Business Intelligence Report for Clara Toys Ltd

Julian Hatwell

MSc Student in Business Intelligence

Birmingham City University

Julian.Hatwell@mail.bcu.ac.uk

# Introduction

This report presents the Business Intelligence (BI) proposal for Clara Toys Ltd, an internet retailer of high quality, hand-made toys. The management team have expressed that they would like to expand the company and need to reduce the uncertainty and risk around certain strategic decisions:

* The company needs to move to new premises. Should this be to an industrial unit, a new shop or both?
* How to capitalise on recent growth in overseas markets?
* Is expanding the product range to include older children or adults a credible option?

A Business Intelligence solution can help with these questions. By directly analysing the company’s data (and possibly some other sources), actionable insights can be discovered that will support the management team’s decisions.

# Methodology

The proposed strategy is to use the available business data (as described in the customer brief) to create analyses (in the form of charts and tables) that help to answer the business questions above. These questions are thought to be the priorities for the management team.

It is assumed that there are additional data sets readily available or discoverable with relative ease to support this strategy. Some additional items will be described in text only, and are assumed to be secondary priorities. These items can be raised in priority and included as a project extension, should the management team decide.

The strategy will look at several key areas:

1. Analysis of Order data by year to validate the recent growth that has prompted the expansion plans. It is expected to reveal strong year on year growth to support all other assumptions.
2. Analysis of Customers to compare UK and overseas customer buying behaviour to support the decision between opening a local shop and/or focussing on internet sales.
3. Analysis of Product data in terms of categories & suppliers to look at the profitability of different items and possible extensions to the range.
4. Analysis of Product data in terms of numbers of units with their bulk and size. This will be used to perform an Inventory forecast to help plan the move to a new industrial unit.

# Required Data

Clara Toys Ltd have withheld their commercial data prior to making a final decision. Therefore, it has been necessary to make assumptions (see Appendix A) about data that is held by the company and to create some test data sets.

These test data sets are described in the tables below, with an explanation of how they are used to answer business questions. Also described is how to collect this data. A sample is provided for reference and there is an assumed data schema/dictionary for Clara Toys Ltd e-commerce system in Appendix B.

Additional data sets are also discussed and listed below which could be used to enhance the BI solution and might be included in the project with the agreement of the management team.

|  |  |
| --- | --- |
|  | |
| How this data can be used to answer strategic business questions | This data can be analysed to show annual revenues and/or monthly and seasonal trends as well as segmenting by Order Source. |
| How to collect the data | Merge the Order and Customer data from Clara Toys Ltd e-commerce system. |

|  |  |
| --- | --- |
|  | |
| How this data can be used to answer strategic business questions | This can show trends over the years with statistics to compare each year to the previous year Domestic and International segments. |
| How to collect the data | This dataset combines Order and Customer data from Clara Toys Ltd e-commerce system. |

|  |  |
| --- | --- |
|  | |
| How this data can be used to answer strategic business questions | This is used to look at profitability of individual Products, Categories, Age Groups and Suppliers. The notional inventory cost is used for warehouse capacity planning. |
| How to collect the data | Merge the Order, OrderItem and Product data from Clara Toys Ltd e-commerce system. |

|  |  |  |
| --- | --- | --- |
| Additional Data (Optional Project Extensions) | How this data can be used to answer strategic business questions | How to collect the data |
| Historical Products and Pricing | Making use of the historical product information as it changes over the years would give a more accurate account of popular and profitable products. The information could be tied to campaigns and product launches to determine the effectiveness of these campaigns.  Tracking of rates of inflation over the years will offer the best comparison with previous years’ revenues. | This data set is assumed to be part of the e-commerce system |
| e-newsletter Responses | This data should comprise the recipients of each newsletter, dates that newsletters are sent out, products advertised in each mail out and response rates. These responses are further characterised by both email responses, and indirect measures such as web-searches for the specific products. | If this data is not already held by the company, they may consider an email marketing service that can automate many of the tasks involved by providing data on responses, deletes after opening, deletes without opening and bounced emails. |

|  |  |  |
| --- | --- | --- |
| Customer demographic | The customer information could be enriched with information about how many children there are and what ages they are as well as the age of the mother.  The relationship of the purchaser to the children (parent, grandparent, aunts, uncles or family friends) would help to predict future purchasing patterns and potential for products for older children.  For UK customers, Customer City, County or Region could also be used to decide the best location for a shop. A map plot showing which areas have the highest concentrations of customers (and potential customers from the mailing list) would be useful for this. | If this information is not already held in the customer database, the company should update their online registration form to request it. The could also try to engage their mailing list recipients with a survey or questionnaire. Discounts and prizes could improve the response rates.  The company could also form a customer focus group to get more qualitative feedback about the new product ideas they want to explore. |
| Social Media Data | Social Media activity relating to specific marketing campaigns e.g. hashtags and retweets could provide information on the effectiveness of marketing campaigns and whether these translate into real returns on investment. | If the company doesn’t collect this already, the various social media platforms all have interfaces which can be accessed to download information such as “as all the tweets with the hashtag #claratoys” |

|  |  |  |
| --- | --- | --- |
| Web-traffic | Product searches can be linked to seasonal marketing activities and e-newsletter mailshots to understand if marketing campaigns are translating in to return on investment.  Abandoned orders can be analysed to determine if there are any problems with the web-site usability which might be improved to increase sales. | This data is assumed to be part of the e-commerce system or could be downloaded from the Internet Service Provider who hosts the commercial web-site. If not, the company web-site should undergo a development project to use e.g. Google Analytics |
| Competitor Analysis / Supplier Trends | Data from a variety of sources could help answer the question about whether to expand the product range.  Products sold by other retailers could give some idea of any gaps in the existing range.  Third party market research may provide insight into unmet customer demand both in the UK and countries the company would like to penetrate. | A web based search of competitor e-commerce sites and conversations with their existing network of suppliers could provide further information. |

