DICE ANALYTICS

Agenda

- Task is to segment the Telcom users market
- Exploratory Data Analysis
- Collinearity Removal
- Elbow Analysis
- Clustering
- Profiling
- Product Recommendations

DICE ANALYTICS

Problem Statement

Market segmentation enables companies to better serve customers by identifying and marketing different products and services to different consumer sets. In this problem, following task were to achieve:

- Segmentation of the Customers
- Presentation of segments visually in detail
- Suggestion of various Campaigns which can been thrown to specific clusters.

DICE ANALYTICS

Assumptions

- Number of Observations: 100
- Number of Variables: 474
- Sample of 100 Rows isn't significant.
- However due to availability of the Data, it is being considered as a significant sample.
- Data Analysis and Clustering is done on the basis of Data Sample of 100 observations

DICE ANALYTICS

DICE ANALYTICS

Few insights from the Data

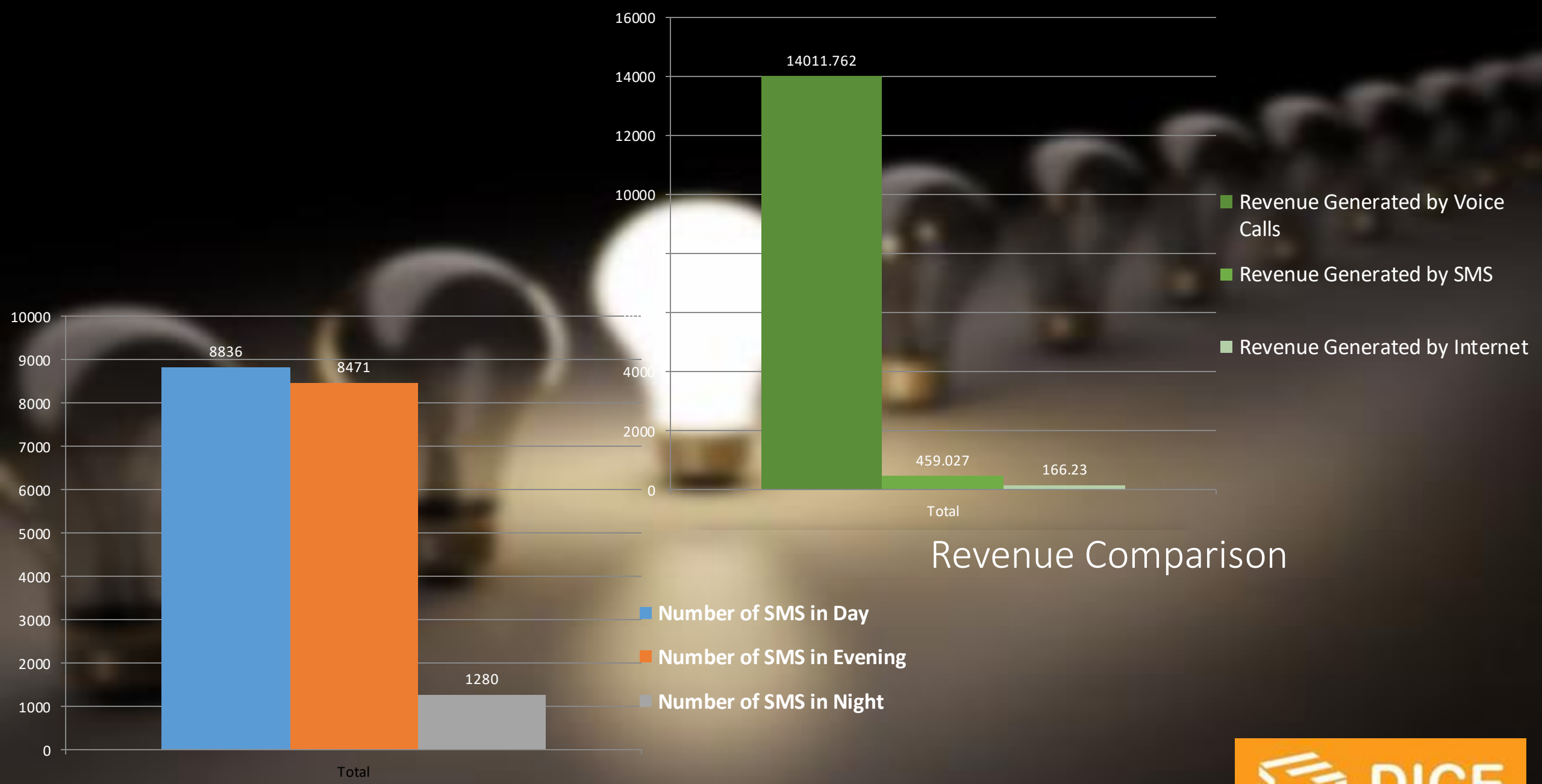
Most of the users are call users

Most of the Calls are being done at Evening Time

Usage of SMS is in greater in Day Time

95% of the revenue is generated by Calls*

*Stats of PTA's reports are different (70% are call users and 30% are SMS and internet users)



