

Data and Text Mining

MCDA5580

Master of Science in Computing and Data Analytics

Assignment-1 Report

Submitted by:

Vivekanand Boopathy	A00425792
Parijat Bandyopadhyay	A00430847
Kothai Kannappan Murugappan	A004727876

Submitted to:

Dr. Pawan Lingras



**SAINT MARY'S
UNIVERSITY SINCE 1802**

One University. One World. Yours.

Table of Contents

- **Executive Summary**
- **Data Summary**
- **Product Analysis**
 1. Data Model for products
 2. Data Cleaning
 3. Data Normalizing
 4. Selecting Number of Clusters
 5. Data De-normalizing and Cluster column addition
 6. Data Analysis
 7. Product Profiling
- **Customer Analysis**
 8. Data Model of the customers
 9. Descriptive Analysis
 10. Data Cleaning
 11. Data Normalizing
 12. Selecting Number of Clusters
 13. Data De-normalizing and Clustering
 14. Data Analysis
 15. Customer Profiling

Executive Summary

Sobeys Inc. is a retail franchise that operates over several parts of Canada .Sobeys wants to analyze the data that it has collected through its Point of sales system for providing better services to customer and to decide the departments it needs to invest in for more profit.As part of this , Sobey's has given the historical retail transactions data recorded in 2015 i.e. from 1st January 2015 to 14 th September 2015to a team of analysts to provide insights for its operation.

Data Summary

The data from the Sobey's has been loaded in the MySQL database name dataset01.The tables that we are using to create our target tables are

- sales219: Containing the transaction data of its customer over several outlets.
- Customer: Containing the data about the customers.
- Items: Contains the information of all the items that Sobeys has in its inventory.

Total No of Products: 32591

Total No of Customers:44469

Total Revenue: \$13645221

Time span in Consideration: 814 days (2015-01-01 to 2005-09-14)

Product Analysis:

1. Data Model for Products:

From the “sales219” table the following columns have been extracted for product analysis.

The Top 2000 products have been chosen based on the total revenue earned by each product.

Column Name	Description
ITEM_SK	UNIQUE PRODUCT SERIAL NUMBER
TOTAL_REVENUE	TOTAL REVENUE GENERATED FROM AN INDIVIDUAL ITEM
BASKETS	NO OF DISTINCT TRANSACTION OF A PARTICULAR PRODUCT
DISTINCT_CUSTOMERS	NO OF DISTINCT CUSTOMERS PURCHASED A PRODUCT
AVERAGE_PRICE	AVERAGE PRICE OF A PRODUCT

PRODUCT_NAME	NAME OF A PRODUCT
PRODUCT_CAT	IN WHICH CATEGORY A PARTICULAR PRODUCT BELONGS

Top 2000 Products data structure and summary:

```
'data.frame': 2000 obs. of 7 variables:
 $ ITEM_SK          : int 11740941 11740923 11680016 11610106 11686823 ...
1143 11685694 11740964 12518517 11696675 ...
 $ TOTAL_REVENUE   : num 126516 78940 72299 59210 55806 ...
 $ BASKETS          : int 87545 26762 11766 6200 16244 8602 4472 9287 4771
4207 ...
 $ DISTINCT_CUSTOMERS: int 16445 7151 290 2272 5800 3515 2505 3424 1252 2320
...
 $ AVERAGE_PRICE    : num 1.4 1.44 5.94 9.49 3.13 ...
 $ PRODUCT_NAME     : Factor w/ 114 levels "", "Alternative", ...
36 75 21 75 1 8 ...
 $ PRODUCT_CAT      : Factor w/ 22 levels "", "Books & Maga", ...
5 19 17 19 1 17 ...
```

TOTAL_REVENUE	BASKETS	DISTINCT_CUSTOMERS
Min. : 1327	Min. : 137.0	Min. : 5.0
1st Qu.: 1662	1st Qu.: 362.0	1st Qu.: 168.0
Median : 2285	Median : 555.5	Median : 249.0
Mean : 3832	Mean : 982.5	Mean : 402.8
3rd Qu.: 3685	3rd Qu.: 904.2	3rd Qu.: 400.0
Max. : 126516	Max. : 87545.0	Max. : 16445.0
AVERAGE_PRICE	PRODUCT_NAME	PRODUCT_CAT
Min. : 0.4206	Milk Produc: 90	Grocery : 470
1st Qu.: 3.0240	Other Fruit: 82	Produce : 367
Median : 4.0942	Refrigerate: 74	Dairy : 319
Mean : 4.5170	Bread Produ: 72	Meat : 215
3rd Qu.: 5.6612	Yogurt : 71	Home Meal Re: 117
Max. : 10.0000	Cooking Veg: 66	Deli : 116
	(Other) : 1545	(Other) : 396

In top 2000 products, there are 114 types of products and 22 types of categories.

The following steps 2 to 5 have been followed for data preparation and clustering:

2.Data Cleaning:

The extracted data has been cleaned to remove any outliers, errors and unwanted data. In that process

ITEM_SK = 11740941, has been found as an outlier and removed from the data set and a cleaned data table has been prepared.

3.Data Normalizing:

Next, the data has been normalized to bring different data to a common scale.

4.Selecting Number of Clusters:

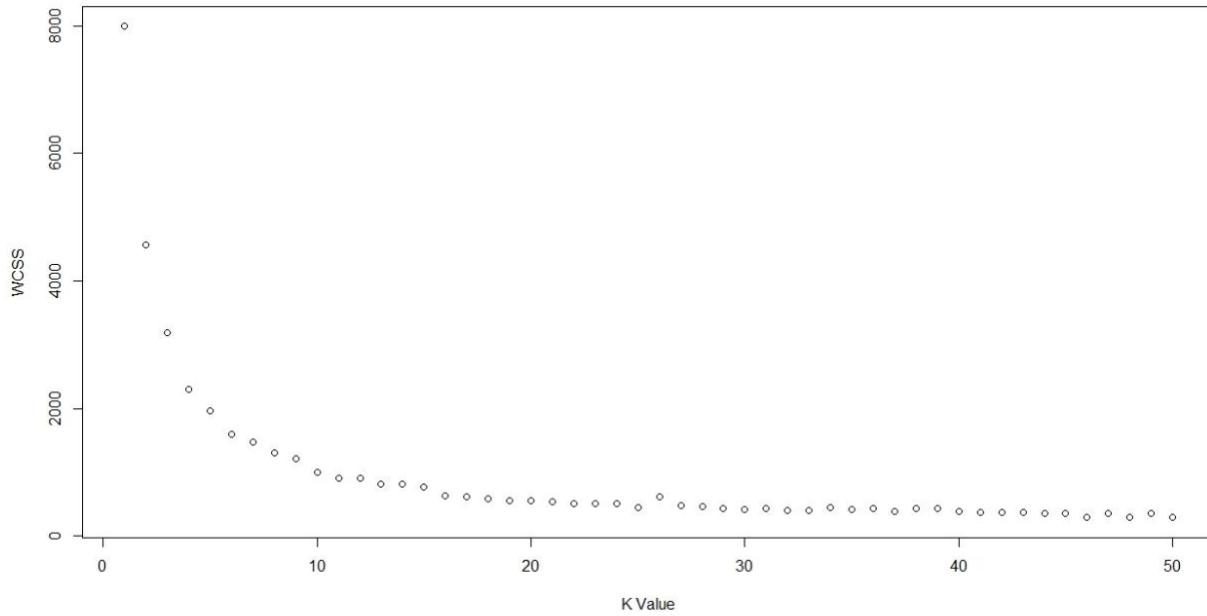
After scaling, the number of clusters are selected based on value of WCSS for a number of clusters.

The following table shows how the WCSS values dropped from cluster no=1 to cluster no=50. K=5 i.e. 5 no of clusters have been considered as form k=6 onwards, the rate of reduction of WCSS gets reduced.

[1]	" k "	-	" WCSS "	-	"
[1]	"1"	" - "	"7992"		
[1]	"2"	" - "	"4566.427"		
[1]	"3"	" - "	"3195.832"		
[1]	"4"	" - "	"2300.133"		
[1]	"5"	" - "	"1972.112"		
[1]	"6"	" - "	"1592.037"		
[1]	"7"	" - "	"1454.052"		
[1]	"8"	" - "	"1270.739"		
[1]	"9"	" - "	"1098.391"		
[1]	"10"	" - "	"1008.128"		
[1]	"11"	" - "	"1082.874"		
[1]	"12"	" - "	"902.177"		
[1]	"13"	" - "	"858.6"		
[1]	"14"	" - "	"829.208"		
[1]	"15"	" - "	"803.62"		
[1]	"16"	" - "	"751.811"		
[1]	"17"	" - "	"599.515"		
[1]	"18"	" - "	"695.96"		
[1]	"19"	" - "	"680.002"		
[1]	"20"	" - "	"564.69"		
[1]	"21"	" - "	"506.197"		
[1]	"22"	" - "	"518.753"		
[1]	"23"	" - "	"636.628"		
[1]	"24"	" - "	"467.919"		
[1]	"25"	" - "	"495.991"		
[1]	"26"	" - "	"447.221"		
[1]	"27"	" - "	"442.306"		
[1]	"28"	" - "	"432.459"		
[1]	"29"	" - "	"442.825"		
[1]	"30"	" - "	"433.77"		
[1]	"31"	" - "	"415.86"		

```
[1] "32"      " -"      "427.579"
[1] "33"      " -"      "412.583"
[1] "34"      " -"      "409.93"
[1] "35"      " -"      "396.378"
[1] "36"      " -"      "391.105"
[1] "37"      " -"      "389.923"
[1] "38"      " -"      "389.781"
[1] "39"      " -"      "384.664"
[1] "40"      " -"      "379.026"
[1] "41"      " -"      "386.735"
[1] "42"      " -"      "399.3"
[1] "43"      " -"      "333.975"
[1] "44"      " -"      "392.981"
[1] "45"      " -"      "362.001"
[1] "46"      " -"      "367.914"
[1] "47"      " -"      "368.326"
[1] "48"      " -"      "364.012"
[1] "49"      " -"      "328.607"
[1] "50"      " -"      "384.927"
```

