

# SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

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

Amelia Van Schalkwyk, Frans Kanfer, Sollie Millard

USERS PROGRAM





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

Amelia Van Schalkwyk, Frans Kanfer, Sollie Millard

University of Pretoria

## ABSTRACT

Understanding customer behaviour profiles are of great value to companies. Customer behaviour is influenced by a multitude of factors, 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 behaviour 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 spending segments of customers and latent spending profile using a clustering approach. Businesses benefit in the way the behavioural spend model is utilised. It can be used for market segmentation, where each cluster is seen as a target marketing segment, leads optimisation, 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 behaviour of customers makes it ideal for companies driving retention and loyalty in their customers.

## METHODS

### Data preparation

- The data obtained was from the transactional cheque card spend of 6000 unique customers for a given month. These customers had an income ranging from R700k to R1,000k.
- Three text columns were identified as significant in the data.
- The three text columns and all transactions for the month were concatenated to obtain a single column of text per customer.

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

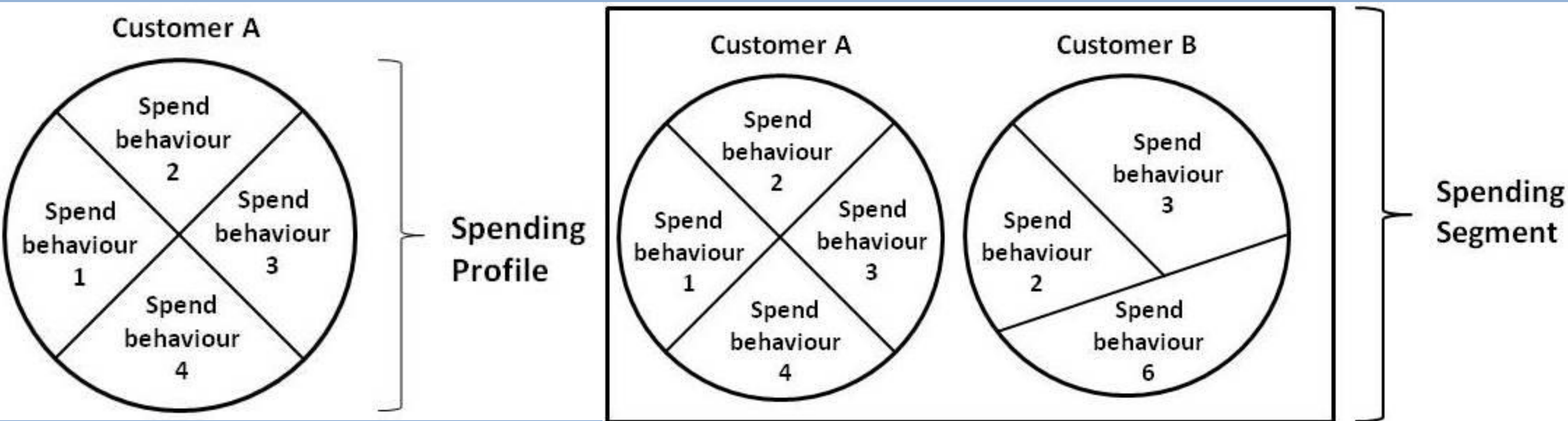
## METHODS

Customer document	Trans_Text
1	Description1 Category1 Type1 Description1 Category1 Type1 Description1 Category1 Type1

- 3 textual input structures were determined:
  - The first textual input structure contained only Trans\_Desc
  - The second textual input structure contained Trans\_Desc and Trans\_Cat
  - The third textual input structure contained Trans\_Desc, Trans\_Cat and Trans\_Type
- The third textual input structure was discarded as the results indicated redundancy.
- In terms of text mining and for the purpose of SAS® text miner, a single customers concatenated set of transactions was viewed as a document and the words within that document viewed as the terms. In other words a collection of spending terms make up the customer's spending document.

