

Department of Computer Science and Engineering

High value customers identification for an E-Commerce company

Date: 28 May 2021

Mohana Naga S Vyshnavi Medisetty : 17WH1A0567

Shanmitha Paruchuri : 17WH1A05B4

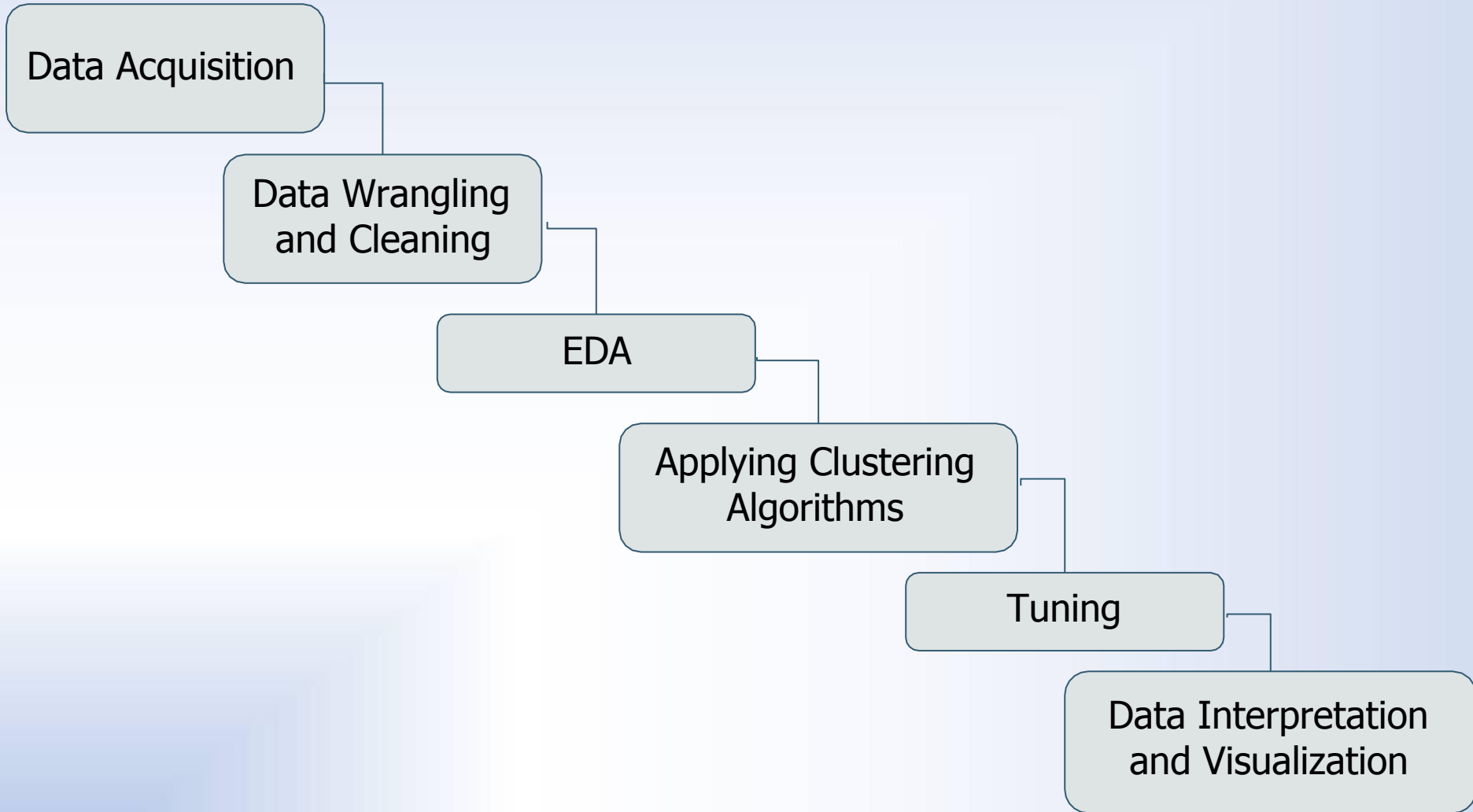
Bhadavath Vandana : 17WH1A0562

Internal Guide : Dr. Ganti Naga Sathish
Designation : Professor

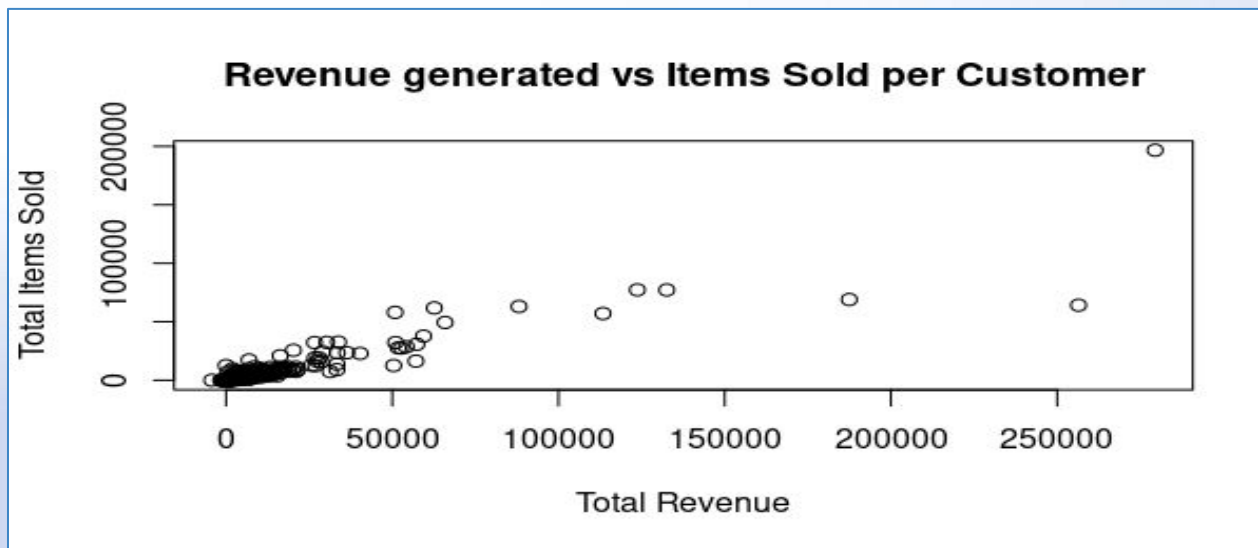
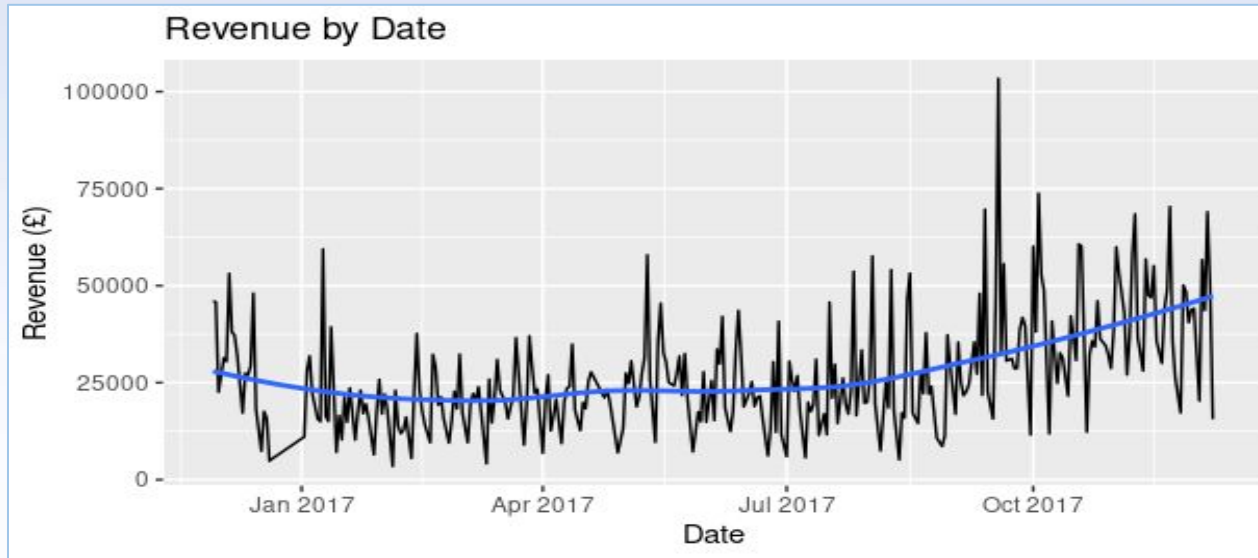
Abstract

The ability of any business to understand the needs of each of its customers will provide greater customer support in providing targeted customer services and developing customized customer service plans. This understanding is possible through *structured customer service*. Each segment has customers who share the same market features. Big data ideas and machine learning have promoted greater acceptance of automated customer segmentation approaches. Clustering algorithms like k-means and Hierarchical clustering can be used for this purpose. The goal is to better understand customers' spending and ordering habits with the following features: Number of products ordered, average return rate and total spending. In this project, we aimed at identifying the right number of customer segments, providing the number of customers who are highly valued and identifying the clustering algorithm that gives maximum accuracy.

