- Explore how clustering can be used to help businesses learn more about their customers.

This is a required discussion and will count toward course completion.

Now that you've had practice utilizing the cluster algorithms in **Colab** (https://mo-pcco.s3.us-east-1.amazonaws.com/MO-PCDS/module4/activity-name-2_starter.zip) , write a short report explaining your main findings using either the hierarchical or k-means clusters. Choose one of the following three functions: product development, advertising, or customer interactions and sales, and explain how your clustering analysis could be applied in this area.

For additional practice, we encourage you to copy the code used in the Colab activity, apply it to a new dataset and run the hierarchical or k-means clustering algorithms.

Be sure to read the statements posted by your peers. Engage with them by responding with thoughtful comments and questions to deepen the discussion.

Suggested Time: 20 minutes

Rubric: Discussion 4.3

Criteria	Exceeds expectations	Meets expectations	Below expectations
Thoughtful and complete response to the question(s)	**4 pts** Fully responds to the question(s), post is supported by connections to the reading and real-life examples, and post makes additional connections to the field of data engineering with novel ideas, critical thinking, or extensive application of how to use the topic in future work.	**3 pt** Fully responds to the question(s), and post is supported by connections to the content or real-life examples.	**0 pts** Partially responds to the question(s), or connections to the content are missing or vague.
Engagement with the learning community	**2 pt** Posts thoughtful questions or novel ideas to multiple peers that generate new ideas and group discussion.	**1.5 pt** Asks questions or posts thoughtful responses to generate a single peer's response.	**0 pts** No responses to peers or posts minimal or vague responses to peers that do not motivate a

			response (e.g., "I agree.").

Search entries or author

Unread

Subscribed

Reply

(https:/

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 21, 2024

Function: Product Development,

Through hierarchical clustering, we identified distinct groupings within our dataset that depict relationships between different product function attributes.

Provided dataset was comprised of four features, representing different product attributes, which varied significantly in scale. On the initial display the observation was that Col A far out-scaled the other three columns, B,C and D. To address this problem we explored preprocessing the data through 3 normalization techniques, Max Scaling, Min-Max, Standardization.

Results:

Max Scaling, in where we divided each column by its maximum value helped standardize each column relative to itself, where before they were all compared to Col A.

In Min-Max scaling by subtracting the minimum value of the feature and then dividing by the range, we shifted all values so that the minimum value becomes 0.

Both these techniques compressed the range but did not center the data around zero as the Standardization technique did.

In standardization, z-score normalization, where we subtracted the mean of each column from each of its respective data points and dividing the result by the standard deviation of the column, we effectively scaled dataset to have a mean of 0 and a standard deviation of 1. By bringing all the columns to the same magnitude, normalizing the distribution around zero, we are able to conduct a like-for-like comparison.

Because this was a relatively small dataset and we do not have prior knowledge of the cluster group size we went on to plot dendrograms for all three normalization techniques using hierarchical clustering. The observation was that the Z-Score Standardization observed the most significant dissimilarity between all three clusters.

Evaluation metrics were applied to the other two normalization techniques with very similar and mixed signals for the three normalization techniques (Table 1). Hierarchical cluster dendograms were then plotted for all three normalized datasets (Figures 1-3). A higher separation was observed when applying the Z-score (Figure 3), determined by the longer distance of between group 1 and the other groups.

For Z-Score Standardization the evaluation results were:

- The Silhouette Score of 0.414 indicates moderate cluster definition but room for improvement.
- The Calinski-Harabasz Index at 6.809 suggests clusters are not highly dense nor well-separated.
- A Davies-Bouldin Index of 0.268 points towards reasonable separation.

The evaluation metrics for Z-score Standardization suggest that the clusters are not highly distinct and pointing out that there is room for improvement. We can explore more robust scaling methods or if desired, remove outliers as future steps.

Findings Summary:

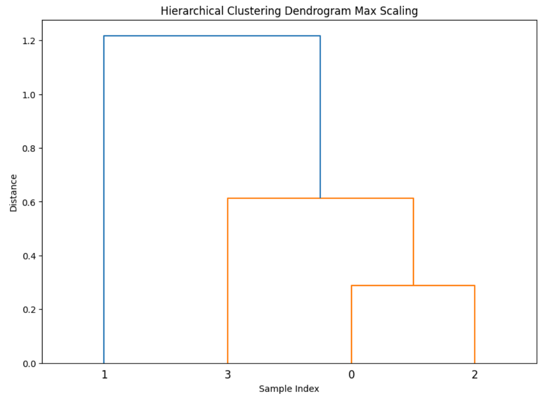
More detailed insights were found thanks to this Hierarchical cluster dendrogram (Figure 3) depicting associations between products and segregation. For example when observing similarities, some products grouped together at a lower threshold, 2 on the y axis, depicted by the orange edges in the dendrogram. These groupings suggest similarities which can be used to categorize product types and appeal to similar customer segments.

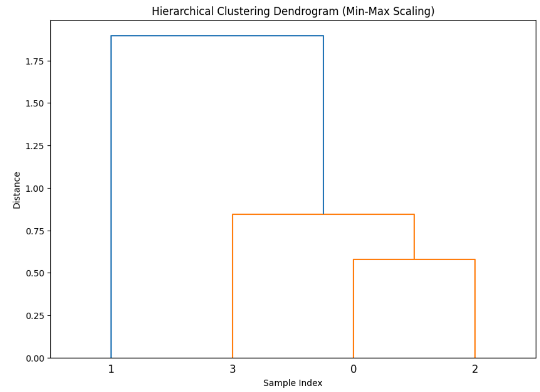
In terms of segregation, as the threshold increased, more diverse products merged into a more separated cluster. The similarities here can be used to cater to more niche markets.

Given these preliminary insights, further analysis might involve refining the scaling process, possibly removing outliers, and exploring different clustering configurations. This iterative approach could uncover more definitive patterns that would support strategic decisions in product development, enhancing the appeal and differentiation of our offerings in the marketplace.

Figure 1 **Figure 2**

Figure 3





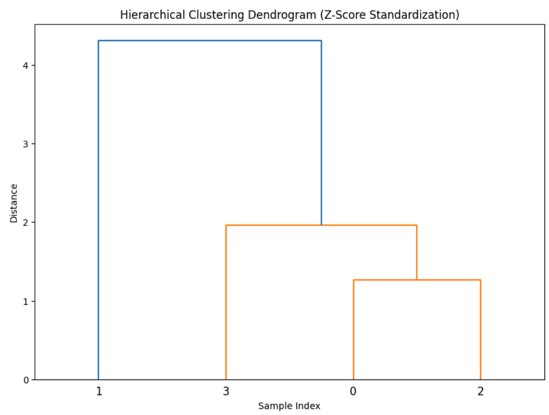


Table 1

Results for StandardScaler: Silhouette Score: 0.4138058710265374 Calinski-Harabasz Index: 6.809100386860259 Davies-Bouldin Index: 0.2682114363825345

Results for MinMaxScaler: Silhouette Score: 0.41287221612605 Calinski-Harabasz Index: 6.831911422617602 Davies-Bouldin Index: 0.26806371938807905

Results for MaxAbsScaler: Silhouette Score: 0.41363670483876525 Calinski-Harabasz Index: 6.456120618122842 Davies-Bouldin Index: 0.27186437697995697

Edited by **Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)** on Apr 24 at 2:49am

Reply (2 likes)

(http

Manjari Vellanki (https://classroom.emeritus.org/courses/9054/users/231480)

Apr 23, 2024

Hi Roy-

Thanks for explaining in detail and providing additional details with dendrograms.

Reply (1 like)

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 23, 2024

Thanks Manjari. It was also an explanation to myself :)

Reply

(http

Diego Milanes (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/228518)

Apr 23, 2024

Hi Roy

Excellent post! Do you know how the results compare between the hierarchical clustering wrt those from k-means? Which are the pros and cons of each of those methods for this special case?

thank you again for the nice explanation!

Diego

Reply

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 23, 2024

Hi Diego,

Thanks! For 4 columns with the provided data, with not much information on the features, I think HC was the optimal choice for initial data exploration. The elbow method suggested 2 clusters for this dataset that would be used for K means and I saw HC results as providing a more detailed and informative story with more depth on the relationships, similarities, distances, and some insight on the separation observed with group 3. K means would be a subsequent analysis to consider after getting some sense of the relationship between features as I alluded to for future steps, but for exploratory in this case I saw HC as better fit with a bit more information to provided vs clustering in 2 groups. Thank you for your comments!

Reply

(http

[Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)](https://classroom.emeritus.org/courses/9054/users/233864)

Apr 24, 2024

Thanks for the great example, it helped me understand the effect normalization has on clustering and the decision process.

Invaluable.