### The behavioural spend model

- The aim of this model was to obtain a customers spending profile and spending segment using an SVD clustering approach. A spending profile is made up of multiple spend behaviours and a spending segment is the group that each customer is segmented to, containing customers that spend in a similar manner.
- The spending profiles are determined by identifying similar spend terms and bucketing these terms into groups called spend behaviours through the use of SAS® text miners Text Topic node.
- The spending segments are determined by bucketing the customers into groups based on their spending terms, this is done through the use of SAS® text miners Text Cluster node.





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

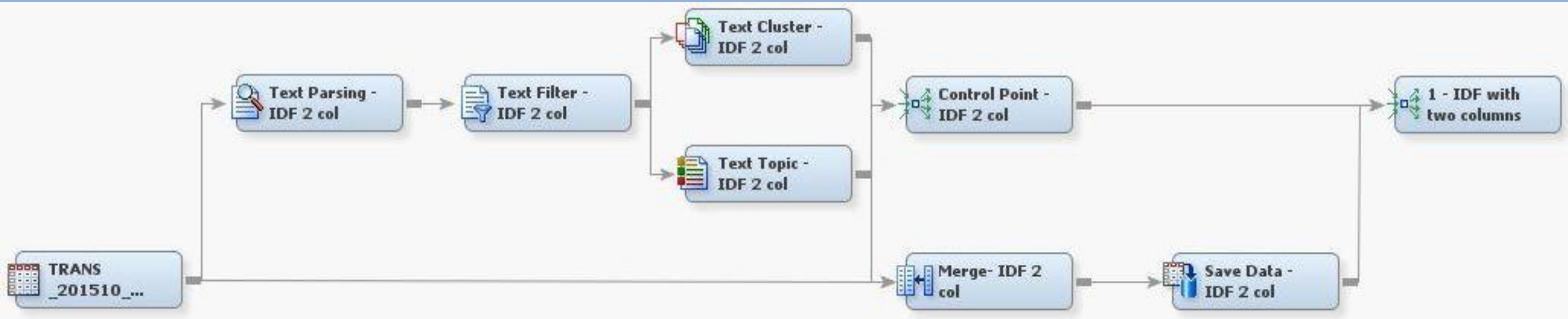
Amelia Van Schalkwyk, Frans Kanfer, Sollie Millard

University of Pretoria

## METHOD (CONTINUED)

### SAS Text Miner

- 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.



- The transactional text for this behavioural spend model 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. Consider the spending term shop, this term would not be used as a verb, i.e. to shop, it would always refer to the noun, i.e. a shop, in this transactional text.
- For the term weighting property in the text filter node, inverse document frequency (IDF), Entropy and no weight were used.
- The entropy weighting was discarded as the results indicated redundancy.
- The 2 primary nodes used to obtain the results were the text topic and text cluster nodes

### Final Structure

The spending profile and spending segment results were determined for the below 4 input sets:

- IDF1 - The first textual input structure using IDF.
- Noweight1 - The first textual input structure using no weight.
- IDF2 - The second textual input structure using IDF.
- Noweight2 - The second textual input structure using no weight.

## RESULTS

To illustrate the results, customer document 3998 was selected.

### Spend profile using IDF1

- 25 spend behaviours were determined using SAS® text miner. Customer document 3998 displayed 4 of the 25 behaviours. These 4 behaviours make up this customers spending profile for IDF1.

- Approximately 80% of the spend behaviour in IDF1 are related to an area, the other 20% suggest a greater relation to merchants.
- Spend behaviour 4 contains spending terms associated with the north western side of the city of Johannesburg.
- Spend behaviour 25 doesn't show a well-defined area, the only area mentioned is Glen Marais. However when considering the additional spend terms in the topic viewer option, the following terms were revealed, Kempton, Greenstone, Glenacres, Norkem and Bonaero. These terms all represent areas in the North Easterly side of Johannesburg.
- Spend behaviours 8 and 15 are more merchant specific, with 15 identifying mall specific merchants, 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.

