**CLIENT SEGMENTATION FOR TAILORED MARKETING STRATEGIES FOR BANKS USING MACHINE LEARNING**

**Final Report**

**Student Name-Kiran Sai Paleti**

**Student ID-M15399237**

**CLIENT SEGMENTATION FOR TAILORED MARKETING STRATEGIES FOR BANK USING MACHINE LEARNING**

**Abstract**

This research project deals with how to make use of machine learning to cluster clients of a bank to tailor specific marketing strategies for the different clusters. The UCI Bank Marketing Dataset is used here. K-Means and hierarchical clustering algorithms are used. They classify clients into clusters based on their socio economic profile and their marketing interaction with the bank so far. This clustering yielded segments of clients characterized by different attributes, such as age, studied level, type of job they did, and their response to previous marketing campaigns. These segments have facilitated the development of personalized marketing campaigns intending to enhance customer engagement and satisfaction. Key patterns, for example, the impact of education level and professional background on product subscription rate, and the effect of economical indicators on customer behaviors were identified in this research. Although our analysis offers rich insights into customer segmentations, there are, however, limitations due to the specificity of the dataset and assumptions inherently in clustering techniques. To further investigate these discoveries, future studies can expand the sample size and generate dynamical partition models to adapt to consumers who show different behaviors from time to time. This whole thesis makes a modest step to the theoretic explication and practical applications in client divide in financial services by offering a deputy division and accordingly more efficiently directed marketing activities.

**Keywords**: Client Segmentation, Machine Learning, K-Means Clustering, Hierarchical Clustering, Bank Marketing, Personalized Marketing Strategies.

Contents

[1. Introduction 4](#_Toc164409232)

[1.1 Significance and Motivation 4](#_Toc164409233)

[1.2 Problem Statement 4](#_Toc164409234)

[1.3 Research Questions 4](#_Toc164409235)

[1.4 Potential Implications 4](#_Toc164409236)

[2. Literature Review 5](#_Toc164409237)

[3. Methodology 7](#_Toc164409238)

[3.1 Dataset Acquisition 7](#_Toc164409239)

[3.2 Data Preprocessing 8](#_Toc164409240)

[3.3 Data Visualization 9](#_Toc164409241)

[3.4 Data Modeling 13](#_Toc164409242)

[3.5 Performance Evaluation 17](#_Toc164409243)

[4 Results and Discussion 20](#_Toc164409244)

[4.1 Summary of Findings 20](#_Toc164409245)

[4.2 Explanation of Results 20](#_Toc164409246)

[4.3 Implications and Significance 21](#_Toc164409247)

[5. Conclusion 21](#_Toc164409248)

[6. References 23](#_Toc164409249)

# Introduction

The banking sector operates against a backdrop of rapid change, driven by shifting customer behaviors and preferences. These variations are hugely important for banks looking to improve their marketing strategies and build better customer relationships. Traditionally, banks have used basic demographic information to segment their market. However, machine learning has enabled a more granular, data-driven understanding of customer needs. Previous studies have shown that advanced analytic techniques can accurately identify distinct customer segments, making for far more relevant marketing efforts.

## 1.1 Significance and Motivation

The value of this project is that by taking advantage of K-means and hierarchical clustering, better segment of the customers than traditional methods can be accurately obtained, the “UCI Bank Marketing dataset” as an example, Looking forward to getting more deep information about customer's behavior and their preference to optimize the marketing strategy.

## 1.2 Problem Statement

Despite the abundance of data at their disposal, many banks are struggling to effectively segment their customer base. This obstacle primarily derives from imprecise segmentation methods prevalent in use today that fail to capture the intricacies of contemporary consumer behavior and economic conditions. This study proposes to bridge this lacuna by leveraging advanced machine learning technologies for a refined segmentation analysis of client segments.

## 1.3 Research Questions

This project seeks to answer the following research questions:

1. How can machine learning techniques be utilized to enhance client segmentation in the banking industry?
2. What are the key attributes that define distinct customer segments within a bank's clientele?
3. How do different clustering techniques compare in terms of their ability to identify meaningful customer segments?