Design

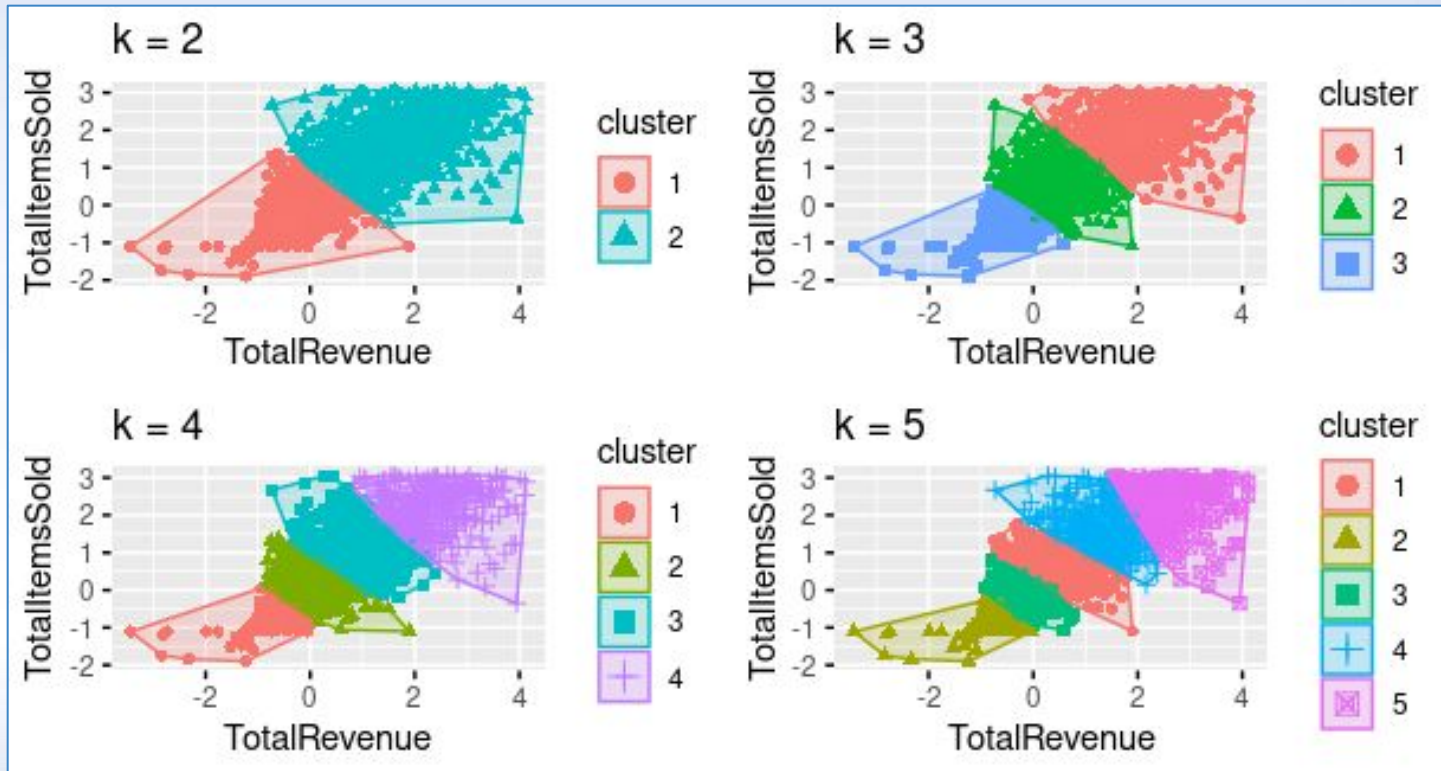


Data Exploration



Result

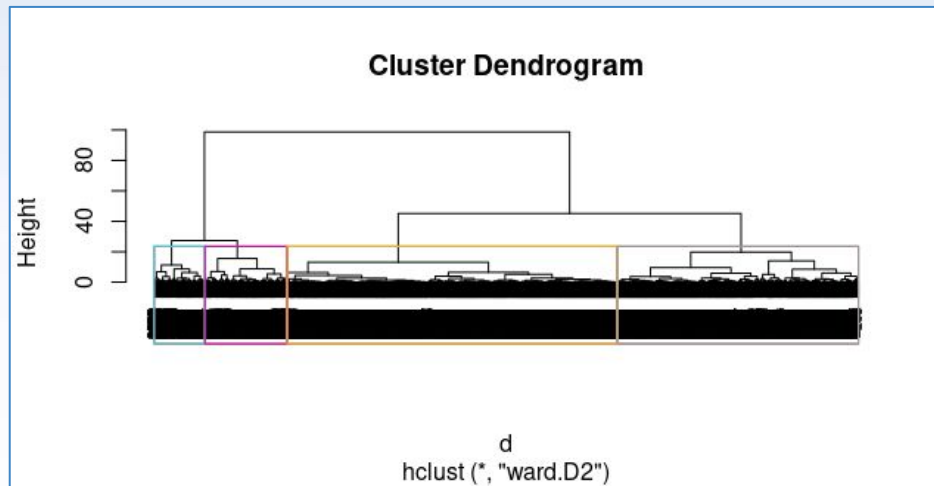
K-MEANS CLUSTERING



By using Elbow method which is used to determine optimal number of clusters, we concluded that in case of K-means, the optimal number of clusters for the ecommerce data is **3**

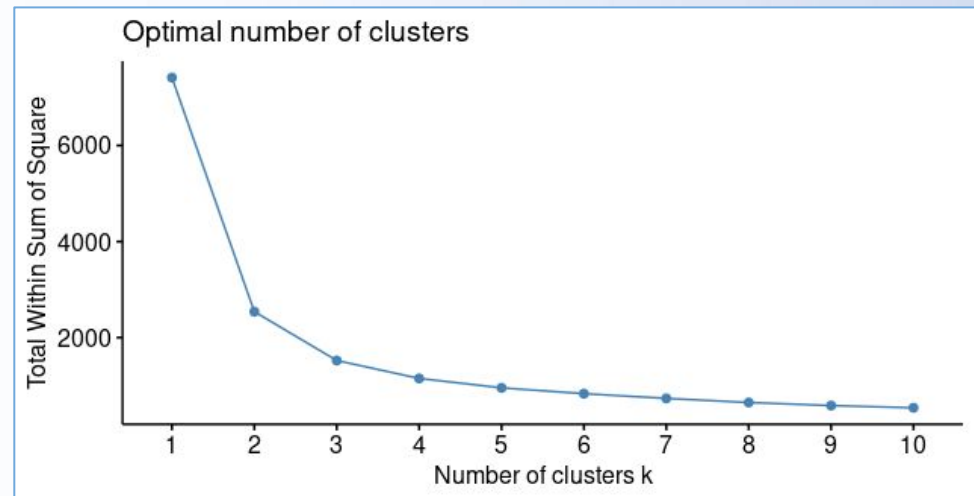
Result

HIERARCHICAL AGGLOMERATIVE CLUSTERING



Agglomerative coefficient, which measures the amount of clustering structure found (values closer to 1 suggest strong clustering structure) can be used to identify stronger clustering structures. The Ward's method is identified as the strongest clustering structure of the four methods assessed with agglomerative coefficient **0.9997**

Similar to how we determined optimal clusters with k-means clustering, we executed approaches for hierarchical clustering. We concluded that the optimal number of clusters for the ecommerce data is 3. Therefore, there are 700 customers who are highly valued when identified using Hierarchical clustering algorithm.



Conclusion



We conclude that the optimal number of clusters for the ecommerce data is **3** with **527** customers identified as highly valued obtained.

References

- [1] Blanchard, Tommy. Bhatnagar, Pranshu. Behera,Trash. (2019). Marketing Analytics Scientific Data: Achieve your marketing objectives with Python's data analytics capabilities. S.I: Packt printing is limited
- [2] Griva, A., Bardaki, C., Pramadari, K.,Papakiriakopoulos, D. (2018). Sales business analysis: Customer categories use market basket data. Systems Expert Systems, 100, 1-16.
- [3] Hong, T., Kim, E. (2011). It separates consumers from online stores based on factors that affect the customer's intention to purchase. Expert System Applications, 39 (2), 2127-2131.
- [4] Hwang, Y. H. (2019). Hands-on Advertising Science Data: Develop your machine learning marketing strategies... using python and r. S.I:Packt printing is limited
- [5] Puwanenthiren Premkanth, -Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC.|| Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1.
- [6] Puwanenthiren Premkanth,-Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC.|| Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1.
- [7] By Jerry W Thomas. 2007. Accessed at:www.decisionanalyst.com on July 12, 2015.
- [8] Jianfu, L., Jianshuang L., Huaqing H. (2011). A Simple and Accurate Approach to Hierarchical Clustering. Journal of Computational Information Systems, 7(7), 2577--2584.

Thank you