Reply (1 like)

(https:/

[Roman Jazmin (https://classroom.emeritus.org/courses/9054/users/225803)](https://classroom.emeritus.org/courses/9054/users/225803)

Apr 22, 2024

As an advertising agent I would use the K-Means clustering method since I would already know and factor in, for a given population sample, demographic factors like (1) gender, (2) age group, (3) location, (4) income, and (5) education. Through these factors, I can cluster people to determine their spending habits on what products they buy most and how much they are willing to pay for any products. From this analysis my company can create customized ads that will influence them to buy whatever we show them, or we can generate a marketing report and sell it to other interested parties.

Reply (1 like)

(http

[Lee Lanzafame (https://classroom.emeritus.org/courses/9054/users/231975)](https://classroom.emeritus.org/courses/9054/users/231975)

Apr 23, 2024

good example, i wonder where you could get this demographic info from, in Australia we have something called the ABS that has census data updated every 5 years, i wonder if there is something more current and localised.

Reply

(http

[Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)](https://classroom.emeritus.org/courses/9054/users/229552)

Apr 24, 2024

Hi Roman,

Good example. I am just curious to how you will determine how much they are willing to pay for products? Is this feature being included in the K-Means cluster or in an analysis afterwards.

Reply

(https://

Manjari Vellanki (https://classroom.emeritus.org/courses/9054/users/231480)

Apr 22, 2024

Clustering analysis is a useful procedure to identify patterns and relationships within complex datasets. It involves in grouping the data points into subgroups called clusters by using algorithms. Though the categorical/classification variables are available in data, which only classifies the data based on certain criteria (like age >50 or age<50). Clustering aids in grouping datapoints based on their similarities and differences allow users to gain insights into underlying structure of data.

Why customer segmentation is important? To derive business decisions based on historic data, a part from data collection, an in depth analysis is essential and Clustering analysis is one of the functional procedure by segmenting customers based on behavior or characteristics by identifying the similarities or patterns from datapoints. These characteristics or behaviors can enclose a wide range of factors, such as demographics (age, gender, income), psychographics (lifestyle, values, interests), purchase history, geographic location, and more. By customer segmentation can gain a clearer understanding of who they are, what they want, and how they interact with business.

To achieve this, it is essential to follow certain steps:

1. Identify and define the objective of this analysis.
2. Gathering relevant data required for analysis is one of the important steps as data might come from various sources like surveys, feedback forms, purchase history and social media interactions and more. Capturing this data and organize these into datasets by assigning to most reasonable variables. Deriving and defining these variables is much important, like deriving categorical or classification variables when required and adding existing dataset.
3. Once the data is collected, it is essential to figure out how to handle the missing data. This we can achieve either by ignoring the missing data as part of analysis or by assigning most reasonable value considering historic data. Data accuracy and consistency is vital.

Normalize the numerical data by performing most suitable Normalization techniques if needed.

4. Analyzing the data that has been collected through identifying the patterns, similarities and differences which leads to creation of customer segments.
5. Identifying the variables that are used to perform clustering analysis.
6. Identifying the most suitable clustering model like Kmeans or Hierarchical clustering techniques and number of clusters by elbow method and silhouette score to gain certain informed business decisions by considering below points:
   - Data Characteristics
   - Cluster Shape
   - Scalability
7. Interpret the results by using visualization tools and Clustering profiling.

Possible outcomes:

- Enhanced customer retention.
- Improved product development.
- Efficient resource allocation.
- Enhanced customer engagement.
- Precision in targeting.

Reply

(http

Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)

Apr 24, 2024

Very structured approach, thank you for sharing.

Reply

(https:/

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Apr 22, 2024

i would choose product development, clustering customers based on behaviors, or their preferences the company would gain knowledge of what kind of products are more favorable and popular.

This can help in enhance the product features to satisfy specific customers cluster, making sure that meet their needs that hence increase sales.

understand the willingness to pay for different customer clusters for better pricing strategies.

Reply (1 like)

(http

Mariana Flores (https://classroom.emeritus.org/courses/9054/users/237198)

Apr 24, 2024

Hi Yossr, so nice to connect on the discussion board again. Great post, cluster analysis across product development to identify features most favorable of the target audience can have a significant positive business impact. I agree with you in adapting the product to fit market demand and audience segments.

Cluster analysis and the real-world application in product development is remarkable - thank you for sharing.

Reply

(https:/

Jignesh Dalal (https://classroom.emeritus.org/courses/9054/users/229173)

Apr 23, 2024

Clustering Analysis Report: Application in Customer Interactions and Sales

Introduction

In the ever-evolving landscape of retail banking, understanding customer behaviors through transaction data can significantly enhance customer interactions and sales strategies. This report outlines the findings from a k-means clustering analysis applied to a synthetic dataset, highlighting how clustering can be leveraged to improve customer service and sales tactics in the banking sector.

Methodology

The synthetic dataset created includes variables such as transaction amount, transaction type (ATM withdrawals, POS transactions, and online purchases), customer age, and the number of transactions per day. K-means clustering was chosen for its efficacy in grouping data based on similar transaction behaviors. The analysis aimed to identify distinct groups or clusters within the transaction data, which could reveal patterns indicative of different customer needs and preferences.

Clustering Process

The clustering process began with the necessary preparation of the synthetic dataset to ensure the accuracy of our analysis. The data was standardized, meaning each variable, such as transaction amounts and frequency of transactions, was adjusted to have zero mean and unit variance. This standardization is crucial as it allows different attributes to contribute equally to the analysis, avoiding any bias towards variables with inherently higher magnitudes or variances. Following the data preparation, the k-means algorithm was implemented to segregate the data into distinct groups. The choice of the number of clusters was carefully considered; ultimately, four clusters were chosen as the optimal number to effectively represent the diverse behaviours in the dataset without complicating the model excessively.

Key Findings

The key findings from our k-means clustering analysis of the synthetic dataset provide a detailed breakdown of customer transaction behaviors across four distinct clusters:

1. **High-Value Transactions (Cluster 1)**: This group primarily consists of older customers engaging in large but infrequent transactions. These patterns suggest potential interest in high-value financial products like investments or large loans, highlighting an opportunity for targeted financial planning services.

2. **Frequent Small Transactions (Cluster 2)**: Dominated by younger customers, this cluster's frequent, small-scale transactions reflect routine daily spending. This behavior indicates a strong potential for products focused on short-term financial management, such as overdraft facilities or high-interest savings accounts aimed at younger demographics.

3. **Mixed Transactions (Cluster 3)**: A diverse group in terms of transaction types and amounts, customers in this cluster demonstrate a wide range of financial needs and behaviors. This variability suggests a versatile approach in marketing, where multiple product types, from credit cards to savings plans, could be cross-sold depending on individual customer profiles identified within the cluster.

4. **Online and Mobile Transactions (Cluster 4)**: Comprising tech-savvy individuals who prefer digital transactions, this cluster represents an ideal target for digital banking solutions, tech-forward banking apps, and online-exclusive offers that cater to their preference for digital interactions.

Application in Customer Interactions and Sales

The insights derived from the k-means clustering can be directly applied to enhance customer interactions and refine sales strategies in several ways:

- **Personalized Marketing and Product Offers:** Understanding the specific transaction behaviors and preferences of different clusters allows the bank to tailor marketing messages and product offers. For instance, customers in Cluster 1 might be targeted with investment opportunities or high-value loan products, whereas those in Cluster 2 could be offered overdraft protections or micro-saving tools.
- **Enhanced Customer Service:** Service approaches can be customized based on the preferred transaction methods of different clusters. For example, Cluster 4 customers would benefit from enhanced mobile banking support and digital service options.
- **Cross-Selling Opportunities:** By analyzing transaction types and frequencies, sales teams can identify cross-selling opportunities tailored to the needs of each cluster. Customers in Cluster 3, who exhibit varied transaction behaviors, might be receptive to a broader range of banking products.

Conclusion

The application of k-means clustering to transaction data offers substantial benefits for banking institutions looking to enhance their customer interactions and sales strategies. By understanding the distinct patterns and preferences of different customer groups, banks can not only increase their sales effectiveness but also significantly improve customer satisfaction and retention.

Reply (1 like)

(http

Ahmad Abu Baker (https://classroom.emeritus.org/courses/9054/users/234460)

Apr 24, 2024

Hello Jignesh,

I thoroughly enjoyed reading your comprehensive clustering analysis report on its application in customer interactions and sales within the retail banking sector. Your detailed methodology and key findings provide valuable insights into leveraging K-means clustering for enhancing customer service and refining sales strategies.

The approach of utilizing a synthetic dataset with variables such as transaction amount, transaction type, customer age, and transaction frequency demonstrates a robust methodology for understanding customer behaviors. Standardizing the data prior to clustering ensures the accuracy and reliability of the analysis, while the choice of four clusters effectively captures the diverse behaviors present in the dataset.