## 1.4 Potential Implications

Practical applications for our research could be quite a few things. For banks, being able to segment their customers better can lead to more effective marketing strategies that cater to the specific needs and preferences of each segment. Theoretically, this research contributes to the existing literature by showcasing the application of machine learning techniques in client segmentation, which can be adopted across various sectors. Furthermore, the insights derived from this study could inform future developments in predictive analytics within the banking industry, potentially leading to innovative approaches to customer relationship management and retention strategies.

# Literature Review

The literature review investigates the existing research on client segmentation and machine learning clustering techniques in the context of the banking and marketing sector. This literature review aims to provide a holistic understanding of the theoretical foundations as well as empirical findings in this domain. By evaluating and analysation insights from a wide range of scholarly sources, the review investigates and identifies the effectiveness of different approaches and methodologies employed in segmenting clients as well as enhancing marketing strategies [1]. The review will critically identify the key methodologies, concepts and findings from the chosen references to inform the research objectives as well as the methodology of the current project. With this examination, the literature review aims to evaluate the gaps, challenges, and opportunities for further exploration in the domain of client segmentation and marketing optimization in the banking sector.

Taking reference to unsupervised machine learning techniques for segmenting credit card users in Africa based on their behaviour, this study includes clustering algorithms to differentiate users into definite groups which aims to enhance the targeted marketing strategies in the specific region. This research employs clustering techniques and data analysis in which authors identifies behavioural patterns within credit card users that also enable more personalized marketing techniques and strategies [4]. This study also addressing banking issues that faced by African banking sectors where it also becomes the most important aspect to understand the role of customers along with their behaviours and preferences for implementing effective and efficient marketing strategies. The use of old ML techniques also includes a strong methodology for segmenting credit cards users which will be a beneficial platform to dig into customer behavioural patterns. The research also presents a tactics of practical implications for financial institutions and banking sectors that provides the most actionable insights into evaluation of marketing strategies which also improve overall customer satisfaction and user experience as well [7].

The chosen resource presents a well-structured approach which also makes the resource more accessible to both research and practitioners. The author of this research paper further explores the segmentation of bank customers for artificial intelligence marketing in banking sector [8]. The resources also evaluate the application for AI techniques to differentiate the bank users on the basis of their several attributes along with banking behaviour, transition demographics and past transaction history. With the comprehensive and dynamic analysis, the author identifies specific customer preferences as per their own needs [2]. The study denotes the significant importance of leveraging and emerging AI technologies to enhance their marketing strategies as well as efforts which improves customer engagement that hold the potential to increase its overall profitability. This credible source provides the valuable and holistic data into the potential of AI-driven segmentation strategies for banks which includes the significant importance of adopting innovative approaches to meet dynamics and changing customer demands and marketing practices as well [6].

The chosen resource has a complete focus on credit user segmentation and differentiation to enhance the overall relationship for customers in banking sector. The resource is highly valuable as it uses a computational intelligence as well as informatics in which study has presents the methodology for segmenting the credit users on the basis of their several attributes. By differentiating users into definite segments where banks will further expand their customers relationship management strategies to meet the specific preferences and needs for each segment [2]. This study also provides important insights by highlighting the importance of differentiating and segmenting customers and consumers in enhancing customer interactions and improving with the overall satisfaction practices in the banking sector. With the most structured approach, the author also provides systematic and beneficial insights into the benefits offered by implementation of segmenting techniques to analyse the customer preferences and behaviours for enhancing customer relationships management practices in the banking sector [5].