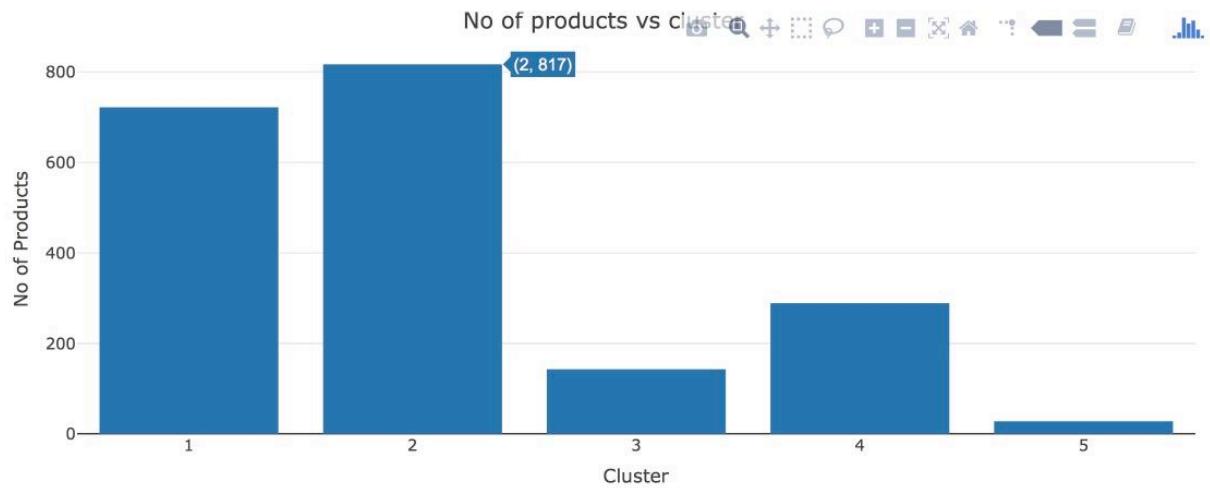
The following figure shows K vs WCSS plot.



5.Data De-normalizing and Cluster column addition:

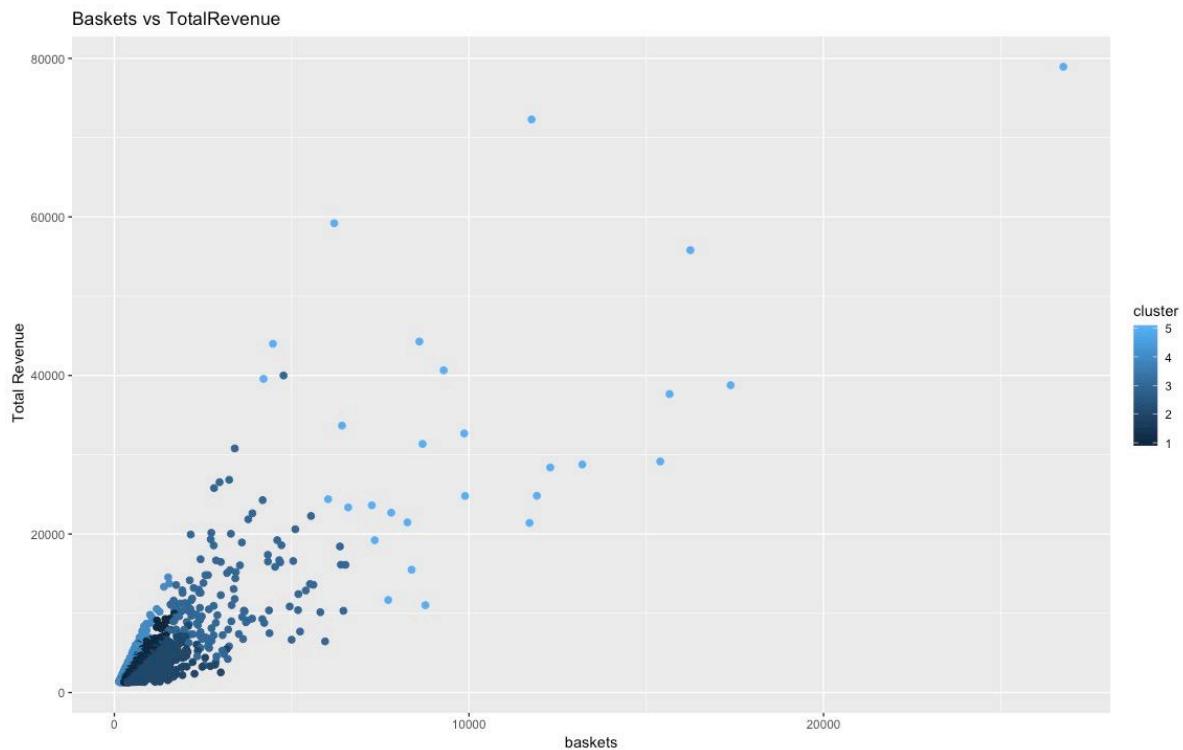
After number of clusters is decided. The data is again denormalized to get actual centroids of the clusters and cluster information is appended to the cleaned product table.

The following table shows the cluster sizes i.e. how many products are there in each cluster:

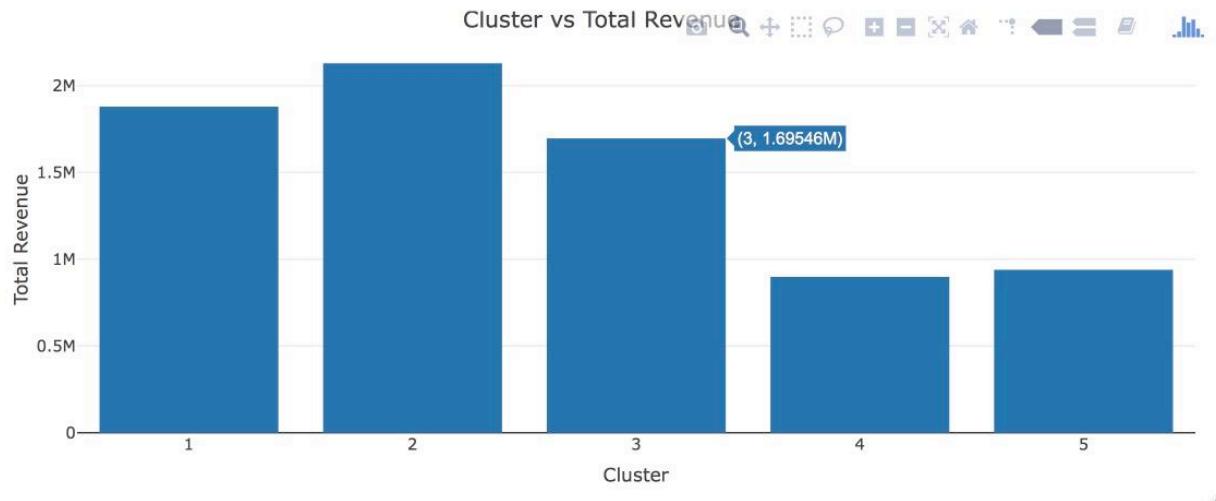


6.Data Analysis:

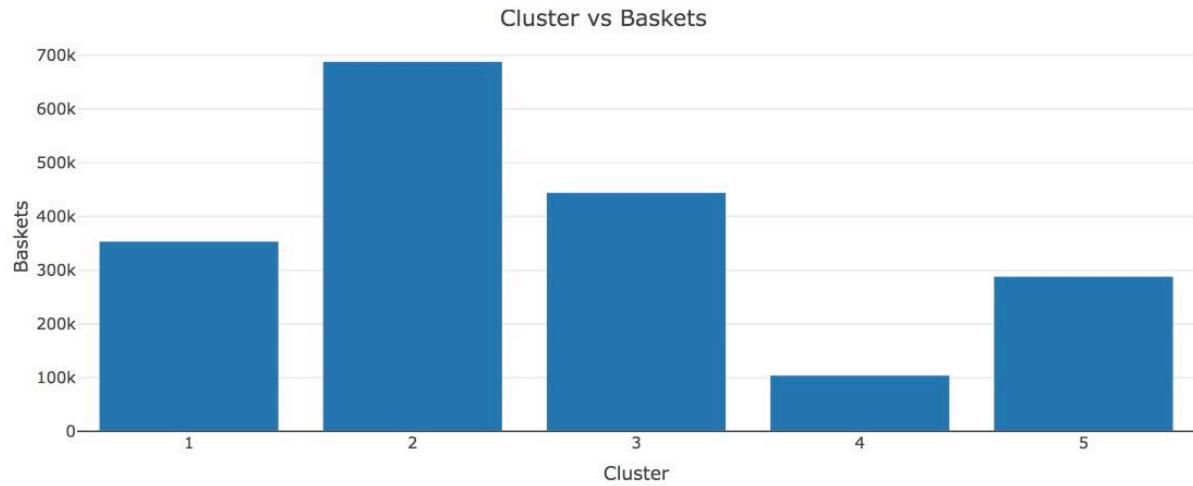
The following figure shows Basket vs Revenue cluster plot .