Your key findings shed light on the distinct customer segments identified through clustering, ranging from high-value transactions to online and mobile transactions. Each cluster represents unique customer preferences and behaviors, offering actionable insights for personalized marketing, enhanced customer service, and cross-selling opportunities.

I particularly appreciate the emphasis on personalized marketing and product offers tailored to the specific needs of each customer segment. Your insights into enhancing customer service based on preferred transaction methods and identifying cross-selling opportunities underscore the potential for maximizing customer satisfaction and retention.

In conclusion, your report highlights the significant benefits of applying K-means clustering to transaction data in retail banking. By understanding customer behaviors and preferences at a granular level, banks can optimize their sales strategies, improve customer interactions, and ultimately drive business growth.

Thank you for sharing your insights, and I look forward to future discussions on this topic.

Best regards

Ahmad Baker

Reply

(http

Chris Cosmas (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/226607)

Apr 25, 2024

Hello Jignesh,

Thank you for the presented scenario it provide valuable real world examples on the applicability of clustering. In regards to your 3rd cluster would it be more useful to increase the number of clusters to isolate differing profiles and have a more accurate representation of the customers within?

This could maybe enable the decision maker to formulate a better idea in regards to the products which could be targeted to customers.

Reply

(https:/

Lee Lanzafame (https://classroom.emeritus.org/courses/9054/users/231975)

Apr 23, 2024

As a sales manager I've come up with a short description to better explain to the sales team areas they should focus on when approached by a potential customer.

We conducted k-means clustering and had 3 clusters

- The 3 clusters had 311, 191 and 291 surveys. We identified the following personas:
  - high performance – fast zippy car
  - long distance drivers – comfort and safety
  - fuel savers – reliability and cheap to run

Some facts about the people we surveyed

- Most people that take the survey are male
- Reliability is more important than comfort
- Technology is more important than interior
- People who are after a powerful car should go for the sporty model

Going forward when approached by a customer it's important to determine which of the above categories they belong to that way you can focus the inventory and test drives with the highest success rate first.

Reply (1 like)

(http

Shahrod Hemassi (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/224267)

Apr 23, 2024

Hi Lee. Thanks for your report. I appreciate your analysis of the survey data and your recommendations for investment going forward.

Reply

(http

Ricardo Anaya (https://classroom.emeritus.org/courses/9054/users/228915)

Apr 24, 2024

Good info

woud Price vs income would also be needed? among thr 3 categories to have a better match with the customer?

Reply

(https:/

Shahrod Hemassi (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/224267)

Apr 23, 2024

The following is my report on a survey of 793 customers who had purchased a vehicle from us within the past three months. These customers were asked to rate several different aspects of their purchased automobile as either "important" (1) or "not important" (0). Various aspects were rated such as driving properties, interior, technology, comfort, reliability, handling, power, consumption, sporty, and safety. We also gathered a couple pieces of demographic information in gender and household.

In our analysis, we used the k-means method as our clustering algorithm. We also utilized the “elbow method” to determent the optimal number of clusters. Because we had multidimensional data, we needed to use the Principal Component Analysis algorithm to be able to plot this data in 2 dimensions for our analysis.

Based on our data analysis, we have insight into our new product development roadmap. We have identified that technology is very important to our customers. Reliability and safety are also very important to our customers. Going forward, we should increase our marketing focus on these attributes of our products. Comfort and interior are not as important and we can likely reduce our marketing investment on these lesser important attributes.

Reply (1 like)

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 24, 2024

Hi Sharod,

Do you think there could be niche markets for comfort and interior where these less important features might be more valued?

Edited by **Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)** on Apr 24 at 1:58am

Reply

(http

Jignesh Dalal (https://classroom.emeritus.org/courses/9054/users/229173)

Apr 24, 2024

Your use of k-means clustering to analyze customer preferences in automotive attributes is very insightful, particularly your methodical approach with the elbow method and PCA for multidimensional data visualization. It's clear that your findings—emphasizing technology, reliability, and safety as key customer priorities—offer valuable guidance for both product development and marketing strategies.

Given these insights, how do you envision integrating these priorities into your upcoming vehicle models? Are there specific technological innovations or safety features that are now a higher priority? Also, considering that comfort and interior features were deemed less critical, how might you adjust these elements to maintain customer satisfaction without significant investment, potentially reallocating resources towards more critical features?

Moreover, how could these insights reshape your advertising strategies? Would there be a shift towards campaigns that highlight the tech-savvy and safety features to appeal to those particular demographics? This targeted approach could refine your marketing efforts and better align with customer values.

Reply

(http

Mariana Flores (https://classroom.emeritus.org/courses/9054/users/237198)

Apr 24, 2024

Hi Shahrod, so nice to meet you. Great post, k-means clustering with principal component analysis to identify product features and create a product development roadmap tailored to the target market is such a business impactful manner to implement cluster analysis. I'm with you in tailoring the product to target audiences and segments.

Cluster analysis and the real-world application in product development is remarkable - thank you for sharing.

Reply

(https:/

Diego Milanes (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/228518)

Apr 23, 2024

(disclaimer: the mentioned colab file doesn't have any information about clustering but about normalization. Thus, I'm not sure what this assessment is about, and I'm not sure the function we are supposed to take corresponds to which case. It also doesn't make much sense to run the algorithms in the 4 events data set)

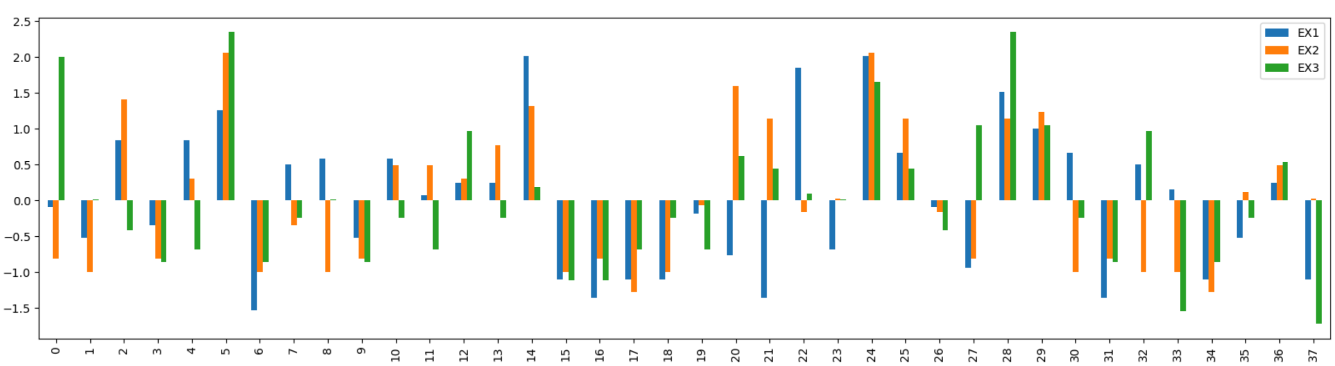
Introduction: Clustering analysis is a powerful tool for identifying patterns and grouping data points into clusters based on similarity metrics. This grouping provides information that informs businesses' decisions via the categorisation of population, anomaly detection, social analysis, among others.

Findings: We have worked with two clustering algorithms: k-means and hierarchical clustering. The approach and philosophy behind each method are different. K-means requires knowing the number of clusters in advance and computes centroids recursively, while the hierarchical does not require a specific number of clusters in advance and works by grouping data recursively; the number of clusters is achieved in both cases depending on the desired dissimilarity. In the k-means method, the plot of dissimilarity vs. number of clusters determines the desired clustering level. This is known as the elbow plot. If data present several features, a principal-component analysis can be driven reduce the system's dimensionality and provide additional insight to the segmentation. While hierarchical clustering is capable of determining differently shaped clusters, k-means tend to work better under linear-separated clusters. However, the latter is computationally less expensive since it scales differently for large data samples. Therefore, the choice of the method depends on the problem.

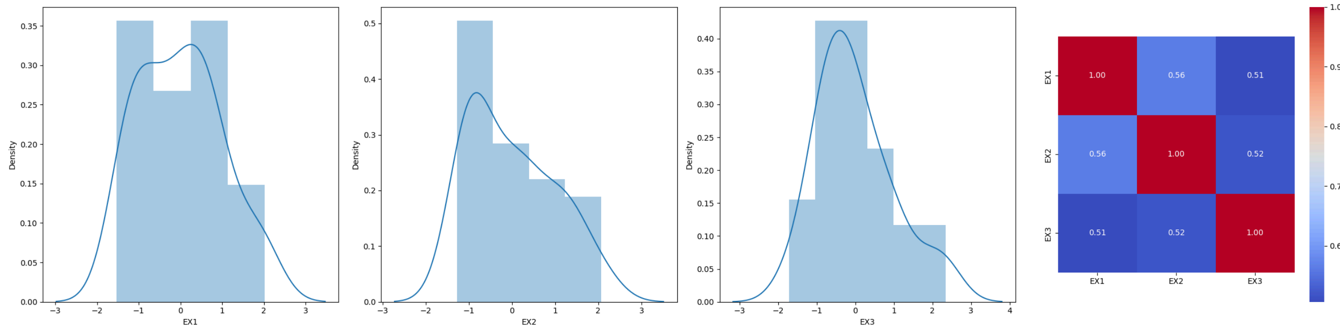
Use of clusterization in advertising: Customers can be segmented into groups with similar characteristics. Understanding these similarities helps identify the needs of each segment, allowing for specific targeting market campaigns. In addition to this, personal messaging or product recommendations are potential actions once the corresponding segments are determined.

