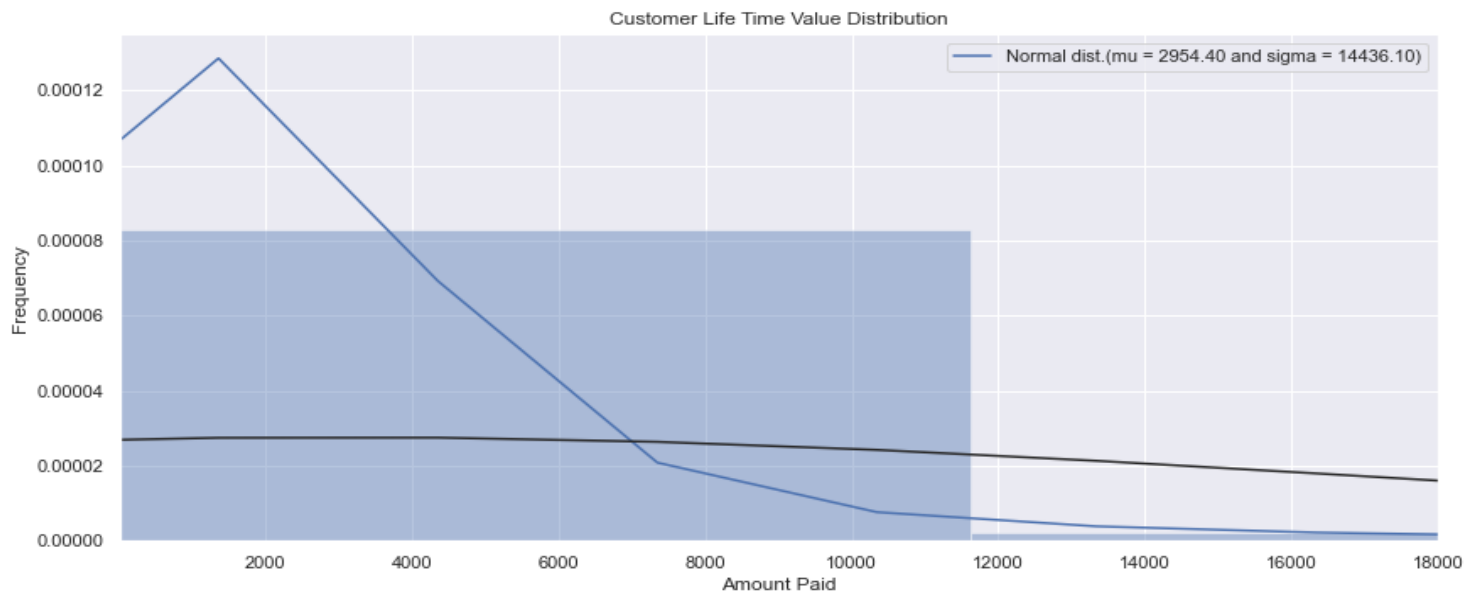


# EDA Project

By Amitava Roy

**Q- Discuss the customer's lifetime with respect to the given dataset.**

**A-** Customer value or ***Customer Lifetime Value (CLV)*** is the total monetary value of transactions/purchases made by a customer from a business over his entire lifetime. Here the lifetime means the time period till the customer purchases from the business before moving to its competitors. For this assignment, we have considered the time period given in the dataset as the lifetime of the customer.



**Q- Can customers be segmented into different categories? If yes then perform analysis on the same and also propose categories. If not, then explain why?**

**A-** The customers can be segmented on the basis of their **CUSTOMER LIFETIME VALUE (CLV)**. All the customers have been sorted according to the total expenditure by them in the span of two years. The customers lying below the **50 percentile** mark are categorized as **Class C**. Individual contribution from the customers in this class is not significant. However since a majority of the customers lie in this class, the overall contribution from this class is greatly significant for this shop. The ones lying in the **50-80 percentile** have been regarded as **Class B** and these customers are backbone of the economy of the shop. That means they should be maintained in case of loss of **Class A** and the **Elite** customers.

**Class A** customers lie between **80-95 percentile** and the **Elite** customers lie in the **top 5 percentile**. Their preferences should be noted and required changes should be brought in the items of the shop and this will motivate them to invest furthermore into the shop.

