Paper 1267-2017

Behavioural spend modelling of cheque card data using SAS® Text Miner

Amelia Van Schalkwyk, Frans Kanfer, and Sollie Millard, University of Pretoria.

ABSTRACT

Understanding customer behavior profiles is of great value to companies. Customer behavior is influenced by a multitude of elements-some are capricious, presumably resulting from environmental, economic, and other factors, while others are more fundamentally aligned with value and belief systems. In this paper, we use unstructured textual cheque card data to model and estimate latent spending behavioral profiles of banking customers. These models give insight into unobserved spending habits and patterns. SAS® Text Miner is used in an atypical manner to determine the buying segments of customers and the latent buying profile using a clustering approach. Businesses benefit in the way the behavioral spend model is used. The model can be used for market segmentation, where each cluster is seen as a target marketing segment, leads optimization, or product offering where products are specifically compiled to align to each customer's requirements. It can also be used to predict future spend or to align customer needs with business offerings, supported by signing customers onto loyalty programs. This unique method of determining the spend behavior of customers makes it ideal for companies driving retention and loyalty in their customers.

INTRODUCTION

Businesses are driven by their ever changing goals and aspirations which may differ from large corporations to sole traders, however they all have one very significant commonality, aiming to sell their products and services. Whether companies are selling groceries, bank accounts or fitness classes, their products are their underlying structure. For this reason it is of paramount importance to each company's continuous success to sell their products efficiently. In the fluctuating uncertainty that arises from environmental, political and economic change, it is essential for companies to understand the needs, constraints and aspirations of their customers. To do this it is necessary to delve deeper into the behaviour of customers, this type of modelling can be known as a form of behavioural modelling that stems from fields such as social and behavioural sciences. The term behaviour is fairly broad, for example it can range from how often a customer travels to what time of day they spend their money. Consumer behaviour falls within the fields of social and behavioural sciences as it is a component of social behaviour. Krishna, [3], defines consumer behaviour as “A discipline that encompasses all processes involved in acquiring, using and disposing of products, services and ideas. As a discipline it is also concerned with the results of such acquisitions, uses and disposals”. This paper focuses primarily on the acquisition side of consumer behaviour, more specifically the purchasing/buying/spending behaviour of the customers and not the psychological reasons behind acquisition.

Behavioural spend modelling is one of the many possible consumer behaviours that can be modelled, it is the process of detecting and understanding patterns in the spend data of customers. These identified patterns assist in determining behavioural segments and spending profiles. Interpreting the model and recognising whether customers spend more on groceries, holidays or restaurants, etc., in other words whether they spend on luxury items or necessities, assists in understanding the customers lifestyle and unique needs.

SAS® text miner is used in an atypical manner to determine the spending segments and latent spending behaviour of customers using cheque card data. The latent spending behaviours of a customer are in turn used to determine the customers spending profiles. The cheque card data is transformed in such a way that the result consists of a combination of spending terms making up a customer’s spending document. This assists SAS® text miner in capturing all the spend in each customer document as behaviours. The results provided in this paper suggest the use of SAS® text miner as a tool for modelling behavioural spend is a successful one, the customer selected at random and analysed produces results indicating a lavish lifestyle, this could identify a certain set of products uniquely suited to this customers lifestyle. This

customer is also identified as belonging to a specific segmentation of spend, this could assist in identifying leads if similar customers were sought out. The capabilities of the spending profile and customer segmentation are endless if utilized correctly.

BEHAVIOURAL SPEND MODEL

This consists of 3 sections, the first describes the data used to fit the behavioural spend model. The data in its original format had to be transformed to obtain data usable for SAS® text timer. This data was manipulated to obtain three alternate sets of input structures for the model, this was done to obtain, view and compare a variety of possible spending profiles and segments. The second section describes the objectives of the model and highlights the differences between a customer's spending profile and their spending segment. Finally, the last section displays and discusses the results of the model.

DATA PREPARATION

The data used for the behavioural spend model consists of cheque card data for one month. This data contains 6000 unique customers with an income ranging from R700k to R1,000k p.a. The data was cleaned and studied, which identified three transactional text columns of interest. The first column contained the description of the transaction that occurred, the second contained the category of the transaction and the last contained a broad transaction store type. An example of the column structures can be seen in Table 1: Example of transactional data.

Trans_Desc	Trans_Cat	Trans_Type
MICROSOFT *STORE	Computer Software Stores	Miscellaneous Stores
C*SILVER MOUNTAIN SPUR	Eating Places, Restaurants	Miscellaneous Stores
WOOLWORTHS GRASSY PARK	Grocery Stores, Supermarkets	Retail Store

Table 1: Example of transactional data

It is interesting to note that the transaction descriptions occasionally record the area the transaction took place in. For example, consider row three of Table 1: Example of transactional data, this transaction took place at a retail store called Woolworths in Grassy Park, which is a suburb in the province of the Western Cape.

The original data contained multiple transactional entries per customer, due to the independent capture and storage of each transaction. For this reason it was necessary to summarise the data to obtain a single unique text column per customer, accomplished through the concatenation of the three text columns across all transactions per customer. The before and after view of the concatenation process can be observed in Table 2: Data layout before concatenation and Table 3: Data layout after concatenation.