Conclusion: Clustering analysis offers valuable insights into different steps and stages of business based on the behaviour and preferences of the actors involved. By leveraging clustering techniques, businesses can optimize their marketing strategies, improve customer engagement, and drive sales growth, ultimately enhancing overall performance and competitiveness in the market.

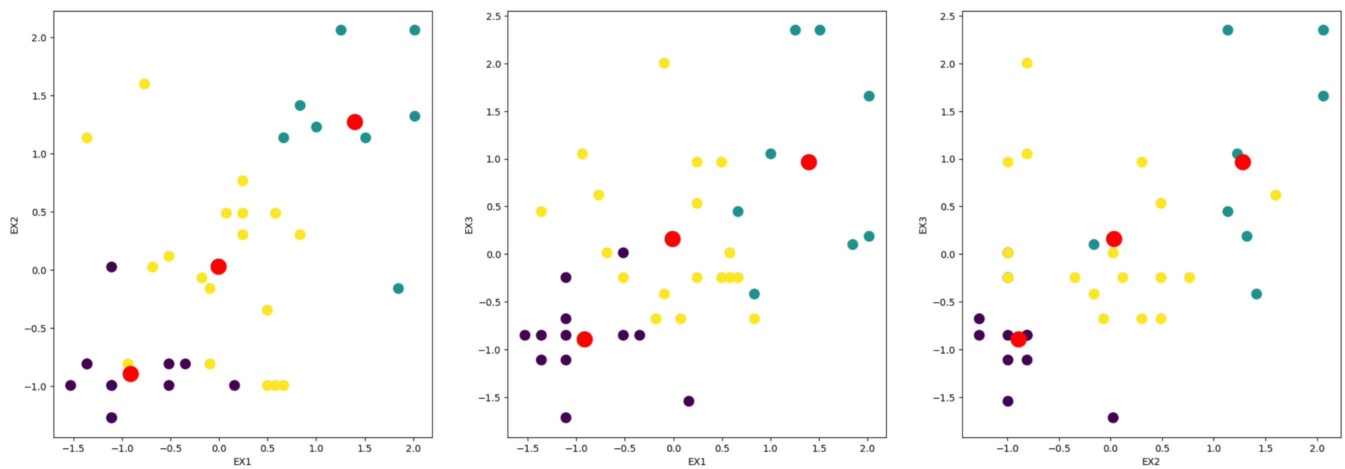
An example from academia: I have the grades for 3 tests for 38 students in my course. I want to categorize in high-, average- and low-academic performance. The standardised data for the 3 tests looks like this:



Here, the most positive is the value, the better the grade. The standardised histograms, as well as the correlation among the variables, are shown next:



As observed, the correlation in general is of about 0.5 among all variables. An Elbow analysis has confirmed the number of categories and the k-means algorithm shows the following results, when comparing the three different pairs of variables



In the above plot, the red dots represent the centroids for the 3 clusters. The 3 performance clusters that we expected are now evident. The plot suggests taking action w.r.t. the violet population, which, if the trend continues, will not approve the course. Special learning strategies must be applied here. The yellow population will be monitored and followed for difficulties, while advanced material can be released to the green population since it shows a better performance.

Reply (1 like)

(http

Turki Alghusoon (https://classroom.emeritus.org/courses/9054/users/229165)

Apr 23, 2024

Hi Diego,

Great report and very good example. I like how you interpreted the findings at the end and incorporated it into suggested actions. Do you think PCA could be used instead of the 3 clustering plots in this scenario? or do you think the 3 plots give additional insights?

Thank you.

Turki

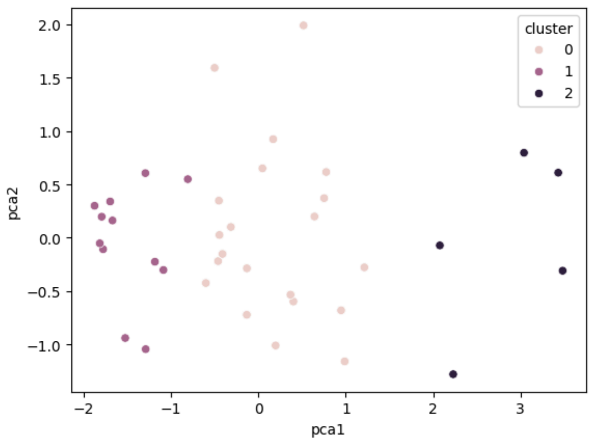
Reply

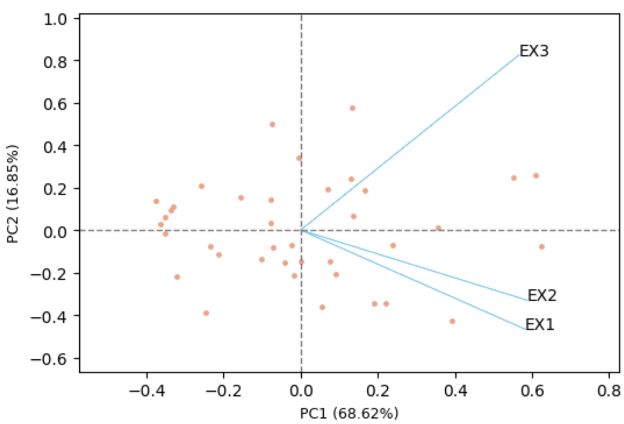
(http

Diego Milanes (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/228518)

Apr 23, 2024

Hi Turki, I've just tried the PCA to reduce dimensionality. There you can separate the clusters by lines which allows a better separation. The three tests look important, which they are, and perhaps the third one is statistically different from the others (smaller std deviation as observed in the standardised distributions). Below the result of the pca tests.





Reply

(http

Chris Cosmas (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/226607)

Apr 25, 2024

Hello Diego,

Very impressive visualisations. I am still having trouble grasping how the PC1 plot would be read and understood. Is importance read based on the X axis or Y axis? Why do the features skew the way they do?

Reply (1 like)

(http

Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)

Apr 24, 2024

Thanks, Diego, for sharing the graphs and visuals.

Could you please advise on how to profile PC clusters? please feel free to point me in the right direction.

Edited by Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864) on Apr 24 at 5:05pm

Reply

(http

Timothy Andrew Ramkissoon (https://classroom.emeritus.org/courses/9054/users/226697)

May 1, 2024

Diego,

The plot and graphs used added a nice touch to your response and was able to drive your point home. Good Job.

Reply

(https:/

Turki Alghusoon (https://classroom.emeritus.org/courses/9054/users/229165)

Apr 23, 2024

Here are my main finding K-Mean clusters in product development:

K-means clustering is a powerful tool to consolidate and stratify end-user profiles and preferences. In addition, the elbow method increases the effectiveness of K-means clustering as it spares the product owners from having to guesstimate the optimal number of clusters. K-means is particularly useful in the 2 following areas:

- For end-user profiles, K-means could help product managers consolidate their end-users into distinct categories based on spending behavior, income, lifestyle data…etc. once those categories are defined, product managers can more easily identify the target segments and the optimal strategy to target each segment.
- For preferences: product managers can use K-means clustering to organize the end users into clusters based on user's feedback related to product features. Those clusters can then guide the design of the different combination of features to target each cluster. For example, a furniture manufacturer using K-means clustering could determine that their customers preferences fall into 2 main categories: luxury furniture and baseline economy. Using that insight, the manufacturer can then develop 2 lines of products that directly address the needs of both clusters. Such an approach would increase the likelihood of higher sales and increased customer satisfaction without having to compromise on either product line.

Reply (1 like)

(http

Dawn Prewett (https://classroom.emeritus.org/courses/9054/users/233112)

Apr 24, 2024

Really nice succinct summary. I like the idea of product managers using clustering to help organize their end users. Having worked with many project managers, I know this is a pain point for them. Though, I admit I was thinking of product management in regards to software development and how product managers help influence the development and trajectory of the product.