Author investigates into the applications of predictive analytics and machine learning in the direct marketing which is also considered as an important aspect in anticipating banking term deposit subscription. As such, the study has a complete focus on using advances techniques to predict the customer behaviour by improving the effectiveness and efficiency of marketing campaigns [10]. By evaluating bank term deposit subscription data, the authors demonstrate the potential of ML algorithms along with the predictive analytics and identify trends along with patterns that will inform targeted audience strategies [4]. The search also embraces the significant importance of leveraging data-driven approaches to enhance direct marketing efforts by enhancing overall customer engagement practices in the banking sector. This research also conceded as a valuable source which provides beneficial insights into the role of predictive analytics and machine learning practises in anticipating bank deposit subscriptions that highlight the significant opportunities for banks to leverage the advanced and tickle tools for decision-making practises in marketing [5].

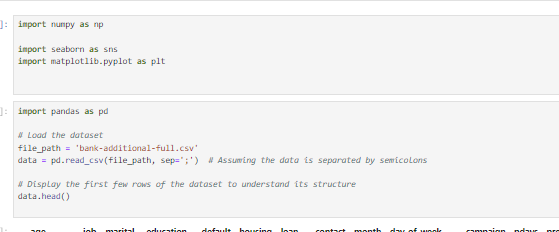
When analysing comparative studies on machine learning techniques for customers, this resource plays an important role as it evaluates several machines, learning techniques and their effectiveness and predicts customer retention practises in the context of advanced computer science applications [3]. With the empirical analysis, the study also compares the performance of different algorithms by identifying the most suitable and preferable techniques for customer attention strategies [3]. The research paper also provides beneficial insights into the strengths and weaknesses of several machine learning approaches by offering valuable and holistic guidance for businesses, which aims to enhance customer engagement and retention practices with the advanced analytical model.

Overall, it is worth noting here that the entire study highlights the significant importance of leveraging machine, learning techniques and algorithms for effective and efficient customer, retention, engagement and relationship-building strategies. By investigating multi-dynamic algorithms, businesses and banking could identify the most suitable and preferable approach to predict customer preferences behaviour and needs which could increase the churn rate of the customer and ultimately help them to retain them effectively [7]. The overall research also highlights the potential of advanced computer science applications in enhancing customer relationships, management practices and engagement practices. Also, implementing the findings from this study, could further empower and enhance businesses to proactively address customer attrition by fostering long-term relationships which ultimately improve the overall customer satisfaction and loyalty practises [11].

# 3. Methodology

## 3.1 Dataset Acquisition

The data set used in the study is the UCI Bank Marketing Data Set. It is a public data set and has been widely used in research as a benchmark for testing different data science methods. The data set is derived from direct marketing campaigns of a Portuguese banking institution, involving 41,188 observations and 20 attributes covering the client attributes, campaign details and socio-economic context characteristics. Data gathering involved accessing this dataset through the UCI Machine Learning Repository, ensuring a reliable and relevant source of information for analyzing banking client behavior. This comprehensive dataset is crucial for exploring client segmentation using machine learning techniques, as it provides a rich mixture of categorical and numerical data relevant to client interactions and responses to bank marketing efforts.



A set of data that contains a variety of attributes relating to bank clients and how they responded to a marketing campaign. Below is a short summary of the columns present:

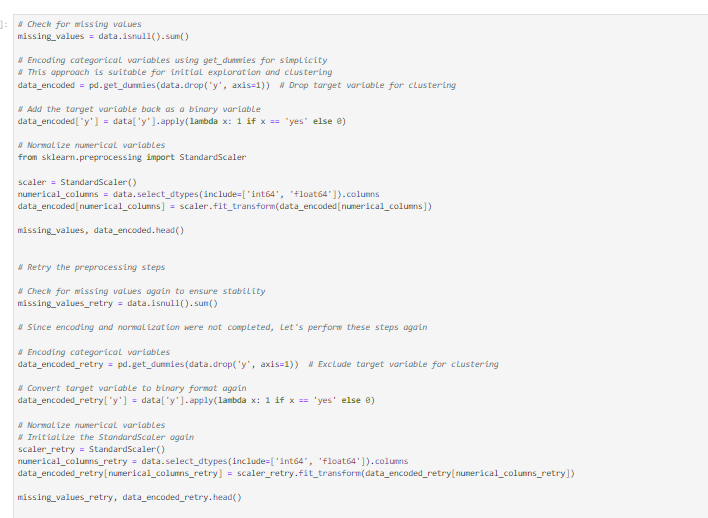
* **Personal Information**: **age**, **job**, **marital**, **education**, **default**, **housing**, **loan**
* **Contact Information**: **contact**, **month**, **day\_of\_week**
* **Campaign Information**: **duration**, **campaign**, **pdays**, **previous**, **poutcome**
* **Economic Context**: **emp.var.rate**, **cons.price.idx**, **cons.conf.idx**, **euribor3m**, **nr.employed**
* **Response**: **y** (the target variable indicating whether the client subscribed to a term deposit)