Customer document	Trans_Desc	Trans_Cat	Trans_Type
1	Description1	Category1	Type1
1	Description2	Category2	Type2
1	Description3	Category3	Type3

Table 2: Data layout before concatenation

Customer document	Trans_Text
1	Description1 Category1 Type1 Description2 Category2 Type2 Description3 Category3 Type3

Table 3: Data layout after concatenation

Three slightly altered textual input structures were obtained using the above method of concatenation. The first contained only the first column of text, the transaction descriptions or Trans_Desc, see Table 4: Data layout after concatenation of text column 1, this is the first textual input structure. The second contained the first and second columns of text, the transaction descriptions and transaction category, or Trans_Desc and Trans_Cat, see Table 5: Data layout after concatenation of text columns 1 and 2, this is the second textual input structure. Finally, the third contained all three text columns of interest, the transaction descriptions, transaction category, and transaction type, or Trans_Desc, Trans_Cat and Trans_Type, as in Table 3: Data layout after concatenation, this is the third textual input structure. These three input structures were obtained to determine whether different results would be attained through the inclusion of the extra text columns.

Customer document	Trans_Text
1	Description1 Description2 Description3

Table 4: Data layout after concatenation of text column 1

Customer document	Trans_Text
1	Description1 Category1 Description2 Category2 Description3 Category3

Table 5: Data layout after concatenation of text columns 1 and 2

In terms of text mining and for the purpose of SAS® text miner, a single customers concatenated set of transactions will be viewed as a document and the words within that document will be viewed as the terms. In other words a collection of spending terms makes up the customer's spending document. Figure 1: Example of the spending terms within the second textual structure from customer document 3998 displays all the spending terms within spending document 3998, in the case of the second textual input structure.[1]

[1] Each of the 6000 customer documents are identified by a number between 1 and 6000. Customer document 3998 simply refers to the 3998th document in the corpus.

Customer document	Trans_Text
3998	Uber BV Taxicabs/Limousines PICK N PAY DAINFERN SQU Grocery Stores, Supermarkets STER KINEKOR FESTIVAL Motion Picture Theaters PICK N PAY DAINFERN SQU Grocery Stores, Supermarkets PICK N PAY DAINFERN SQU Grocery Stores, Supermarkets WAXIT Barber and Beauty Shops WESTPACK EXPRESS LONEHI Home Supply Warehouse Stores CELLTECH LONEHILL Telecommunication Equipment Including Telephone Sales PIZZA HUT DAINFERN Fast Food Restaurants WOOLWORTHS LONEHILL 306 Grocery Stores, Supermarkets CHECKERS HYP FOURW 2-73 Grocery Stores, Supermarkets SASOL NORTHRIDING Service Stations (with or without ancillary services) ELNA FOURWAYS Furniture, Home Furnishings and Equipment Stores, Except STER KINEKOR - FOURWAYS Motion Picture Theaters HAPPY ME FOURWAYS Eating Places, Restaurants AMCI PIZZERIA Fast Food Restaurants EAGLE FALLS STEAK RA H Eating Places, Restaurants ANAT DAINFERN Eating Places, Restaurants Uber BV Taxicabs/Limousines DEBONAIRS DOWERGLEN Fast Food Restaurants MCDONALDS FOURWAYS Eating Places, Restaurants MCD Fourways (75) Eating Places, Restaurants FESIVAL MALL ICE ARENAA Bands, Orchestras, Entertainers PICK N PAY DAINFERN SQU Grocery Stores, Supermarkets TOYS R US FOURWAYS Hobby, Toy, and Game Shops PNA DAINFERN Stationery, Office Supplies, Printing and Writing Paper BP MONTE Service Stations (with or without ancillary services) THE GROVE ICE RINK Bands, Orchestras, Entertainers Halfway Fourways Automobile and Truck Dealers (New and Used) Sales, Servi FESIVAL MALL ICE ARENAA Bands, Orchestras, Entertainers MCD Fourways (75) Eating Places, Restaurants Uber BV Taxicabs/Limousines Estoril Books & Station Book Stores WIMPY FOURWAYS Fast Food Restaurants MR PRICE SPORT- FOURWAY Family Clothing Stores Uber BV Taxicabs/Limousines FLOYD'S 99 BARBERSHOP Barber and Beauty Shops THE GREEN HOUSE 31896 Caterers STER KINEKOR - FOURWAYS Motion Picture Theaters Cltx Kingfisher S S TIO Miscellaneous Food Stores -Specialty, Markets, Convenien Uber BV Taxicabs/Limousines AMCI PIZZERIA Fast Food Restaurants PNA DAINFERN Stationery, Office Supplies, Printing and Writing Paper PICK N PAY DAINFERN SQU Grocery Stores, Supermarkets ANAT DAINFERN Eating Places, Restaurants THE GREEN HOUSE 31896 Caterers DIS-CHEM THE MALL FOUR Drug Stores and Pharmacies DIS-CHEM DAINFERN - CKE P Opticians Uber BV Taxicabs/Limousines MCD Fourways (75) Eating Places, Restaurants MCD Stonebridge (289) E Eating Places, Restaurants PNP EDEAN MEADOW Grocery Stores, Supermarkets JK CRAZY STORE - MODDER Gift, Card, Novelty and Souvenir Shops WOOLWORTHS MAROUN SQUAR Grocery Stores, Supermarkets FRUIT AND VEG CITY FOUR Miscellaneous and Specialty Retail Stores MR PRICE SPORT- FOURWAY Family Clothing Stores SPORTSMANS WAREHOUSE FO Sporting Goods Stores BP MONTE Service Stations (with or without ancillary services) BP MONTE Service Stations (with or without ancillary services) DIS-CHEM CRESTA EMV 44 Drug Stores and Pharmacies Clicks Dairnfern Drug Stores and Pharmacies MCD Fourways (75) Eating Places, Restaurants THE GREEN HOUSE 31896 Caterers MUGG AND BEAN NORTH GATE Eating Places, Restaurants BP MONTE Service Stations (with or without ancillary services) HOME ETC - FOURWAYS Miscellaneous Home Furnishing Specialty Stores WIMPY FOREST HILL Eating Places, Restaurants AMCI PIZZERIA Fast Food Restaurants BP EDENVALE NORTH S/ST Service Stations (with or without ancillary services) HOME ETC - FOURWAYS Miscellaneous Home Furnishing Specialty Stores PICK N PAY DAINFERN SQU Grocery Stores, Supermarkets