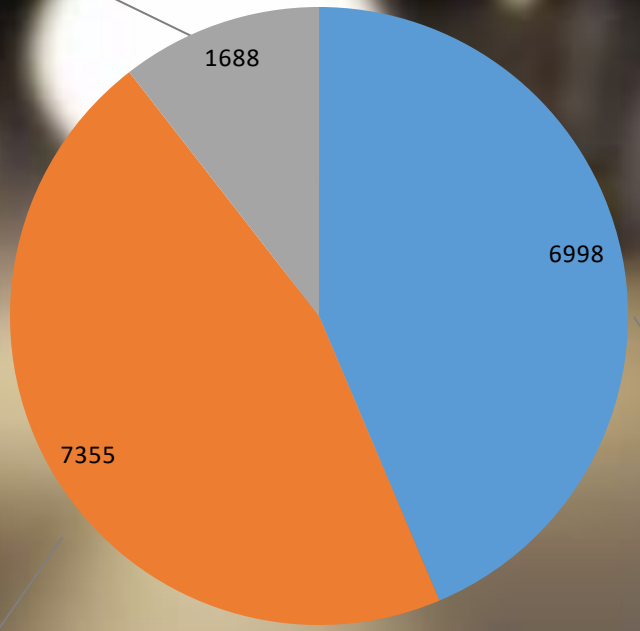
Revenue Comparison

Number for SMS in various shifts

DICE ANALYTICS

DICE ANALYTICS

Amount of Calls in Night
(1,688)



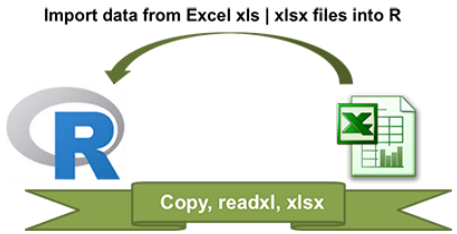
Amount of Calls in Day time
(6,998)

Amount of Calls in Evening
(7,355)

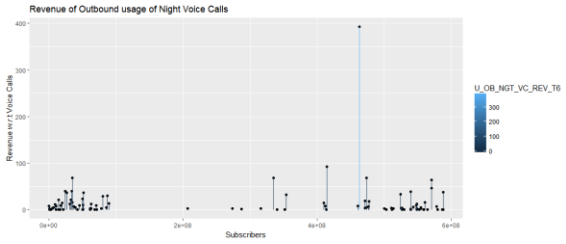
Number of Calls in various shifts

System Architecture

READING AND CLEANING THE GIVEN DATA

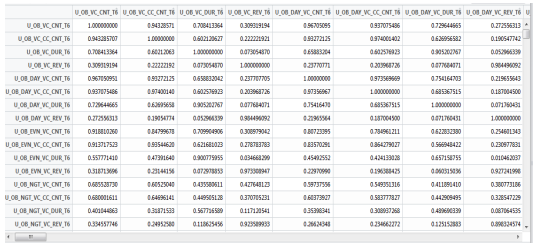


PREPROCESSING FOR VISUAL ANALYSIS OF DATA



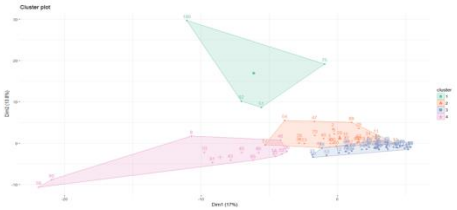
DATA ANALYSIS

FEATURE ENGINEERING & COVARIANCE MARIX GENERATION

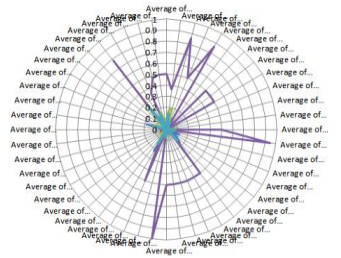


FEATURE ENGINEEIRING & MIN MAX NORMALIZATION

SEGMENTATION THROUGH K MEANS CLUSTERING & VISUALIZATION OF SEGMENTS



RESULTS FOR DATA VISUALIZATION & SEGMENTATION



DATASET USED FOR THE SYSTEM: TELENOR SAMPLE DATA

OBSERVATIONS: 100

DICE ANALYTICS

Data Preprocessing Steps

- Exploratory Data Analysis
- Outliers Detection
- Variables with constant values (0) were removed
  (Variables: 381)
- Correlation matrix generated (381*381)
  - Removed collinearity at a cutoff value of 0.9
  - Removed the variables with NA values