Reply

(http

Ricardo Anaya (https://classroom.emeritus.org/courses/9054/users/228915)

Apr 24, 2024

in product development

-understanding customer preferences and segmenting products based on similar characteristics.] it is key

that is what I placed as response, but thanks to your post, I have also think to add "customer feedback" as an entry point thanks for bringing it up

Reply

(https:/

Mariana Flores (https://classroom.emeritus.org/courses/9054/users/237198)

Apr 23, 2024

Clustering is the process of extracting features from and grouping unlabeled data through unsupervised learning algorithms based on data similarities. A cluster is a collection of data points grouped together based on certain similarities forming similar subgroups. The notion of similarity and dissimilarity is vital both in terms of clusters and across clusters when evaluating outcome features. Due to the assumption that similar objects behave and respond in a similar manner to different interventions. Two common clustering algorithms are Hierarchical Clustering and K-Means Clustering.

Hierarchical Clustering is based on an iterative classification method which treat each data point as an individual cluster then combines data points together through a bottom-up approach to create clusters to form a Dendrogram tree with the groupings. The Dendrogram tree is a hierarchy of partitions with the ability to subjectively decide a partition by truncating the tree a given level based on predefined constraints.

K-Means Clustering is an algorithm which based on the number of clusters (k) randomly creates centroids and assigns each data points to the nearest centroid then works to recalculate the k centroids as the average of their assigned observations and repeats iteratively until convergence.

Cluster analysis is an analytical technique that can enhance product development by uncovering data signals to create products that meaningfully resonate with customers and guide the development process. A detailed understanding of product-market fit is vital in identifying customer needs and designing products that fulfill those needs. K-Means Clustering can help identify features and increase product appeal by creating personas that are representative of each cluster then prioritizing those areas that matter most to the target market at every stage of the product development process. Leading to an increased likelihood of product success in the market and by empowering firms to stay ahead of the competition.

Reply

(http

Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)

Apr 24, 2024

Great explanation. Thank you Mariana

Reply

(https:/

Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)

Apr 24, 2024

Clustering method could be used to determine the effectiveness of different sales & interaction methods in customer service. For example calling people or emailing them and applying different sales techniques and testing it's effectiveness by results of final sales, calls back, engagement, etc.

Different methods of sales could be tested on the X Axis Vs Final Sales on the Y Axis and identify clusters that are relevant to understand the effectiveness of different sales actions/methods.

Hierarchical clustering can be used to identify distinct groupings and understand relationships between different sales methods and outcomes/sales conversion. K-Means Clustering can help identify features and increase product appeal by creating personas that are representative of each cluster then prioritizing those areas that matter most to the target market

Edited by **Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)** on Apr 24 at 2:40am

Reply (1 like)

(https:/

Dawn Prewett (https://classroom.emeritus.org/courses/9054/users/233112)

Apr 24, 2024

During my tenure at a call center, I had access to the massive amounts of data that were constantly being collected - including the data collected via the customer sentiment surveys. While a common end-of-call feature at most call centers, ours involved 2-3 questions that customers responded to using their touchpad and a 1-5 scale. While these sought to provide general feedback on agent performance and overall company perception, their effectiveness was limited.

Using the hierarchical cluster algorithm presents a promising avenue for improvement. Since this method groups data points based on similarities, it could reveal patterns within customer feedback data. Since each survey question assesses a different aspect of the customer service experience, this algorithm could expose both correlations and discrepancies between the different questions.

Based on my knowledge of the data we collected, I would posit that we'd see two primary clusters: satisfied and dissatisfied customers. Despite what everyone thought were carefully crafted questions, strong emotions, particularly negative ones, can override emotions regarding the individual questions. This can hinder our ability to pinpoint specific customer concerns, despite targeted questioning.

However, hierarchical clustering may reveal a third cluster containing both negative and positive sentiment. This cluster would indicate that the questions are, at least in those cases, working as expected. On the other hand, the absence of a distinct "middle ground" cluster would suggest that customers often view both entities as one and the questions aren't working as expected. While I would not expect this cluster to be as dense as the other two, it should not be absent.

Leveraging the insights provided by using this algorithm would lead to improved insight into the customer's experience and could aid in more meaningful refinement of the survey questions while providing some much needed feedback to aid in developing targeted strategies to improve service quality.

Reply (1 like)

(https:/

Ricardo Anaya (https://classroom.emeritus.org/courses/9054/users/228915)

Apr 24, 2024

write a short report explaining your main findings using either the hierarchical or k-means clusters.

Hierarchical vs. K-Means Clustering:

Hierarchical Clustering:

Creates a hierarchy of clusters (like branches in trees).

Better to use for small datasets.

Specially Useful when you don’t know the number of clusters in advance.

K-Means Clustering:

Divides data into a fixed number of clusters.

best for large datasets.

Requires defining the number of clusters beforehand.

It is faster but less flexible than hierarchical clustering

Choose one of the following three functions: product development, advertising, or customer interactions and sales, and explain how your clustering analysis could be applied in this area.

Product Development:

Clustering in Product Development:

-understanding customer preferences

segmenting products based on similar characteristics.

Application:

Use clustering to group similar products together based on features such as price, features, form factor or design.

Identify product specificatrion gaps or areas where new features are missing and can be added.

Optimize product portfolios by analyzing which products are most popular within specific customer segments.

Example:

A tech company wants to launch a new smartphone.

By clustering existing smartphones based on features:

camera quality in Mega pixels

number of cameras,

battery life,

screen size

memory Size

The company can identify gaps in the market and design a product that address specific customer needs.

Reply

[https:/](https://classroom.emeritus.org/courses/9054/users/233864)

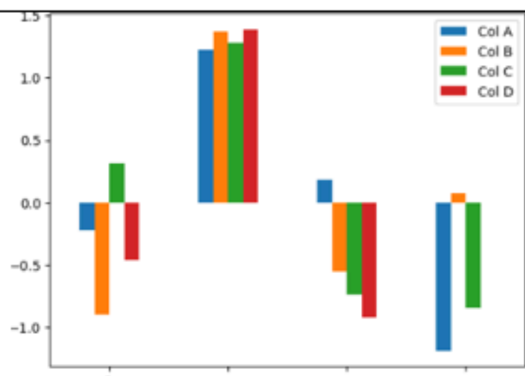
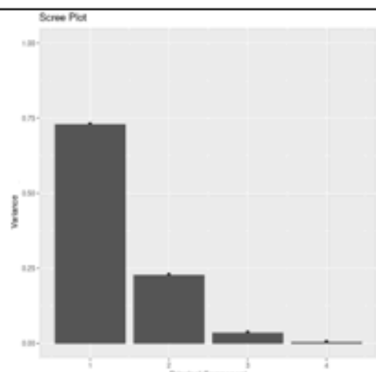
Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)

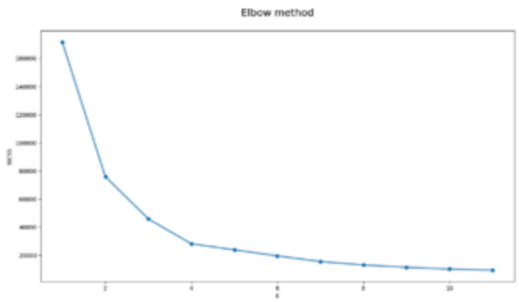
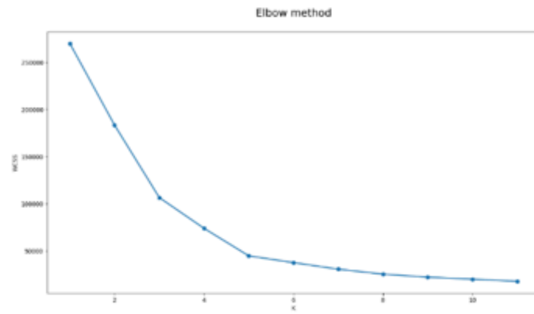
Apr 24, 2024

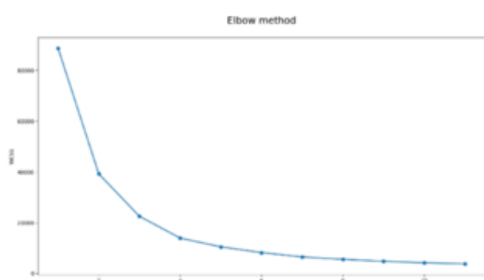
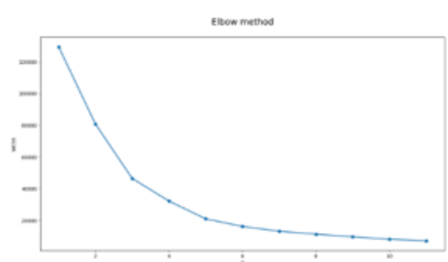
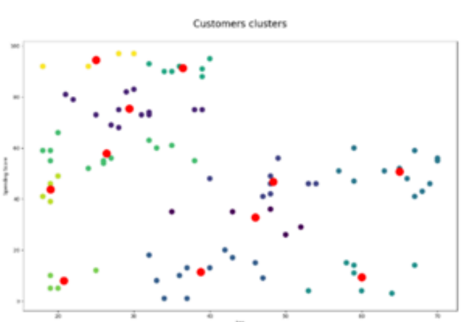
Case Brief