Figure 1: Example of the spending terms within the second textual structure from customer document 3998

BEHAVIOURAL MODEL

There are two main objectives to this paper. The first is to obtain the spend profile linked to each customer and the second is to obtain the customer's unique segment identifying other customers of similar spend. Although these two objectives may seem the same, they are in fact obtained in different manners and utilised for entirely different reasons. The spend profiles are obtained by analysing the text and identifying similar terms, the similar terms are clustered to produce a predetermined number of groups, these groups are the spending behaviours of the customers. Customers are not restricted to a single spending behaviour, they can possess multiple spending behaviours dependent on the terms contain in their transactional text. The particular combination of these spending behaviours makes up each customers spending profile. In terms of text mining these groups of spending behaviour can be regarded as the latent or hidden topics within the data, where each customer has a behavioural profile consisting of multiple latent behavioural topics. See Figure 2: The layout of a customer's spend behaviours and spending profile to view the layout of a customer's spending behaviours and spending profile.

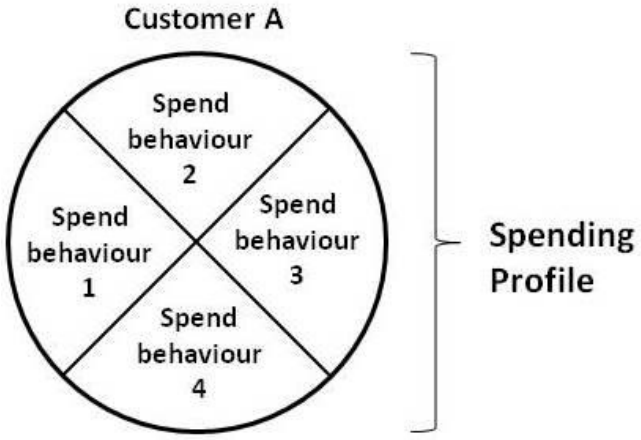

Figure 2: The layout of a customer's spend behaviours and spending profile

The second objective is obtained by clustering the customers based on the terms in their transactional text, this can be seen as a manner of grouping customers based on their spending behaviour however the clustering does not occur specifically based on the spend behaviours but on the text, as in the first objective. This produces a select number of groups containing customers of similar spending behaviours, these groups are the spending segments of the customers. Each customer can only belong to one spending segment. See Figure 3: The layout of a spending segment for the layout of a spending segment.