	U_OB_VC_CNT_T6	U_OB_VC_CC_CNT_T6	U_OB_VC_DUR_T6	U_OB_VC_REV_T6	U_OB_DAY_VC_CNT_T6	U_OB_DAY_VC_CC_CNT_T6	U_OB_DAY_VC_DUR_T6	U_OB_DAY_VC_REV_T6	U
U_OB_VC_CNT_T6	1.000000000	0.94328571	0.708413364	0.309319194	0.96705095	0.937075486	0.729644665	0.272556313	
U_OB_VC_CC_CNT_T6	0.943285707	1.00000000	0.602120627	0.222221921	0.93272125	0.974001402	0.626956582	0.190547742	
U_OB_VC_DUR_T6	0.708413364	0.60212063	1.000000000	0.073054870	0.65883204	0.602576923	0.905202767	0.052966339	
U_OB_VC_REV_T6	0.309319194	0.22222192	0.073054870	1.000000000	0.23770771	0.203968726	0.077684071	0.984496092	
U_OB_DAY_VC_CNT_T6	0.967050951	0.93272125	0.658832042	0.237707705	1.00000000	0.973569669	0.754164703	0.219655643	
U_OB_DAY_VC_CC_CNT_T6	0.937075486	0.97400140	0.602576923	0.203968726	0.97356967	1.000000000	0.685367515	0.187004500	
U_OB_DAY_VC_DUR_T6	0.729644665	0.62695658	0.905202767	0.077684071	0.75416470	0.685367515	1.000000000	0.071760431	
U_OB_DAY_VC_REV_T6	0.272556313	0.19054774	0.052966339	0.984496092	0.21965564	0.187004500	0.071760431	1.000000000	
U_OB_EVN_VC_CNT_T6	0.918810260	0.84799678	0.709904906	0.308979042	0.80723395	0.784961211	0.622832380	0.254601343	
U_OB_EVN_VC_CC_CNT_T6	0.913717523	0.93544620	0.621681023	0.278783783	0.83570291	0.864279027	0.566948422	0.230977831	
U_OB_EVN_VC_DUR_T6	0.557771410	0.47391640	0.900775955	0.034668299	0.45492552	0.424133028	0.657158755	0.010462037	
U_OB_EVN_VC_REV_T6	0.318713696	0.23144156	0.072978853	0.973308947	0.22970990	0.196388425	0.060315036	0.927241998	
U_OB_NGT_VC_CNT_T6	0.685528730	0.60525040	0.435580611	0.427648123	0.59737556	0.549351316	0.411891410	0.380773186	
U_OB_NGT_VC_CC_CNT_T6	0.680001611	0.64696141	0.449505128	0.370705231	0.60373927	0.583777827	0.442909495	0.328547229	
U_OB_NGT_VC_DUR_T6	0.401044863	0.31871533	0.567716589	0.117120541	0.35398341	0.308937268	0.489690339	0.087064535	
U_OB_NGT_VC_REV_T6	0.334557746	0.24952580	0.118625456	0.923589933	0.26624348	0.234662272	0.125152883	0.898324574	

Covariance Matrix of Variables

- New dataset formed and ready for elbow analysis (variables: 143)
- Standardization was done with min max Normalization (0-1)

INTRODUCTION- What is clustering?

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset share some common trait.

K-Means Clustering is chosen in our problem to affectively identify the segments/clusters for checking the behavior of customers as well as visualization of segments.

DICE ANALYTICS

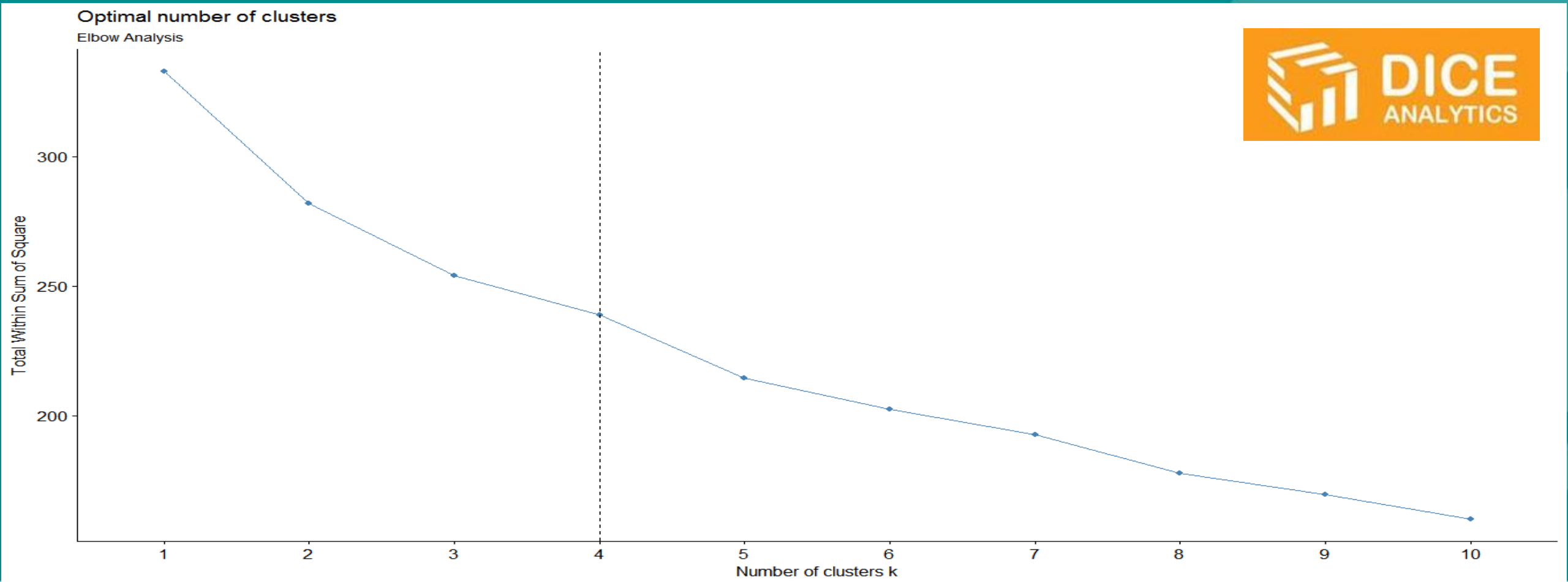
K-MEANS CLUSTERING

- K-means clustering is used which is an algorithm to classify or to group the objects based on attributes/features into K number of groups.
- K is a positive integer number.
- The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid.