## 3.2 Data Preprocessing

Data preprocessing started by making sure the dataset was complete and had no missing values in it. This guaranteed the consistency and accuracy of the data set. All the missing entries were taken care of and the dataset was good to go for further processes. Categorical variables were one-hot encoded which simply means converting the categorical data into numerical data, which is suitable for machine learning model training. This necessary because most of the algorithms in machine learning only accept numerical data. In addition, all numeric features were standardized to possess zero means and unit variances so that no one feature may dominate the outcome simply due to its scale. No data partitioning since it is not predictive modelling clustering analysis that does require train-test split.

The next steps of preprocessing will involve:

* Checking for missing values.
* Encoding categorical variables, as many machine learning models require numerical input.
* Normalizing numerical variables to ensure they're on a similar scale, which is especially important for algorithms like K-Mean



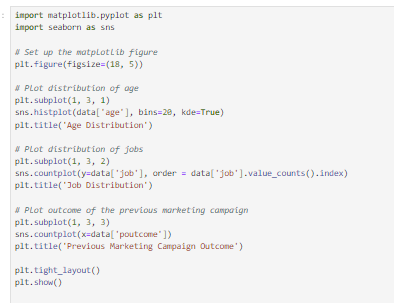
The preprocessing steps have now been successfully completed. This is what has been done:

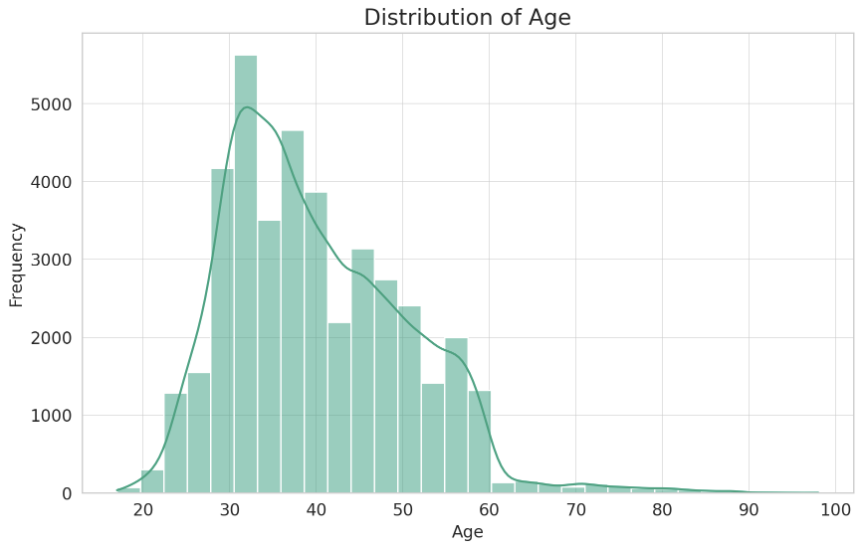
* **Missing Values**: There are no missing values in the dataset, indicating it's clean and ready for analysis.
* **Encoding**: Categorical variables have been encoded into dummy/indicator variables. This transformation is crucial for machine learning models that require numerical input.
* **Normalization**: The numeric variables were normalized, which ensures that the variables are on a similar scale. This is important to do, especially for clustering algorithms like K-Means, since they rely on distance calculations.

