Marketing Analytics and Data Mining

Shakeel Raja (shakeelraja@hotmail.com)

ABSTRACT

This report summarizes the data mining and related experimentation conducted on a large sales dataset of an anonymized UK based online retailer, to identify consumer buying behaviours and support managerial decision making towards future sales and marketing strategies. The report gives an introduction to the role of marketing analytics within BI domain and some analytical questions currently asked by marketing experts. Candidate techniques have been selected for the purpose of analysis and testing, followed by data pre-processing steps. The report presents a statistical and visual analysis with findings of applying clustering and association mining algorithms to marketing data in order to identify customer behaviour and to aid product promotions. Results show a great potential of these techniques towards developing intelligent marketing analytics systems leading to productive marketing strategies for data driven organisations.

1. INTRODUCTION

Business Intelligence (BI) is a set of technology driven strategies and processes that deals with application of advanced analytical techniques on business data and decision automation to provide corporate executives with critical insights into core operations and a high degree of control over key performance areas. [1]. BI systems provide historical, current and predictive analysis capabilities with features like advanced reporting, OLAP, digital dashboards, data mining and warehousing and process visualisation etc. Business Data analytics can provide unprecedented benefits to data driven organisations and significantly increases the likelihood of improvement in their decision making capabilities [2]. In this regard, BI and big data analytics equipped with data mining and machine learning tools have become increasingly popular in data driven organisations for advanced analysis and exploitation of vast amounts of business data in order to gain competitive advantage. [3].

Marketing analytics practices form a large part of BI and are centred on gathering and consolidating customer interaction and sales data for actionable analytical insights leading to the development of intelligent and reactive marketing strategies. According to latest CMO survey, data-driven companies spend about 6.7 percent of their marketing budget and this is expected to grow to more than 11% in next three years [4]. The main challenge faced by marketing professionals is to quantify the impact of their sales and marketing strategies, which can further improve decisions about future strategies and plan investments into marketing analytics. Marketing Analytics aims to provide solutions to these challenges by providing analytical insights into plethora of historical and current customer and sales data. This results as marketing intelligence and attempts to answer strategic analytical questions.

The objective of this report is to provide an analytical marketing framework using different data mining approaches to achieve customer and product insights from historical sales data of real products. The outcome of this report would provide a model for identifying customer segments according to their buying behaviours and a market basket analysis for sales and inventory planning. Following two analytical questions have been addressed in this regard:

KNOW THY CUSTOMER

Q1. Can we use data mining techniques on past transactions to understand and value our customers' buying behaviours individually and group them in segments?

Understanding customers buying behaviour helps identify customer segments from most valuable to least valuable. This knowledge can be used towards marketing campaigns, customer acquisition, customer retention and controlling customer churn rate with reactive marketing strategies.

SALES OPTIMIZATION

Q2. Can we use data mining techniques on historical sales data to identify associations between products or services likely to be bought together in future?

Knowledge of products likely to be sold together could be a valuable commodity while developing future sales and promotion strategies. This can also be beneficial towards inventory planning and channel optimization and leads to automated product recommendations for customers.

2. OBJECTIVES

This analysis exercise is to apply marketing analytics techniques to real-world sales and customer data in order to understand the buying behaviours of customers over a period of one year and to identify potential for improved sales and promotion practices. Following set of objectives has been identified for this analysis:

1. Identification of a sales dataset offering ample amounts of data needed for analysis.
2. Formulating and analysis strategy and selection of appropriate software tools.
3. Data pre-processing to eliminate and correct data that not suitable for analysis.
4. Applying RFM calculations needed for segmentation of customers.
5. Evaluating K-means and hierarchical clustering (unsupervised) for customer segmentation along with hyper-parameter selection i.e. number of segments.
6. Applying association mining to perform a market basket analysis
7. Presenting the results of findings graphically and statistically

3. DATASET DESCRIPTION

The "Online Retail Dataset" from UCI machine learning repository [5] used in this experiment contains 541909 transactions between 1/12/2010 to 9/12/2011 for a UK based online retail store specialising in all-occasion gifts, mainly to other online and high street retailers. The company also uses Amazon as an intermediary to sell its products. The company has its primary clientele in UK but also has clients in other parts of Europe. There are a total of 8 variables stored in the dataset against each transaction as follows:

- Nominal: "InvoiceNo", "StockCode" and "CustomerID" as unique identifiers for transaction, product and customer respectively. "Description" and "Country" as text based descriptions for product and country of customer.
- Numeric: "Quantity" as number of units sold with "UnitPrice" as price per unit item.
- Date/Time: "InvoiceDate" as timestamped information of date and time of the transaction.

The huge number of transactions presented in this dataset makes it ideal for application of data mining techniques to study customers' buying behaviours and product associations.

4. ANALYSIS STRATEGY

Following diagram [Fig. 1] summarizes the analysis strategy used in this exercise. The strategy is mainly based around performing necessary steps to meet the analytical objectives defined earlier. Further details on individual steps are provided in detail later in the report.