In spite of being the part of major expenditure made in the shop, these top two classes can afford to buy from other shops, hence can leave the shop whenever they want. And if they do the company is going to rely upon **Class B** and a few **Class A** customers.

Percentile	CLV
50	865.6
80	2909.86
95	9371.709999999998

100	580987.04
-----	-----------

Q- How would you define a loyal customer?

A- We can define loyal customers based on the different number of days a customer visited the shop. This would represent the frequency and interest of the customers. According to this trend, the customer with **Customer ID 14911.0** has visited the shop for **255** distinct days and hence is important for the shop.

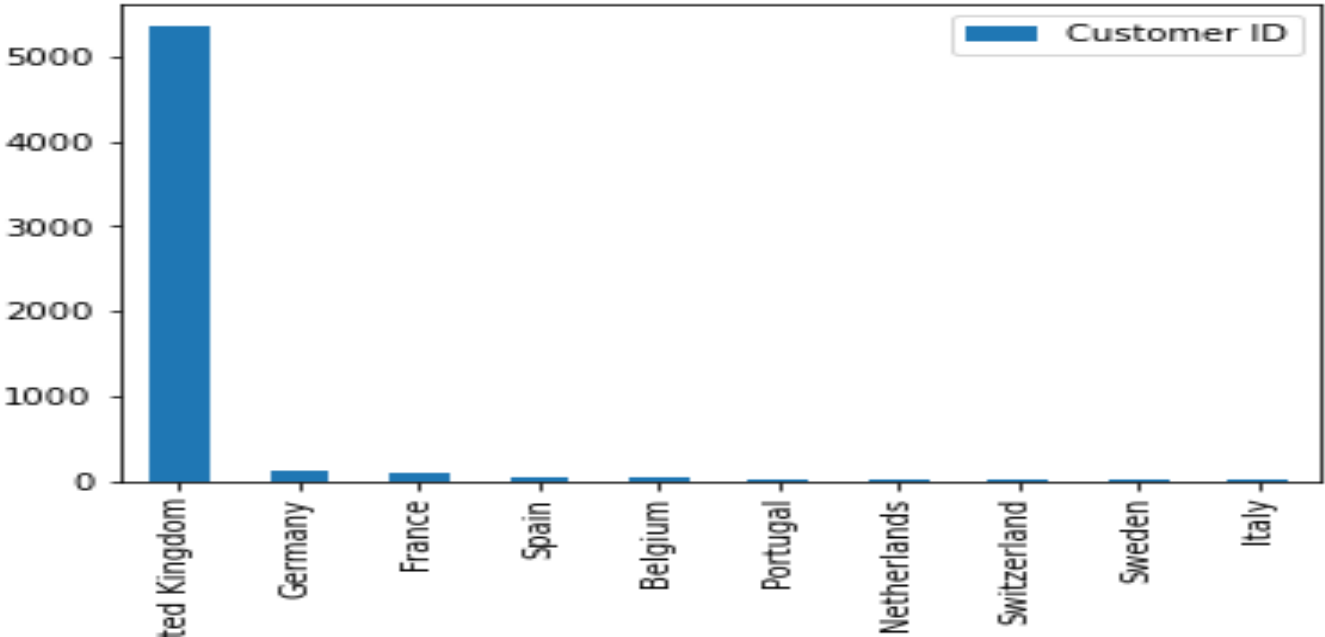
The second variable that can be used to compare is the number of invoices corresponding to a customer which would denote the number of bills made for a particular customer. Now, this trend shows that the same customer with **CustomerID 14911.0** has made **398** distinct invoices and hence performed so many transactions resulting in a huge amount of profit for the shop.

We have used both unique(dates) and unique(invoices) for sorting the data with higher priority given to unique(dates).

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Country	Date	Time
Customer ID									
14911.0	398	2550	2746	46	398	135	1	255	280
12748.0	337	2287	2377	65	338	121	1	203	242
15311.0	208	864	947	58	208	81	1	196	154
17841.0	211	2116	2269	48	211	81	1	193	164
14606.0	192	1262	1381	39	194	71	1	181	153
...	...	...	...	...	...	...	...	...	...
18269.0	1	7	7	3	1	5	1	1	1
18274.0	1	11	11	6	1	8	1	1	1
18279.0	1	13	13	6	1	4	1	1	1
18284.0	1	28	28	10	1	10	1	1	1
18285.0	1	12	12	6	1	12	1	1	1
5881 rows x 9 columns									

Q- Which country is more beneficial to the shop?