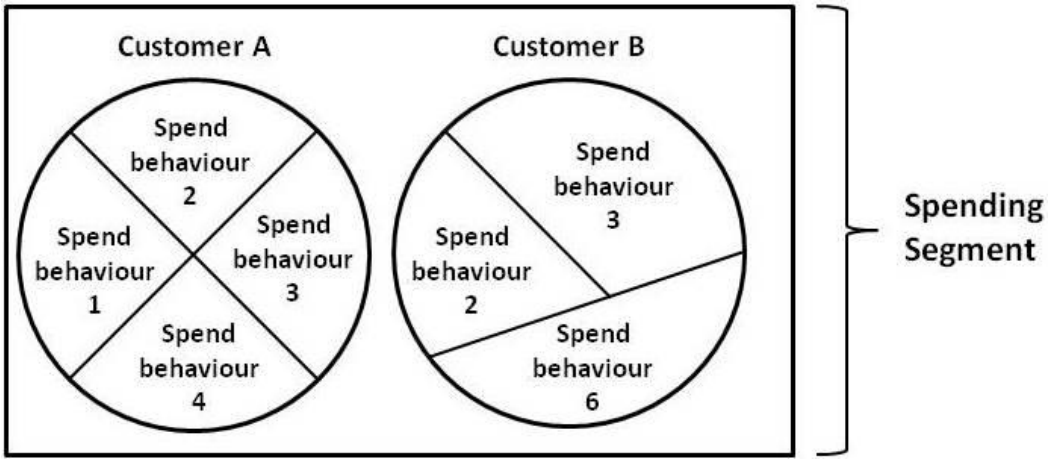

Figure 3: The layout of a spending segment

SAS® text miner was used to analyse this transactional text in an unorthodox manner with the aim of obtaining results to achieve the above objectives. SAS® text miner uses nodes to analyse textual data based on a predetermined set of properties, see [1], [2] for additional information on these properties. SAS® text miner has been used on data containing natural language in sentiment analysis and topic identification. However the transactional text in this paper was gathered as an unstructured unconventional collection of spend terms with no natural language structure. For this reason certain natural language properties were not selected, capitalising on the capabilities of SAS® text miner. For example, the parts of speech property in the text parsing node was not selected due to the fact that terms in the transactional text are a set of spending descriptions. Consider the spending term “shop”, this term would not be used as the verb, i.e. to shop, it would always refer to the noun, i.e. a shop, in the transactional text. One of the properties considered in SAS® text miner was the term weighting property in the text filter node, the weights taken into account were inverse document frequency (IDF), Entropy and no weight. Both IDF and entropy assign a lower weight to terms that appear in more customer documents, the two weightings differ in the sense that IDF considers the number of times the term has appeared in all documents, whereas entropy looks at the number of times the term has appeared in each document. The 2 primary nodes used to obtain the results for the above objectives were the text topic and text cluster nodes. The text topic node extracts topics from the transactional text, identifying the latent spend behaviour in the data, due to the existence of multiple behavioural spend topics per customer document the combination of these spend behaviours make up a customer behavioural spend profile. The text cluster node segments customers of similar spend to a unique segment based on the transactional text. For a more detailed discussion of the selected properties see the appendix.

RESULTS

The set of 3 textual input structures were run through the text mining process using the 3 different weighting properties described in subsection BEHAVIOURAL MODEL, this yielded 9 results. The results however displayed very similar findings for the second and third textual input structures, as well as for the

IDF and entropy weightings. For this reason the third textual structure and the entropy weighting were discarded. The 4 remaining sets of results were observed, for the ease of discussion the first textual input structure using IDF will be called IDF1, the first textual input structure using no weight will be called Noweight1, the second textual input structure using IDF will be called IDF2 and finally the second textual input structure using no weight will be called Noweight2. The use of the IDF weighting was important as it displayed intriguingly different results to when no weight was used.

After evaluating the 4 results it was apparent that each contributed to a different area of behavioural spend. The results using IDF1 suggest a topographical behaviour due to the fact that more than 80% of the spend behaviours were related to an area. Noweight1 suggested a behaviour more aligned to specific merchants given that more than 80% of the spend behaviours were related to particular merchants. The results using IDF2 displayed broader spend behaviours, however it additionally contained an overlap with the topographical behaviour using IDF1, this might be due to the differences between the first and the second textual input structures, or to the fact that IDF gave larger weightings to the terms that appeared less frequently. Similarly, the results using Noweight2 displayed broader spend behaviours with an additional overlap of merchant specific behaviours as seen in Noweight1. It is important to note that a customer may be assigned a specific spending behaviour where the spending terms within that behaviour are not necessarily present in the customers document, this is due to the fact that the spend behaviours have been determined through a methodology that clusters a selection of similar terms.

To illustrate the results, customer document 3998 was again selected. See Figure 1: Example of the spending terms within the second textual structure from customer document 3998. This customer document contained a large amount of text, confirming that manually sifting through the spending terms of 6000 customers would be an overwhelming and tedious task.

The 4 results displaying the 4 spend profiles of customer document 3998 can be viewed in Figure 4: IDF1 - First textual input structure with IDF for customer document 3998 Figure 5: Noweight1 - First textual input structure with no weight for customer document 3998 Figure 6: IDF2 - Second textual input structure with IDF for customer document 3998 and Figure 7: Noweight2 - Second textual input structure with no weight for customer document 3998, respectively. The spend behaviours that are present for customer document 3998 are highlighted in yellow and contain a 1 in the “spend behaviour present indicator” column. The 4 results displaying the 4 spending segments of customer document 3998 can be viewed in Figure 8: Spending segments of customer document 3998.

The below subsections are structure as follows, from *Spend profile for IDF1* to *Spend profile for Noweight2* discuss the results of each of the 4 spending profiles using IDF1, Noweight1, IDF2 and Noweight2. Customer document 3998's combined spending profile is discussed in *Overall combined spend profile for customer document 3998* and finally the results of each of the 4 spending segments using IDF1, Noweight1, IDF2 and Noweight2 is discussed in *Spending segments*.