# The Analysis

A programme has been created using SAS software to demonstrate the various tables and charts that can be produced. Please see Appendix E for code and comments.

|  |  |
| --- | --- |
| Annual Revenue Analysis | |
|  | Total Annual Revenue appears to show a very healthy growth trend.  Points have been added at each year and are red where growth was slower than the previous year. |
|  | There was very modest peak in median and mean order totals in 2011. Even though the most common order total languished at the minimum, this led to the unprecedented revenue growth which was not replicated the following year.  It is recommended that the management team review their 2011 marketing plans for inspiration on how to have another winning year. |
| The SGPlot Procedure | The number of orders annually has levelled off in 2016. This requires management attention because the mean and median order totals have held steady for the last 10 years. All the growth has been through increasing the customer base and not through increasing customer value. If orders tail off and no mitigation is in place, this could put the growth strategy at risk. |

|  |  |
| --- | --- |
| Customer Revenue Analysis | |
| |  |  |  |  | | --- | --- | --- | --- | | **Customer Country** | **N** | **Mean** | **Sum** | | **Belarus** | 1 | £100.00 | £100.00 | | **Belgium** | 1 | £823.50 | £823.50 | | **Bosnia and Herzegovina** | 266 | £274.36 | £72,979.50 | | **France** | 80 | £252.75 | £20,220.00 | | **Kazakhstan** | 55 | £283.15 | £15,573.00 | | **Moldova** | 287 | £286.64 | £82,265.50 | | **Norway** | 1 | £201.50 | £201.50 | | **Poland** | 1 | £113.50 | £113.50 | | **Russia** | 235 | £274.18 | £64,431.50 | | **Spain** | 1 | £30.00 | £30.00 | | **Turkey** | 136 | £272.21 | £37,020.00 | | **United Kingdom** | 7491 | £270.87 | £2,029,123.00 | | **Vatican City (Holy See)** | 2 | £290.75 | £581.50 | | This table shows the number of orders, mean order total and the sum (total of totals) for all the orders by customer country over the last 10 years.  Standard deviations are not shown because they can’t be calculated for single orders (as occurs in several countries shown) and aren’t very useful for highly skewed data, as is the case with order totals (see Appendix C for a discussion about this).  Although the bulk of revenue comes from UK customers, Clara Toys Ltd products are also very popular in some Eastern European countries with relatively modest sales to France too. Perhaps these are markets where traditional toys are still highly valued.  The management team may wish to further investigate this phenomenon by asking these customers directly what they like about the products, and possibly forming a focus group or running a survey through the mailing list. |
|  | This simple bar chart shows the relative contribution to revenue from the domestic (UK) and international sales. On the face of it, the domestic market has always been the strongest and is still growing. |
| The SGPlot Procedure | This Year over Year trend line shows the importance of paying close attention to the detail. The calculation used is:  (this year’s revenue / last year’s revenue) - 1  The international sales were very erratic in the early years, but the smaller number of orders (approx. 30) makes it statistically unstable. In more recent years, international sales appear to be growing steadily at a much faster rate; over 20% per year compared to just under 10% for domestic sales. |
| Pie chart of DomInt | These pie charts for years 2007 and 2016 should make the point very clear. International sales have doubled as a share of revenue over the years. Could they double again in another 10?  The management team may want to look at their warehousing and international shipping costs to ensure they can maximise the opportunity this presents. |
|  | This plot shows the change in customer buying behaviour, just for the domestic market. Internet sales mirrors the strong growth trend in revenues.  Notice that 10 years ago the customer channels were evenly split between internet, telephone and mail order. The latter two channels have dropped away over the same 10 year period. The company might consider abandoning their printed catalogue altogether.  The management team will need to look for other evidence to support opening a local shop. The dominance of internet orders and the rapid growth of international sales can only suggest expanding the online business. |

|  |  |
| --- | --- |
| Product Profitability Analysis | |
| The SGPlot Procedure | This Pareto style chart shows total profits for each product over all the years in order of profitability. The series line shows the cumulative profits and the reference lines shows the product at which 90% of all profits are derived. Anything to the right of the vertical has not been delivering much value to the business. |
| The SGPlot Procedure | This is a similar plot which shows total number of items sold, instead of profits. The reference lines here also identify the 90% of all orders (includes everything to the left of the vertical).  Using both plots, the management team can see that some products are poor all round performers, e.g. *Doris the Daisy, Animals Colouring Book, Clatterbox* and *Click Chick*  Other products show very interesting patterns such as *Ultra Deluxe Crayon Set* which has mediocre profitability but is the most popular purchase. Perhaps it is added to many orders as a “stocking filler.”  Before expanding the range, the management team may wish to review the current product catalogue and drop the poorest performing products. This will allow warehouse space and other resources to be optimised before introducing riskier products that are untested in their market. |