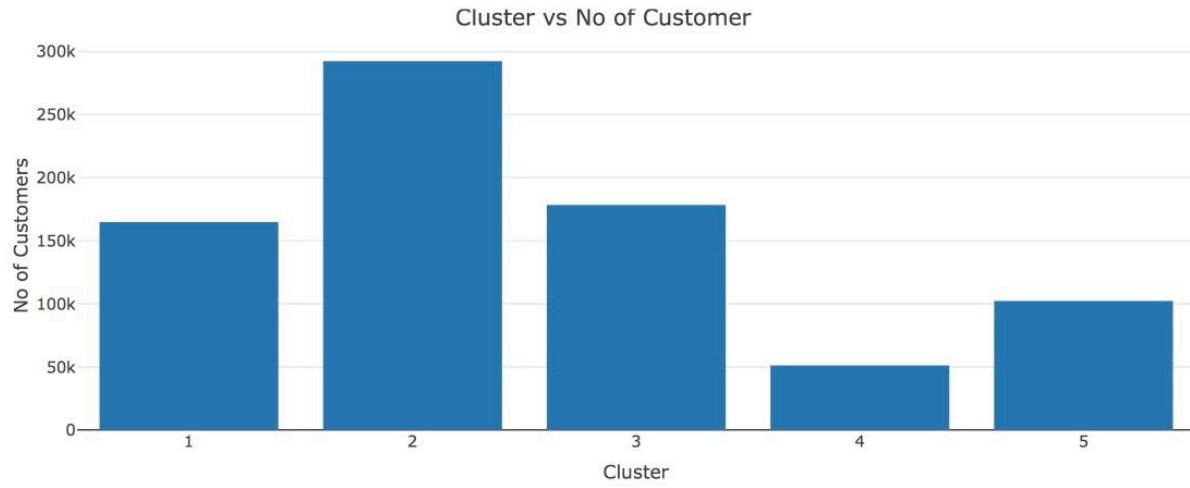
The following plot shows that cluster 2 generates highest revenue.



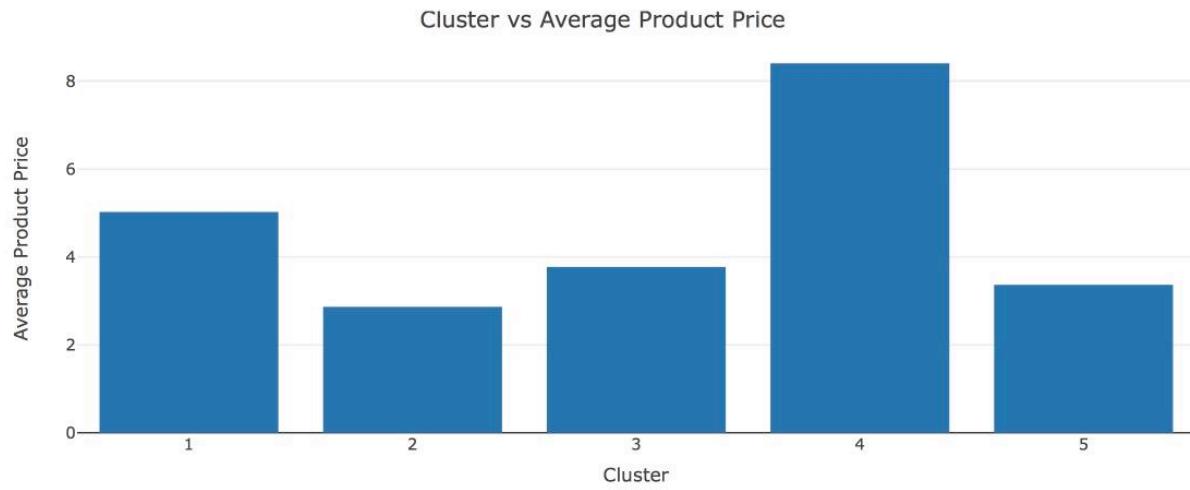
According to the following figure cluster-2 also has highest number of baskets.



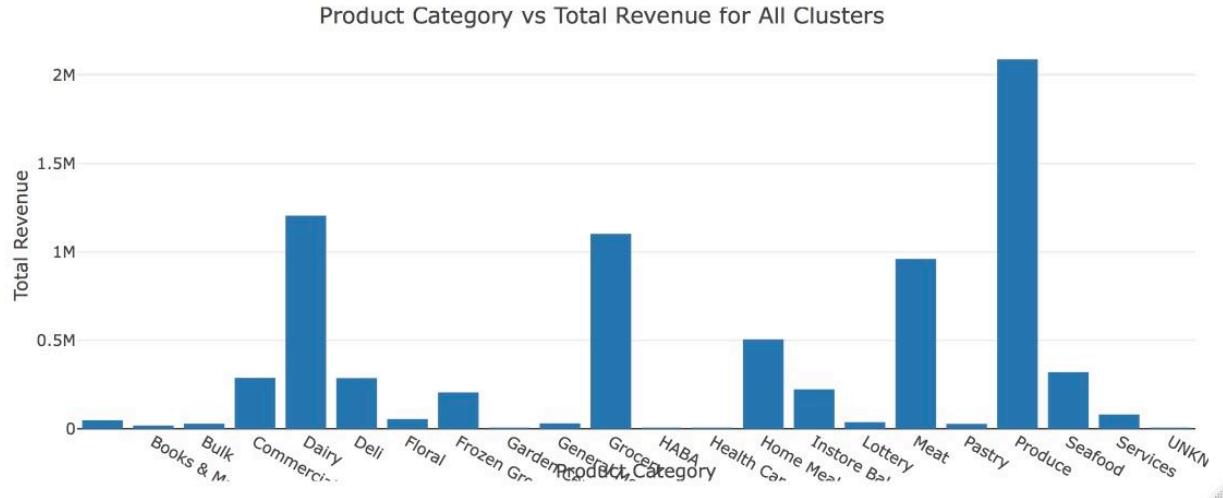
The following figure shows that Cluster-2 has also been purchased by most customers, followed by cluster-3 and cluster-1.



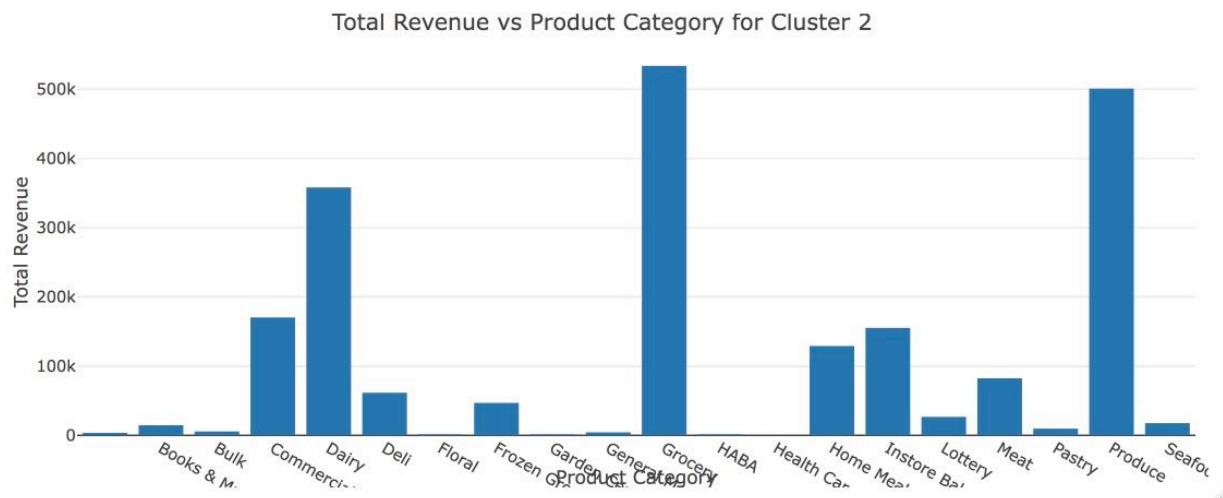
The average product price is highest for cluster-4.



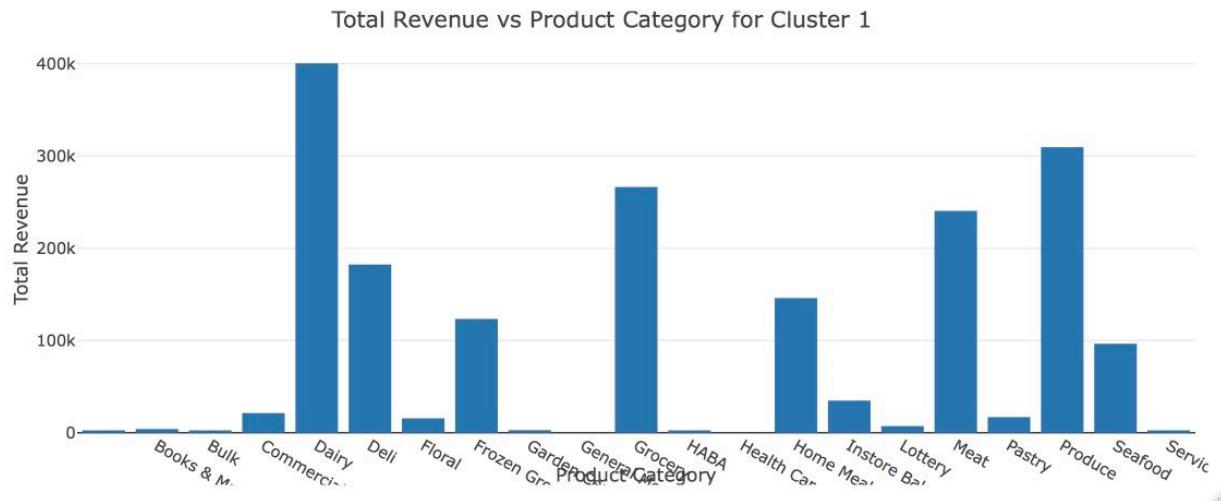
The following file shows that Produce has the highest contribution to total revenue among all the product categories.