Spend profile for IDF1

It was previously noted the results using IDF1 identified the presence of topographical spending behaviour, this is confirmed when considering the results in Figure 4: IDF1 - First textual input structure with IDF for customer document 3998. Spend behaviour 1 contained the spending terms Zambesi, Montana, Wonderboom and Sinoville, these are all suburbs of close proximity in the north of Pretoria, in the province of Gauteng. Although these terms are not directly related to the spend of the customers, they are terms relating to the area the spend occurred in and are therefore still considered spending terms. Spend behaviour 2 identified the city of Nelspruit as the most important spending term as it was ranked first in the list of terms. After further investigation, the terms I'langa, Riverside and Westend were all recognised as malls in Nelspruit. Spend behaviour 3 revealed the spending terms Phakalane, Sebele and Botswana, these terms are related to spend in Botswana. The remaining spend behaviours reveal similar topographical trends. Customer document 3998 was assigned 4 topographical spend behaviours for IDF1, namely spend behaviours 4, 8, 15 and 25, this can be viewed in Figure 4: IDF1 - First textual input structure with IDF for customer document 3998. Spend behaviour 4 contained the terms Fourways, Douglasdale, Olivedale and Northgate, these are suburbs of close proximity in the North Western side of the city of Johannesburg. The spending terms displayed in spend behaviour 25 don't show a well-defined area, the only area mentioned is Glen Marais. However when considering the entire list of spending terms

in spend behaviour 25This can be viewed in the topic view option under the text topic node.2, the following additional terms are revealed, Kempton, Greenstone, Glenacres, Norkem and Bonaero. These terms all represent areas in the North Easterly side of Johannesburg. Although 80% of the spend terms in these spending behaviours are related to topography, the other 20% suggest a greater relation to merchants. Spend behaviour 8 and 15 contain these specific merchant spending terms more than the topographical spending terms. One of the spending terms within spend behaviour 8 is the suburb of Rosebank in Johannesburg, however the other spending terms contain a variety of merchants from uber to iTunes to cafes. Spend behaviour 15 again suggests merchants, however these merchants appear to be mall specific, such as the clothing stores Mr. Price and Edgars, the pharmacies Dischem and Clicks, and other retail stores such as Woolworths, Checkers and Pick n Pay. Although the topographical spend behaviour is present in the majority, more than 80%, of the spend behaviours in IDF1, spend behaviours 8 and 15 verify it is not the only behaviour considered. The combination of these 4 spend behaviours makes up the spending profile of customer document 3998.

Spend behaviour	Customer document	Spending terms within the spend behaviour	Spend behaviour present indicator
1	3998	+zambesi,+montana,tollgate,+wonderboom,+sinoville	0
2	3998	+nelspruit,+ilanga,+riverside,+westend,+cross	0
3	3998	phakalane,+airport,+sebele,+junction,+botswana	0
4	3998	+fourways,+douglasdale,+olivedale,+northgate,+jhb	1
5	3998	canal,+walk,+waterfront,century,+cavendish	0
6	3998	+centurion,+jean,+avenue,+raslouw,+red	0
7	3998	+menlyn,+woodland,+glen,+faerie,+atterbury	0
8	3998	bv,uber,+bill,itunes,+rosebank	1
9	3998	+westville,+pavilion,+gateway,+lucia,+la	0
10	3998	+beacon,+bay,+vincent,+hemingways,+gonubie	0
11	3998	dalpark,+east,gosforth,+boksburg,+carnival	0
12	3998	+plaza,+toll,+mainline,+tongaat,mvoti	0
13	3998	+rustenburg,+waterfall,+platinum,+brit,+mainline	0
14	3998	+clearwater,+cresta,+florida,+roodepoort,+jhb	0
15	3998	+woolworths,+dis-chem,+mall,+Mr Price,+clicks	1
16	3998	google,+king,supercell,+mobile,+burger	0
17	3998	walmer,+park,+pe,+greenacres,+newton	0
18	3998	+polokwane,kranskop,cycad,+savannah,+mall	0
19	3998	+tyger,tygervalley,+gate,cape,durbanville	0
20	3998	+secunda,+family,+star,+sasol,+sec	0
21	3998	+sandton,+woodmead,+midrand,fam,+station	0
22	3998	+middelburg,+trac,+witbank,+hill,+diamond	0
23	3998	+bloemfontein,+shoprite,+loch,+logan,+n1	0
24	3998	+somerset,+waterstone,+west,+bay,+strand	0
25	3998	+fran,+glen,+family,+marais,+Pick n Pay	1

Figure 4: IDF1 - First textual input structure with IDF for customer document 3998

Spend profile for Noweight1

The 25 spend behaviours identified in Noweight1 contained primarily spending terms related to merchants. Terms such as fourways, mall, clicks, family, bv, uber and bill are represented in the spend

[2] This can be viewed in the topic view option under the text topic node

behaviours in both IDF1 and Noweight1. The spending profile of customer document 3998 using Noweight1 can be viewed in Figure 5: Noweight1 - First textual input structure with no weight for customer document 3998. Behaviours 2, 8 and 18 were identified as the spend behaviours present for this customer document and therefore make up their spending profile. Consider spend behaviour 2, within the top 10 spending terms identified, 4 petrol stations and 3 Johannesburg suburbs were identified, for this reason it is possible to suggest that this spend behaviour could represent a behaviour related to travel or more specifically travel around the city of Johannesburg. It could additionally signify the customers need to refuel at certain petrol stations in the area. Spend behaviour 8 identified a variety of terms that might indicate retail shopping and/or motor vehicle maintenance. Spend behaviour 18's top 10 spending terms contained Uber, Bv, Station, Sandton, Rosebank, Park, Gautrain and could be indicative of travel using public transport as Sandton, Rosebank and Park are Gautrain stations and uber bv refers to the transport company.