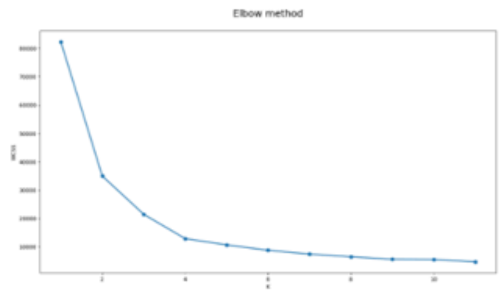
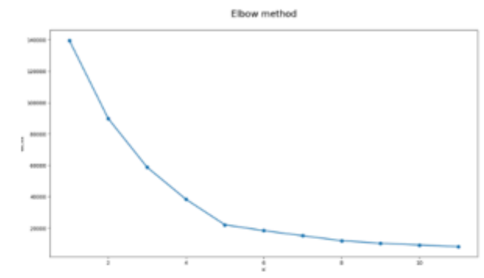
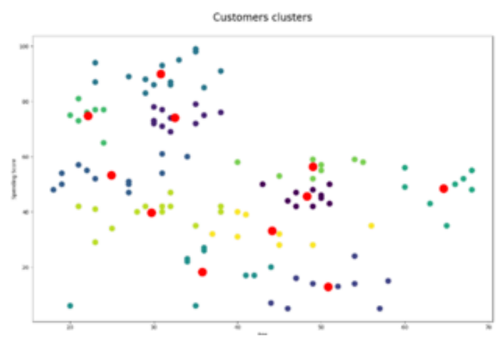
- For an NGO, identify the donor clusters that are most likely to
1. Subscribe to a monthly donation scheme.
2. Increase their monthly donation.
- Each observation has 12 features.
- Some data is missing (e.g. age for some donors)

Clustering Process

Step	Graph (illustrative)	Outcome
1.Data Preparation		Address missing data, coding …etc.
2.Scaling Techniques Standardization of Donations		-Establish a common universal scale (standard deviation unit) -show which side of the mean (x-axis) is each data point (donor).
3.Calculate Covariance		matrix of 12 features Covariance values (-1 &1)
4.Apply Principal Component Analysis (Scree Plot)		-determine PCs variation -identify the most important ingredient (feature) within each PC

Principal Component features		Age vs Donation	Annual Income vs Donation
		A	B
5.Script Plot (all Donors)	1		
6.KMeans Clustering (4 clusters)	2		

7.Assess Simpsons Paradox			
7.1Script Plot (Male)	3		
7.2KMeans Clustering (4 clusters)	4		

Principal Component features		Age vs Donation	Annual Income vs Donation
		A	B
8.1Script Plot (Female)	5		
8.2KMeans Clustering (4 clusters)	6		

Recommendation

- Since gender does not seem to have a significant effect on the clusters, the marketing campaign can be designed to target all genders.
- Based on the Script Plot the number of clusters is 4 (*elbow*). This seems to be consistent across both analysis features and gender.
- Clustering Spending scores based on age
- Selection of the clustering approach is contingent on the business objective (marketing campaign) and the expected return on investment, which will require bilateral discussion with the marketing department.
- The key initial observation is that annual Income vs Donation amount (cell B2) is worthy of further analysis, especially after excluding the outliers. This would significantly affect the WCSS (cell B1).

Limitation

- Steps 5 and onward are made on the assumption that *Age* and *Annual income* are the 2 features most (positively) correlated to *Donation*.
- Ability to track and assess behaviour change over time (e.g. what are the triggers that elicited change in donation patterns)
- Further information on practically applying PCA to determine the key features, is needed.

Inquiry!

To what degree would KMeans produce different clusters than AHC if both used the same dimensions?

Edited by Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864) on Apr 24 at 5:09pm

Reply

(http

Priscilla Annor-Gyamfi (https://classroom.emeritus.org/courses/9054/users/226376)

Apr 26, 2024

Great post. I love all the graphical illustrations.

Reply

(http

Koffi Henri Charles Koffi (https://classroom.emeritus.org/courses/9054/users/208039)

May 1, 2024

hi Haithman , I love the detailed explaination and the graph , it give more inside.
I after google , got this one from AI , I could find this answer to your inquiry

The degree to which KMeans produces different clusters than Agglomerative Hierarchical Clustering (AHC) when both algorithms use the same dimensions depends on several factors, including the nature of the data, the number of clusters specified, the initial configuration of centroids or cluster seeds, and the distance metric used to measure similarity between data points. Here's a breakdown of these factors:

1. **Data Nature:** KMeans and AHC may produce different clusters if the underlying data has different shapes or distributions. KMeans tends to create spherical clusters around centroids, while AHC forms clusters based on hierarchical relationships between data points. If the data has non-linear or complex structures, AHC may capture these patterns more effectively than KMeans.
2. **Number of Clusters:** The number of clusters specified as input to the algorithms can significantly influence the clustering results. KMeans requires the user to specify the number of clusters beforehand, whereas AHC can automatically determine the optimal number of clusters based on the dendrogram. If the chosen number of clusters differs between KMeans and AHC, the resulting clusters will likely vary.
3. **Initial Centroid Configuration:** KMeans is sensitive to the initial placement of centroids, as it iteratively updates cluster assignments based on the nearest centroid. Different initializations of centroids can lead to divergent clustering outcomes. AHC, on

the other hand, does not rely on initial seeds and constructs clusters based on hierarchical relationships between data points.

4. **Distance Metric:** The choice of distance metric used to calculate similarity between data points can impact clustering results. Common distance metrics include Euclidean distance, Manhattan distance, and cosine similarity. KMeans typically uses Euclidean distance, while AHC can accommodate various distance metrics. If different distance metrics are used between the two algorithms, the resulting clusters may differ.

In summary, while KMeans and AHC may produce similar clusters under certain conditions, such as well-separated spherical clusters and consistent initializations, they are inherently different algorithms with distinct approaches to clustering. Therefore, it's not uncommon for them to yield different clustering results, especially when applied to complex or high-dimensional datasets. Conducting sensitivity analyses, experimenting with different parameters, and assessing cluster stability can help evaluate the robustness of clustering solutions generated by both algorithms

Reply

(https:/

Ahmad Abu Baker (https://classroom.emeritus.org/courses/9054/users/234460)

Apr 24, 2024

Hello everyone,

Kindly find below my report. We leveraged K-means clustering for customer segmentation in advertising strategy. Hope you enjoy the read. Looking forward to your replies and wonderful insights.

Leveraging K-means Clustering for Customer Segmentation in Advertising Strategy

Introduction:

In this report, we explore the application of K-means clustering in advertising strategy, specifically for customer segmentation. Customer segmentation plays a crucial role in advertising as it enables businesses to tailor their marketing efforts to different groups of customers based on their preferences, behavior, and characteristics.

Methodology:

We utilized K-means clustering algorithm to segment customers into distinct groups based on their attributes and interactions with the advertising campaigns. The dataset consisted of various features such as demographic information, browsing history, purchase behavior, and response to advertisements.

Findings:

Through K-means clustering analysis, we identified several distinct customer segments with unique characteristics and preferences:

1. **Tech-savvy Millennials:** This segment comprises young individuals who are highly engaged with technology-related advertisements. They exhibit a preference for innovative products and are likely to respond positively to campaigns focusing on cutting-edge technology.

2. **Value-conscious Shoppers:** This segment consists of price-conscious customers who prioritize discounts and promotions. They are more likely to respond to advertisements offering discounts, coupons, or special offers.

3. **Luxury Enthusiasts:** This segment includes high-income individuals who value luxury and premium products. They are attracted to advertisements that emphasize exclusivity, quality, and prestige.

4. **Family-oriented Consumers:** This segment consists of families with children who are interested in products and services catering to family needs. They respond well to advertisements highlighting family-friendly features and benefits.

Implications:

By understanding the distinct preferences and behaviors of each customer segment, businesses can tailor their advertising strategies to effectively target and engage with their audience. For example:

- Targeted Messaging: Businesses can create personalized advertisements that resonate with each customer segment's preferences and interests.
- Channel Optimization: By identifying the preferred channels and platforms of each segment, businesses can allocate their advertising budget more efficiently.
- Product Development: Insights from customer segmentation can inform product development strategies, helping businesses create products that meet the specific needs and preferences of each segment.

Edited by **Ahmad Abu Baker (https://classroom.emeritus.org/courses/9054/users/234460)** on Apr 24 at 3:26pm

Reply

(http

STEPHEN HUTSON (https://classroom.emeritus.org/courses/9054/users/233645)

Apr 24, 2024

Great response Ahmad! I really liked the distinct types of customer segments you provided here given that you can clearly determine types of follow up approaches the business can take to increase engagement with these audiences, since they all appear to have pretty distinct interests and would respond to different types of marketing campaigns.

Reply

(http

Priscilla Annor-Gyamfi (https://classroom.emeritus.org/courses/9054/users/226376)

Apr 26, 2024

Great post highlighting on how advertising strategy can be enhanced using the clustering algorithm.

Reply

(https:/

STEPHEN HUTSON (https://classroom.emeritus.org/courses/9054/users/233645)

Apr 24, 2024

Some of the main findings from using K-means clustering from this module is that we're able to specify a desired number of clusters, and the algorithm is able to assign our outputs to these clusters which emphasizes the importance of having an understanding of the business process when applying this methodology. Some other important observations includes techniques like normalizing the data to get better insights into results in the context of the larger population of observations, as well as methods like the elbow curve for determining the optimum hyper parameter for how many clusters we'd want to include. If I were to approach using K-means for customer interactions and sales, I think K-means would be really helpful in looking at different segments of customers and how certain demographic features like age, income levels, geographic location, as well as historical information like how often they shop with us would be useful in helping us cluster these customers together to identify populations that we could look to target with specific marketing strategies to increase interactions with customers we could identify as populations who could become more frequent customers, or look at populations we may want to offer discounts and promotions to based on past spending activity and interests.

Reply

(https:/

Chris Cosmas (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/226607)

Apr 25, 2024

In this module we've learned about clustering and it's benefits. Especially in grouping target observations. Observations are units which are comprised of many variables, an observation can be a transaction, or an individual which each can be described with many different variables.

K-mean clustering tries to group data points on a plot based on two chosen parameters. This aims to categorize different groups with the most similar characteristics.

Clustering can be applied to sales data to identify different customer segments for which different promotional material could be used or could be incentivised to purchase more by being offered varying programs such as point systems, cashback programs, free items at a certain purchase level, and so on.

In this assignment I focused on sales data which was provided it to us earlier in the module which aims to quantify Spending Score compared to different variables. The previous assignment focused on the Age as a parameter, I performed another analysis based on the Annual Income of individuals.

I first visualised the data to get a sense of it's contents. I created a bar chart which showed a high concentration of individuals who earn between 50 and 75 k a year.

I then moved on to the computation of the k-means clusters.

firstly, I had to create an Elbow Method plot which shows the Within-Cluster Sum of Square per cluster. This method allows us to choose the number of clusters for which we are most comfortable with the level of dissimilarity. Dissimilarity calculates the square distance between each observation and the closest centroid at the current loop. The centroid is the the average of all observations on the plot. As the centroids increase the WCSS also decreases. The Elbow Method provides the user with information on the level of specificity he wishes to continue with. If the K chosen is too small large dissimilarities will be observed, if the K is too small the clusters might be too specific and similar clusters might be split. Another issue with too many clusters is it offers diminishing returns, the similarity or dissimilarity level does not change much as a larger k is chosen. The clusters are formed in a way which minimizes the total WCSS so that each observation is grouped with it's closest centroid.

After reviewing the Elbow Method plot computed I chose to continue with 5 different clusters. The k_means function is the fit to the data set and the cluster centroids coordinates are computed. The data and the centroids are then fed into a plot with different clusters carrying different colours.

The conclusions I received are the following:

We can see three different spending habits even though we can see 5 distinct clusters.

There are two distinct groups with an average Annual Income of 88.2 K and 26.3K but both seem to have a very similar spending score at 17.1 and 20.91 respectively.

We can also see aa concentrated group of individuals who receive 55.2 K on average per year which has a more moderate spending score of 49.51.

We can also see another two clusters with very similar high spending score but very differing income averages. One has an income of 86.53 K with a spending score of 82.12 and the other has an average income of 25.72 with a spending score of 79.91.

Even though we can see widely differing income levels we can see similar spending behaviour presented by different groups.

Reply

(http

Priscilla Annor-Gyamfi (https://classroom.emeritus.org/courses/9054/users/226376)

Apr 26, 2024

Great post Chris. I love how thorough your presentation is. It is enlightening.

Reply

(http

Swati Sharma (https://classroom.emeritus.org/courses/9054/users/236938)

Apr 30, 2024

Hello Chris: Great explanation! quick question, how would you recommend utilizing the insights gained from clustering analysis to inform sales strategies and promotional campaigns for different customer segments?

Reply

(https:/

[Priscilla Annor-Gyamfi (https://classroom.emeritus.org/courses/9054/users/226376)](https://classroom.emeritus.org/courses/9054/users/226376)

Apr 26, 2024

Clustering Analysis for Customer Interactions and Sales

Introduction:

In this report, we discuss how clustering analysis is applied in customer interactions and sales. Our goal is to understand customer behavior and preferences through clustering algorithms, specifically K-means clustering. By clustering customers into distinct groups, businesses can tailor their engagements and marketing strategies to better meet the needs of each customer group.

Methodology:

Utilizing the K-means clustering analysis on a customer dataset comprising various customer characteristics such as purchase history, demographic information, and engagement with marketing channels, I was able to identify homogeneous groups of customers based on their similarities in their behavior and preferences.

Main Findings:

The analysis revealed some unique customer groups, each with unique characteristics and preferences:

1. **Price-Conscious Shoppers:** This segment consists of customers who prioritize affordability and make use of discounts and promotions. They are often sensitive to price changes and their purchase decisions are based on value for their money.
2. **Brand Loyalists:** Customers in this segment show strong loyalty to specific brands or products. They are less price-sensitive and are more likely to purchase from familiar brands, even at higher price points.
3. **Impulse Buyers:** The customers in this segment make spontaneous purchases mostly driven by their emotions. They are often targeted by impulse-driven marketing strategies such as limited-time offers or flash sales.
4. **Research-Oriented Consumers:** This segment comprises of customers who extensively research before buying any brand or product. They highly consider product reviews, recommendations, and detailed product information before making any purchase.

Application in Customer Interactions and Sales:

We could apply this clustering algorithm through:

1. **Targeted Marketing Strategies:** Tailoring promotional offers, advertisements, and loyalty programs to cater to the unique needs and preferences of each customer segment.
2. **Product Assortment:** Optimizing product selection and placement within stores to align with the preferences of different customer segments.
3. **Customer Lifetime Value:** Identifying high-value customer segments for prioritized retention efforts and personalized service.
4. **Market Expansion:** Identifying untapped market segments or geographic areas for potential expansion or targeted marketing campaigns.

In conclusion, clustering analysis helps generate great insights into customer segmentation, causing businesses or companies to optimize their customer engagements and sales strategies. A more personalized experience is established with both existing and even potential customers when there is a better understanding of the unique needs and demands of every customer segment.

Reply

(https:/

Todd Engle (https://classroom.emeritus.org/courses/9054/users/228910)

Apr 29, 2024

In the Jupyter assignment, I observed the compiling of 792 records of customer survey data broken down by male and female respondents. Each respondent scored different aspects that led to their decision to buy the vehicle. Each vehicle aspect was scored using a boolean score of 'Important' or 'Not Important'. The data was then viewed using a dendrogram to see how the data clustered followed by a heat map. The heat map was difficult to read, but I was able to manage. I then was able to display the data using the elbow method and it showed that it was best to start with two clusters. These two clusters, were then displayed in a scattergram.

As a mental exercise, I chose to analyze how I would use the k-means clusters for customer interactions and sales. For the clustering data I would focus on specific types of customer demographics: Age, income level, location, occupation, marital status, and family size. I would also want to look at other key data related to the purchase.

- **Vehicle Purchase History**: Car models purchased, price range, financing options used, trade-in details. This data reveals customer preferences and buying power.
- **Service History**: Frequency of maintenance visits, parts replaced, service costs. This helps understand customer loyalty and potential after-sales needs.
- **Marketing Interactions**: Channels used to interact with the manufacturer (website visits, social media engagement, dealership visits). This reveals how customers learn about and

engage with the brand.

My intention would be to identify distinct customer segments based on their demographics, vehicle preferences, brand interaction, and service needs. I would expect the data to yield clusters representing some distinct customer segments: Luxury car enthusiasts, budget-conscious families, eco-conscious buyers etc. Each segment will likely have different car preferences, service needs, and preferred communication channels.

The marketing department could use the clustering data to develop marketing strategies tailored to each segment's interests and preferred channels. For instance, reaching luxury car enthusiasts through high-end auto shows and magazines, while budget-conscious families might be targeted with online ads highlighting fuel efficiency and family-friendly features.

The Sales teams could train salespeople to better understand the needs and preferences of each customer segment, allowing for more personalized sales experiences and improved satisfaction.

Reply

(https:/

Swati Sharma (https://classroom.emeritus.org/courses/9054/users/236938)

Apr 30, 2024

Hello All: This report explores the use of clustering analysis to understand customer interactions and sales. By employing clustering algorithms, we aim to reveal distinct customer segments based on purchasing behavior, engagement patterns, and demographics.

We analyzed customer interaction and sales data, including demographics, purchase history, and website interactions. Using the k-means clustering algorithm, we segmented customers to identify common traits among different segments.

key customer segments:

1. **High-Value Customers**
2. **Occasional Buyers**
3. **Bargain Hunters**
4. **Tech Enthusiasts**

Application:

- **Product Development:** Tailor offerings to meet segment preferences.

- **Advertising:** Improve ad relevance and engagement with personalized campaigns.
- **Sales Strategies:** Enhance customer interactions and drive revenue.

Clustering analysis provides actionable insights into customer interactions and sales, empowering businesses to optimize strategies and maximize revenue.

Edited by Swati Sharma (https://classroom.emeritus.org/courses/9054/users/236938) on Apr 30 at 11:32pm

Reply

(http

Mhelissa Yayalar (https://classroom.emeritus.org/courses/9054/users/233590)

May 2, 2024

Hi Swati,

I resonate with your application of clustering in customer interactions and sales. The examples of key customer segments that you provided are good ones. I had similar analysis for Targeted Marketing. If I may add, using K-means clustering to your customer segments, we can find suitable targeted campaigns for each segments. Specifically, once you determine the optimal number of clusters (using elbow method), then we can apply the K-means to relevant features, such as frequency of the buyers. Then, after we analyze the clusters, we can further understand the behavior of each segments to determine how we can tailor sales campaigns.

-my

Reply

(https:/

Koffi Henri Charles Koffi (https://classroom.emeritus.org/courses/9054/users/208039)

May 1, 2024

Clustering analysis for Customer interaction and Sale

In our report we will focus on using hierarchical clustering to group customer and their interaction with different products and services . by grouping customers into separate clusters , for business to gain valuable insights from their behaviors , such as their purchase habit , their preferences, the spending score , that can guide marketing teams to target the right customer with the right product or services.

Process

We use Hierarchical clustering as a clustering technique to organize observations (customer and their characteristics) into a hierarchy of clusters based on their similarity or what principal characteristic that they have in common. The dataset contain customer interaction or relation with the product or the service offered by the company.

Result of the finding after clustering

- We can identify 3 clusters:

- most valuable customer : the customer that subscribe to more services or purchase more services or spend more

- Occasional Customer : customer who occasionally subscribe to services

- Low spending customer : customer who rarely makes a purchase .

Recommandation and conclusion

The marketing team can target the customer in cluster 1 and cluster 2 (Occasional Customer , Low spending customer) , understand their real need , propose them a product or a service that

really fit their need.

Cluster analysis is a powerful and important way to understand customer behavior and address their need based on their similarity . it help marketing team to target a group of customers with different similarity to be addressed separately . it help to customize marketing strategic base on the cluster defined

Reply

(https:/

Timothy Andrew Ramkissoon (https://classroom.emeritus.org/courses/9054/users/226697)

May 1, 2024

Introduction

As an Asset Integrity Manager, I recognize the critical role that data-driven insights play in optimizing our operations. In this report, I will focus on how clustering analysis—specifically hierarchical and K-means clustering—can enhance our product development efforts.

Business Context

Product development is a multifaceted process that involves designing, prototyping, testing, and launching new products or improving existing ones. Our goal is to create innovative, reliable, and cost-effective solutions for our clients in the oil and gas industry. Clustering can provide valuable insights at various stages of this process.

Hierarchical Clustering

Hierarchical clustering can group our suppliers based on their performance metrics (e.g., delivery time, quality, cost).
By identifying clusters of reliable suppliers, we can strengthen our supply chain and reduce production delays. Clustering can help us categorize materials based on their properties (e.g., corrosion resistance, durability). We can choose materials from clusters that align with our project requirements, ensuring optimal performance and longevity.

K-means Clustering

K-means can identify clusters of defective products based on quality metrics (e.g., yield, defect rate). We can investigate the characteristics of each cluster to pinpoint root causes and improve manufacturing processes. K-means can group products by demand patterns (e.g., high, medium, low). By aligning inventory levels with demand clusters, we can reduce excess stock and minimize shortages. K-means can segment customers based on preferences (e.g., product features, pricing). We can tailor product variants to specific customer clusters, enhancing customer satisfaction.

Benefits and Recommendations

- Resource Allocation: Clustering guides resource allocation—whether it's investing in R&D for specific product lines or optimizing production schedules.
- Risk Mitigation: Identifying high-risk clusters (e.g., defect-prone products) allows targeted quality control efforts.
- Efficiency Gains: Clusters reveal bottlenecks, enabling process improvements.
- Market Insights: Customer clusters inform market segmentation and product positioning.

Conclusion

Clustering analysis empowers our product development team to make informed decisions, enhance product quality, and drive innovation. By leveraging both hierarchical and K-means clustering, we can unlock hidden patterns and optimize our product portfolio.

Edited by **Timothy Andrew Ramkissoon (https://classroom.emeritus.org/courses/9054/users/226697)** on May 1 at 4:17pm

Reply

(https:/

Mhelissa Yayalar (https://classroom.emeritus.org/courses/9054/users/233590)

May 2, 2024

In my analysis, the two clustering functions: Hierarchical Clustering and K-Means Clustering usages or application are depended on the dataset and objectives of the analysis. Below are key insights on why two techniques were applied.

Hierarchical Clustering

- Provides a bottom-up view by merging individual data points into larger groups step by step.

- Enable flexible grouping, as it doesn't require specifying the number of clusters beforehand.

- Dendrogram visualization that allows you to see the relationships between clusters.

K-Means Clustering:

- Provides a top-down view by assigning data points to a fixed number of clusters (k) based on their proximity to cluster centers.

- The data set is equally sized clusters.

- The main challenge is choosing the right value of k (number of clusters) beforehand.

If we consider these clustering techniques for application in Customer Interactions and Sales, consider the following:

Hierarchical Clustering:

- Customer Segmentation: Use hierarchical clustering to segment customers based on their region or purchase history.

- Personalization: Create personalized marketing campaigns for different customer segments identified by the hierarchy.

- Sales Strategies: Tailor sales strategies (e.g., discounts, tier pricing) based on the specific needs of each customer group.

K-Means Clustering:

- Customer Profiling: Profile customers into distinct clusters (e.g., loyal customers).

- Targeted Marketing: Design targeted marketing efforts for each cluster. For example, send promotions to loyal member or engage non-member buyers with loyalty campaigns.

- Optimize Inventory: Use K-Means to optimize inventory management by understanding demand patterns across different customer segments.

In summary, both clustering methods have their advantages for customer interactions and sales. Hierarchical Clustering provides a visual hierarchy, while K-Means offers efficiency and equal-sized clusters. The choice depends on the dataset size and business goals.

Reply

(https://

Isabella Tockman (https://classroom.emeritus.org/courses/9054/users/207395)

May 13, 2024

It's very important to understand what customers like and how they behave. This helps businesses create better marketing plans and sell more effectively. Using a liquor Business survey data, and applying K-means clustering, we can find out about different groups of customers and serve them better.

Methodology

- Data Collection: We gathered information from a survey that asked people about their favorite drinks, how often they buy, and how happy they are with their purchases.
- Data Preprocessing: We cleaned up the data by fixing missing values and making sure everything was in a format that we could analyze.
- Clustering Process: We used K-means clustering to sort the data into distinct groups of customers. We chose the best number of groups by looking for a balance between simplicity and detail in our analysis.

From our K-means Clustering we found four main types of customers:

- Cluster 1: Budget Buyers - These customers look for the best deals and like buying cheaper brands. They love sales and discounts.
- Cluster 2: High-end Drinkers - These folks prefer top-quality drinks and don't mind spending more for premium brands.
- Cluster 3: Adventure Seekers - This group likes to try new and different drinks. They're always on the lookout for the latest and greatest.
- Cluster 4: Health-minded Drinkers - These customers want drinks that are better for their health, like low-alcohol options or organic drinks.

Application in Customer Interactions and Sales

1. Personalized Marketing: Knowing what each group likes helps us create ads and promotions just for them. For example:
   - Budget Buyers might like emails about discounts and special deals.

- High-end Drinkers could be invited to exclusive events where they can try new products first.

2. Better Customer Service: We can offer things that make shopping better for each group.
    - Adventure Seekers might enjoy events where they can learn more about different drinks.
3. Suggestions for Products: We can suggest products that each group might like.
    - Health-minded Drinkers would be interested in new, healthier drink options.

Using K-means clustering to look at our survey data has really helped us understand our customers better. This lets us target our ads better, improve our customer service, and ultimately sell more. In the future, we could add more types of data, like comments from social media, to get even better at understanding and predicting what our customers want.

Reply