A- After plotting Country vs total number Customer ID, it is clear that the country **United Kingdom** has proven to be most beneficial to the shop.

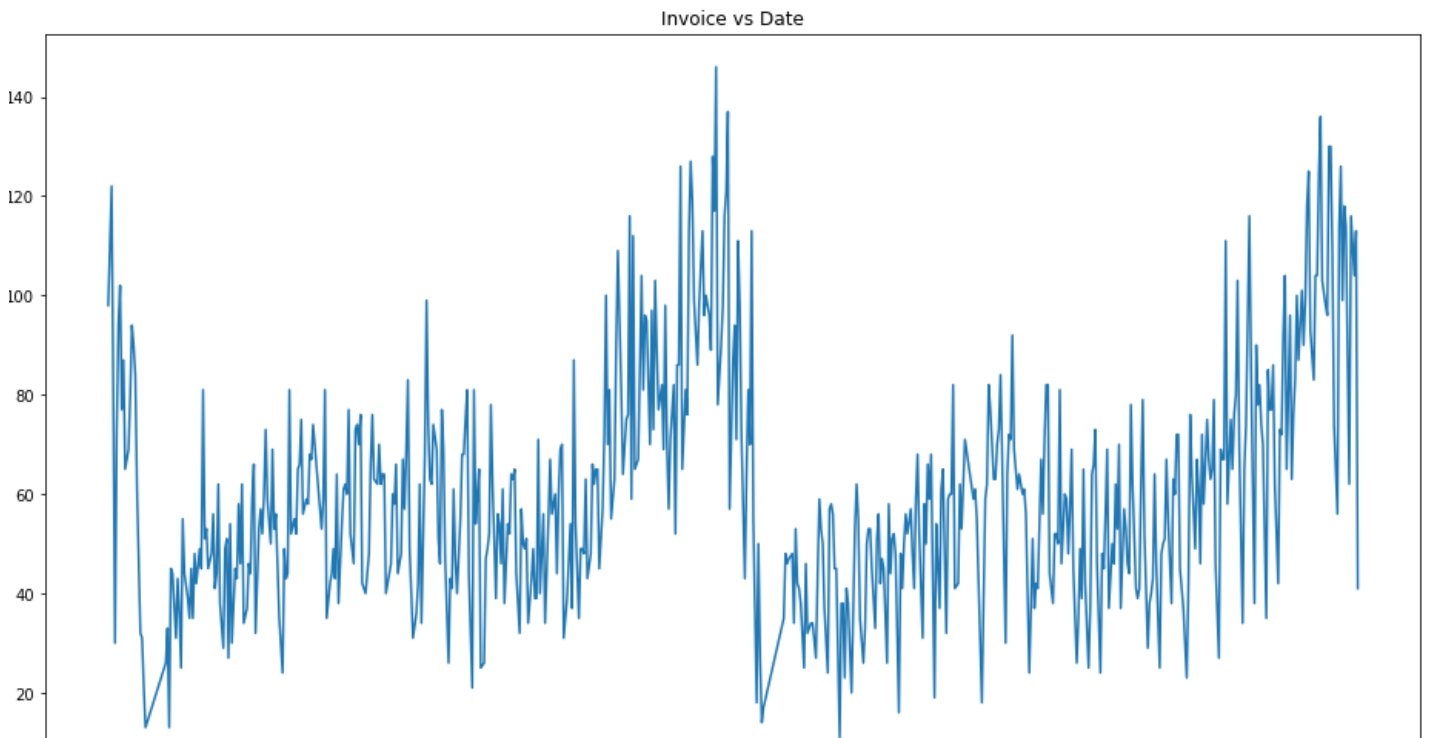
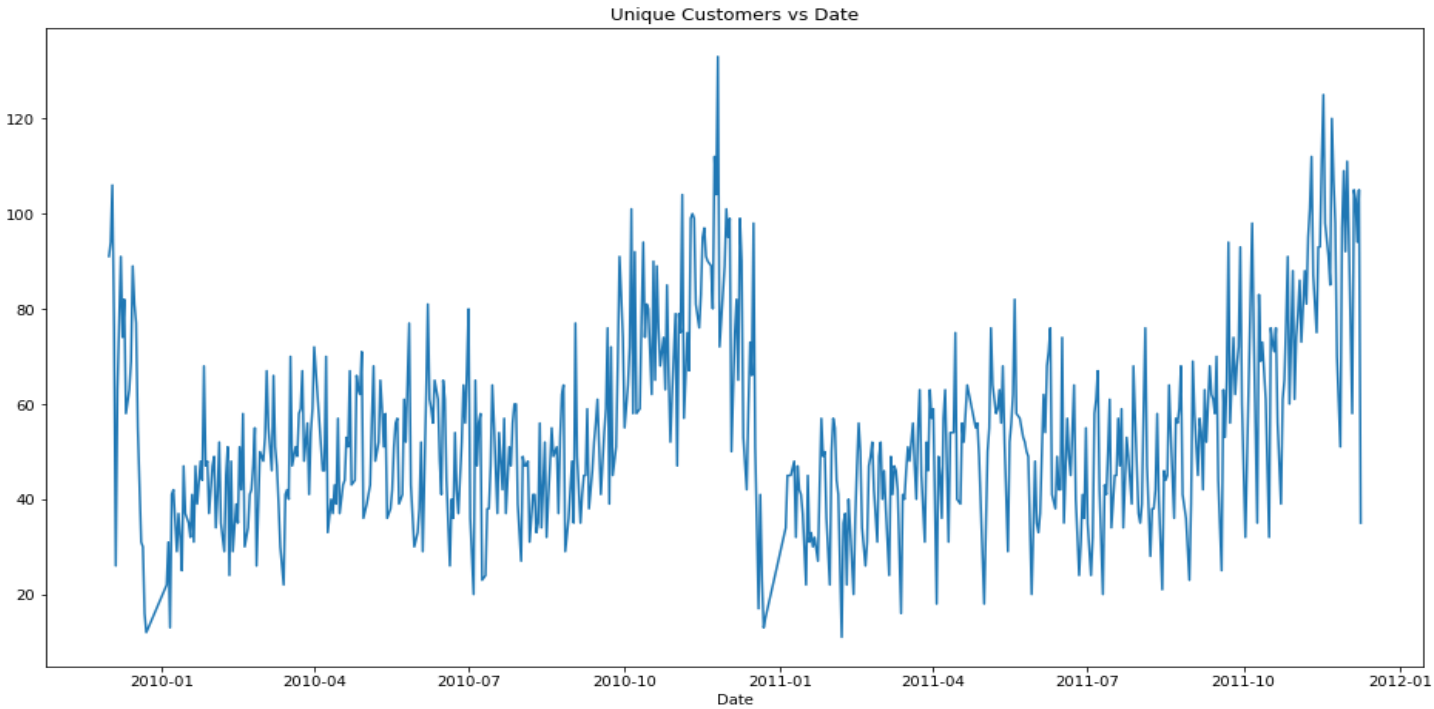