DICE ANALYTICS

Elbow Analysis for optimal number of K

Elbow method is followed to validate the number of clusters. Elbow method ran k-means clustering on the data for a range of values of *k* (from 1 to 10), and for each value of *k* calculated the sum of squared errors (SSE).



K-Means Clustering Algorithm

- **Step 1:** Begin with a decision on the value of k = 4 (number of clusters from Elbow Analysis) .

- **Step 2**: Placed an initial partition that classifies the data into k clusters. Assigned the training samples by the following steps:

  1. Took the first k training samples as single - element clusters

  2. Assigned each of the remaining (N-k) training samples to the cluster with the nearest centroid. After each assignment, recomputed the centroid of the gaining cluster.

DICE ANALYTICS

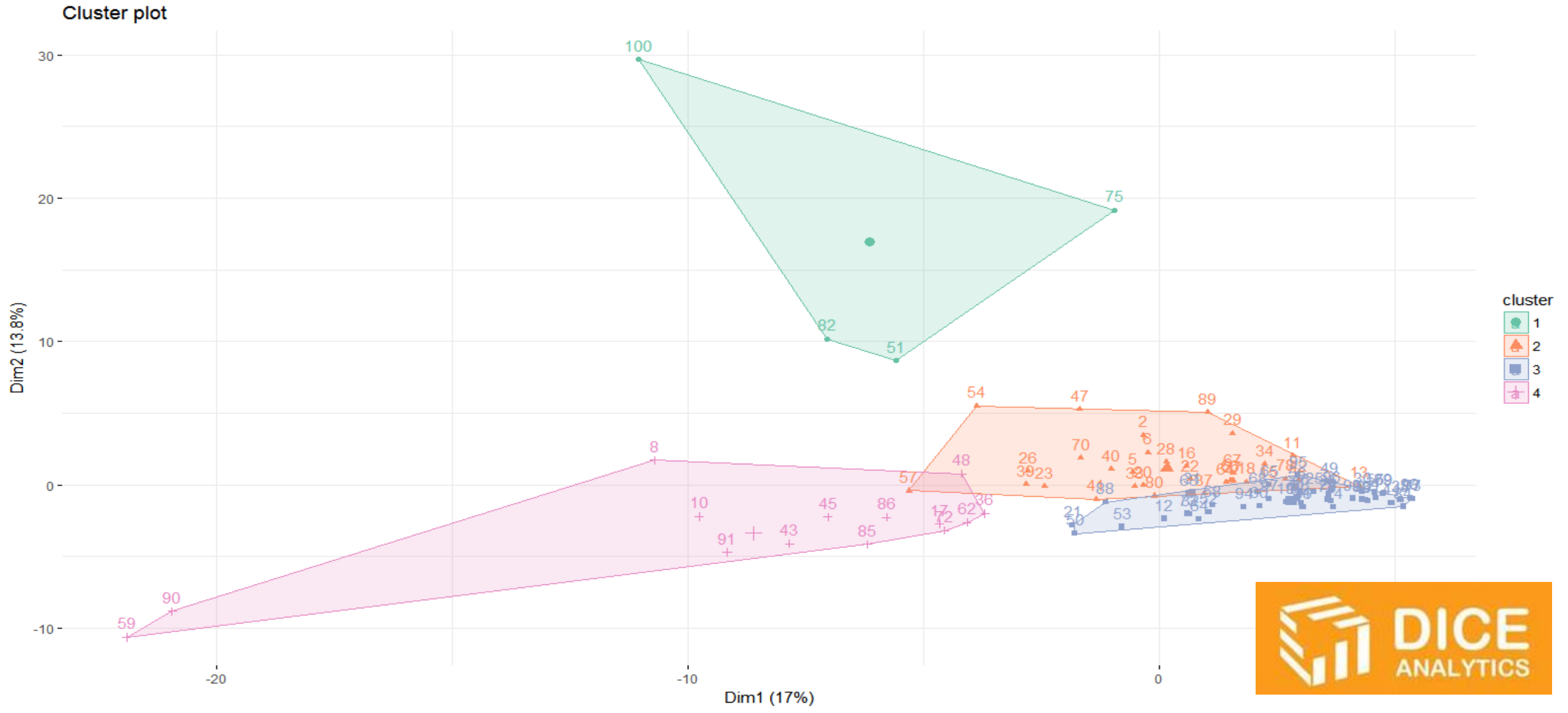
K-Means Clustering Algorithm contd..

- **Step 3:** Took each sample in sequence and compute its distance from the centroid of each of the clusters. If a sample is not currently in the cluster with the closest centroid, switched this sample to that cluster and update the centroid of the cluster gaining the new sample and the cluster losing the sample.