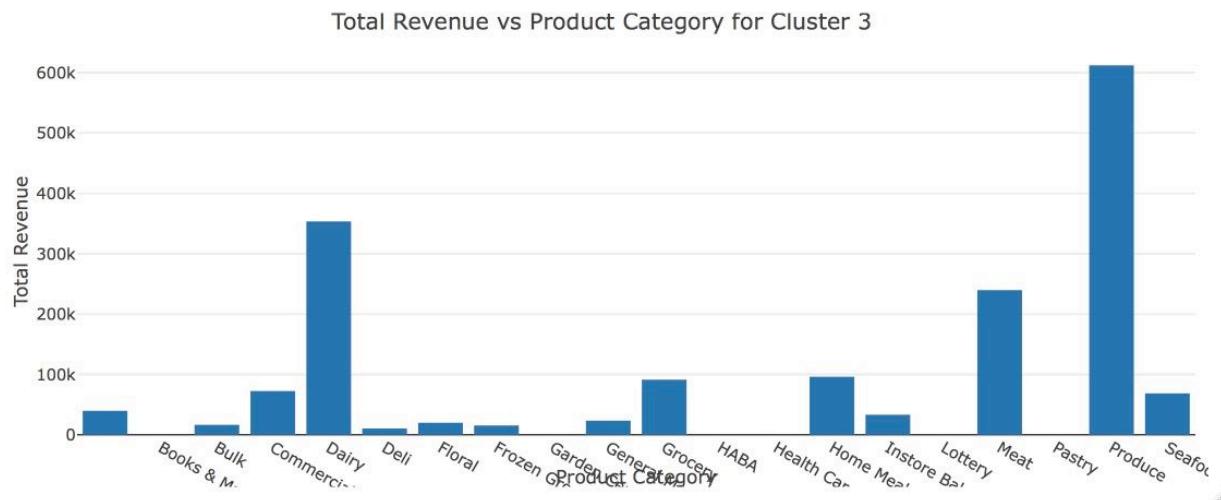
The following plot shows that grocery and produce have the highest contribution to total revenue in product cluster-2.



The following plot shows that dairy and produce have the highest contribution to total revenue in product cluster-1.



The following plot shows that produce has the highest contribution to total revenue in product cluster-3.



7. Product Profiling:

The following table shows description of different products and recommendations.

Product Segment	Description	Recommendation
Cluster 2: The Money Cow	<ul style="list-style-type: none"> Highest revenue generating products (Grocery and produce are the main contributors) Highest sold products Lowest average price 	<ul style="list-style-type: none"> Products are already in demand: Maintaining enough stocks is recommended.
Cluster 1: The Bulk Runner	<ul style="list-style-type: none"> Medium revenue generator (Produce is the main contributor) Third highest selling product Lower average price 	<ul style="list-style-type: none"> Deals on bulk purchases should be increased.
Cluster 3: Fast Movers	<ul style="list-style-type: none"> Medium revenue generator (Dairy and Produce are the main sources) Second highest selling product Products having higher average price Higher customers 	<ul style="list-style-type: none"> Fast moving products: Maintain sufficient stocks
Cluster 5: Needs Attention	<ul style="list-style-type: none"> Lower revenue generator Lower customers Lower sales Medium average price 	<ul style="list-style-type: none"> Deals (e.g. BOGO) should be given on individual item sales to make customers prefer these products.
Cluster 4: High Stakes	<ul style="list-style-type: none"> Lowest revenue generator Least selling products Products having highest average price 	<ul style="list-style-type: none"> Products should be advertised and deals should be given. Only optimum stock should be kept to avoid risk of loss in revenue.

Customer Analysis:

For performing customer analysis, we consolidate the top 2000 customers from the sales219 table and customer tables for creating our final table customer cluster

8.Data Model of the Customer

Following is schema of the “customercluster” table for our analysis:

Column Name	Description
CUSTOMER_SK	Customer ID for each customer
TOTAL_No_Of_PURCHASE	Sum of the total quantity purchased by the customer
TOTAL_REVENUE	Total money spent by the customer
RECENCY	Gives the difference between the last visited date of the customer and current date
TOTAL_No_of_VISITS	Total transaction made by the customer
DISTINCT_PRODUCTS	Total distinct product bought by the customer
AVERAGE_SPEND	Average money spent per visit by the customer
CITY_NM	City of the purchase

9.Descriptive analysis

The following is the summary of the top 2000 customers in our tables. We do this in order to understand the range and distribution of values on various variables.

Summary of Top 2000 Customers:

TOTAL_No_Of_PURCHASE	TOTAL_REVENUE	RECENCY	TOTAL_No_of_VISITS	DISTINCT_PRODUCTS
Min. : 214.0	Min. : 978	Min. : 109.0	Min. : 5.0	Min. : 43.0
1st Qu.: 337.0	1st Qu.: 1124	1st Qu.: 111.0	1st Qu.: 22.0	1st Qu.: 169.0
Median : 406.0	Median : 1331	Median : 242.0	Median : 34.0	Median : 207.0
Mean : 1277.8	Mean : 3960	Mean : 181.7	Mean : 225.2	Mean : 234.1
3rd Qu.: 536.2	3rd Qu.: 1758	3rd Qu.: 244.0	3rd Qu.: 50.0	3rd Qu.: 256.0
Max. : 1626564.0	Max. : 4804678	Max. : 293.0	Max. : 369037.0	Max. : 28468.0
AVERAGE_SPEND				
Min. : 4.861				
1st Qu.: 28.977				
Median : 42.633				
Mean : 52.274				
3rd Qu.: 67.379				
Max. : 273.919				

10.Data Cleaning

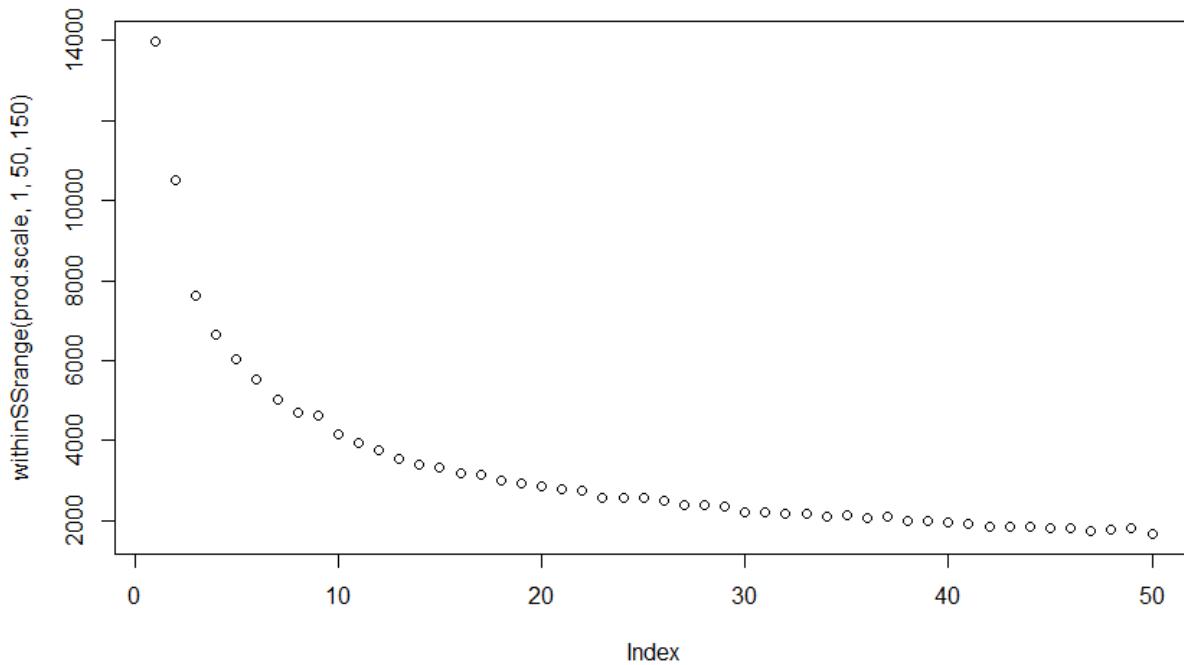
From the top 2000 customer that we extracted, we observe that customer responsible for the highest revenue was customer with CUSTOMER_SK 1. We find that total revenue of CUSTOMER_SK 1 is very high compared to the rest of the data. We also find from the customer table the CUSTOMER_SK 1 corresponds to a Non Member customer of Sobeys. So, in order for avoid biasing of a very high data point, we are considering this data point as outlier and ignore it for rest of our analysis.

11. Data Normalizing

The data has been scaled to a uniform scale for bringing data to a normalized format.

12.Selecting number of clusters

Below is the elbow curve plotted for different cluster size, from observation 6 is point where the curve changes abruptly. 6 is selected as the number of clusters by K Means algorithm for our problem.