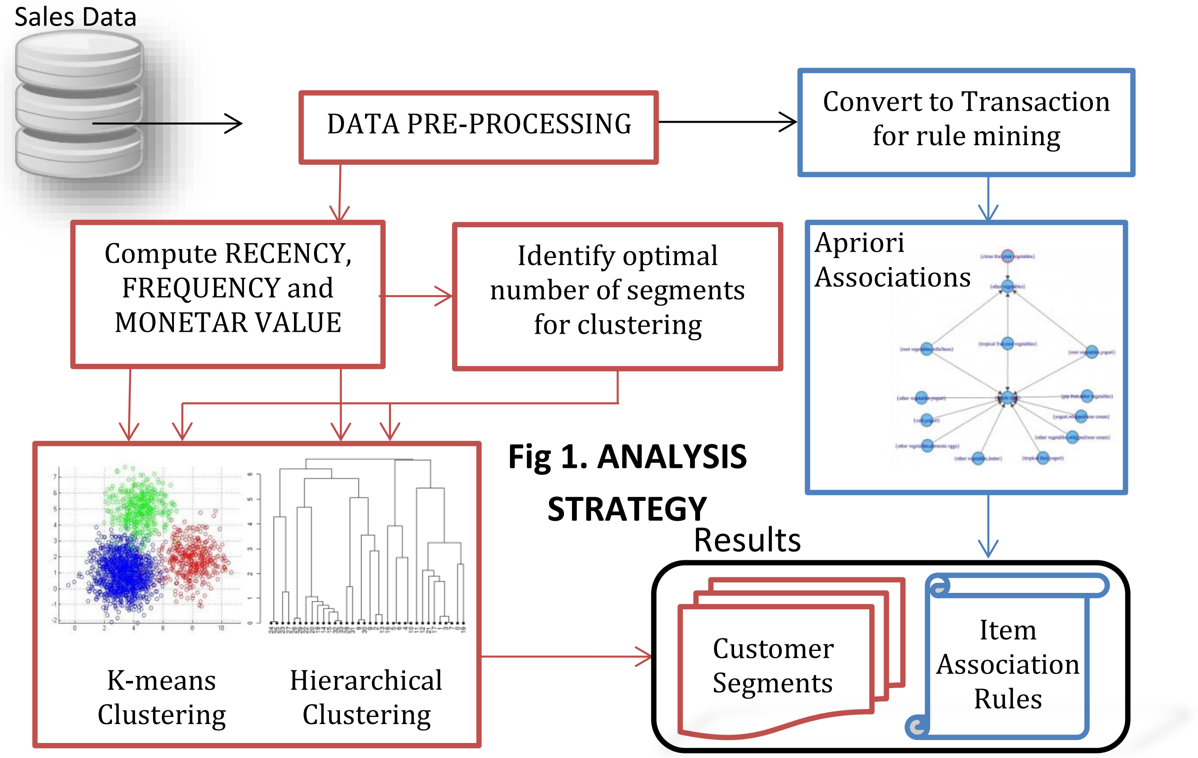


Fig 1. ANALYSIS STRATEGY

5. TOOLS

Following tools were selected for the analysis based on the suitability for the task, past experience and availability.

- **Python** was the primary tool of choice for this experiment due to its versatility and robustness. numpy, Pandas, ScikitLearn and Seaborn libraries were imported into python environment to aid data pre-processing, clustering evaluation, implementation and visual analysis of the dataset. Both the hierarchical clustering and K-means clustering experiments were performed in the python environment.
- **R statistical programming** environment was chosen for its popular implementation of Apriori association algorithm, with the arules and arulesviz libraries which also come packaged with superior graphical capabilities.
- Results from clustering and association were plotted using **Plotly** library. This proved exceptionally beneficial towards visual analysis of clusters to identify optimal customer segments and association rules with necessary dimensions.

6. DATA PRE-PROCESSING

Although the chosen dataset contains a wealth of information in terms of number of transactions, our initial data exploration highlighted the need for some preprocessing in order to make it suitable for clustering and association. Initial data wrangling steps included following:

- "CustomerID" contained 135080 whereas "Description" contained 1454 null values. For customer centred RFM clustering, rows with missing "CustomerID" were dropped as RFM calculation is based on individual customers IDs. These values, however, were considered for association mining which was based on "InvoiceNo". Other variables had no missing values.
- Some "Quantity" and accompanying "UnitPrice" values appearing in negative indicated damaged/returned items and refunds and were removed while preparing the data for clustering.
- A new variable "Sales" was generated from "UnitPrice" and "Quantity" to reflect the total revenue generated as a result of selling a particular product in a transaction.
- Appropriate categorical, numerical and text types data types were assigned to all variables.

The resulting dataset had a total of 397924 records with 4339 unique customers, 3665 unique products.