Spend behaviour	Customer document	Spending terms within the spend behaviour	Spend behaviour present indicator
1	3998	+spar,+tops,kfc,+woolworths,+fran	0
2	3998	+bp,+city,+fourways,+sasol,+jhb	1
3	3998	+Pick n Pay,+family,+fran,fam,+clicks	0
4	3998	+bp,+checkers,+shoprite,+service,+food	0
5	3998	+liquor,+store,+tops,+city,+family	0
6	3998	+plaza,+toll,+mainline,+trac,gosforth	0
7	3998	+total,+checkers,+caltex,+family,absa	0
8	3998	+mall,+bp,+motor,+family,+clicks	1
9	3998	cape,+walk,canal,+centre,+city	0
10	3998	+airport,+junction,phakalane,+pty,+sebele	0
11	3998	+park,+family,walmer,+checkers,+spar	0
12	3998	+mcdonalds,kfc,+king,+nandos,+total	0
13	3998	+motor,+shoprite,+caltex,kfc,+checkers	0
14	3998	+checkers,+hyper,+city,+clicks,+mall	0
15	3998	+builder,+warehouse,+centre,+mall,+city	0
16	3998	+centurion,+jean,+avenue,+red,sbsa	0
17	3998	+fran,+park,+mall,+liquor,+east	0
18	3998	uber,bv,+cafe,+sandton,+bill	1
19	3998	+sasol,absa,+engen,kfc,+clicks	0
20	3998	+market,+food,+lover,+city,+hyper	0
21	3998	+Mr Price,+bay,+clicks,+dis-chem,+home	0
22	3998	+hyper,+glen,+menlyn,+bp,+engen	0
23	3998	+woolworths,+dis-chem,+mall,+jhb,emv	0
24	3998	+station,+service,+city,+family,+motor	0
25	3998	+shoprite,+mall,+city,kfc,+game	0

Figure 5: Noweight1 - First textual input structure with no weight for customer document 3998

Spend profile for IDF2

IDF2 identified a broader picture of how a customer spends, Figure 6: IDF2 - Second textual input structure with IDF for customer document 3998 displays the spend behaviours identified in IDF2. As previously mentioned, these results contain a combination of broader merchant categories and areas. IDF2 identified 6 spend behaviours making up the spending profile of customer document 3998. Spend behaviours 12, 20 and 25 represent the topographical spending behaviour. Spend behaviour 12 containing Fourways, Douglasdale, Olivedale, Sandton, depicts an almost identical topographical spending behaviour to that identified in spend behaviour 4 in IDF1. It is interesting to note that spend

behaviours 11 and 14 in IDF1 are also very similar to spend behaviours 20 and 25 in these results, however in the former these spend behaviours were not included for customer document 3998 and in the latter results they were. This could indicate that there were certain spending terms in the former behaviours that weren't present in spend behaviours 20 and 25 in IDF2 and visa versa. Spend behaviour 4 again identified the spending terms related to public transport, for example uber, taxicab, etc., and spend behaviour 15 contained spending terms that were related to shopping and the pampering of a person or a home, for example in addition to the spending terms displayed in Figure 6: IDF2 - Second textual input structure with IDF for customer document 3998, the terms gift card, novelty, souvenir, furnishing, beauty, were also present but of a lower rank. Lastly, spend behaviour 19 indicated a direct relation to entertainment, the spending terms cinema, theatre, sports clubs, country clubs, music, etc. were all present within this spend behaviour.

Spend behaviour	Customer document	Spending terms within the spend behaviour	Spend behaviour present indicator
1	3998	+shoprite,+clothe,+family,+mall,+woman	0
2	3998	+toll,+fee,+bridge,+plaza,+mainline	0
3	3998	phakalane,+junction,+sebele,+airport,+botswana	0
4	3998	bv,taxicabs,+limousine,uber,+bill	1
5	3998	commuter,+transportation,includ,gautrain,suburban	0
6	3998	+bet,gaming,+lottery,+casino,+chip	0
7	3998	alcoholic,taverns,+cock,+drink,+beverage	0
8	3998	+hotel,+lodge,+resort,+motel,+reservation	0
9	3998	canal,+walk,cape,+tyger,+n1	0
10	3998	google,business,+classify,elsewhere,+king	0
11	3998	automated,+disbursement,financial,+institution,+cash	0
12	3998	+fourways,+jhb,+sandton,+olivedale,+douglasdale	1
13	3998	+build,+hardware,+material,lumber,+builder	0
14	3998	+glen,+menlyn,+woodland,+faerie,+atterbury	0
15	3998	+home,+shop,+clothe,+woolworths,+gift	1
16	3998	+bay,+beacon,+vincent,+hemingways,walmer	0
17	3998	+westville,+gateway,+pavilion,+lucia,+umhlanga	0
18	3998	+centurion,+jean,+avenue,+raslouw,+red	0
19	3998	+motion,+picture,+theater,ster,+kinekor	1
20	3998	+fran,+east,+boksburg,gosforth,dalpark	1
21	3998	+nelspruit,+ilanga,+trac,+riverside,+middelburg	0
22	3998	+meat,+freezer,+locker,provisioners,+butchery	0
23	3998	+liquor,+package,+wine,+beer,+tops	0
24	3998	+zambesi,+montana,tollgate,+wonderboom,+sinoville	0
25	3998	+clearwater,+roodepoort,+cresta,+florida,+west	1