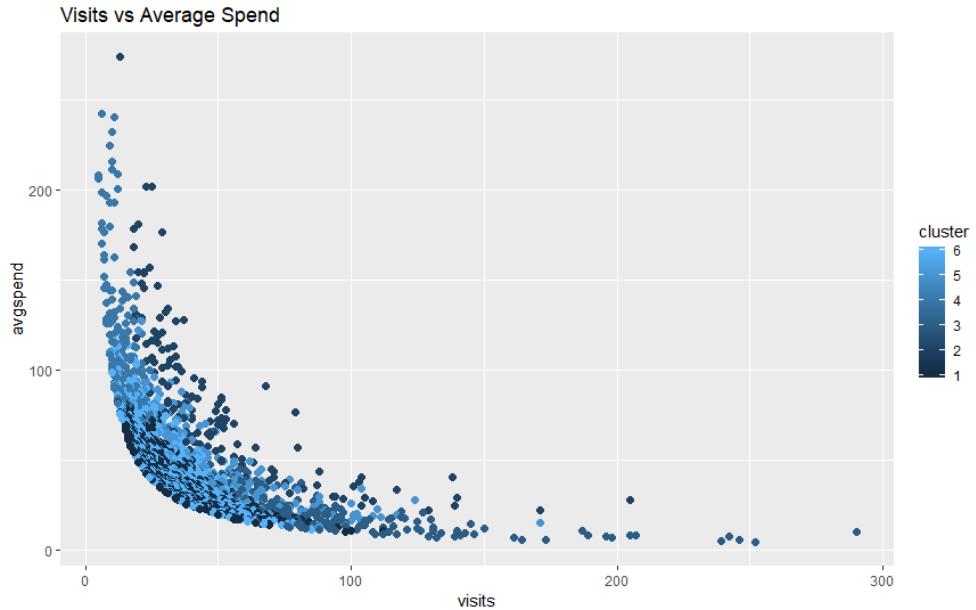


13: Data Denormalization and clustering

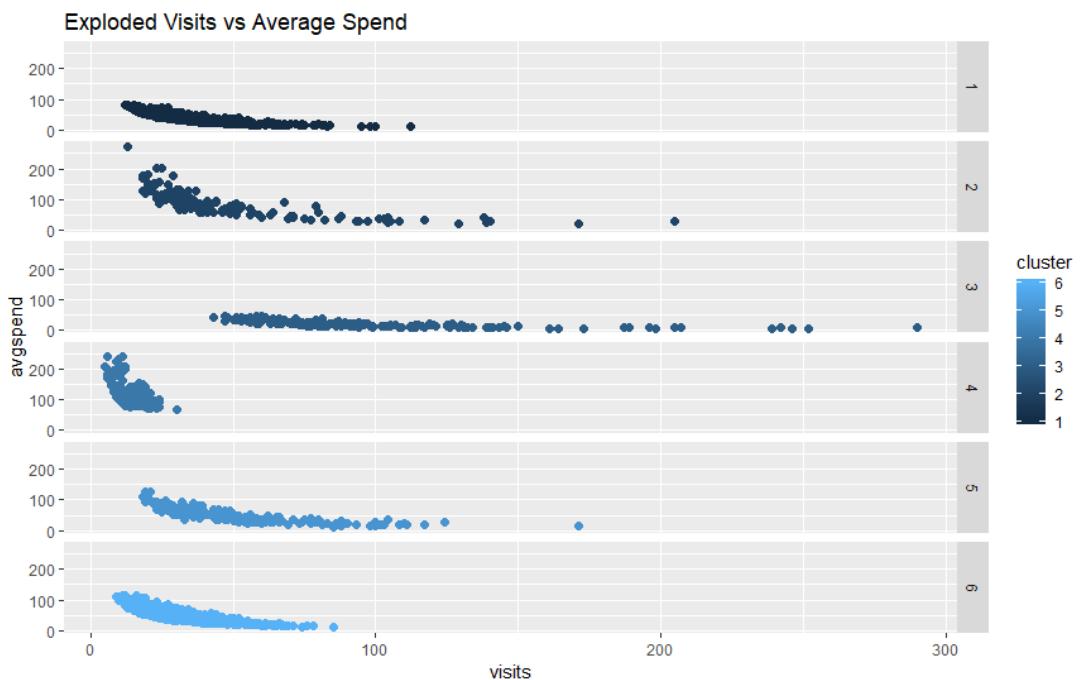
After the number of clusters is decided, data is again denormalized to get actual centroids of the clusters and data along with the clustering information in output to a CSV file.

14: Data Analysis

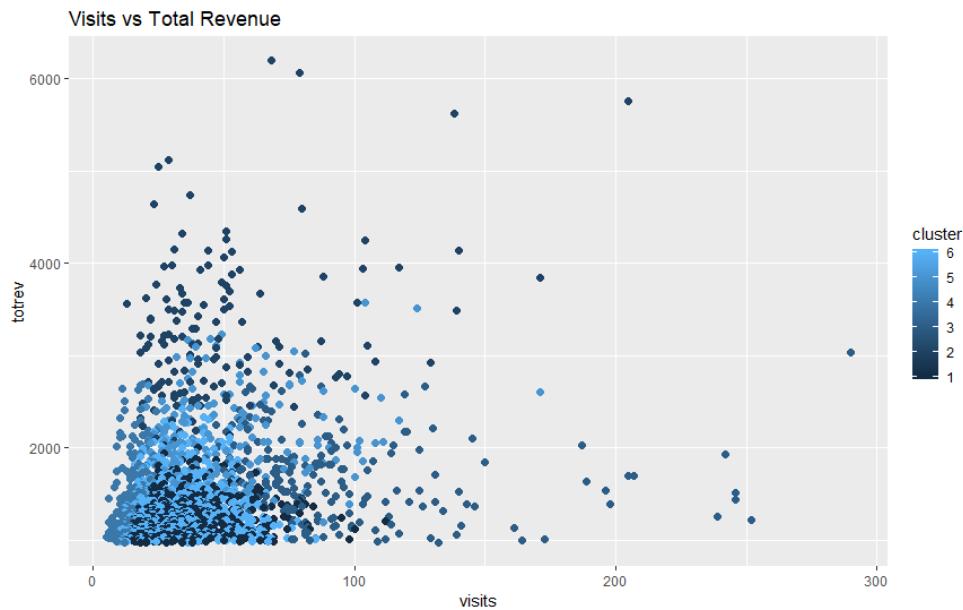
Visits vs Average Spend



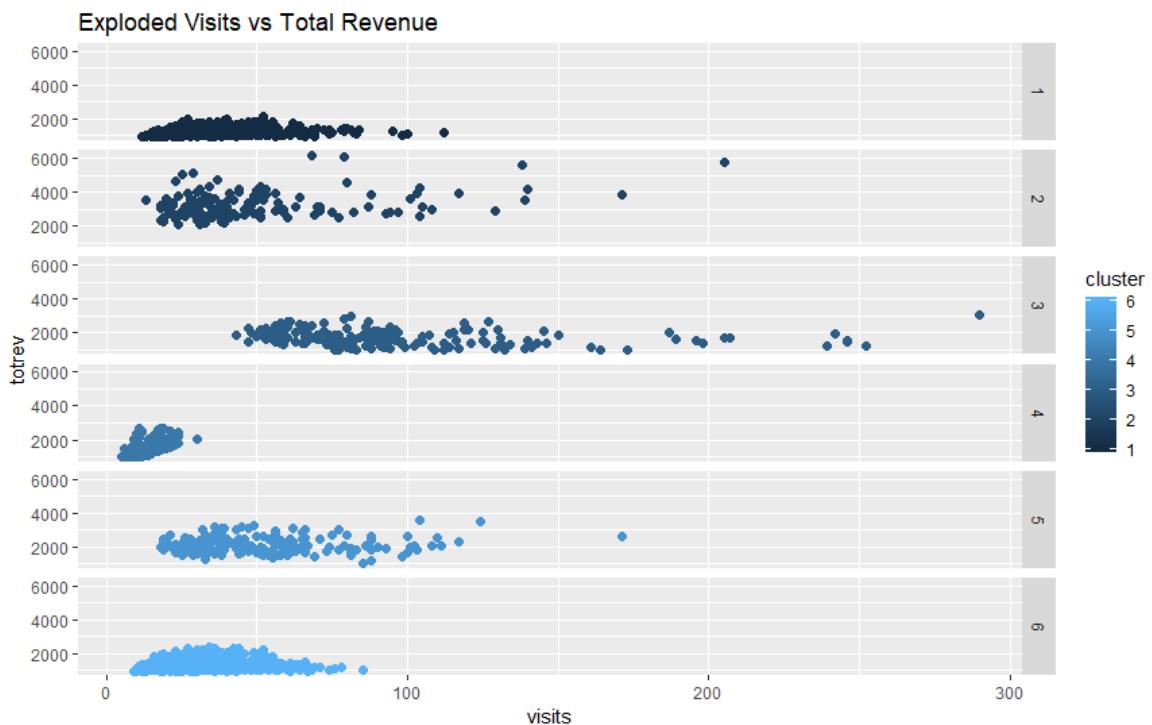
On plotting visits and average spend in ggplot, we analyze the characteristics of each of the cluster. For example the cluster 5 , has lower number of visits but higher average spend .The characteristic of the clusters have been recorded in the results table.We also use the exploded plots to analyze clearly the clusters which is not possible in the above plot.



Total number of visits vs Total Revenue



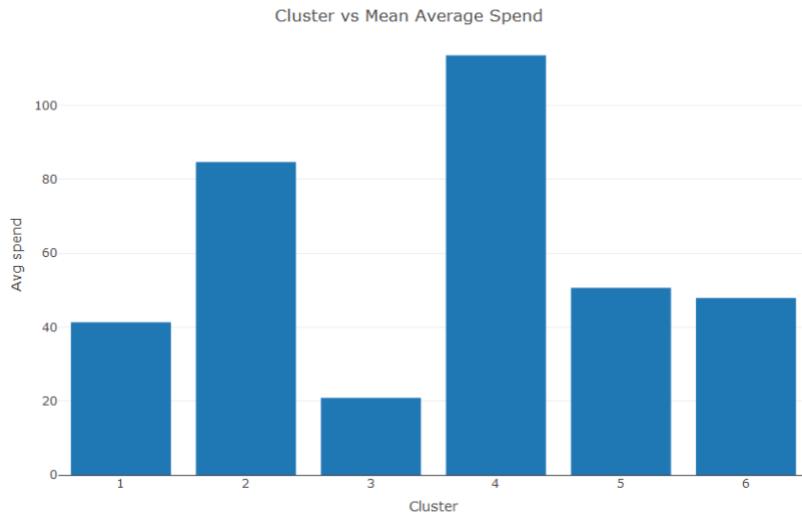
On plotting visits and total revenue in ggplot, we analyze the characteristics of each of the cluster. For example, the cluster 5, has lower number of visits and lower total revenue production. The characteristic of the clusters has been recorded in the results table. We also use the exploded plots to analyze clearly the clusters which is not possible in the above plot.