|  |  |
| --- | --- |
| Pie chart of AgeGroup | This pie chart shows the profits generated by each product range a.k.a. Age Group.  It’s a very even split. The management team might not be able to make use of this information on its own. |
| Pie chart of Category | Profits by Category may be a bit more insightful. *Play Environment* and *Early Learning* are both very profitable categories. These include toys such as *Dozy Faces Night Light* and *Count With Me*. These are top performers on the Pareto charts.  *Baby Teethers* and *Activity* (e.g. Colouring Books) are not doing so well. Apart from the *Ultra Deluxe Crayon Set* these products aren’t adding much value and possibly take up a significant amount of workshop time and warehouse space. |
| Pie chart of SupplierName | This final pie chart shows the profits by Supplier. *Totness Toy Company* has pole position while *Wooden Toy Supplies* is in last place.  The management team will need to weigh up profitability of buying products compared to making them in house. This could be done as a project extension, looking at the margins and manufacturing costs. This would help the team decide whether to deepen the relationship with suppliers or try to design new products for in house manufacture that address the most popular categories. |
|  | This mosaic plot shows the breakdown of sales of each product over the years by Category and Supplier. The area of each rectangle is proportional to the units sold.  The management team can see that Baby Teethers are their niche product, not supplied by any other manufacturer. On the other hand, Companion category toys that are made in house don’t sell very well and could perhaps be discontinued so as not to compete with Totness Toy Company in this category.  Early Learning products from Tradition Toy Wholesale Co. are the top sellers of all but that’s all they produce. The management team could review the competition in this category with Wooden Toy Supplies and try to get a more favourable cost price. |

|  |  |
| --- | --- |
| Inventory Analysis | |
| The SGPlot Procedure | This chart shows the monthly inventory “cost” trend. This has been calculated based all the orders each month, the size of the product and whether a product is made in the workshop or bought in from suppliers (arriving already in its packaging).  Inventory will move in and out of warehouse all the time, so this is intended to be an approximation and to understand past, present and potential future positions.  The pink line is the total cost and is very “wiggly” because of the seasonality of the business, having a lot of sales concentrated around Christmas and other holidays. The black and green lines are moving average and moving max respectively. These give a clearer picture of the long-term trends. The moving max line has been shaped to indicate that available inventory is needed some months before the busiest seasons.  Inventory is a notional or relative cost, rather than a financial one. The management team should understand from this chart that their requirement for inventory capacity has approximately doubled over 10 years. If they’re intending to implement a growth strategy which includes more overseas shipping, they might assume that this rate will increase sharply and they should plan for extra space accordingly.    They may also benefit from finding extra space to rent during the busiest months, rather than trying to keep enough reserve space when their needs fluctuate so much during the year. |

# Summary and Conclusions

Based on the test data, there is a track record of year on year growth to support investment in an ambitious growth plan. Attention to the detail revealed additional strategies the management team may wish to consider.

There was no obvious support for opening a shop. Online and international sales are the engines of growth.

Before expanding the product range, the company may wish to terminate a small number of poor performers. Inventory demands on the warehouse space have doubled in ten years so streamlining the range may be a sensible choice before developing a plan to identify exactly which new product lines to develop. Further work to identify new lines is suggested as a project extension. There are various possibilities available, assuming the required additional data can be gathered with relative ease.

It is expected that the company’s real data can be plugged into the SAS programme without complication and a new set of charts and explanations be delivered to the management team, should they choose to go ahead with the project.

# Technical Appendices

## Appendix A – Assumptions

### Product Catalogue

CTL trade in Toys and have 3 main product lines, categorised by target age group.

Target age groups are 0-17m (Babies), 18m-3y (Toddlers) and 3y-5y (Kiddies).

Products sell at between GBP 30 – 100 and have a cost to manufacture or buy from other suppliers. Prices are set while trying to balance the following influencing factors:

* Compares well to similar products from competitors.
* Target profit margin per item sold > 15%
* Marketing prefers round numbers. Prices should be to the nearest 50p increment.

It’s not always possible to satisfy all three of these and so some discretion by Sales & Marketing team is applied.

Products are of different sizes and some are manufactured in house while others are sourced from suppliers. Each product bears a different notional inventory cost which has been modelled as a points based system. In house manufactured products “cost” less to inventory as they are stored without packing materials, unlike externally sourced products which are held in the bulky packaging they ship with. Cost of shipping from supplier to warehouse is already built into the product cost.

## Appendix B – Data Dictionary

|  |  |
| --- | --- |
| Data Set | Notes |
| Order Table   |  |  |  | | --- | --- | --- | | **Variable** | **Type** | **Len** | | custId | Num | 8 | | id | Num | 8 | | month | Num | 8 | | sourceId | Num | 8 | | total | Num | 8 | | year | Num | 8 |   Order Source Table   |  |  |  | | --- | --- | --- | | **Variable** | **Type** | **Len** | | id | Num | 8 | | name | Char | 10 | | custId is a Foregn Key to the Customer table  The order source table holds the source categories: (*internet, phone, mail order*) |
| Customer Table   |  |  |  | | --- | --- | --- | | **Variable** | **Type** | **Len** | | country | Char | 23 | | email | Char | 30 | | id | Num | 8 | | name | Char | 8 | |  |
| OrderItem Table   |  |  |  | | --- | --- | --- | | **Variable** | **Type** | **Len** | | orderId | Num | 8 | | price | Num | 8 | | productId | Num | 8 | | orderId and productid are foreign keys to their respective tables. |
| Product Table   |  |  |  | | --- | --- | --- | | **Variable** | **Type** | **Len** | | InventoryCost | Num | 8 | | ageGroupId | Num | 8 | | categoryId | Num | 8 | | cost | Num | 8 | | id | Num | 8 | | margin | Num | 8 | | name | Char | 27 | | originId | Num | 8 | | price | Num | 8 |   AgeGroup, Category and Origin Tables   |  |  |  | | --- | --- | --- | | **Variable** | **Type** | **Len** | | id | Num | 8 | | name | Char | 10 | | margin is a calculated field:  (price – cost) / price  The age group table holds the product age groups: (*babies, toddlers, kiddies*)  The categories table holds the product categories: (*baby teethers, early learning, etc*)  The origin table holds the product origin: (*in house manufacture, supplier 1 name, supplier 2 name, etc*) |