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, +tonga, +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, tyger valley, +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	+someset, +waterstone, +west, +bay, +strand	0
25	3998	+fran, +glen, +family, +marais, +Pick n Pay	1



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

Amelia Van Schalkwyk, Frans Kanfer, Sollie Millard

University of Pretoria

## RESULTS (CONTINUED)

### Spend profile using Noweight1

- The spend behaviours identified in noweight1 contained primarily merchants. Customer document 3998 displayed 3 of the 25.
- Spend behaviour 2 and 18 suggested transport related behaviours. 2 indicated fuel and gas stations and 18 indicated a relation to public transport since Uber, Bv, Station, Sandton, Rosebank, Park, Gautrain are contained in the top 10 spending terms. 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
2	3998	+bp,+city,+fourways,+sasol,+jhb	1
8	3998	+mall,+bp,+motor,+family,+clicks	1
18	3998	uber,bv,+cafe,+sandton,+bill	1

### Spend profile using IDF2

- The spend behaviours using IDF2 display broader merchant categories as well as areas.
- The spending profile of customer document 3998 displayed 6 of the 25 spending behaviours when IDF2 was used.
- The customers spend profile using IDF2 contains certain spend behaviours displayed in the spend profile using IDF1.
- Spend behaviours 12, 20 and 25 indicate area related spend behaviours.
- Spend behaviour 4 suggests travel related to public transport.
- Spend behaviour 15 suggests the pampering of a person or a home, since additional spending terms such as gift card, novelty, souvenir, furnishing, beauty were also present but of a lower rank.
- Spend behaviour 19 indicated entertainment due to 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
4	3998	bv,taxicabs,+limousine,uber,+bill	1
12	3998	+fourways,+jhb,+sandton,+olivedale,+douglasdale	1
15	3998	+home,+shop,+clothe,+woolworths,+gift	1
19	3998	+motion,+picture,+theater,ster,+kinekor	1
20	3998	+fran,+east,+boksburg,gosforth,dalpark	1
25	3998	+clearwater,+roodepoort,+cresta,+florida,+west	1

## RESULTS (CONTINUED)

### Spend profile using Noweight2

- This spend profile once again indicated pampering and public transport, however it additionally indicated a tendency to spend at restaurants and retail stores like fruit and veg city.
- It is interesting to note the presence of the spending term “eat” in the transport spend behaviour. This might suggest public transport specially related to the restaurant and food industry, for example uberEATS or other meal delivery services.

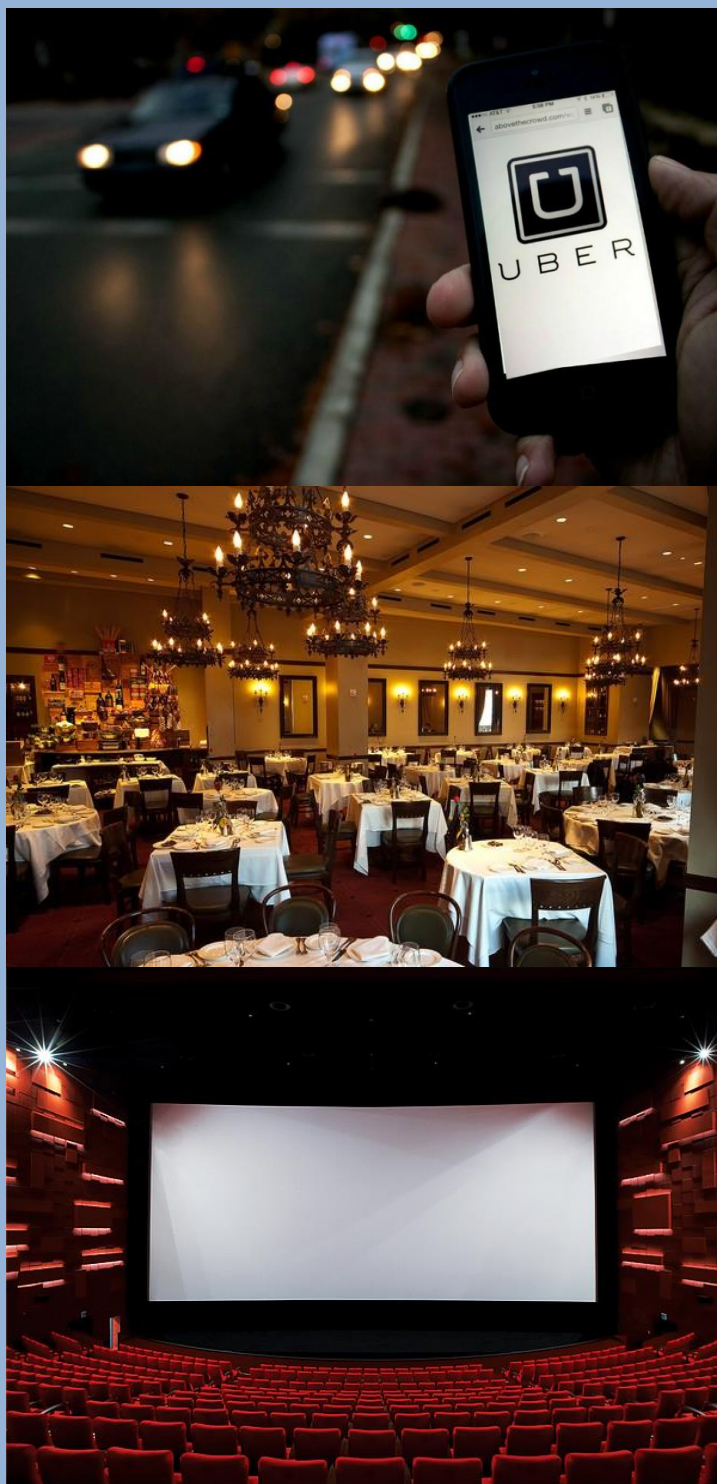
Spend behaviour	Customer document	Spending terms within the spend behaviour	Spend behaviour present indicator
5	3998	+eat,+place,+restaurant,+mall,+spur	1
18	3998	+shop,+supply,+gift,+beauty,+card	1
21	3998	taxicabs,+limousine,uber,bv,+eat	1
25	3998	+retail,+city,+home,+fruit,+veg	1