7. ANALYSIS STAGE 1: RFM CLUSTERING

RFM analysis is a classical segmentation tool used to calculate customers' value to the business. The idea of marketing segmentation was first introduced in [6] as a way to divide customer base into subsets of customers who behave in a similar fashion. RFM Modelling and analysis is thus a behaviour based methodology used to analyse the buying behaviour of customers for future strategies. Traditional RFM analysis offered supervised segmentation based on quantiles of customers' recency, frequency and monetary value (R, F and M) using Pareto's principle. During recent years, data mining techniques mainly clustering has been vastly applied to transaction data to intelligently create unsupervised RFM market segmentation [7][8]. Clustering algorithms belong to the category of unsupervised machine learning algorithms used to discover patterns and structure in the given data without making any distinction between attributes. Cluster analysis finds groups of data that are homogeneous yet different from other clusters. These methods provide a robust and effective way of segmentation analysis and exploitation.

For unsupervised clustering, K-means and hierarchical clustering algorithms have been widely adopted for combining observed examples into clusters based on R, M and F attributes. K-means clustering takes in k as the number of centroids for the clusters and repetitively associates nearest examples to each centroid forming k clusters. Hierarchical clustering, on the other hand, is a two-step process where the data first gets compressed as sub-clusters. These sub-clusters are then hierarchically and progressively merged to form larger clusters.

RFM MATRIX NORMALIZATION

First step towards RFM analysis was the creation of a matrix with three new variables for each customer namely Recency, Frequency and Monetary value as given below:

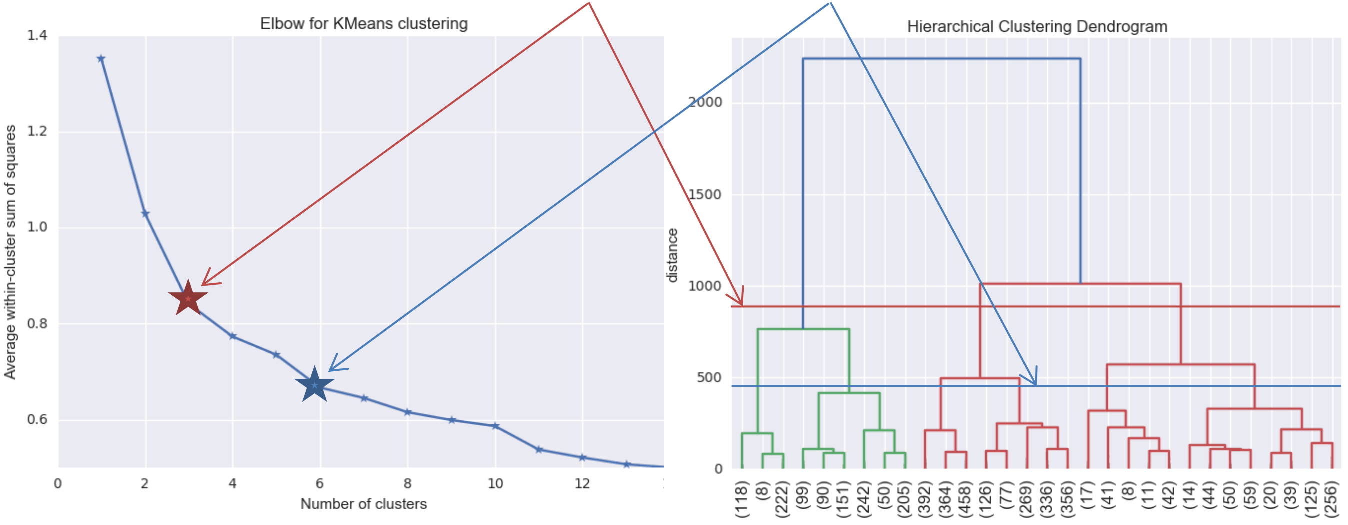
Recency :	Total Time since a customer's last purchase.
Frequency:	Total number of purchases by a customer.
Monetary value:	Total amount spent by a customer in all transactions.

The key objective behind the development of RFM matrix was to allow clustering algorithms to segment customers based on similarities between these attributes. From the histograms shown in Appendix A.1, It was found that the calculated RFM matrix had some extreme outliers in frequency and monetary value attributes. Clustering algorithms, however, are generally very sensitive to outliers and attributes with varying scales as described in [9]. The outliers instances were therefore removed using SciPy's stats package by leaving out the values that existed outside the 3rd standard deviation. Also, the R, M, F attributes were normalized on a scale of 0 – 100 using min-max normalisation with scikit-learn's pre-processing package in Python. Final dataset for RFM analysis had 4,290 customers

NUMBER OF CLUSTERS

Next phase was deciding the number of suitable segments as a hyper-parameter to clustering algorithms. In a real world scenario, this would have been done with input from marketing executives having in-depth knowledge of their clientele. However, for this experiment, it was decided to use standard visual methods including "Elbow method" and "Dendrogram" approaches for identifying optimal number of segments from the data.

Elbow method plotted the internal sum of square distance against number of clusters following K-means clustering. It could be seen that the line graph formed a shallow curve with an elbow appearing at k=3 and again at k=6 clusters. This made it challenging to visually decide an optimal number for k . Looking at the dendrogram used in hierarchical clustering showing distance among clusters on y-axis and revealed that at k=3, there was risk of losing too much customer information as 3 clusters could not compensate for the variance in customer behaviours. Iterating through this process a number of times with values of k from 3 – 9, and looking at the output revealed that with k=6, data was divided into much more balanced clusters while preserving the variance in recency, frequency and monetary value attributes of customers.