**Q- Is there any seasonality in data? Explain with supportive evidence.**

**A-** Yes there is significant seasonality observed in the data. We plotted two graphs,

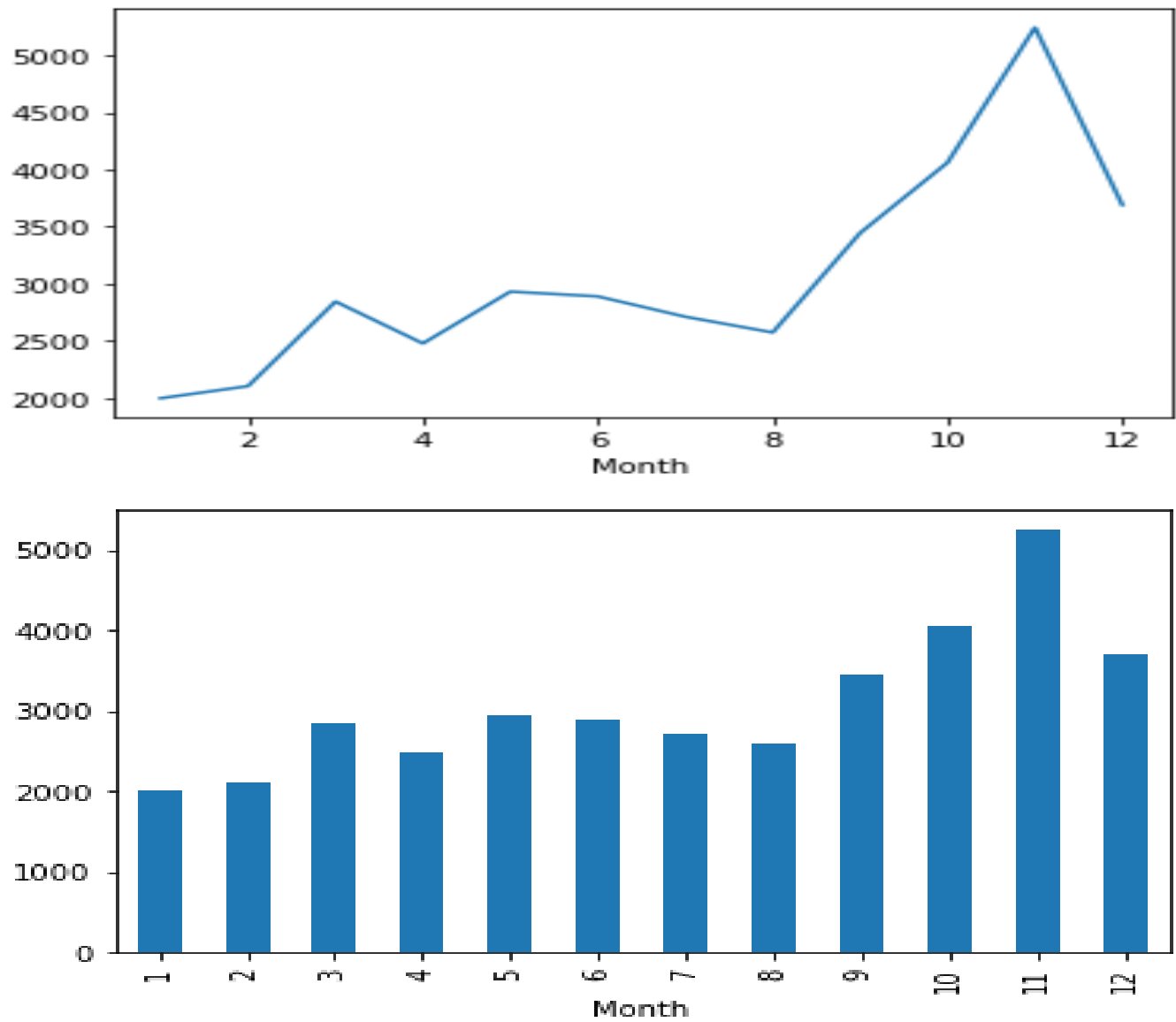
- 1. Number of Unique Customers vs Date
- 2. Number of Unique Invoices vs Date

By observing both the graphs we can conclude that the sale rises for the **winter season** every year (**October to December**) then a sudden drop in the month of **January**. Then we observed a gradual increase in sales for the **summer season (April to September)** followed by a slight drop for the **Autumn season**. At last, a sharp increase in **October** followed by a peak in **November**.



Q- What is the most popular time of the year based on this sales data?

A- The most popular time of the year is the **winter sale** i.e. the period from **October to December**. Plotting a bar plot of the number of invoices vs month showed the maximum number of sales around the month of **November**. This trend in the data can be explained by the fact that it is the festive season during these months and customers will buy a lot of items and lead to profit for the shop.



Q- Which item has been bought the most throughout the year?

A- After grouping the description column and adding all the quantities for the same products, it is clear that **“WORLD WAR 2 GLIDERS ASSID DESIGNS”** is the most bought item throughout the year with a count of **105425**.