Figure 6: IDF2 - Second textual input structure with IDF for customer document 3998

Spend profile for Noweight2

The spend behaviours in Noweight2 can be viewed in Figure 7: Noweight2 - Second textual input structure with no weight for customer document 3998. These spend behaviours be categorised more easily than the spend behaviours in IDF1, Noweight1 and IDF2. It is immediately noticeable that spend behaviour 1 is related to supermarkets, spend behaviour 2 to road travel and maintenance, spend behaviour 3 to cash withdrawals and other financial institutions, etc. Spend behaviour 5 and 20 both revealed food related spending behaviours however 5 suggested a relation to restaurants whereas spend behaviour 20 suggested a relation to fast food. Although different results were produced in IDF2 and Noweight2 due to the differently weighted spend terms, spend behaviour 18 in Noweight2 strongly resembles 15 in IDF2, this

behaviour again hints at a pampering behaviour. The fact that this customer is willing to pamper themselves, their home or someone else due to the “gift” and “card” spending terms is indicative of a more lavished spending lifestyle. Spend behaviour 21 again identified the travel spending behaviour, however it strongly suggests the public transport behaviour over the likelihood to refuel their own car due to the absence of spend behaviour 6. It is also interesting to note the spend term “eat” in behaviour 21, this might suggest a type of public transport specifically related to the restaurant and food industry, for example it could imply uberEATS or other meal delivery services. Lastly, spend behaviour 25 contained spend terms such as Fruit and Veg indicating a healthier grocery store, and the spend term home which again might indicate home papering or spend on furniture.

Spend behaviour	Customer document	Spending terms within the spend behaviour	Spend behaviour present indicator
1	3998	+supermarket,+grocery,+store,+Pick n Pay,+spar	0
2	3998	+toll,+bridge,+fee,+plaza,+mainline	0
3	3998	+disbursement,financial,automated,+institution,+cash	0
4	3998	+liquor,+wine,+package,+beer,+tops	0
5	3998	+eat,+place,+restaurant,+mall,+spur	1
6	3998	+service,+station,ancillary,+motor,+bp	0
7	3998	+convenienc,+market,specialty,miscellaneous,+food	0
8	3998	+woolworths,+park,+dis-chem,+beauty,+jhb	0
9	3998	+clothe,department,+woman,+man,+mall	0
10	3998	+hotel,+lodge,+resort,+central,+reservation	0
11	3998	+meat,+freezer,+locker,provisioners,+butchery	0
12	3998	+bar,alcoholic,taverns,+cock,+drink	0
13	3998	+hardware,+supply,+build,+builder,+warehouse	0
14	3998	+spar,+tops,+fast,+grocery,+supermarket	0
15	3998	elsewhere,+classify,business,google,+service	0
16	3998	+pharmacy,drug,+clicks,+dis-chem,emv	0
17	3998	+family,+clothe,+woolworths,phakalane,+airport	0
18	3998	+shop,+supply,+gift,+beauty,+card	1
19	3998	+include,+ticket,+casino,+chip,+bet	0
20	3998	+fast,+food,+restaurant,kfc,+mcdonalds	0
21	3998	taxicabs,+limousine,uber,bv,+eat	1
22	3998	+checkers,+hyper,+park,+grocery,+supermarket	0
23	3998	+sale,+automobile,+dealer,+truck,+motor	0
24	3998	+park,+mall,+sasol,+spar,+centurion	0
25	3998	+retail,+city,+home,+fruit,+veg	1

Figure 7: Noweight2 - Second textual input structure with no weight for customer document 3998

Overall combined spend profile for customer document 3998

Customer document 3998 has multiple spend profiles depending on whether IDF1, Noweight1, IDF2 or Noweight2 is considered. When observing all 4 spend profiles together, it becomes easier to understand how this customer spends. The 4 spend profiles indicate that the majority of this customer spends occurs in the north western and eastern sides of the city of Johannesburg. Their profiles additionally indicate the frequency of travel, either through personal or public transport, however their spend profiles strongly suggest it is a combination of both. Due to certain spend behaviours in IDF2 and Noweight2, this customer displays the lifestyle of someone who is considerably well-off. These results imply a tendency to pamper oneself or ones home, frequently go to restaurants as opposed to doing grocery shopping and participating in forms of entertainment on a regular basis. The combinations of restaurants and public transport might be indicative of a lifestyle associated to nightlife, however the absence of spend

behaviours 4 and 12 in Noweight2, suggests this customer does not use public transport for liquor and bar purposed but prefers to travel by public transport for other reasons.