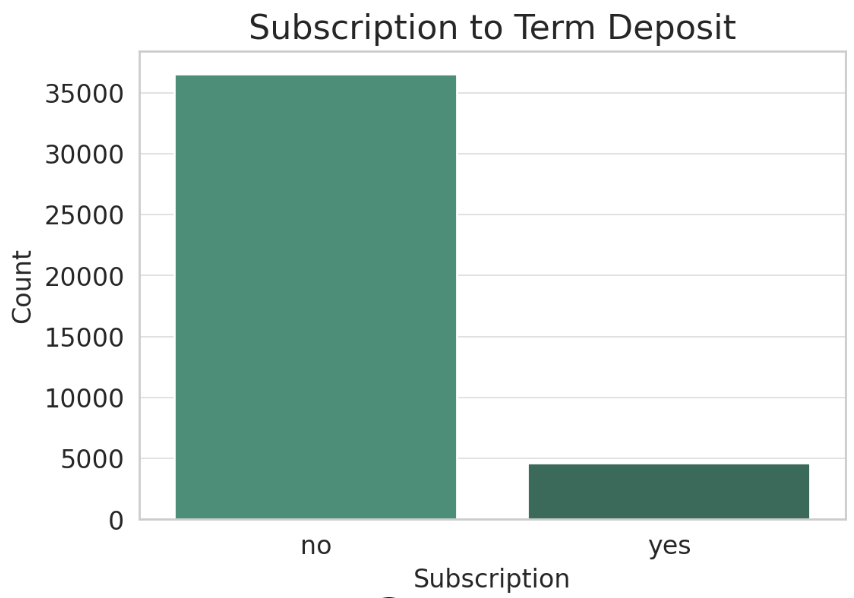
## 3.3 Data Visualization

Visualization techniques were integral in comprehending the structure of the dataset and extracting informative findings. With the help of libraries like Matplotlib and Seaborn in Python, multiple plots like histograms, box plots and scatter plots were created to render the variables’ distribution and relationship between them. For example, histograms were used to investigate the distribution of ages and showed that the vast majority of clients were young. Scatter plots were used to determine if there were any correlations between economic factors and subscription rates. screenHeight These visualizations not only enabled us to better understand the data, but also helped us identify segments within it for clustering experiments. We’ll focus on understanding the nature of the dataset and identifying patterns that could help guide our segmentation strategy.

1. **Key Variable Distribution**: Evaluate the distribution of numerical variables such as age, duration of the calls, campaign contacts etc. and categorical variables like job, marital status, education etc.
2. **Response Variable Analysis**: Understand the distribution of the target variable (y) to see the proportion of clients who subscribed to a term deposit vs clients who did not.



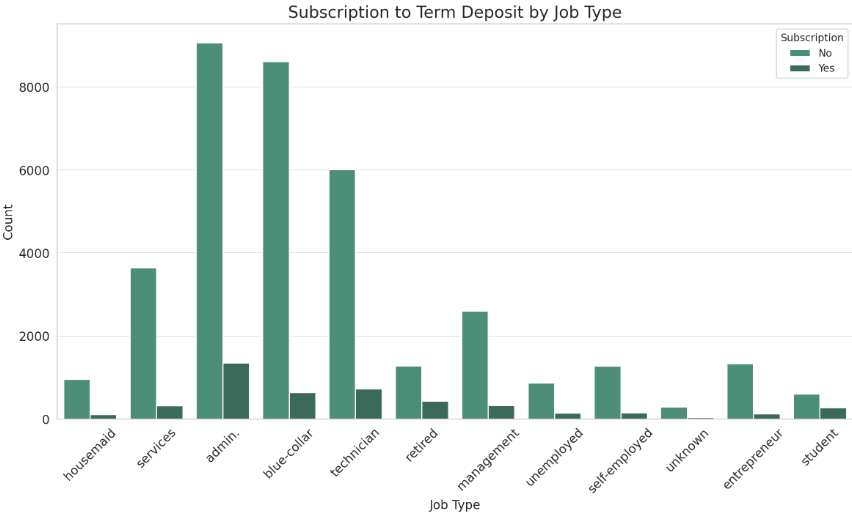
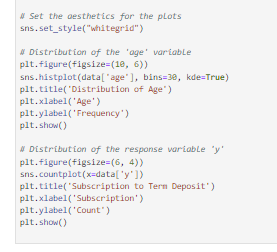