- **Step 4 .** Repeated step 3 until convergence is achieved, that is until a pass through the training sample causes no new assignments.

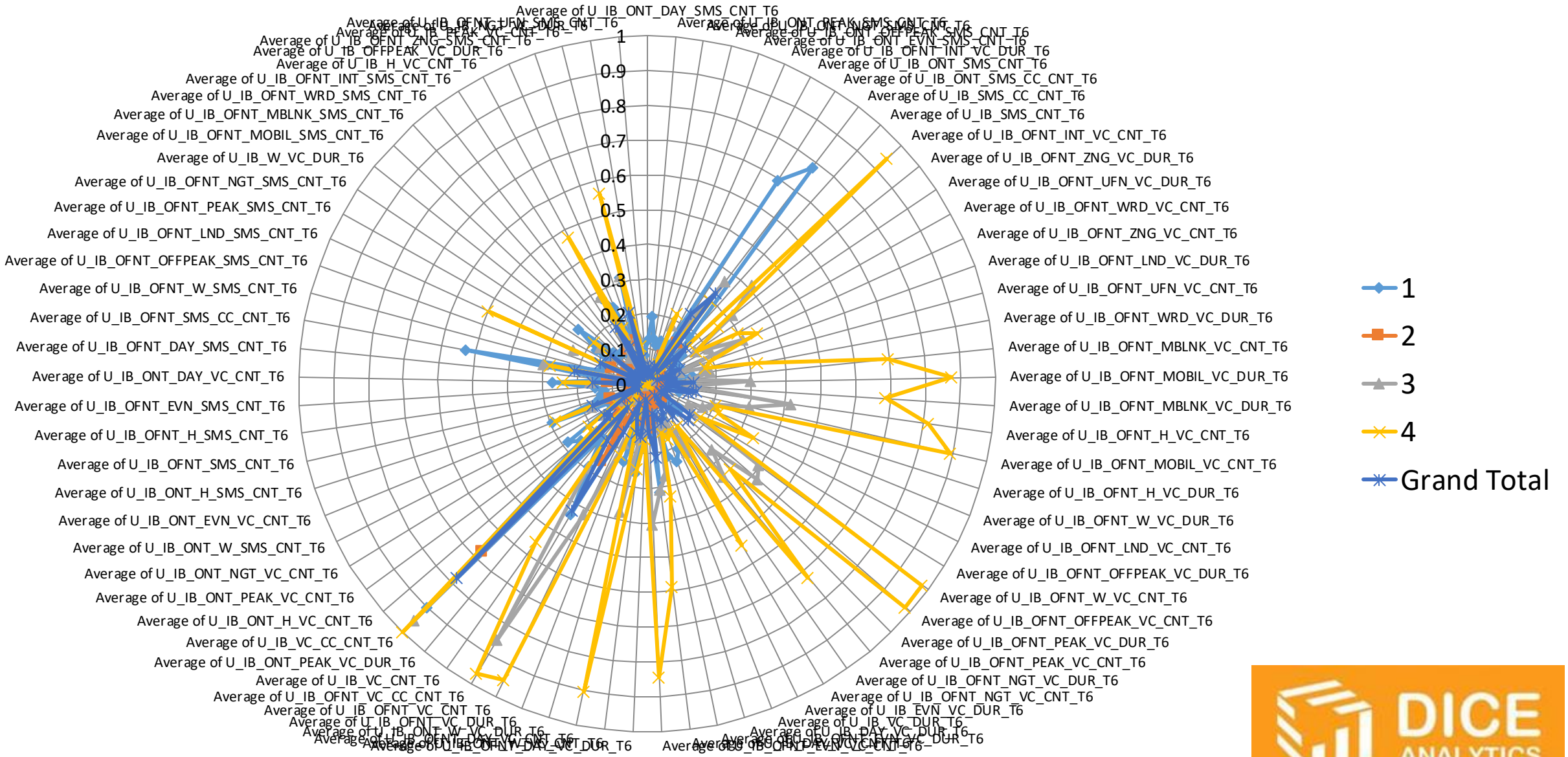
- After implementation of K-Means on the dataset, we get this cluster representation

