PROJECT REPORT

CUSTOMER SEGMENTATION

COURSE:FOUNDATION OF DATA SCIENCE

MANVENDRA SHARMA : AM.EN.U4CSE19034
MAYANK KUMAR SHAKYA : AM.EN.U4CSE19035
SOURAV CHINDARMONY : AM.EN.U4CSE19053
CHLLA VENKATA SAI KRISHNA:AM.EN.U4CSE19064

TABLE OF CONTENTS

ABSTRACT:

In this project, we will implement customer segmentation in R. Whenever you need to find your best customer, customer segmentation is the ideal methodology.

In this project, DataFlair will provide us the background of customer segmentation. Then we will explore the data upon which we will be building our segmentation model. Also, in this data science project, we will see the descriptive analysis of our data and then implement several versions of the K-means algorithm.

INTRODUCTION

Customer Segmentation is one the most important applications of unsupervised learning. Using clustering techniques, companies can identify the several segments of customers allowing them to target the potential user base. In this project, we will make use of ***K-means clustering*** which is the essential algorithm for clustering unlabeled dataset.

What is Customer Segmentation?

Customer Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits.

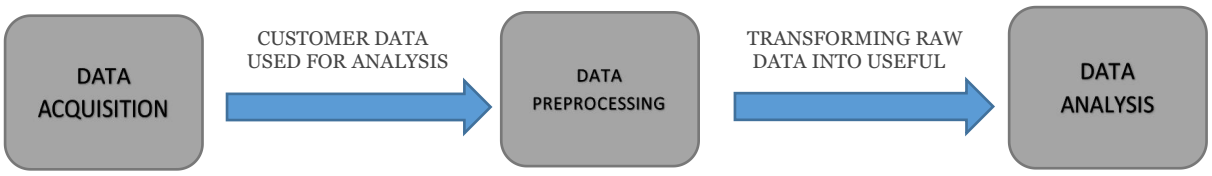
Companies that deploy customer segmentation are under the notion that every customer has different requirements and require a specific marketing effort to address them appropriately. Companies aim to gain a deeper approach of the customer they are targeting. Therefore, their aim has to be specific and should be tailored to address the requirements of each and every individual customer. Furthermore, through the data collected, companies can gain a deeper understanding of customer preferences as well as the requirements for discovering valuable segments that would reap them maximum profit. This way, they can strategize their marketing techniques more efficiently and minimize the possibility of risk to their investment.

The technique of customer segmentation is dependent on several key differentiators that divide customers into groups to be targeted. Data related to demographics, geography, economic status as well as behavioral patterns play a crucial role in determining the company direction towards addressing the various segments.

METHOD

In the first step, we will perform data exploration. We will import the essential packages required for this role and then read our data. Finally, we will go through the input data to gain necessary insights about it.

3.1 FLOWCHART



3.2 DATASET

The dataset contains 4 columns (Gender,Age,Annual Income,Spending Score)

3.3 ABOUT R STUDIO

RStudio is an Integrated Development Environment (IDE) for R. It consists of a console, syntax highlighting editor that supports direct code execution, as well as tools for plotting, history, and debugging and workspace management.

3.4 LIBRARIES USED

Specifically, we made use of a clustering algorithm called K-means clustering.

gridExtra-Provides a number of user-level functions to work with "grid" graphics, notably to arrange multiple grid-based plots on a page, and draw tables.
grid-The grid package in R implements the primitive graphical functions that underliethe ggplot2 plotting system.
NbClust- NbClust package provides 30 indices for determining the number of clusters and proposes to user the best clustering scheme from the different results obtained by varying all combinations of number of clusters, distance measures, and clustering methods.
Factoextra-factoextra is an R package making easy to *extract* and *visualize* the output of exploratory multivariate data analyses.
Ggplot2-ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

3.5 ALGORITHM

1. In the first step, we will perform data exploration. We will import the essential packages required for this role and then read our data. Finally, we will go through the input data to gain necessary insights about it.

2. We will create a barplot and a piechart to show the gender distribution across our customer_data dataset.

3. Plot a histogram to view the distribution to plot the frequency of customer ages.

4. We will create visualizations to analyze the annual income of the customers.

5. Analyse the spending score of customer.

6. we made use of a clustering algorithm called K-means clustering.

7. Collecting insights from all visualisation