure 10 shows how individual products steal from each other. Each row represents a product which is on promotion. The columns represent products which are affected by the promotion.

- Red cells indicate an excellent sales boost.
- Orange cells indicate a good sales boost.
- White cells indicate relatively unchanged sales.
- Blue cells indicate a steal.

SALES BOOSTS & STEALS	EFFECT ON PRODUCT								
	PREMIUM A SQ CUT BREAD PREMIUM	PREMIUM A SQ CUT THICK	PREMIUM A WHITE BREAD MEDIUM	PREMIUM A WHITE BREAD THICK	PREMIUM B PREM WHITE BREAD MEDIUM	PREMIUM B PREM WHITE BREAD THICK	SUPERMARKET BRAND BASIC WHITE BREAD	SUPERMARKET BRAND WHITE BREAD MEDIUM	SUPERMARKET BRAND WHITE BREAD THICK
PRODUCT ON PROMOTION									
PREMIUM A SQ CUT BREAD PREMIUM	147.379	131.408	158.357	157.379	79.4681	86.4211	85.1042	93.6458	93.3684
PREMIUM A SQ CUT THICK	147.465	131.096	158.478	156.644	80.3226	86.8817	85.3684	93.5789	93.4409
PREMIUM A WHITE BREAD MEDIUM	149	131.304	158.125	158.169	79.4624	86.3441	84.7872	93.4043	93.2609
PREMIUM A WHITE BREAD THICK	147.379	131.096	157.959	156.689	80.1031	86.5625	85.1021	93.8776	93.5417
PREMIUM B PREM WHITE BREAD MEDIUM	101.765	90.8511	79.8	96.5385	116.731	131.275	83.4615	93.2692	91.7647
PREMIUM B PREM WHITE BREAD THICK	102	90.6522	79.7959	96.8627	116.471	131.275	83.3333	93.5294	92

KEY	
149	Major Sales Boost (> 135 % of non-promotion sales)
131.096	Good Sales boost (105-135 % of non-promotion sales)
102	Relatively Unchanged Sales (90-105 % of non-promotion sales)
79.7959	Steal (< 90 % of non-promotion sales)

Figure 10 Matrix showing steal interactions

TTN Supermarkets Ltd offer multiple promotions. In the period in question either all "Premium A" products are in promotion, or all "Premium B" products are on promotion. It can be seen that when a product is on promotion, the sales of that product are enhanced (as expected). The following steal effects can be seen:

- All promotional products steal from "Supermarket brand basic white bread"
- There is significant steal from "Premium B" to "Premium A"
- There is some steal from "Premium A" to "Premium B"

3.6 Final results

In the final phase we can see the final results of the data mining analysis and their consequences for the ordering process.

Figure 11 shows how our discoveries have influenced and changed the ordering algorithms. In this chart:

- The blue bars indicate the 'standard' order quantities
- The red bars indicate the original 'on-promotion' order quantities
- The yellow bars indicate the new 'on-promotion' order quantities

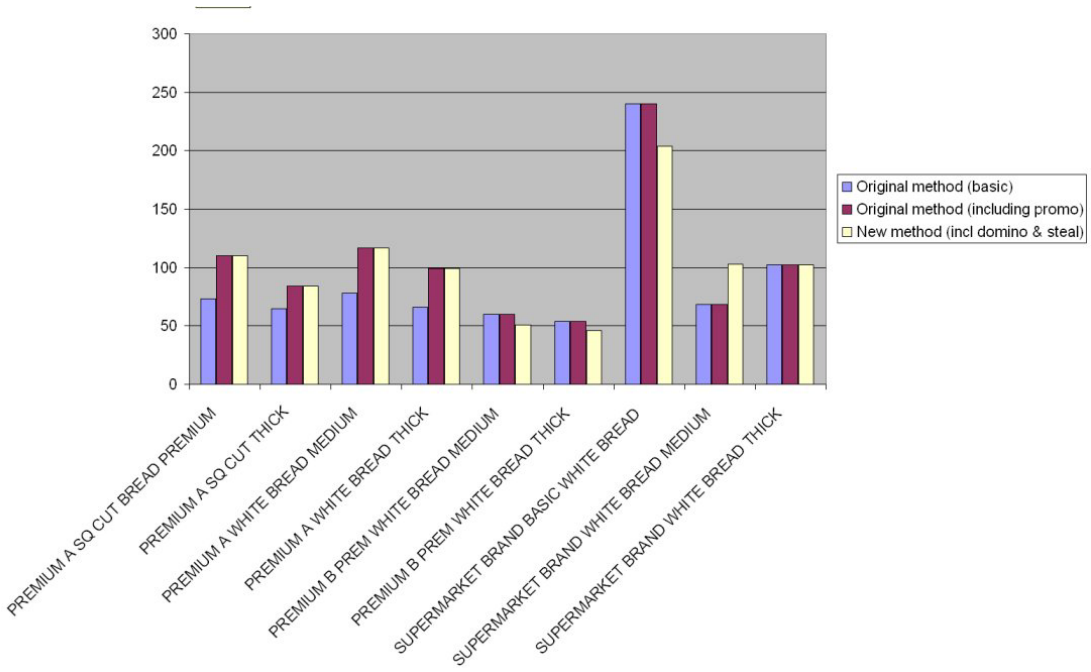


Figure 11 Impact on ordering process

In Figure 11, the four "Premium A" products are on promotion. The order quantities for these products are increased in the same way as they were before the analysis. However:

- Order quantities for the two "Premium B" products and for "Supermarket brand basic white bread" are reduced due to steal effects. The result is reduced wastage for these products.
- The order quantity for "Supermarket brand white bread medium" has been increased as this has been identified as the critical domino. The result is increased availability for bread products.

4 Conclusions

This report has demonstrated the power and benefits of data mining. In particular, it has shown how "TTN Supermarkets Ltd" have achieved ROI through reducing wastage costs and increasing revenue due to increased sales and increased customer satisfaction.