Clusters from K-means Clustering Algorithm on given Data



DICE ANALYTICS

Segments of Users



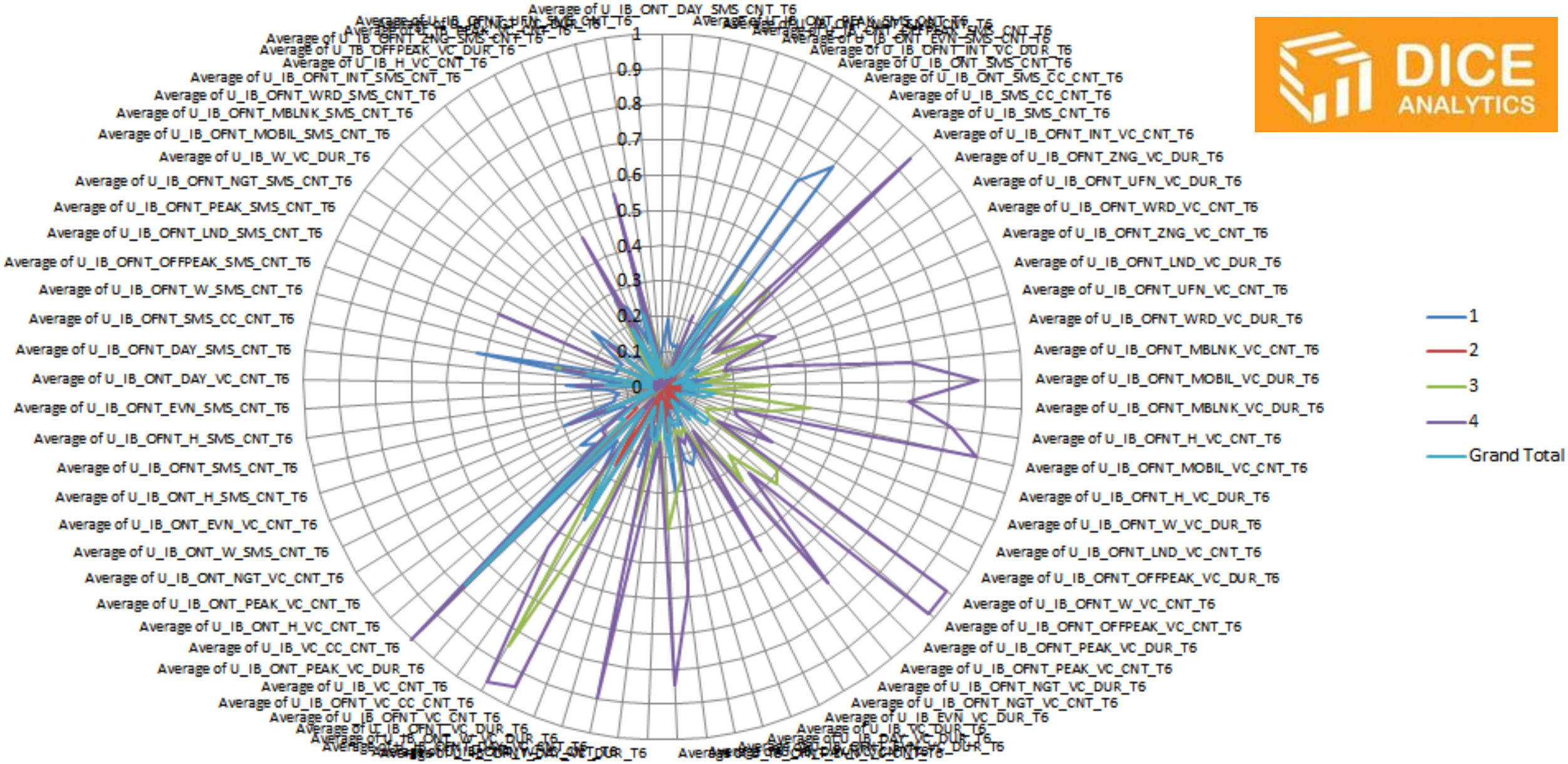
DICE ANALYTICS

Segments Details

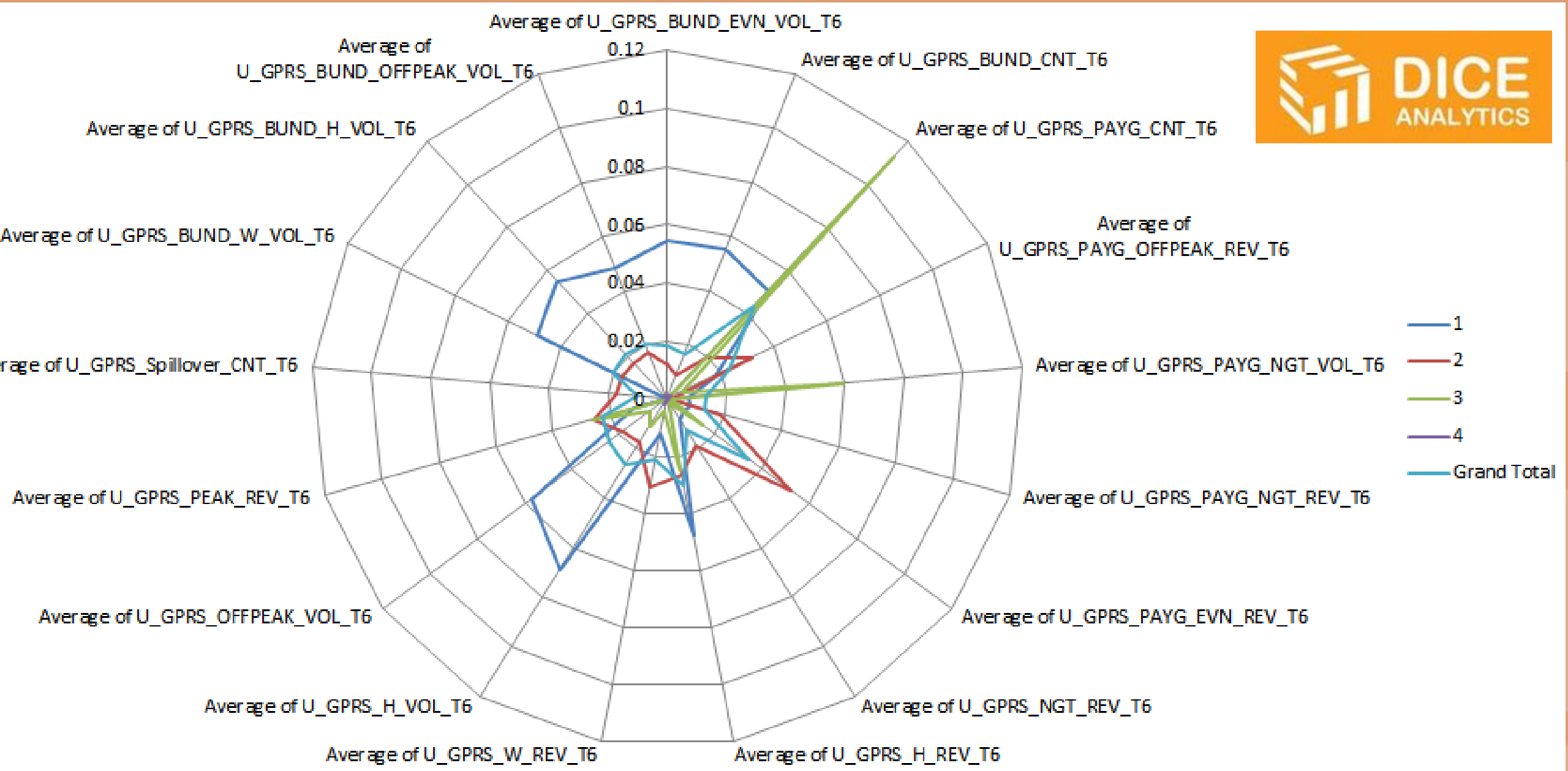
In order to read and understand the clusters, all features are separated from each other. Mainly three segments were found which were creating impact.