At k = 3, we get into a risk of losing too much customer information as seen in dendrogram

At k = 6, dendrogram shows nicely formed clusters while avoiding big jumps along y axis

Fig 2. The Elbow Method and dendrogram used for visual identification for optimal number of clusters to identify customer segments. The optimal range of clusters appeared between 3 and 6. Eventually, 6 clusters were chosen to ensure retention of buying behaviours in clusters.

Cluster Analysis

After running the clustering algorithm with k=6, it was found that understanding customer segments using 2D graphs and statistics proved to be very challenging. Most of the information differentiating clusters was lost while trying to graph these using simpler methods. To avoid this and to clearly view the varying behaviour of the customers in terms of R, F and M values, it was decided to map the clusters with attributes values as a 3 dimensional scatter plot where X= Receny, Y= Frequency and Z = Monetary value. Cluster labels were used to colour customer examples into calculated clusters. This proved a valuable way to visually analyse customer segments.

Our resulting RFM matrix with cluster values appended was exported to plotly mapping application. This allowed on-screen rotation of the 3D plot facilitating the view from different angles to study the clusters in more detail. Interactivity with 3D plots offered by plotly also helped studying the individual customers in each cluster/segment. A screenshot of results is shown in fig 3 below.

K-Means Clustering Results

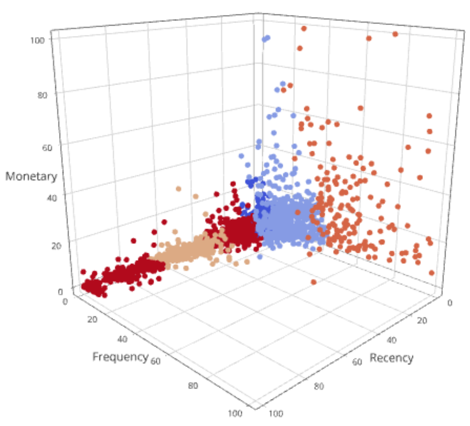
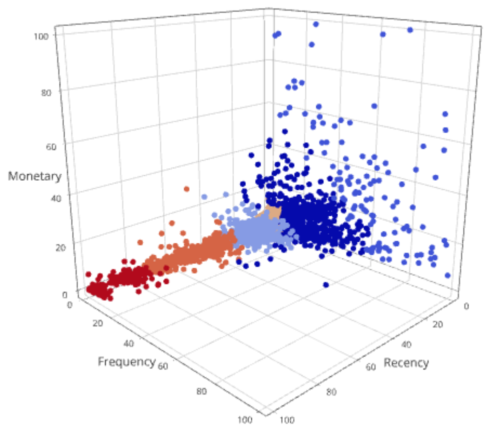


Fig 3.
Both clustering algos. show similar results. Some examples were thought to be better clustered with hierarchical clustering. Results of h-clustering were used for further analysis of customer behaviour.

Hierarchical Clustering results



Average RFM values for K-Means clusters

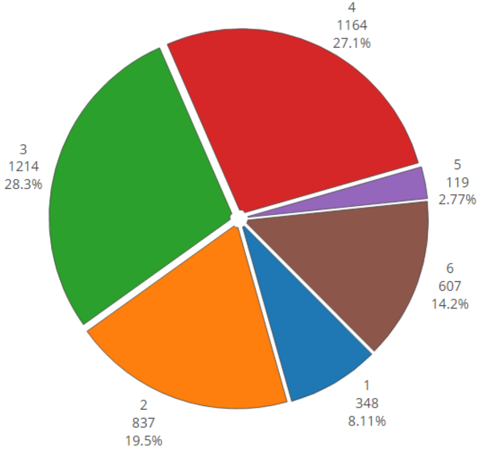
Cluster	recency	frequency	monetary	count
0	23.554581	5.242268	2.948104	891
1	82.695343	3.004433	1.425712	496
2	3.831482	58.93964	26.55983	175
3	51.816897	4.238634	1.978700	592
4	5.931290	25.42796	12.47734	582
5	6.423441	6.607336	3.278482	1553

Average RFM values for Hierarchical clusters

Cluster	recency	frequency	monetary	count
1	87.714554	3.078328	1.308276	348
2	53.226287	3.912586	1.950713	837
3	4.899718	6.866251	3.345026	1214
4	19.209207	6.280012	3.151891	1164
5	2.958073	60.18476	36.09133	119
6	5.013007	27.82766	12.181241	607

Both clustering algorithms provided similar classification of customers into segments. It was however noticed that K-means clustering created a large cluster (cluster 5) of customers having low values of R, M and F attributes. This cluster reflected new casual customers who could be turned into loyal customers in future through targeted marketing. Cluster 6 showed customers with slightly higher value of R and similar F and M. This could be a reflection of less recent casual customers who were on the verge of leaving the business and reactive strategies must be put in place for their retention. Hierarchical clustering algorithm's clusters 3 and 4 reflected the same idea, however, customers were split into these segments more efficiently. Due to this, Hierarchical clustering was chosen to further study the customer segments and mean behaviour of each segment in terms of R. F and M to allocate customer personas.