There are numerous combinations of spending behaviours that make up a customers spending profile. Another example could be the combination of spend behaviour 3 in IDF1, which contains areas related to Botswana, and the public transport behaviour, this could indicate using the public transport to get to the airport. A spending profile consisting of spend behaviour 1 in Noweight2, the supermarket behaviour, and other grocery shopping spend behaviours across IDF1, Noweight1 and IDF2 might indicate a customer whose lifestyle is centred around spending on necessities over luxury items.

Spending segments

Figure 8: Spending segments of customer document 3998 displays the spending terms and associated spending segment that customer document 3998 has been assigned to using IDF1, Noweight1, IDF2 and Noweight2. When IDF1 was considered, customer document 3998 was assigned to spending segment 10, this segment contains all customers that behave similar to customer document 3998 in their spend, in other words these customers all contain similar spend terms. Spending segments 9, 2 and 1 were assigned to customer document 3998 for Noweight1, IDF2 and Noweight2, respectively. These segments again contain other customers that behave in a similar manner to customer document 3998 with regards to there spend. These segments can be used for market segmentation and lead optimisation in campaigning or as a decision making tool to drive many significant business choices.

These results are all extremely useful in identifying a customers lifestyle and spending patterns.

IDF1		
Customer document	Spending segment	Spending terms within the spending segment
3998	10	+bryanston +douglasdale +fourways +jhb +nicol +northgate +sandton +cross +cr +rosebank hyp +square +road +city +bp
Noweight1		
Customer document	Spending segment	Spending terms within the spending segment
3998	9	+checkers +fran +hyper +park +veg +fruit +family +'Pick n Pay' +dis-chem +liquor emv +store +mcdonalds +market +express
IDF2		
Customer document	Spending segment	Spending terms within the spending segment
3998	2	+limousine bv taxicabs uber +bar alcoholic taverns +classify +beverage +cock +drink elsewhere cape business +beauty
Noweight2		
Customer document	Spending segment	Spending terms within the spending segment
3998	1	+convenienc +fast +market +woolworths kfc drug +checkers +pharmacy ancillary +station +eat +bp specialty +'Pick n Pay' +place

Figure 8: Spending segments of customer document 3998

CONCLUSION

SAS® text miner is a useful tool for determining customer spending segments and latent spending profiles. The suggested approach using SAS® text miner to determine the spending behaviour of customers is a unique and innovative method that produces satisfactory results. These results identified a customer living in Johannesburg with a high spending tendency towards travel, restaurants and entertainment. This customers profile is indicative of a luxurious spending lifestyle. The spending segment this customer was assigned to can be used to identify customers containing similar spending terms and would therefore have similar spending behaviours.

This behavioural profiling and customer segmentation is beneficial for businesses in such a way that when utilised correctly it can be used for market segmentation to cluster customers into target marketing segments. The results can be used to compile the correct product offering to the correct customers, this would increase product take up and in turn boost the revenue of the business. Segmenting customers into clusters of similar spend can assist in identifying possible fraudulent cases, identifying customers

who spend often and those that are high risk. Understanding a customers spend profile can assist in classifying a customer as a luxury or necessity spender. It can additionally assist in identifying the current behaviour of customers with the goal of driving that behaviour to align with the business strategy, which is often the case in loyalty programs.

REFERENCES

[1] Goutam Chakraborty, Murali Pagolu, and Satish Garla. *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS*. SAS Institute, 2014.
[2] SAS Institute Inc. *Getting started with sas text miner 12.1*. SAS Institute Inc., Cary, NC, 2012.
[3] Rajneesh Krishna. *Consumer behaviour*. Oxford University Press, 2014.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Amelia Van Schalkwyk
+27 82 495 8250
amelia.vanschalkwyk@gmail.com

APPENDIX

The properties in SAS text miner were selected based on the following explanations. In the text parsing node, the parts of speech and the noun group properties were both not selected. This was done due to the fact that the data was not structure as full sentences but rather as unstructured, concatenated descriptions. Additionally certain parts of speech, punctuation and numbers were ignored, and stemming was set to yes. The synonyms list in the text parsing node was updated after using the filter viewer in the text filter node, there were many manually selected synonyms as the data contained multiple retailers of the same name, for example, the retailer pick n pay was written as PNP, pick n` pay, pick and pay and many other alterations. In the text filter node, spell check was set to yes and the minimum number of documents was left to the default. The term weight in the text filter node was altered to find the optimal results, since without a weighting each term was treated as equally important. The weights considered and tested were IDF, Entropy and no weight. These results were discussed in the BEHAVIOURAL MODEL section. The text topic node groups the terms into “topics” such that every customer document would have multiple spend behaviours. The following default options in the text topic node were selected, single term topics 0, multi term topics 25 and correlated terms set to no. The text cluster node on the other hand grouped the customer documents into similar groups, where one document could only belong to one cluster. In the text cluster node the SVD resolutions were selected to be low due to the fact that the data contained many terms, by selecting low there was a high loss of information, however the dimensions were greatly reduced. Expectation-Maximisation was selected as the cluster algorithm and all other properties were set to the defaults.