1. Out Bound Features
2. In Bound Features
3. Internet Features

Segment for Internet Users

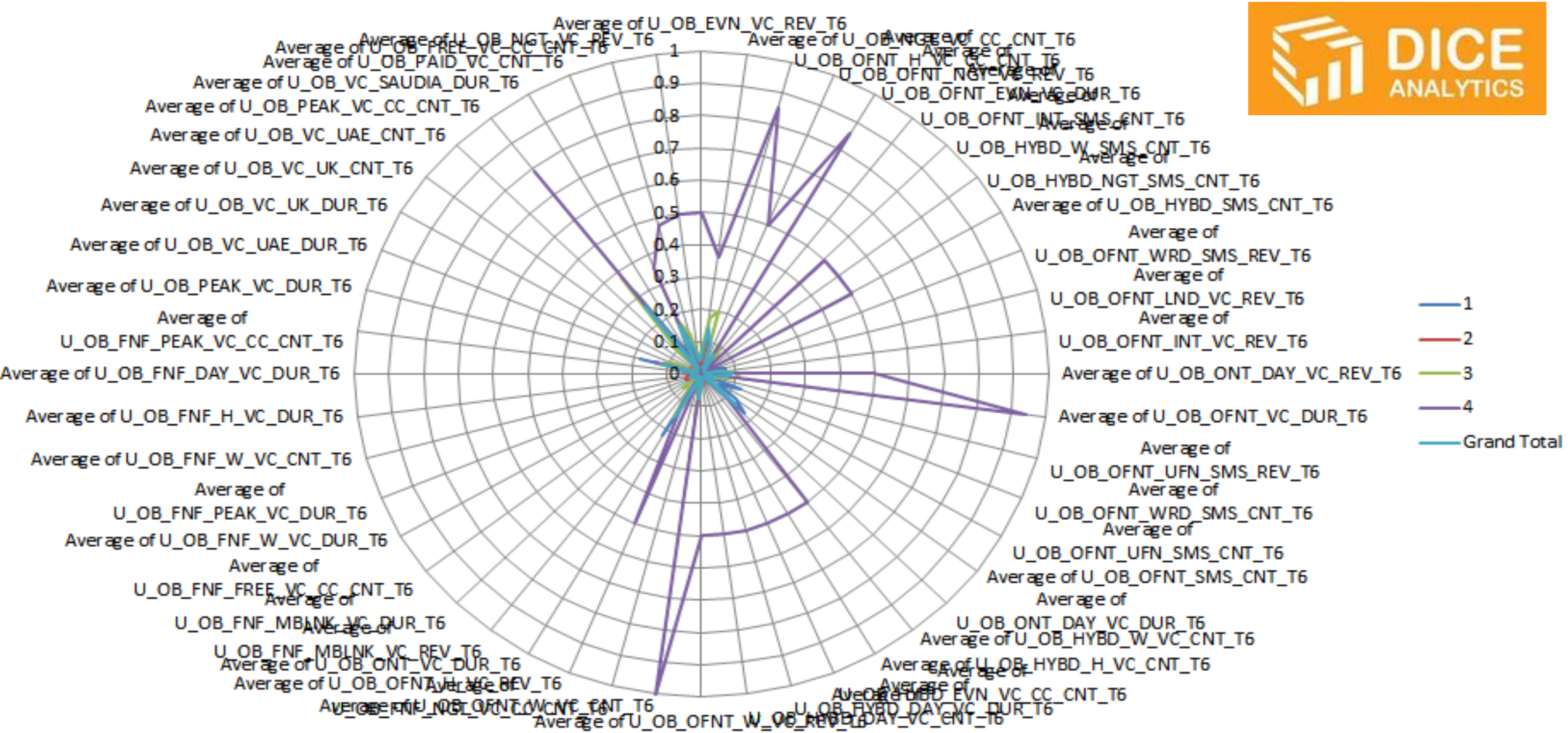


Visualization for Inbound Features



DICE ANALYTICS

Visualization for Internet Features



Visualization for Out bound Features

Description of Segments

Out Bound Features

Cluster 1
Outbound usage of voice calls in evening is low
Outbound usage of voice calls in night is low
Outbound duration of off net voice calls in evening is low
Outbound usage of voice calls during peak times is low
Outbound usage of on-net voice calls duration is moderate

Cluster 2
Outbound usage of voice calling circle count at peak hours is low
Outbound usage of voice calls in night is low
Outbound duration of on net voice calls in day is low
Outbound usage of off net voice calls in week days is low
Outbound usage of FNF voice calls duration on week days is low

Cluster 3
Outbound usage of off net helpline voice calls counts is low
Outbound usage of off net call counts on week days is low
Outbound usage of voice calls to UAE is low
Outbound usage of voice calls during peak times is moderate
Outbound usage of voice call on day count and it's revenue is low

Cluster 4
Outbound usage of off net helpline voice calls counts is high
Outbound usage of voice calls in evening is high
Outbound usage of off net week days voice call counts is high
Outbound usage of voice calls during peak times is high
Outbound usage of off net evening voice calls duration is high

Heavy Off net Outbound Users

Internet Features

Cluster 1
Usage of internet bundle during evening is low
Usage of internet helpline and revenue is low
Usage of internet in peak time and revenue is low
Usage of internet helpline volume consumption is low
Usage of internet on week days is low

Cluster 2
Usage of internet bundle during evening is low
Usage of internet for Pay as you go during off peak is low
Usage of internet for Pay as you go during evening is low
Usage of internet for Pay as you go during night is low
Usage of internet bundle for Pay as you go during day is low
Usage of internet helpline on week days is low

Cluster 3
Usage of internet for Pay as you go is low
Usage of internet for Pay as you go in night is low
Usage of internet for Pay as you go in evening is low
Usage of internet helpline revenue is low
Usage of internet helpline during peak time revenue is low

Cluster 4
Usage of internet bundle during evening is too low
Usage of internet for Pay as you go during off peak revenue is low
Usage of internet helpline volume is vey low
Usage of internet helpline volume is vey low
Usage of internet during night is too low

Limited Internet Users

In Bound Features

Cluster 1
Inbound usage of on net peak SMS is low
Inbound usage of SMS count is high
Inbound usage of off net Warid call count is low
Inbound usage of off net voice call count is high
Inbound usage of off net SMS count is high

Cluster 2
Inbound usage of SMS count is low
Inbound usage of Day Voice Call count is low
Inbound usage of on net Day Voice Call count is low
Inbound usage of peak voice call count is low
Inbound usage of on net Day Voice Call count is low

Cluster 3
Inbound usage of on net Day Voice Call count is high
Inbound usage of off net Voice Call count is high
Inbound usage of off net Helpline Voice Call count is high
Inbound usage of off net Evening Voice Call count is high
Inbound usage of off net SMS count is high

Cluster 4
Inbound usage of off net Voice Call count is very high
Inbound usage of off net off peak Voice Call count is very high
Inbound duration of off net voice calls in evening is very high
Inbound usage of off net Voice Call count is very high
Inbound usage of off net Helpline Voice Call count is high

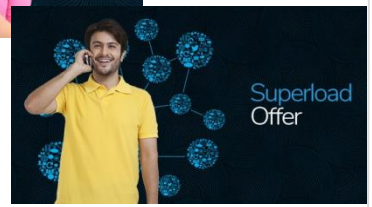
High on net and off net Inbound Users

telenor Products Portfolio

Call

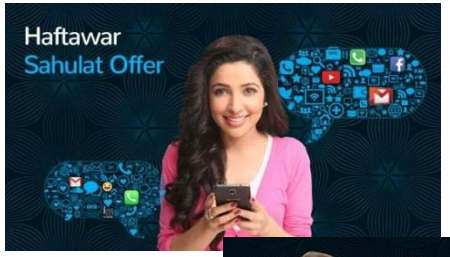
Off-net

On-net



SMS

Internet



Campaigns for Segments

Out Bound Users

Easy Card for Internet

- Rational
- Purpose for suggesting easy card is that they can be loyal customers.

Internet Users

Easy Card for Internet

Weekly Bundle

- Purpose for suggesting easy card/weekly bundle for internet users is to encourage them to use start utilizing high speed 3G/4G internet.

In Bound Users

TELENOR GOOD TIME OFFER

Purpose for suggesting TELENOR GOOD TIME OFFER offer will motivate inbound customers to use the on net minutes for their regular use.

THANK YOU!
ANY QUESTIONS?