Description	
WORLD WAR 2 GLIDERS ASSTD DESIGNS	105425
WHITE HANGING HEART T-LIGHT HOLDER	91877
PAPER CRAFT , LITTLE BIRDIE	80995
ASSORTED COLOUR BIRD ORNAMENT	78445
MEDIUM CERAMIC TOP STORAGE JAR	77916
...	
BLACK DIAMOND CLUSTER EARRINGS	1
ENGLISH ROSE TEA FOR ONE GIFT BOX	1
F FAIRY POTPOURRI CUSHIONS SUMMER	1
AMBER CRYSTAL DROP EARRINGS	1

Q- What are the most and least sold items in the top 5 countries with the most number of sales?

A- The data has been grouped by countries and then by items. Then the total quantity of each item has been calculated. The entire data has been sorted according to it. Following is a list of some of the most and least sold items in each country.

UNITED KINGDOM

MOST SOLD ITEMS

Description	Quantity
WORLD WAR 2 GLIDERS ASSTD DESIGNS	96736
WHITE HANGING HEART T-LIGHT HOLDER	84446
PAPER CRAFT , LITTLE BIRDIE	80995
MEDIUM CERAMIC TOP STORAGE JAR	76919
ASSORTED COLOUR BIRD ORNAMENT	72577

LEAST SOLD ITEMS

Description	Quantity
PINK HEART CHRISTMAS DECORATION	1
RED HEART CANDY POP LIGHTS	1
PINK BABY GIRLS PHOTO ALBUM	1
CIRCUS PARADE LUNCHBOX	1
PINK STRIPES SHOULDER BAG	1

GERMANY

MOST SOLD ITEMS

Description	Quantity
ROUND SNACK BOXES SET OF4 WOODLAND	2680
WOODLAND CHARLOTTE BAG	2570
ASSORTED COLOURS SILK FAN	2100
POSTAGE	2051
RED TOADSTOOL LED NIGHT LIGHT	1796

LEAST SOLD ITEMS

Description	Quantity
TEA COSY RED STRIPE	1
TOILET METAL SIGN	1
PADS TO MATCH ALL CUSHIONS	1
IVORY EMBROIDERED QUILT	1
ZINC BOX SIGN HOME	1

## FRANCE

### MOST SOLD ITEMS

Description	Quantity
SET/6 FRUIT SALAD PAPER CUPS	7176
SET/6 FRUIT SALAD PAPER PLATES	7044
POP ART PEN CASE & PENS	5196
MULTICOLOUR SPRING FLOWER MUG	4992
BLACK SILVER FLOWER T-LIGHT HOLDER	4752

### LEAST SOLD ITEMS

Description	Quantity
ABSTRACT CIRCLES SKETCHBOOK	1
LETTER "L" BLING KEY RING	1
VINTAGE LEAF CHOPPING BOARD	1
SILVER CHRISTMAS TREE BAUBLE STAND	1
BUBBLE GUM CHUNKY KNITTED THROW	1

## SPAIN

### MOST SOLD ITEMS

Description	Quantity
FRENCH WC SIGN BLUE METAL	1788
LOVE GARLAND PAINTED ZINC	1728
ASSORTED COLOUR BIRD ORNAMENT	1009
WOODEN HAPPY BIRTHDAY GARLAND	882
MINI WOODEN HAPPY BIRTHDAY GARLAND	864

### LEAST SOLD ITEMS

Description	Quantity
IVY HEART WREATH	1
ACRYLIC GEOMETRIC LAMP	1
BLUE GIANT GARDEN THERMOMETER	1
ABC TREASURE BOOK BOX	1
WOODEN UNION JACK BUNTING	1

## BELGIUM

### MOST SOLD ITEMS

Description	Quantity
DOLLY GIRL LUNCH BOX	572
SPACEBOY LUNCH BOX	560
MINI PAINT SET VINTAGE	540
PACK OF 60 MUSHROOM CAKE CASES	528
ROUND SNACK BOXES SET OF4 WOODLAND	504

### LEAST SOLD ITEMS

Description	Quantity
ROCOCO WALL MIRROR WHITE	1
WHITE PICNIC BARBECUE BUCKET	1
WHITE BAROQUE WALL CLOCK	1
LE GRAND TRAY CHIC SET	1
I LOVE LONDON BABY GIFT SET	1