## Appendix C - A side note about descriptive statistics for the Order Totals

The “average” is a statistic that is very frequently used to describe the central tendency of random data around a specific value, known as the arithmetic mean. However, it is much abused due to the assumption that the underlying data is normally distributed. In the case of CTL order totals, this is not the case.

There is a minimum unit price 30. This gives a minimum order of 30 and increments in discrete units of 30 or more, depending on the specific products selected. The mode (most frequent) value of order totals hovers between 30, 60 and 90, indicating most orders may have one, two or three units of the cheapest products. Also, every month there are a few outliers between 700 and 3000. Clearly there are complex interactions of price and cost-conscious customers that make the distribution very different from the normal. This issue can be visualised in some further graphs:

|  |  |
| --- | --- |
| Distribution of Order Totals | |
| The SGPlot Procedure | The SGPlot Procedure |
| Histogram of Order Total with superimposed normal density and kernel density functions. | Histogram of natural log of Order Total with superimposed normal density and kernel density functions. The 3 modal points seen in the annual trends are visible, but otherwise the normal distribution is a much better approximation |

The arithmetic mean is a lot more meaningful when calculated over the natural logarithm of order totals. However, it is a bit more difficult for a non-mathematician to interpret.

The management dashboard is using the mean and median of log(order total) and then modifying the scale to show the non-transformed value. This is thought to be the best compromise of interpretability and meaningful statistics.

## Appendix D – External References

The following web resources were used to enhance my knowledge of SAS programming:

Advanced use of PROC MEANS:

<http://www.ats.ucla.edu/stat/sas/modules/collapse.htm>

<http://blogs.sas.com/content/sgf/2015/07/17/customizing-output-from-proc-means/>

Use of PROC SQL template for a custom summary function (initialise, accumulate, output):

<http://www.ats.ucla.edu/stat/sas/library/nesug99/ad121.pdf>

Plotting system:

<https://support.sas.com/sassamples/graphgallery/PROC_SGPLOT.html>

<http://support.sas.com/resources/papers/proceedings10/154-2010.pdf>

<http://support.sas.com/resources/papers/proceedings11/095-2011.pdf>

Formats:

<http://support.sas.com/kb/12/970.html>

Computing a moving statistic with PROC EXPAND:

<http://www2.sas.com/proceedings/forum2008/093-2008.pdf>

<http://blogs.sas.com/content/iml/2016/01/27/moving-average-in-sas.html>

<http://support.sas.com/documentation/cdl/en/etsug/68148/HTML/default/viewer.htm#etsug_expand_details19.htm>

## Appendix E – Example SAS Programme

Based on the assumptions and BI strategy described, an example SAS programme has been developed which could form the basis of a fully implemented solution. The program steps acting on similar data or performing similar functions have been grouped together logically and split into several smaller files to improve maintainability and readability.

File: Main0\_styles

\* Create standard colour templates for a professional look;

**proc** **template**;

define style styles.myStyle1;

parent=styles.statistical;

style graphdata1 /

markersymbol='circlefilled'

contrastcolor=red;

style graphdata2 /

markersymbol='circlefilled'

contrastcolor=black;

end;

**run**;

\* Create standard colour templates for a professional look;

**proc** **template**;

define style styles.myStyle2;

parent=styles.statistical;

style graphdata1 /

linestyle=**1**

contrastcolor=cx1B20FF;

style graphdata2 /

linestyle=**1**

contrastcolor=cx4B72FF;

style graphdata3 /

linestyle=**1**

contrastcolor=cx7FA5F6;

end;

**run**;

\* Create standard colour templates for a professional look;

**proc** **template**;

define style styles.myStyle3;

parent=styles.statistical;

style graphdata1 /

markersymbol='circlefilled'

color=cx1B20FF

contrastcolor=cx1B20FF;

style graphdata2 /

markersymbol='circlefilled'

color=cxBBAA33

contrastcolor=cxBBAA33;

end;

**run**;

File: Main1\_datalines

\* sets up basic categories and keywords for the de-normalised data;

libname clara "C:\Users\Julian\OneDrive\Documents\BCU\Business Intelligence\Clara Toys Assignment";

\* data created directly - categories and key words;

**data** work.Agegroup;

input id name $;

datalines;

1 Babies

2 Toddlers

3 Kiddies

;

**data** work.Category;

length name $ **20**;

input id name $;

datalines;

1 Baby\_Teethers

2 Reach\_&\_Grab

3 Early\_Learning

4 Activity

5 Play\_Environment

6 Companion

;