Fig. 4 Hierarchical Clustering Results



Following scale was used to categorize mean RMF values

Very Low	**VL**	Below 10 %
Low	**L**	11- 20 %
Lower Medium	**LM**	20 – 40%
Upper Medium	**UM**	40 – 60%
High	**H**	60 – 80%
Very High	**VH**	80% and Above

This table calculated above was further used to assign "Personas" to each segment based on the scores of their mean R, F and M attributes to help internalize and recognize the customers.

Segment	**Recency**	**Frequency**	**Monetary**	**Persona**
1	VH	VL	VL	LOST CHEAP USTOMER
2	UM	VL	VL	ALMOST LOST CUSTOMERS
3	VL	VL	VL	RECENT CASUAL CUSTOMERS
4	L	VL	VL	CASUAL CUSTOMERS
5	VL	H	UM	REGULAR BIG SPENDERS
6	VL	LM	L	PROMISING CASUAL CUSTOMERS

The resulting table shown above can be used to include different segments into relevant marketing campaigns. This may involve email, social media or regular mail marketing. Also, reactive strategies can be developed for customers who are on the verge of leaving the business and reward schemes for customers who are loyal with potential of spending big in the future. The results proved to the a success towards answering the first analytical question dealing with creating automated customer segmentation for behaviour identification.

ANALYSIS STAGE 2: ASSOCIATION RULE MINING (Market Basket Analysis)

Association Rule Mining (ARM) is a data mining technique, used for identification of interesting associations between attributes of large databases. ARM helps discovery of "rules" using some measure of "interestingness" as shown in [10]. The idea of using ARM for discovering associations in product sales from large transactional databases was first introduced by [11] demonstrating that this technique could be used to examine customer buying patterns by identifying associations among different items that are most likely to be bought together by customers. This idea is knows as Market Basket Analysis (MBA) which can be useful for future sales and marketing strategies including cross-selling and up-selling, catalogue design and product promotions etc.

This analysis was performed to answer our second analytical question addressing the need finding patterns in product sales to improve future sales and promotion strategies based on historical transactional data.

APRIOR ALGORITHM

Apriori Algorithm is used to frequent product mining and learning association rules from given data. It performs by identifying most frequent items in the dataset and extending their relationship to other items forming larger rules of association [12]. Apriori is considered to be one of the most popular algorithms for association rule mining. The evaluation metrics used by apriori algorithm use X and Y for antecedent and consequent products which are likely to be sold together:

SUPPORT: Fraction of transactions containing both X and Y examples.

CONFIDENCE: How often each item in Y appears in transactions that contain X.
SUPPORT(X+Y)/SUPPORT(X)

LIFT: Lift of an association rule is the confidence of rule divided by the expected, assuming mutual independence between items. A lift value > 1 represents that X and Y appear more often than expected and vice versa.

In Order to Apply Apriori algorithm on our dataset, "InvoiceNo" and "StockCode" variables from the original dataset were exported to R programming environment. The variables were converted into a transaction type in R which is highly recommended for efficiency in data mining practices. Apriori implementation in Arules library was applied on resulting transaction for association mining.

ASSOCIATION RULES ANALYSIS

Apriori algorithms identified 981 rules (Fig. 5) with hyper-parameters set as minimum Support = 5% and Confidence = 80%. Also, "minlen" parameter was set to 2 to include all rules containing at least two item associations. The resulting set of rules was cleansed to remove any redundant entries and final results were plotted to investigate the associations further. A graphical summary of these steps is provided below.