Cluster level Analysis

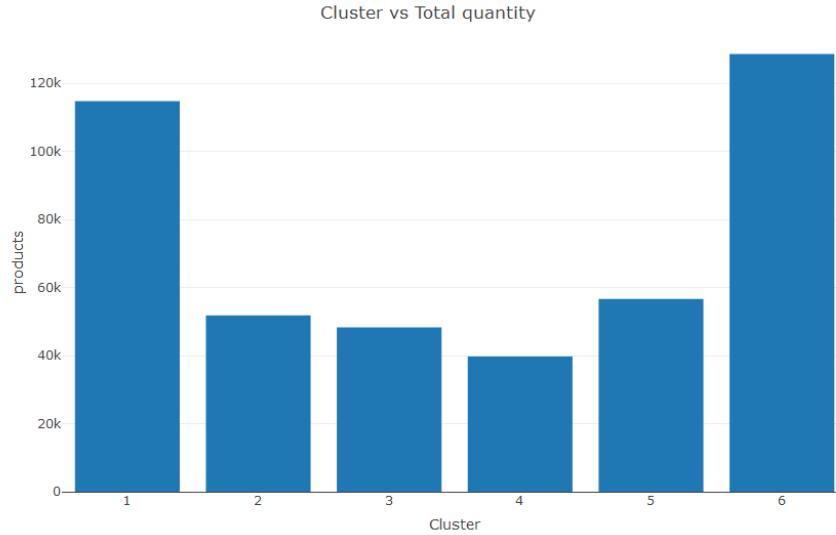
Average Spend

The below bar graph portrays the comparison of Mean Average spend for each cluster. From the visual, we can clearly see that, cluster 4 is above 100 and is leading the other clusters by a huge margin whereas cluster 3 has the lowest average spend with the value dropping just below 20. Clusters 1, 5 and 6 falls almost at the same level (over 40) with not much difference. And cluster 2 follows right behind cluster 4 with the average spend slightly over 80.



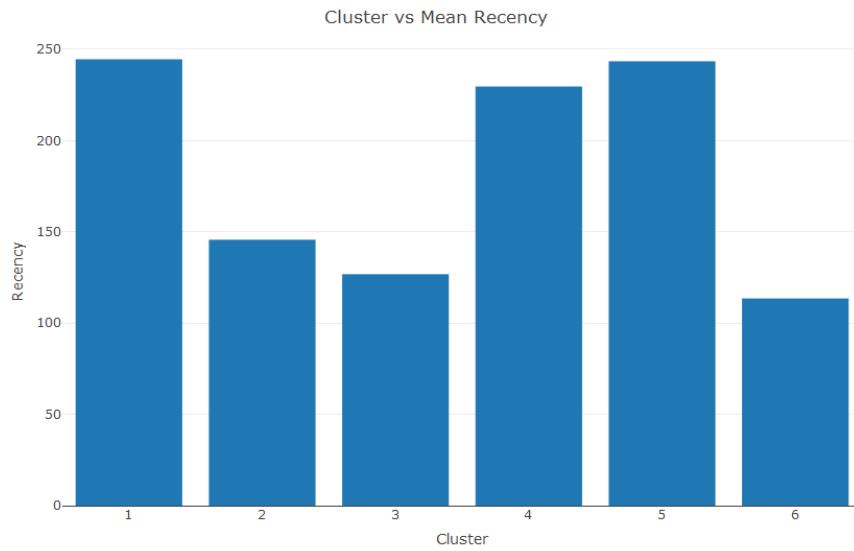
Number of products

The below bar graph portrays the comparison of Total products bought for each cluster. About 2000 customers are grouped into 6 different clusters. Cluster 6 sums up to about 120,000 products and leads other clusters. Cluster 1 is the capable competitor to cluster 6 with total products slightly less than 120k. Other clusters fall almost at the same level.



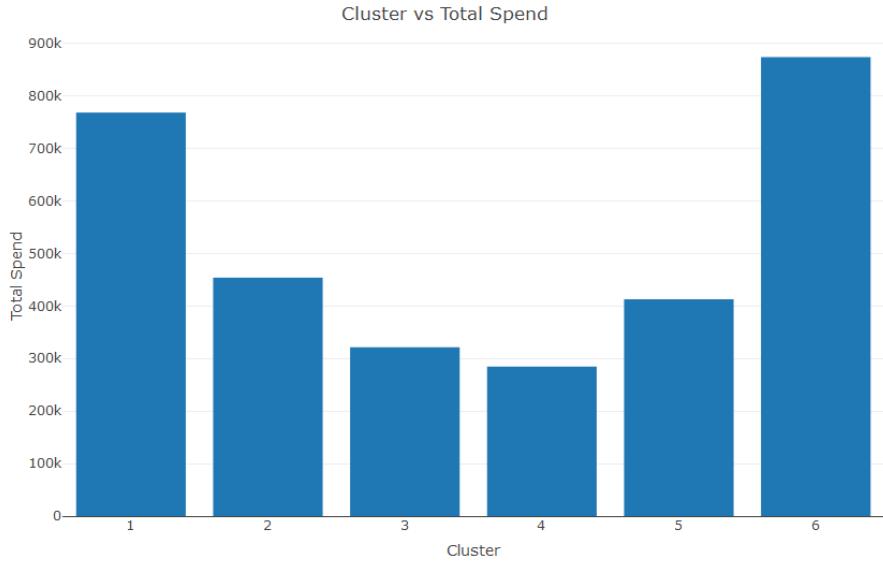
Recency

In the below graph we observe that columns 2,3 and 6 represent very less days meaning that the customers in this cluster have visited the store more recent.



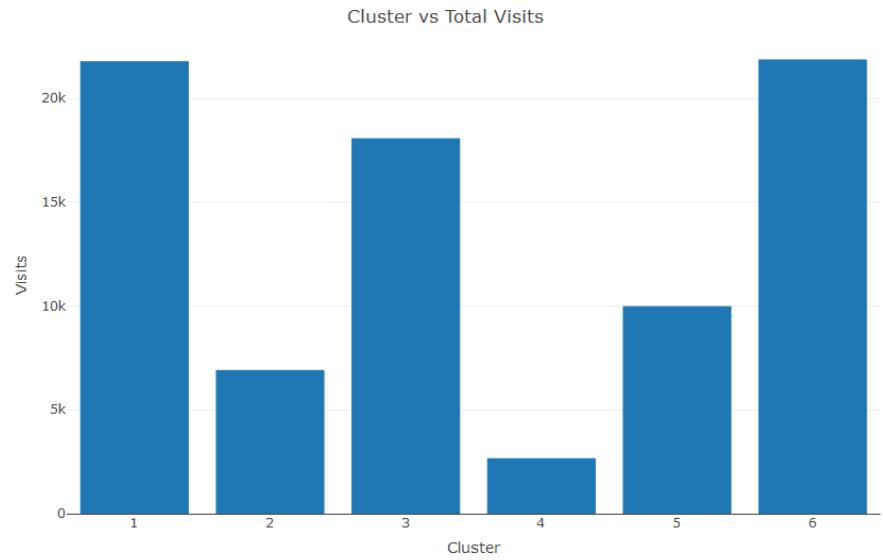
Total Spend

The below bar graph portrays the comparison of Total spend across each cluster. Cluster 6 sums upto about 850,000 and leads other clusters. Cluster 1 is the capable competitor to cluster 6 with total spend slightly less than 800k. Other clusters fall almost at the same level.



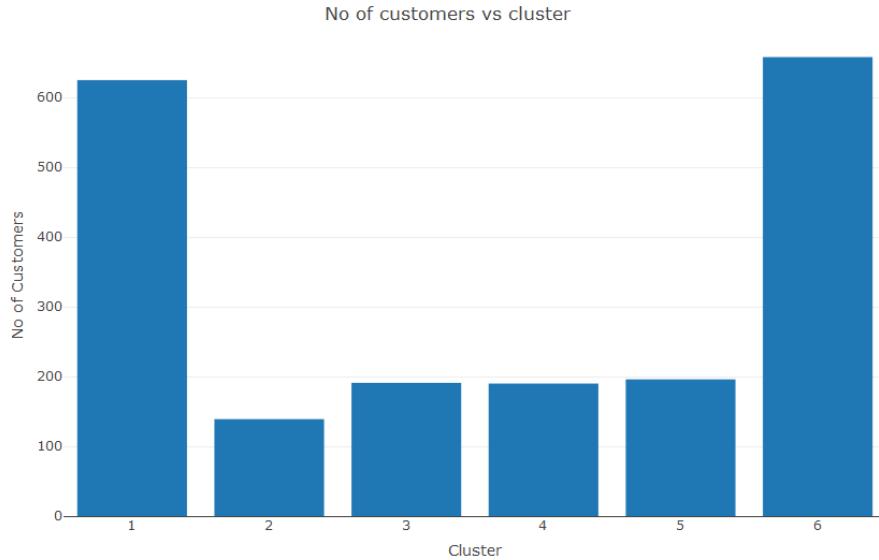
Total Visits

Below is the bar graph showing the total visits across different clusters. Cluster 1 and 6 each sum up to about 20,000 and leads other clusters. Cluster 3 is the capable competitor to cluster 1 and 6 with total visit slightly less than 20k. Clusters 2 and 5 fall almost at the same level with cluster 5 having upper hand.