**data** work.OrderSource;

length name $ **10**;

input id name $;

datalines;

1 mail\_order

2 telephone

3 internet

;

**run**;

libname clara clear;

File: Main2\_JMP

\* loads data created in JMP (and exported as SAS data sets) into work library;

libname clara "C:\Users\Julian\OneDrive\Documents\BCU\Business Intelligence\Clara Toys Assignment";

\* data created in JMP;

**data** work.Origin;

set clara.Origin;

drop address email tel;

**data** work.Product;

set clara.Product;

InventoryCost = **.**;

IF (origin EQ **1**) THEN InventoryCost = size;

IF (origin NE **1**) THEN InventoryCost = size \* **2**;

drop size;

**data** work.Handmade;

set clara.Handmade;

**data** work.HandmadeOrders;

set clara.HandmadeOrders;

**data** work.Maker;

set clara.Maker;

**run**;

libname clara clear;

File: Main3\_csv

\* loads data created in R simulation (and exported as csv) into work library;

libname clara "C:\Users\Julian\OneDrive\Documents\BCU\Business Intelligence\Clara Toys Assignment";

\* data created in R simulations as .csv files;

**data** work.Customers;

length name $ **8** address $ **18** country $ **23** email $ **30** tel $ **8**;

infile "C:\Users\Julian\OneDrive\Documents\BCU\Business Intelligence\Clara Toys Assignment\Customers.csv" dlm=",";

input id firstOrderYear firstOrderMonth name $ address $ country $ email $ tel $;

drop firstOrderYear firstOrderMonth address tel;

**data** work.OrderItems;

infile "C:\Users\Julian\OneDrive\Documents\BCU\Business Intelligence\Clara Toys Assignment\OrderItems.csv" dlm=",";

input month year order productid price;

**data** work.Orders;

infile "C:\Users\Julian\OneDrive\Documents\BCU\Business Intelligence\Clara Toys Assignment\Orders.csv" dlm=",";

input id year month total cust source;

**data** work.Mailinglist;

informat dateSubscribed DDMMYY10.;

length email $ **30**;

infile "C:\Users\Julian\OneDrive\Documents\BCU\Business Intelligence\Clara Toys Assignment\MailingList.csv" dlm=",";

input email $ dateSubscribed;

format dateSubscribed DDMMYY10.;

**run**;

libname clara clear;

File: Main4\_orders:

\* data from combining and summarising data sets that are already loaded;

title; \* clear from any previous runs;

\* master table for denormalised Orders data ;

\* proc sql has been used to reflect the origin of this data in a relational database ;

**proc** **sql**;

create table work.OrdersReportingTable AS

SELECT o.id AS OrderId, o.year AS OrderYear, o.month AS OrderMonth

, o.total AS OrderTotal, log(o.total) AS LogOrderTotal, c.id AS CustomerId

, c.country AS CustomerCountry, os.name AS OrderSource

FROM work.Orders o

INNER JOIN work.Customers c

ON o.cust = c.id

INNER JOIN work.OrderSource os

ON o.source = os.id;

**quit**;

**proc** **print** data=work.OrdersReportingTable (obs=**6**);

format OrderTotal NLMNLGBP30.2;

title 'Orders Reporting Table';

**run**;

\* Order totals and summaries year on year;

**proc** **means** data=work.OrdersReportingTable nway nonobs noprint;

var OrderTotal LogOrderTotal;

class OrderYear;

output out=work.OrdersSummaryYear

mean=AvgOrderTotal AvgLogOrderTotal

median=MedOrderTotal MedLogOrderTotal

mode=ModOrderTotal ModLogOrderTotal

min=MinOrderTotal MinLogOrderTotal

max=MaxOrderTotal MaxLogOrderTotal

sum=TotalOfTotals;

**run**;

\* calculate YoY if revenue growth is greater or less than last year;

**data** work.OrdersSummaryYear;

set work.OrdersSummaryYear;

TotalLY = lag(TotalofTotals);

GrowthAmount = TotalOfTotals - TotalLY;

if GrowthAmount > **0** then vs\_LY\_Revenue = "Increase";

if GrowthAmount <= **0** then vs\_LY\_Revenue = "Decrease";

YoY = (TotalOfTotals / TotalLY) - **1**;

format YoY PERCENT.

AvgOrderTotal MedOrderTotal

ModOrderTotal MinOrderTotal

MaxOrderTotal TotalOFTotals

TotalLY GrowthAmount NLMNLGBP30.2;

**run**;

\* Seasonal Trends month on month;

**proc** **means** data=work.OrdersReportingTable nway nonobs noprint;

var OrderTotal;

class OrderMonth;

output out=work.OrdersSummaryMonth

mean=AvgOrderTotal

median=MedOrderTotal

mode=ModOrderTotal

min=MinOrderTotal

max=MaxOrderTotal

sum=TotalOfTotals;

**run**;

\* Full Time Series Trend;

**proc** **means** data=work.OrdersReportingTable nway nonobs noprint;

var OrderTotal;

class OrderYear OrderMonth;

output out=work.OrdersSummaryFullTrend

mean=AvgOrderTotal

median=MedOrderTotal

mode=ModOrderTotal

min=MinOrderTotal

max=MaxOrderTotal

sum=TotalOfTotals;

**run**;

title; \*clear title;

File: Main4\_orders\_plots\_yoy:

\* call the style sheet for the red & black points;

ods html style=Styles.mystyle1;