### Overall spend profile for customer document 3998

- When observing the all 4 spend profiles together, it becomes easier to understand the how this customer spends.
- These profiles indicate that the majority of this customer spends occurs in the north western and eastern sides of the city of Johannesburg. It additionally suggests frequency of travel, either through personal methods due to the fuel and gas stations or public transport, however their spend profiles strongly suggest it is a combination of both.
- This customer displays the lifestyle of someone who is considerably well off, since 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.

### Spend profile uses

- The spend behaviours and spend profiles of customers can assist in identifying the lifestyle of a customer. It can assist in answering questions such as, is this customer a luxury or a necessity spender? What are the best products suited for this customer in their present situation? Does our product appeal to there lifestyle? How can we market our products now that we better understand our customer?





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

Amelia Van Schalkwyk, Frans Kanfer, Sollie Millard

University of Pretoria

## RESULTS (CONTINUED)

### Spending segments of customer document 3998

- Spending segments 10, 9, 2 and 1 were assigned to customer document 3998 when using IDF1, Noweight1, IDF2 and Noweight2, respectively.
- Many of the spending terms identified for the segments were also observed in this customers spending profiles.
- These segments contain groups of customers with similar spending terms. In other words the customers in the same group behave in a similar manner with regards to there spend.

### Spend segment uses

- These spend segments can be used for market segmentation and lead optimisation in campaigning.
- It can additionally be used as a decision making tool to drive significant business choices.

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

## CONCLUSIONS

SAS text miner is a useful tool for determining a customers latent spending profiles and spending segments. The suggested approach of using SAS text miner to determine the spending behaviour of customers is a unique and innovative method that produces satisfactory results.

The results identified a customer living in the city of Johannesburg with a high spending tendency towards travel, restaurants and entertainment. This spending profile is more indicative of a luxurious lifestyle.

The spending segment of this customer can be used to identify other customers with similar spending terms and therefore similar spending behaviours.

This spend profiling and segmentation is beneficial for businesses if utilised correctly, it can be used for market segmentation to cluster customers into target marketing segments. The results can additionally be used to compile a product offering to suit the correct customers, this would increase product take up and in turn boost the revenue of the business. Segmenting customers into group 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.





# SAS<sup>®</sup> GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL