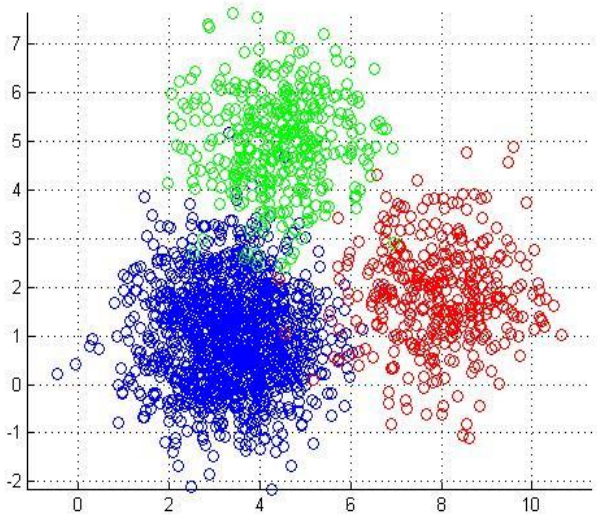

CLUSTERING ANALYSIS

2020

ABSTRACT

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

Wande B. Adeyeye

Project

PROJECT REPORT CLUSTERING ANALYSIS

ONLINE SHOPPER'S INTENTION

BY:

WANDE B. ADEYEYE

E-MAIL:

ADEYEYEWANDE@GMAIL.COM

About This Project:

In this project, we will analyze the Customer's Intentions based on the transactions made online in a one-year duration.

Goals:

The goal of this project is to Cluster the groups using the Bounce Rates to identify which customers are more likely to be the interested (More likely to make a transaction) customers and come to a conclusion to want we can do to make more people purchase.

About the Data:

The Data from Kaggle: here and here

Source:

C. Okan Sakar

Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Bahcesehir University, 34349 Besiktas, Istanbul, Turkey

Yomi Kastro

Inveon Information Technologies Consultancy and Trade, 34335 Istanbul, Turkey*

Dataset and Features Information:

Dataset:

The dataset consists of feature vectors belonging to 12,330 sessions. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period.

Features:

Administrative	Administrative value
Administrative_Duration	The total time spent by visitor on administrative page categories.
Informational	Informational value
Informational_Duration	The total time spent by visitor on Informational page categories.
ProductRelated	Product Related value
ProductRelated_Duration	The total time spent by visitor on Product Related page categories.
BounceRates	Percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session
ExitRates	The percentage that were the last in the session
PageValues	Feature represents the average value for a web page that a user visited before completing an e-commerce transaction
SpecialDay	Indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction. For example, for Valentina's day, this value takes a nonzero value between February 2 and February 12, zero, before and after this date unless it is close to another special day, and its maximum value of 1 on February 8
Month	Month of the year
OperatingSystems	Different types of operating systems used to visit the website
Browser	Different types of browser used to visit the website
Region	Region of the user
TrafficType	Different types of operating systems, browser, region and traffic type used to visit the website
VisitorType	Whether the customer is a returning or new visitor
Weekend	A Boolean value indicating whether the date of the visit is weekend
Revenue	class whether it can make a revenue or not

Cluster Analysis:

The Elbow Method to Find out the Maximum no. of Optimal Clusters

Compute clustering algorithm (e.g., k-means clustering) for different values of k. For instance, by varying k from 1 to 10 clusters.

For each K, calculate the total within-cluster sum of squared error (SSE).

Plot the curve of SSE according to the number of clusters K.

The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters.

Conclusion:

- From the observations, we realized that Customers who actually made the transactions are less likely to Exit or Bounce from the website and the pagevalues is usually higher than the Customers who did not make the purchase
- Based on the Clustering graphs and the tables, it is reasonable to say that the customers who have lower Exit/Bounce Rates with longer stays are more likely to generate profits.
- In order to reduce Exit/Bounce Rates we can introduce some benefits for the costumers such as coupon/promo code or loyalty cards, in doing so more consumers will be encourage to make a purchase and some roamer will visit often and eventually become a regular costumer which will generate more profits.