Cluster Population

Below graph given the population of each of the customer clusters. . About 2000 customers are grouped into 6 different clusters. Cluster 1 and 6 each sums up to about 600,000 and leads other clusters. All other clusters fall almost at the same level at 200,000.



15.Customer Profiling:

From the ggplots and cluster level analysis of the each of the cluster, the common traits of the member of the clusters can be derived. Based on the description the corresponding action that can be taken to target that cluster

From the above graphs, following results can be concluded regarding each of the customer clusters:

<i>Customer Segment</i>	<i>Description</i>	<i>Recommendation</i>
Cluster 1: Lost opportunities	<ul style="list-style-type: none"> • High revenue source • Highest number of visits • High quantity of products sold • Old visitors • Average spend per visit 	Targeted advertising to remind the brand name with competitive deals to drive customers back to store.
Cluster 2: High Potential buyers	<ul style="list-style-type: none"> • Average revenue source • Lower number of visits • Recent visitors to store • High spend per visit 	Provide targeted weekly deals to make customers frequent the store.
Cluster 3: Active customers	<ul style="list-style-type: none"> • Lower revenue source • Higher number of visits • More recent visitors to store • Lowest spend per visit 	Provide combined deals on products to increase the products per visit.
Cluster 4: Impulse Buyers	<ul style="list-style-type: none"> • Lower revenue source • Lowest number of visits • Lowest number of products • Old visitors • Very high spend per visit 	Provide deals on the costly items in the targeted advertisements.

Cluster 5: Informed Customers	<ul style="list-style-type: none"> Average revenue source Average number of visits Higher sales Old visitors Average spend per visit 	Targeted advertising to remind the brand name with bulk purchase deals to drive customers back to store.
Cluster 6: Loyal Customers	<ul style="list-style-type: none"> Highest revenue makers Second highest number of visits Most recent Average spends per visit 	No action needed. Maintain the current strategy with customers.

Note:

1. Kindly refer to the resource folder of R codes and SQL scripts.

NON-DISCLOSURE AGREEMENT

This Non-Disclosure Agreement (the "**Agreement**") is effective as of 23 RD JAN,
2019.

BETWEEN:

Saint Mary's University, an educational institution incorporated under the laws of
the Province of Nova Scotia and having offices at 923 Robie Street, Halifax Nova
Scotia, B3H 3C3

(the "**University**")

AND:

(name) PARIJAT BANDYOPADHYAY, (address)
LOYOLA RESIDENCE, SMU, HALIFAX (the "**Individual**").

WHEREAS:

The Individual ("**Recipient**") is enrolled in _____ at the University ("**Provider**"), and this course requires exposure to confidential information that has been supplied to the University by third parties. The third party organizations do not want their confidential information to be released publicly.

THE PARTIES AGREE AS FOLLOWS:

1. CONFIDENTIAL INFORMATION

Provider will provide Recipient with information and materials concerning **data analytics** which are clearly marked as confidential or proprietary when first disclosed ("Information") and include, without limitation, trade secrets, know-how, show-how, concepts, discoveries, inventions, research or technical data and other proprietary information or material (biological or otherwise). Information may also include information furnished during discussions or oral presentations if it is conspicuously identified as proprietary at the time and then transcribed or confirmed in writing within thirty (30) days, specifically describing what portions of such information is considered to be proprietary or confidential. However, Recipient is under no obligation to maintain the confidentiality of Information which Recipient can show:

- (a) is or subsequently becomes generally available to the public through no act or fault of Recipient;
- (b) was in the possession of Recipient prior to its disclosure by the Provider to the Recipient;
- (c) was lawfully acquired by Recipient from a third party who was not under an obligation of confidentiality to Provider; or

- (d) is required by an order of a legal process to disclose, provided that Recipient gives Provider prompt and reasonable notification of such requirement prior to disclosure; or
- (e) was independently developed by employees, agents or consultants of the recipient who had no knowledge of or access to the discloser's information as evidenced by the recipient's records.

2. OWNERSHIP

The Information is and will at all times remain the exclusive property of Provider and/or its employees and/or a third party as applicable, and nothing in this Agreement grants the Recipient any right, title or interest in or to the Information.

3. NO REPRESENTATION OR WARRANTY

Recipient acknowledges and agrees that the Information is experimental in nature and that any use of the Information by Recipient will be at the sole risk and liability of Recipient. PROVIDER MAKES NO REPRESENTATION OR WARRANTY, WHETHER EXPRESSED OR IMPLIED, WITH RESPECT TO THE INFORMATION, INCLUDING ANY REPRESENTATION OR WARRANTY AS TO ITS ACCURACY, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT ON THIRD PARTY PROPRIETARY RIGHTS. ALSO, PROVIDER WILL NOT BE LIABLE FOR ANY INDIRECT, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGE OR LOSS ARISING FROM ANY USE OF THE INFORMATION BY RECIPIENT EVEN IF PROVIDER HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE OR LOSS.

4. USE

Recipient will not use the Information for any purpose other than to perform academic exercises as instructed by the University. Recipient will not de-compile or reverse engineer the Information or use the Information to develop, or cause to develop, all or part of any process or product whether for internal use or for commercial purposes.

5. PERIOD OF USE

Recipient may use the Information for the purpose set out in section 4 for a period commencing on the date of this Agreement and ending two years thereafter unless terminated earlier by one party upon giving the other party at least 5 days written notice. At the end of such period or at the request of Provider, Recipient will return or destroy all copies of the Information, except that Recipient may provide a sealed copy of the Information to its legal counsel for archival purpose.

6. CONFIDENTIALITY

Recipient will use the same care and discretion to avoid disclosure of the Information as Recipient uses with its own similar information that the Recipient does not wish to disclose. Recipient will use such care and discretion to avoid disclosure for a period of 5 years from the date of this Agreement irrespective of the expiration or earlier termination of the period of use described in section 5.

7. NO WAIVER

No provision of this Agreement will be deemed waived or any breach excused, unless such waiver or consent excusing the breach is in writing and signed by the Provider. A waiver of a provision of this Agreement will not be construed to be a waiver of a subsequent breach of the same provision.

8. ASSIGNMENT

Neither party may assign all or part of this Agreement without the prior written consent of the other party, whose consent will not be unreasonably withheld.

9. ENTIRE AGREEMENT AND COUNTERPARTS

This Agreement contains the entire agreement and understanding of the parties with respect to its subject matter and supersedes all prior proposals, negotiations, agreements, understandings, representations and warranties of any form or nature, whether oral or written, and whether express or implied, which may have been entered into between the parties relating to its subject matter. This Agreement may be signed in counterparts and faxed to the other party or parties, and each counterpart, together with the other counterparts will constitute the entire Agreement.

10. GOVERNING LAW AND JURISDICTION

This Agreement will be governed by and construed under the laws of Nova Scotia and the applicable laws of Canada without reference to its conflict of law rules. Any action or proceeding brought to enforce the terms of this Agreement will be brought in a court in Nova Scotia, and the parties hereby consent and submit to the exclusive jurisdiction of such court.

IN WITNESS WHEREOF the parties have executed this Agreement on the date first written above.

Saint Mary's University
by its duly authorized officer:

Signature: _____
Name: Dr. Adam Sarty
Title: Associate Vice President, Research

Individual

Signature: Parijat Bandyopadhyay
Name: PARIJAT BANDYOPADHYAY

The people responsible for delivering the course agree that they have read and will comply with the terms of this Agreement and agree to ensure that all participants are informed of the obligations under this Agreement.

UNIVERSITY'S SCIENTIST RESPONSIBLE FOR DELIVERING THE COURSE

Signature: _____
Name: Dr. Pawan Lingras
Title: _____

NON-DISCLOSURE AGREEMENT

This Non-Disclosure Agreement (the "Agreement") is effective as of 23rd Jan, 2019.

BETWEEN:

Saint Mary's University, an educational institution incorporated under the laws of the Province of Nova Scotia and having offices at 923 Robie Street, Halifax Nova Scotia, B3H 3C3

(the "University")

AND:

(name) VIVEKANAND BOOPATHY, (address)
LOYOLA RESIDENCE, HALIFAX (the "Individual").

WHEREAS:

The Individual ("Recipient") is enrolled in _____ at the University ("Provider"), and this course requires exposure to confidential information that has been supplied to the University by third parties. The third party organizations do not want their confidential information to be released publicly.

THE PARTIES AGREE AS FOLLOWS:

1. CONFIDENTIAL INFORMATION

Provider will provide Recipient with information and materials concerning **data analytics** which are clearly marked as confidential or proprietary when first disclosed ("Information") and include, without limitation, trade secrets, know-how, show-how, concepts, discoveries, inventions, research or technical data and other proprietary information or material (biological or otherwise). Information may also include information furnished during discussions or oral presentations if it is conspicuously identified as proprietary at the time and then transcribed or confirmed in writing within thirty (30) days, specifically describing what portions of such information is considered to be proprietary or confidential. However, Recipient is under no obligation to maintain the confidentiality of Information which Recipient can show:

- (a) is or subsequently becomes generally available to the public through no act or fault of Recipient;
- (b) was in the possession of Recipient prior to its disclosure by the Provider to the Recipient;
- (c) was lawfully acquired by Recipient from a third party who was not under an obligation of confidentiality to Provider; or

- (d) is required by an order of a legal process to disclose, provided that Recipient gives Provider prompt and reasonable notification of such requirement prior to disclosure; or
- (e) was independently developed by employees, agents or consultants of the recipient who had no knowledge of or access to the discloser's information as evidenced by the recipient's records.