\* plots for yoy trends;

**proc** **sgplot** data=work.OrdersSummaryYear;

title 'Standard Toys';

title2 'Total Annual Revenue';

series x = OrderYear y = TotalOfTotals;

scatter x = OrderYear y = TotalOfTotals / group = vs\_LY\_Revenue;

yaxis label='Total Revenue from Orders';

refline **2011** / axis=x lineattrs=(pattern=LongDash) label = 'Unprecedented growth surge';

**run**;

**proc** **sgplot** data=work.OrdersSummaryYear;

title 'Standard Toys';

title2 'Number of Orders each year';

series x = OrderYear y = \_FREQ\_;

scatter x = OrderYear y = \_FREQ\_ / group = vs\_LY\_Revenue;

yaxis label='Number of Orders';

**run**;

**proc** **sgplot** data=work.OrdersSummaryYear;

title 'Standard Toys';

title2 'Annual Revenue Year over Year';

series x = OrderYear y = YoY;

scatter x = OrderYear y = YoY / group = vs\_LY\_Revenue;

yaxis label='Year Over Year';

refline **2013** / axis=x lineattrs=(pattern=LongDash) label = 'Beginning of stable growth';

**run**;

\* call the style sheet for the blue lines;

ods html style=Styles.mystyle2;

\* plots for summary stats order totals;

**proc** **sgplot** data=work.OrdersSummaryYear;

title 'Standard Toys';

title2 'Order Total Summary Stats Yearly Trend';

series x = OrderYear y = AvgOrderTotal;

series x = OrderYear y = MedOrderTotal;

series x = OrderYear y = ModOrderTotal;

yaxis label='Mean, Median and Mode Order Totals';

**run**;

**proc** **sgplot** data=work.OrdersSummaryYear;

title 'Standard Toys';

title2 'Order Total Summary Stats Yearly Trend';

series x = OrderYear y = AvgLogOrderTotal;

series x = OrderYear y = MedLogOrderTotal;

series x = OrderYear y = ModLogOrderTotal;

refline **2011** / axis=x lineattrs=(pattern=LongDash) label = 'Modest uplift on median/mean';

yaxis grid offsetmin=**0.05** offsetmax=**0.1**

values=(**3.0** **3.4** **4.2** **4.5** **5.0** **5.5**)

valuesdisplay = (" " "GBP30" "GBP60" "GBP90" "GBP150" "GBP250")

label='exp (Mean, Median and Mode log(Order Totals))';

**run**;

\* call the style sheet for the histogram with norm and kern;

ods html style=Styles.mystyle1;

\* histograms of order totals with density functions;

**proc** **sgplot** data=work.OrdersReportingTable;

title 'Standard Toys';

title2 'Distribution of order totals - All years';

histogram ordertotal;

density ordertotal;

density ordertotal / type=kernel;

**run**;

**proc** **sgplot** data=work.OrdersReportingTable;

title 'Standard Toys';

title2 'Distribution of log(order totals) - All years';

histogram logordertotal;

density logordertotal;

density logordertotal / type=kernel;

**run**;

\* reset the title and default style;

title;

ods html style=Styles.Default;

File: Main5\_customers:

\* customers;

title; \* clear any previous title;

\* formatting for the proc means print out;

**proc** **template**;

edit base.summary;

edit mean;

format=NLMNLGBP30.2;

end;

edit median;

format=NLMNLGBP30.2;

end;

edit mode;

format=NLMNLGBP30.2;

end;

edit sum;

format=NLMNLGBP30.2;

end;

edit min;

format=NLMNLGBP30.2;

end;

edit max;

format=NLMNLGBP30.2;

end;

end;

**run**;

\* print the mean and sum by country ;

**proc** **means** data=work.OrdersReportingTable nway nonobs n mean sum;

title 'Customer Orders Totals By Country';

var OrderTotal;

class CustomerCountry;

**run**;

\* print the mean and sum by country and year and output a new dataset;

**proc** **means** data=work.OrdersReportingTable nway nonobs noprint;

var OrderTotal;

class OrderYear CustomerCountry;

output out=work.CustomerSummary

mean=AvgOrderTotal

median=MedOrderTotal

mode=ModOrderTotal

min=MinOrderTotal

max=MaxOrderTotal

sum=TotalOfTotals;

**run**;

\* process just the domestic (UK) customers ;

**data** work.DomesticCustomers;

set work.OrdersReportingTable (where=(CustomerCountry EQ 'United Kingdom'));

**run**;

\* data on domestic trends by channel;

**proc** **means** data=work.DomesticCustomers nway nonobs noprint;

var OrderTotal;

class OrderYear OrderSource;

output out=work.SourceSummary

mean=AvgOrderTotal

median=MedOrderTotal

mode=ModOrderTotal

min=MinOrderTotal

max=MaxOrderTotal

sum=TotalOfTotals;

**run**;

**data** work.SourceSummary;

set work.SourceSummary;

format AvgOrderTotal MedOrderTotal

ModOrderTotal MinOrderTotal

MaxOrderTotal TotalOFTotals

TotalLY GrowthAmount NLMNLGBP30.2;

**run**;

\* data on domestic customers;

**proc** **means** data=work.DomesticCustomers nway nonobs noprint;

var OrderTotal;

class OrderYear;

output out=work.DomesticSummary

mean=AvgOrderTotal

median=MedOrderTotal

mode=ModOrderTotal

min=MinOrderTotal

max=MaxOrderTotal

sum=TotalOfTotals;

**run**;

**data** work.DomesticSummaryEnhanced;

set work.DomesticSummary;

length DomInt $ **13**;

DomInt = 'Domestic';

TotalLY = lag(TotalofTotals);

GrowthAmount = TotalOfTotals - TotalLY;

if GrowthAmount > **0** then vs\_LY\_Revenue = "Increase";

if GrowthAmount <= **0** then vs\_LY\_Revenue = "Decrease";

YoY = (TotalOfTotals / TotalLY) - **1**;

format YoY PERCENT.

