Contents

**Contents ………………………………………………………………..……… (i)**

**List of Figures ..................................................................................................... (iv)**

**List of Tables ...................................................................................................... (v)**

**Abstract ……………………………………...………………………………… 01**

**Keywords .............................................................................................................. 03**

**1. Introduction ………………………………………………….……………... 04**

**2. Literature Review**

**2.1** Customer Segmentation ………………………………………..……………………..... 07

**2.2** Clustering and K-means Algorithm .......................................................................…..... 08

**2.3** RFM Model ……………………………………………………....………………...….. 09

**2.4** Elbow Criterion ……………………………………….………………...…………...… 10

**3. Preliminaries**

3.1 What is RFM Model ? ................................................................................................. 12

3.2 Algorithm for Customer Segmentation …...………………………………………….. 13

3.3 Flowchart and Broader steps for K-means ………………........................................ 14

3.4 Algorithm for Elbow Method .................................................................................... 16

**4. Proposed scheme ............................................................................................ 17**

4.1 Research Methodology Flowchart.................................................................... 18

4.2.1 DFD Level Zero ...................................................................................................... 19

4.2.2. DFD Level One ........................................................................................................ 20

4.3 ER Diagram ................................................................................................................... 21

4.4 Selection of Clustering Algorithm .....…………………………….….……………...… 22

4.5 Proposed Methodology ……...……...…………………………………………………. 23

4.6 Mathematical Model ………………………………………………………………...… 25

4.7 Deciding on the optimum number of Clusters ‘K’ …………………………………….. 26

4.8 Demonstration with a sample dataset …………………………………………………. 26

4.9 Portions of code for our proposed methodology .......................................................... 33

**5. Experimental Results……………………………………………………..…. 41**

5.1 Cluster Orange - Balanced Customer ........................................................................... 41

5.2 Cluster Blue - Pinch-Penny Customers ........................................................................ 41

5.3 Cluster Purple - Normal Customers ............................................................................. 42

5.4 Cluster Red - Spenders ................................................................................................ 42

5.5 Cluster Green - Target Customers ............................................................................... 42

5.6 Outputs ......................................................................................................................... 42

**6. System Requirements**

6.1 Hardwares …………………………………………………………………………...… 51

6.2 Softwares …………………………………………………………………………….... 52

**7. Future Scope …………………………………………………………….…. 53**

**8. Conclusion ………………………………………………………………..… 54**

**9. References ………………………………………………………………...…. 55**

**List Of Figures**

3.3.1. K-means Flowchart ........................................................................................................... 14

3.3.2. Broader Steps of K-Means Algorithm .............................................................................. 15

3.2. Research Methodology Flowchart .................................................................................... 18

4.2.1. DFD Level Zero ................................................................................................................ 19

4.2.2. DFD Level One ................................................................................................................. 20

4.3. E.R. Diagram .................................................................................................................... 21

4.1.1. Original DataSet ................................................................................................................ 22

4.1.2. After Clustering ............................................................................................................... 22

4.5. Steps of Proposed Methodology ....................................................................................... 23

4.7. Elbow Method ................................................................................................................... 26

5.1. Cluster Analysis ................................................................................................................ 41

5.6.1. Recency Distribution Plot ................................................................................................. 45

5.6.2. Frequency Distribution Plot .............................................................................................. 45

5.6.3. Monetary Distribution Plot ............................................................................................... 46

5.6.4. Data Distribution after Data Normalization for Recency ................................................. 47

5.6.5. Data Distribution after Data Normalization for Frequency .............................................. 48

5.6.6. Data Distribution after Data Normalization for Monetary ............................................... 48

5.6.7. Screen Plot of Elbow Method ........................................................................................... 48

5.6.8. Frequency vs. Recency Graph .......................................................................................... 49

**List Of Tables**

5.6.1. Few Entries from our DataSet .......................................................................................... 42

5.6.2. Selecting the Country with maximum number of customers ........................................... 43

5.6.3. Figuring out which attributes have how many number of null entries ............................ 44

5.6.4. Adding an extra column for storing Total Amount Spent ............................................... 44

5.6.5. RFM Scores ..................................................................................................................... 44

5.6.6. Descriptive Statistics of Recency ..................................................................................... 44

5.6.7. Descriptive Statistics of Frequency ................................................................................. 45

5.6.8. Descriptive Statistics of Monetary ................................................................................... 46

5.6.9. RFM Segment Scores ...................................................................................................... 46

5.6.10. R+F+M = RFM Scores ................................................................................................... 47

5.6.11. Determining the Loyalty Level according to RFM Scores ............................................. 47

5.6.12. K-means Clustering Model ............................................................................................. 49

5.6.13. Final Grouping of Customers according to their Loyalty Level ..................................... 50