2. OWNERSHIP

The Information is and will at all times remain the exclusive property of Provider and/or its employees and/or a third party as applicable, and nothing in this Agreement grants the Recipient any right, title or interest in or to the Information.

3. NO REPRESENTATION OR WARRANTY

Recipient acknowledges and agrees that the Information is experimental in nature and that any use of the Information by Recipient will be at the sole risk and liability of Recipient. PROVIDER MAKES NO REPRESENTATION OR WARRANTY, WHETHER EXPRESSED OR IMPLIED, WITH RESPECT TO THE INFORMATION, INCLUDING ANY REPRESENTATION OR WARRANTY AS TO ITS ACCURACY, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT ON THIRD PARTY PROPRIETARY RIGHTS. ALSO, PROVIDER WILL NOT BE LIABLE FOR ANY INDIRECT, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGE OR LOSS ARISING FROM ANY USE OF THE INFORMATION BY RECIPIENT EVEN IF PROVIDER HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE OR LOSS.

4. USE

Recipient will not use the Information for any purpose other than to perform academic exercises as instructed by the University. Recipient will not de-compile or reverse engineer the Information or use the Information to develop, or cause to develop, all or part of any process or product whether for internal use or for commercial purposes.

5. PERIOD OF USE

Recipient may use the Information for the purpose set out in section 4 for a period commencing on the date of this Agreement and ending two years thereafter unless terminated earlier by one party upon giving the other party at least 5 days written notice. At the end of such period or at the request of Provider, Recipient will return or destroy all copies of the Information, except that Recipient may provide a sealed copy of the Information to its legal counsel for archival purpose.

6. CONFIDENTIALITY

Recipient will use the same care and discretion to avoid disclosure of the Information as Recipient uses with its own similar information that the Recipient does not wish to disclose. Recipient will use such care and discretion to avoid disclosure for a period of 5 years from the date of this Agreement irrespective of the expiration or earlier termination of the period of use described in section 5.

7. NO WAIVER

No provision of this Agreement will be deemed waived or any breach excused, unless such waiver or consent excusing the breach is in writing and signed by the Provider. A waiver of a provision of this Agreement will not be construed to be a waiver of a subsequent breach of the same provision.

8. ASSIGNMENT

Neither party may assign all or part of this Agreement without the prior written consent of the other party, whose consent will not be unreasonably withheld.

9. ENTIRE AGREEMENT AND COUNTERPARTS

This Agreement contains the entire agreement and understanding of the parties with respect to its subject matter and supersedes all prior proposals, negotiations, agreements, understandings, representations and warranties of any form or nature, whether oral or written, and whether express or implied, which may have been entered into between the parties relating to its subject matter. This Agreement may be signed in counterparts and faxed to the other party or parties, and each counterpart, together with the other counterparts will constitute the entire Agreement.

10. GOVERNING LAW AND JURISDICTION

This Agreement will be governed by and construed under the laws of Nova Scotia and the applicable laws of Canada without reference to its conflict of law rules. Any action or proceeding brought to enforce the terms of this Agreement will be brought in a court in Nova Scotia, and the parties hereby consent and submit to the exclusive jurisdiction of such court.

IN WITNESS WHEREOF the parties have executed this Agreement on the date first written above.

Saint Mary's University
by its duly authorized officer:

Signature: _____
Name: Dr. Adam Sarty
Title: Associate Vice President, Research

Individual

Signature: *Brunel Patel* _____
Name: VIVEKANAND RAO PATHI

The people responsible for delivering the course agree that they have read and will comply with the terms of this Agreement and agree to ensure that all participants are informed of the obligations under this Agreement.

UNIVERSITY'S SCIENTIST RESPONSIBLE FOR DELIVERING THE COURSE

Signature: _____
Name: Dr. Pawan Lingras
Title: _____

NON-DISCLOSURE AGREEMENT

This Non-Disclosure Agreement (the "Agreement") is effective as of 23rd Jan 2019, 2017.

BETWEEN:

Saint Mary's University, an educational institution incorporated under the laws of the Province of Nova Scotia and having offices at 923 Robie Street, Halifax Nova Scotia, B3H 3C3

(the "University")

AND:

(name) KOTHAI KANNAPPAN MURUGAPPAN, (address)
LOYOLCA RESIDENCE, SMU (the "Individual").

WHEREAS:

The Individual ("Recipient") is enrolled in _____ at the University ("Provider"), and this course requires exposure to confidential information that has been supplied to the University by third parties. The third party organizations do not want their confidential information to be released publicly.

THE PARTIES AGREE AS FOLLOWS:

1. CONFIDENTIAL INFORMATION

Provider will provide Recipient with information and materials concerning **data analytics** which are clearly marked as confidential or proprietary when first disclosed ("Information") and include, without limitation, trade secrets, know-how, show-how, concepts, discoveries, inventions, research or technical data and other proprietary information or material (biological or otherwise). Information may also include information furnished during discussions or oral presentations if it is conspicuously identified as proprietary at the time and then transcribed or confirmed in writing within thirty (30) days, specifically describing what portions of such information is considered to be proprietary or confidential. However, Recipient is under no obligation to maintain the confidentiality of Information which Recipient can show:

- (a) is or subsequently becomes generally available to the public through no act or fault of Recipient;
- (b) was in the possession of Recipient prior to its disclosure by the Provider to the Recipient;
- (c) was lawfully acquired by Recipient from a third party who was not under an obligation of confidentiality to Provider; or

- (d) is required by an order of a legal process to disclose, provided that Recipient gives Provider prompt and reasonable notification of such requirement prior to disclosure; or
- (e) was independently developed by employees, agents or consultants of the recipient who had no knowledge of or access to the discloser's information as evidenced by the recipient's records.

2. OWNERSHIP

The Information is and will at all times remain the exclusive property of Provider and/or its employees and/or a third party as applicable, and nothing in this Agreement grants the Recipient any right, title or interest in or to the Information.

3. NO REPRESENTATION OR WARRANTY

Recipient acknowledges and agrees that the Information is experimental in nature and that any use of the Information by Recipient will be at the sole risk and liability of Recipient. PROVIDER MAKES NO REPRESENTATION OR WARRANTY, WHETHER EXPRESSED OR IMPLIED, WITH RESPECT TO THE INFORMATION, INCLUDING ANY REPRESENTATION OR WARRANTY AS TO ITS ACCURACY, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT ON THIRD PARTY PROPRIETARY RIGHTS. ALSO, PROVIDER WILL NOT BE LIABLE FOR ANY INDIRECT, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGE OR LOSS ARISING FROM ANY USE OF THE INFORMATION BY RECIPIENT EVEN IF PROVIDER HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE OR LOSS.

4. USE

Recipient will not use the Information for any purpose other than to perform academic exercises as instructed by the University. Recipient will not de-compile or reverse engineer the Information or use the Information to develop, or cause to develop, all or part of any process or product whether for internal use or for commercial purposes.

5. PERIOD OF USE

Recipient may use the Information for the purpose set out in section 4 for a period commencing on the date of this Agreement and ending two years thereafter unless terminated earlier by one party upon giving the other party at least 5 days written notice. At the end of such period or at the request of Provider, Recipient will return or destroy all copies of the Information, except that Recipient may provide a sealed copy of the Information to its legal counsel for archival purpose.

6. CONFIDENTIALITY

Recipient will use the same care and discretion to avoid disclosure of the Information as Recipient uses with its own similar information that the Recipient does not wish to disclose. Recipient will use such care and discretion to avoid disclosure for a period of 5 years from the date of this Agreement irrespective of the expiration or earlier termination of the period of use described in section 5.

7. NO WAIVER

No provision of this Agreement will be deemed waived or any breach excused, unless such waiver or consent excusing the breach is in writing and signed by the Provider. A waiver of a provision of this Agreement will not be construed to be a waiver of a subsequent breach of the same provision.

8. ASSIGNMENT

Neither party may assign all or part of this Agreement without the prior written consent of the other party, whose consent will not be unreasonably withheld.

9. ENTIRE AGREEMENT AND COUNTERPARTS

This Agreement contains the entire agreement and understanding of the parties with respect to its subject matter and supersedes all prior proposals, negotiations, agreements, understandings, representations and warranties of any form or nature, whether oral or written, and whether express or implied, which may have been entered into between the parties relating to its subject matter. This Agreement may be signed in counterparts and faxed to the other party or parties, and each counterpart, together with the other counterparts will constitute the entire Agreement.

10. GOVERNING LAW AND JURISDICTION

This Agreement will be governed by and construed under the laws of Nova Scotia and the applicable laws of Canada without reference to its conflict of law rules. Any action or proceeding brought to enforce the terms of this Agreement will be brought in a court in Nova Scotia, and the parties hereby consent and submit to the exclusive jurisdiction of such court.

IN WITNESS WHEREOF the parties have executed this Agreement on the date first written above.

Saint Mary's University
by its duly authorized officer:

Signature: _____
Name: Dr. Adam Sarty
Title: Associate Vice President, Research

Individual

Signature: Karthi -
Name: KOTHAI KANN APPAN MURUGAPPAN

The people responsible for delivering the course agree that they have read and will comply with the terms of this Agreement and agree to ensure that all participants are informed of the obligations under this Agreement.

UNIVERSITY'S SCIENTIST RESPONSIBLE FOR DELIVERING THE COURSE

Signature: _____
Name: Dr. Pawan Lingras
Title: _____