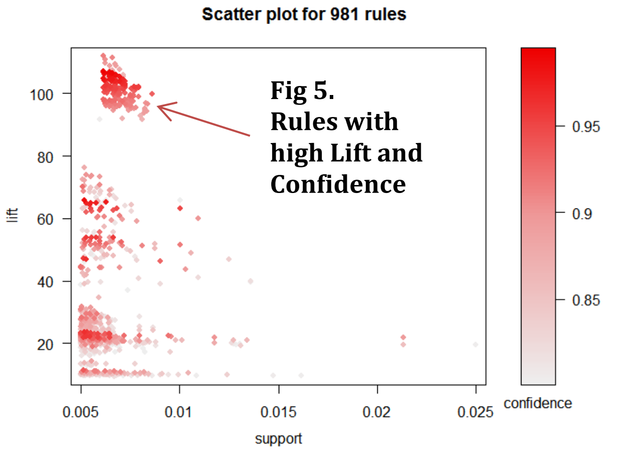


Fig 5. Rules with high Lift and Confidence

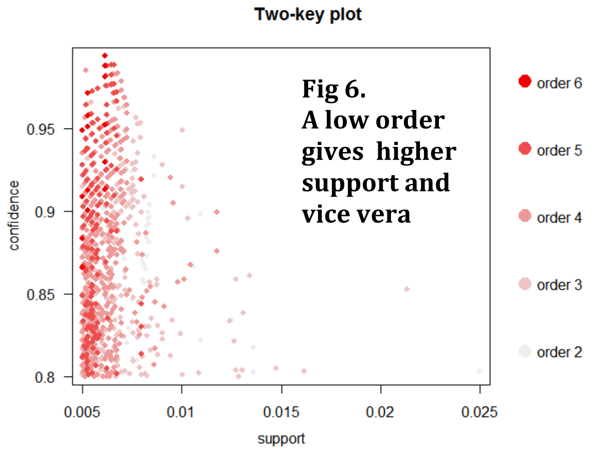


Fig 6. A low order gives higher support and vice vera

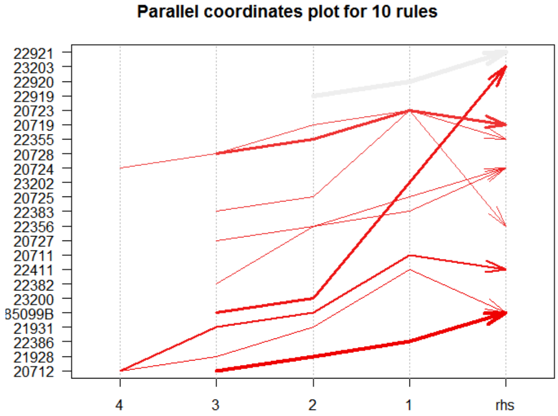
A specialised scatter plot, called the "two-key plot" revealed relationship between Order (number of items in each rule) and support (Fig 6). It was observed that support declines as the number of items in each rule increase. These rules can be further studied to discover frequent items which appear in multiple rules and how they are associated with other items. In order to simplify our analysis, it was decided to select top ten rules from the complete set to further explore how individual products are associated together in groups of 2, 3 and more association.

The parallel co-ordinate plot shown in Fig 7 indicates top 10 rules extracted from the complete set and how different products are associated in groups of up to 5.

A graph based depiction of frequent item and top rule associations can be viewed in Appendix A.2. This insight into product sales can indeed prove to be a valuable asset for a data driven organisation towards developing future sales strategies.

Finally a grouped matrix of top 10 rules, as shown in figure 8 helped better explore the association between different items in terms of changing support. LHS reflects the item which triggers the sale of other items in shopping basket.

The amount of knowledge generated as a result of association mining with apriori algorithms clearly reflects the potential of association mining towards answering the analytical question of performing a market basket analysis on historical data to uncover hidden sales patterns that might be used for future sales strategies.



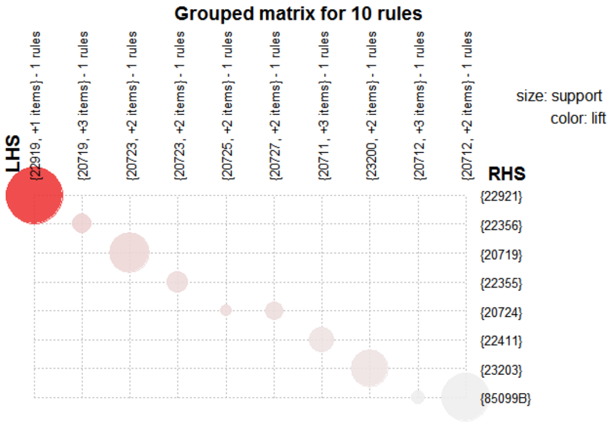


Fig. 8.
Grouped Matrix

9. FINDINGS AND FUTURE WORK

The experimentation and analysis performed in this customer centric marketing intelligence exercise gave promising results in terms of identifying customer buying behaviour and finding the association between products in customers' shopping basket and showed great potential for future experiments.

In the first part of our Analysis, clustering algorithms were successfully used on normalised RFM matrix to group customers into meaningful segmentations and helped assign a persona to each segment. This type of analysis can be further expanded to include different variations of RFM analysis techniques prescribed for specific business types e.g. weighted RFM (assigning market specific weights to RFM) and Temporal RFM (observing segments at different time periods to observe customers moving between segments).

Association Rule mining exercise revealed the likelihood of selling two or more products together while mining through historical sales data. This outcome addressed the initial analytical question dealing with future sales and promotion strategies. Due to anonymization of data, customers' demographic information could not be included in this exercise. Such information (e.g. location, gender, education, profession etc.) can be used to mine much more meaningful and specific associations which may lead to advanced operations like product recommendations, personalisation of services and Just-in-time inventory management.

Based on above findings, it can be confidently said that our objectives for this exercise have been successfully met.