AvgOrderTotal MedOrderTotal

ModOrderTotal MinOrderTotal

MaxOrderTotal TotalOFTotals

TotalLY GrowthAmount NLMNLGBP30.2;

**run**;

\* process just the international customers ;

**data** work.InternationalCustomers;

set work.OrdersReportingTable (where=(CustomerCountry NE 'United Kingdom'));

**run**;

**proc** **means** data=work.InternationalCustomers nway nonobs noprint;

var OrderTotal;

class OrderYear;

output out=work.InternationalSummary

mean=AvgOrderTotal

median=MedOrderTotal

mode=ModOrderTotal

min=MinOrderTotal

max=MaxOrderTotal

sum=TotalOfTotals;

**run**;

**data** work.InternationalSummaryEnhanced;

set work.InternationalSummary;

length DomInt $ **13**;

DomInt = 'International';

TotalLY = lag(TotalofTotals);

GrowthAmount = TotalOfTotals - TotalLY;

if GrowthAmount > **0** then vs\_LY\_Revenue = "Increase";

if GrowthAmount <= **0** then vs\_LY\_Revenue = "Decrease";

YoY = (TotalOfTotals / TotalLY) - **1**;

format YoY PERCENT.;

**run**;

**data** work.DomIntSummaryEnhanced;

set DomesticSummaryEnhanced InternationalSummaryEnhanced;

**run**;

\* data on trends by country;

**proc** **means** data=work.InternationalCustomers nway nonobs noprint;

var OrderTotal;

class OrderYear CustomerCountry;

output out=work.InternationalByCountry

mean=AvgOrderTotal

median=MedOrderTotal

mode=ModOrderTotal

min=MinOrderTotal

max=MaxOrderTotal

sum=TotalOfTotals;

**run**;

\* let's ditch the countries with just a handful of orders;

\* useful to report on trends for the top performing non-UK countries;

**proc** **sort** data = work.InternationalByCountry;

by CustomerCountry;

**run**;

\* this narrow dataset just sums up the number of orders in each country;

**proc** **means** data=work.InternationalByCountry nway nonobs noprint;

var \_FREQ\_;

class CustomerCountry;

output out=work.InternationalOrders

sum=TotalOrders;

**run**;

\* merge it back to the main table and only keep rows that have > 10 orders;

**data** work.InternationalByCountry;

merge work.InternationalByCountry (in = cty)

work.InternationalOrders (in = ord

where = (TotalOrders GT **10**));

by CustomerCountry ;

if cty AND ord;

**run**;

\* combine with Domestic orders and only use the first and latest years - for the pie chart;

**data** work.DomIntTotalsYears;

set work.DomesticSummary;

length DomInt $ **13**;

DomInt = 'Domestic';

format TotalOfTotals NLMNLGBP30.2;

**run**;

**proc** **append** base=work.DomIntTotalsYears data=InternationalSummary;

**run**;

**data** work.DomIntTotalsYears;

set work.DomIntTotalsYears (where=(OrderYear in (**2007** **2016**)));

if DomInt NE 'Domestic' then DomInt = 'International';

**run**;

File: Main5\_customers\_plots\_yoy:

\* call the style sheet for the red & black points;

ods html style=Styles.mystyle3;

\* plots for customer trends;

**proc** **sgplot** data=work.DomIntSummaryEnhanced;

vbar OrderYear / response=TotalOfTotals group=DomInt groupdisplay=cluster;

xaxis display=(nolabel novalues noticks);

yaxis label='Total Revenue from Orders';

keylegend / location=inside position=topleft across=**1**;

**run**;

**proc** **sgplot** data=work.DomIntSummaryEnhanced;

title 'Growth Trend in Domestic vs International Sales over 10 years';

series x = OrderYear y = YoY / group = DomInt;

refline **0.1** / axis=y lineattrs=(pattern=LongDash) label = '10%';

refline **0.2** / axis=y lineattrs=(pattern=LongDash) label = '20%';

**run**;

**proc** **sgplot** data=work.SourceSummary;

title 'Trend in Sales Channel Revenue (Domestic Customers) over 10 years';

series x = OrderYear y = TotalOfTotals / group = OrderSource;

**run**;

**proc** **gchart** data=work.DomIntTotalsYears;

title 'Domestic and International Share of Revenue';

pie Domint / sumvar = TotalOfTotals

group = OrderYear

across = **2**

percent=outside;

**run**;

\* reset the title and default style;

title;

ods html style=Styles.Default;

File: Main6\_products:

\* Master table for denormalised product data, including when ordered;

title; \*clear any previous title;

\* proc sql has been used to reflect the origin of this data in a relational database ;

**proc** **sql**;

title 'Products Reporting Table';

create table work.ProductsReportingTable AS

SELECT oi.order AS OrderID, oi.year AS OrderYear, oi.month AS OrderMonth

, p.id AS ProductId, p.name AS ProductName, p.cost AS ProductCost

, p.price AS ProductPrice, p.margin AS ProductMargin, p.InventoryCost AS InventoryCost

, ag.name AS AgeGroup, cg.name AS Category, og.name AS SupplierName

, og.city AS SupplierCity, og.County AS SupplierCounty

FROM work.OrderItems oi

INNER JOIN work.Product p

ON oi.productid = p.id

INNER JOIN work.Agegroup ag

ON p.Agegroup = ag.id

INNER JOIN work.Category cg

ON p.Category = cg.id

INNER JOIN work.Origin og

ON p.origin = og.id;

**quit**;

\* Products and profits summary;

**proc** **means** data=work.ProductsReportingTable nway nonobs noprint;

title 'Products Ordered';

var ProductPrice ProductCost InventoryCost;

class ProductName;

output out=work.ProfitSummary

sum=TotProdSales TotProdCosts TotInvCost;

**run**;

**data** work.ProfitSummaryEnhanced;

set work.ProfitSummary;

TotProfit = TotProdSales - TotProdCosts;

format TotProfit NLMNLGBP10.2;

**run**;

**proc** **sort** data=work.ProfitSummaryEnhanced;

by descending TotProfit;

**run**;

**proc** **expand** data=work.ProfitSummaryEnhanced out=work.ProfitSummaryEnhanced method=none;

convert TotProfit = CumProfit / transout=(cusum);

format TotProfit NLMNLGBP10.2;

**run**;

\* sort and count by items sold;

**proc** **sort** data=work.ProfitSummaryEnhanced (keep=ProductName \_FREQ\_ rename=(\_FREQ\_ = NumSold)) out=work.ItemSoldSummary;

by descending NumSold;

**run**;

**proc** **expand** data=work.ItemSoldSummary out=work.ItemSoldSummary method=none;

convert NumSold = CumNumSold / transout=(cusum);

**run**;

**data** work.ProductsReportingTable;

set work.ProductsReportingTable;

Profit = ProductPrice - ProductCost;

format Profit NLMNLGBP10.2;

**run**;

File: Main6\_products\_plots:

\* call the style sheet for the red & black points;

ods html style=Styles.mystyle1;

\* plots for product trends;

\* pareto style profitability chart;

**proc** **sgplot** data=work.ProfitSummaryEnhanced;

title 'Total Profit per Product';

series x = ProductName y = CumProfit;

refline **450000**;

refline 'Ultra Deluxe Crayon Set' / axis = x;

scatter x = ProductName y = TotProfit /

markerattrs=(symbol=squarefilled size=**10**px);

**run**;

\* pareto style items sold chart;

**proc** **sgplot** data=work.ItemSoldSummary;

title 'Total Number Sold per Product';

series x = ProductName y = CumNumSold;

refline **36000**;

refline 'Dried Pea Wooden Rattle' / axis = x;

scatter x = ProductName y = NumSold /

markerattrs=(symbol=squarefilled size=**10**px);

**run**;

\* various pies for share of profits;

**proc** **gchart** data=work.ProductsReportingTable;

title 'AgeGroup Share of Profit';

pie AgeGroup / sumvar = Profit

percent=outside;

**run**;

**proc** **gchart** data=work.ProductsReportingTable;

title 'Category Share of Profit';

pie Category / sumvar = Profit

percent=outside;

**run**;

**proc** **gchart** data=work.ProductsReportingTable;

title 'Supplier Share of Profit';

pie SupplierName / sumvar = Profit

percent=outside;

**run**;

\* mosaic frequency plot of units sold;

**proc** **format**;

value $category 'Baby\_Teethers' = 'Baby Teethers'

'Reach\_&\_Grab' = 'Reach & Grab'

'Play\_Environment' = 'Play Environment'

'Early\_Learning' = 'Early Learning';

**run**;

**proc** **freq** data=work.ProductsReportingTable;

tables Category \* SupplierName / nofreq norow nocol chisq plots=all;

format Category $category.;

**run**;

\* reset the title and default style;

title;

ods html style=Styles.Default;

File: Main7\_inventory:

**proc** **means** data=work.ProductsReportingTable nway nonobs noprint;

title 'Inventory By Year and Month';

var InventoryCost;

class OrderYear OrderMonth;

output out=work.InventorySummary

sum=TotInvCost;

**run**;

**proc** **expand** data = work.InventorySummary out = work.InventorySummaryEnhanced;

convert TotInvCost = MovAveInvCost / METHOD = none TRANSFORMOUT = (cmovave (**1** **2** **3** **4** **5** **6** **7** **8** **9**));

convert TotInvCost = MovMaxInvCost / METHOD = none TRANSFORMOUT = (cmovmax **18**);

**run**;

File: Main7\_inventory\_plots:

\* call the style sheet for the red & black points;

ods html style=Styles.mystyle1;

\* inventory trend;

**proc** **sgplot** data=work.InventorySummaryEnhanced;

title 'Inventory Requirments Trends';

series x = TIME y = TotInvCost / transparency = **0.75**;

series x = TIME y = MovAveInvCost;

series x = TIME y = MovMaxInvCost;

xaxis label='Months since Jan 2007';

yaxis label='Notional Inventory Cost: Total with Moving Ave and Max';

**run**;

\* reset the title and default style;

title;

ods html style=Styles.Default;