References

1. Ranjan, Jayanthi. "Business intelligence: Concepts, components, techniques and benefits." *Journal of Theoretical and Applied Information Technology* 9.1 (2009): 60-70.
2. "Big Decisions™". *PwC's Global Data and Analytics Survey,* Global, base: 1,135 senior executives, May 2016

3. Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS quarterly* 36.4 (2012): 1165-1188.
4. “Ten Steps To Better Use Of Marketing Analytics”. *CMO Survey*. https://cmosurvey.org/marketing-analytics/ten-steps-better-use-marketing-analytics/ (Last visited 1/12/2016)
5. “Online Retail Dataset”, *UCI Machine Learning Repository*. https://archive.ics.uci.edu/ml/datasets/Online+Retail. (Last visited 1/12/2016)
6. W. R. Smith, “Product differentiation and market segmentation as alternative marketing strategies,” *The Journal of Marketing*, vol. 21, no. 1, pp. 3–8, 1956.
7. Daqing Chen, Sai Liang Sain, and Kun Guo, “Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining”, *Journal of Database Marketing and Customer Strategy Management*, Vol. 19, No. 3, pp. 197-208
8. M. Namvar, M. R. Gholamian, and S. KhakAbi, “A two phase clustering method for intelligent customer segmentation,” in *Intelligent Systems, Modelling and Simulation (ISMS)*, 2010 International Conference on, 2010, pp. 215–219.
9. Loureiro, Antonio, Luis Torgo, and Carlos Soares. "Outlier detection using clustering methods: a data cleaning application." *Proceedings of KDNet Symposium on Knowledge-based Systems for the Public Sector*. Bonn, Germany. 2004.
10. Piatetsky-Shapiro, Gregory (1991), “Discovery, analysis, and presentation of strong rules”, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., *Knowledge Discovery in Databases,* AAAI/MIT Press, Cambridge, MA.
11. Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* - SIGMOD '93. p. 207. doi:10.1145/170035.170072. ISBN 0897915925.
12. Rakesh Agrawal and Ramakrishnan Srikant. “Fast algorithms for mining association rules in large databases”. *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB, pages 487-499, Santiago, Chile, September 1994.

APPENDIX A

1. RFM MATRIX: removing outliers and normalization effect.

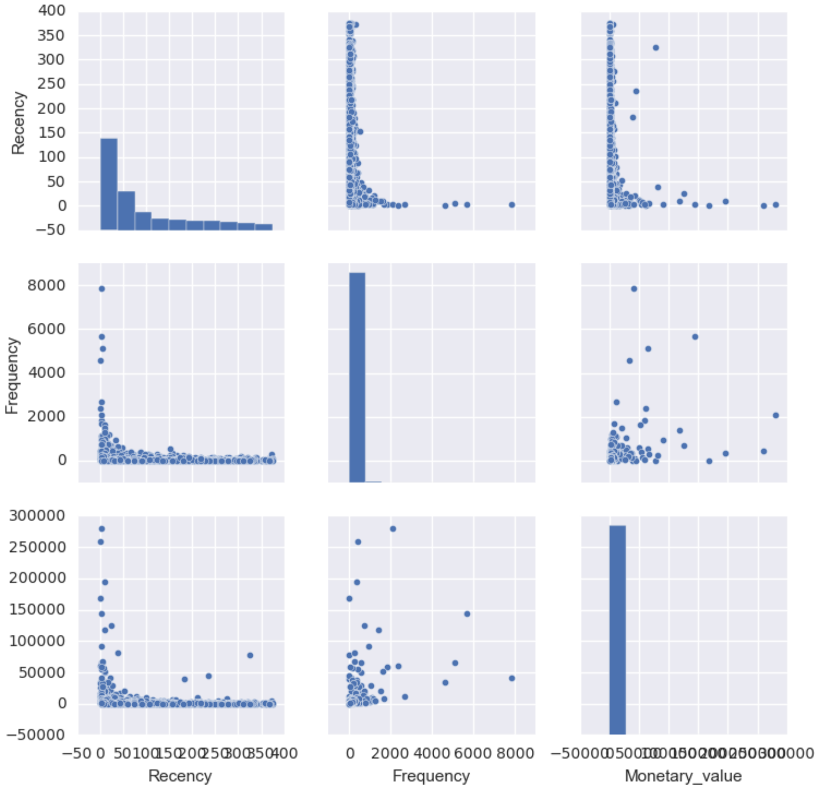


Figure (a) – data with outliers and un-normalized attributes

RFM matrix had some extreme outliers in the Frequency and Monetary value attributes. Also, the data all three attributes had varying scales. This is evident from the set of histograms in the top figure.

These instances highlighted important customers that must be dealt with great care and may provide further insight into buying behaviours

The outliers were removed using sciPy stats package. Data was normalized using min-max scaling on a range from 1 – 100 using scikit-learn preprocessing package.

The figure at the bottom shows a normalized set histograms which provided more meaningful results with clustering techniques applied.

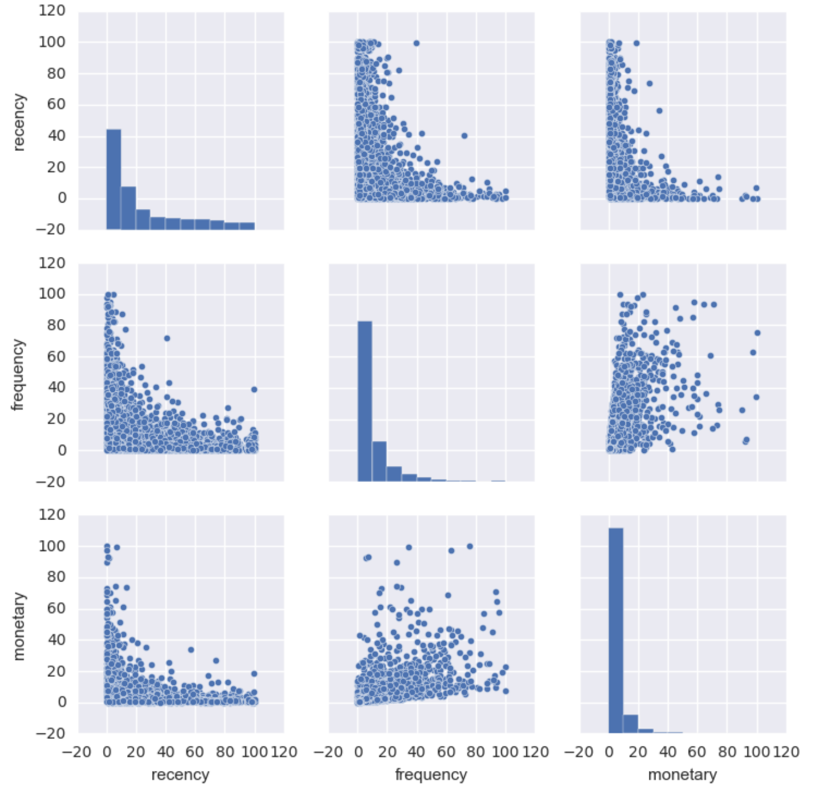
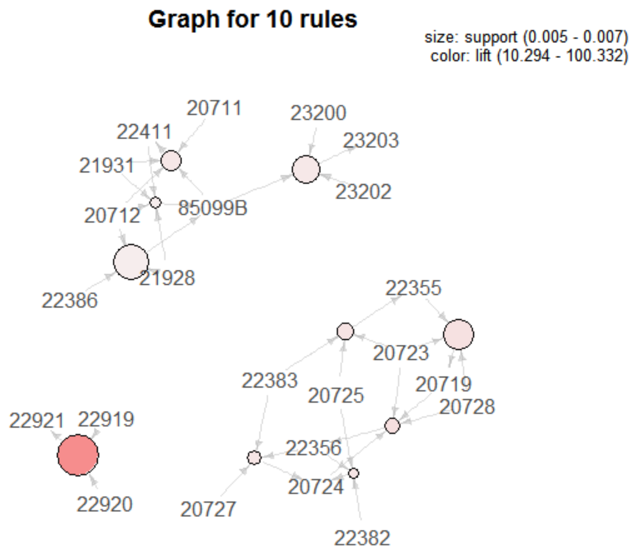


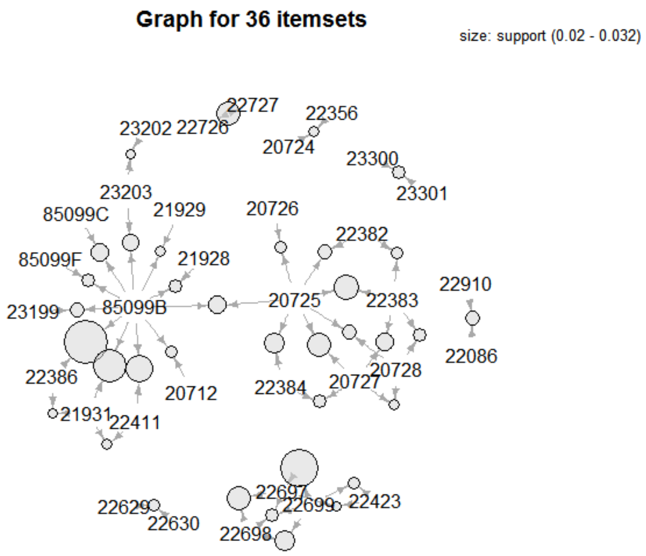
Figure (b) – data with outliers removed and normalized from 1 - 100

2. Graphical Depiction of Item and Rule association from Apriori algorithm

First graph highlights top 36 frequent items from the dataset and identifies relationships between them while showing their associations in terms of support as the size of nodes connecting them together. Second graph shows top ten rules and how different items are repeated within different rules. The size of nodes reflect the support and color intensity reflects the value of lift.

This information can provide a valuable insight fir planning future sales and promotion strategies.


Graph for 10 rules
size: support (0.005 - 0.007)
color: lift (10.294 - 100.332)
20711
22411
21931
20712
85099B
23200
23203
23202
22386
21928
22355
20723
22383
20725
20719
20728
22356
20724
20727
22382
22921
22919
22920


Graph for 36 itemsets
size: support (0.02 - 0.032)
22727
22356
23202
22726
20724
23300
23301
23203
85099C
21929
20726
22382
85099F
21928
23199
85099B
20725
22383
22910
22086
22386
20712
22384
20727
20728
21931
22411
22629
22630
22697
22699
22423
22698