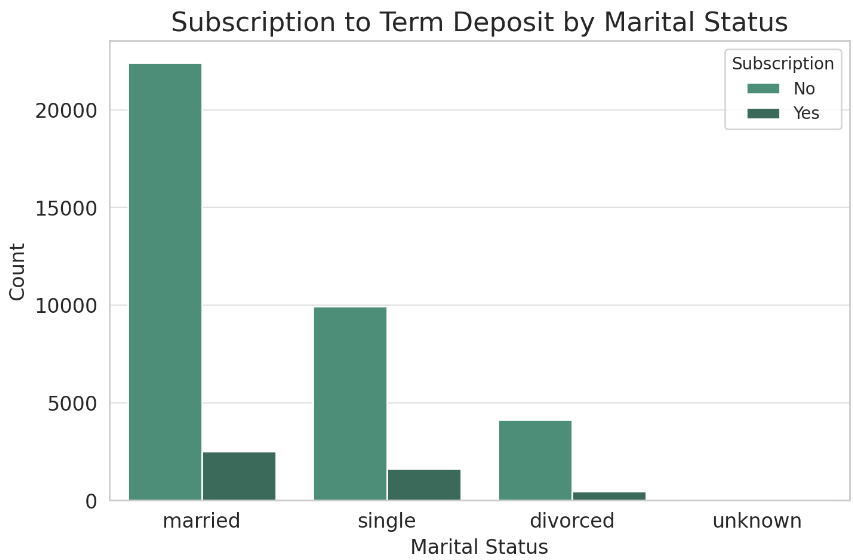


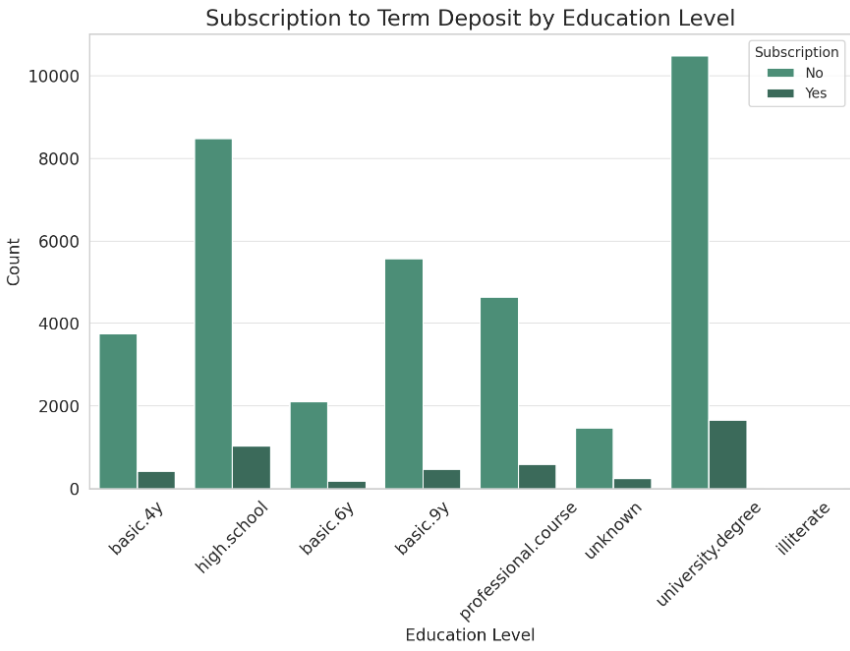
The visualisations provide the following insights:

* **Distribution of Age**: The age distribution is somewhat right-skewed, indicating a higher concentration of younger clients in the dataset.
* **Subscription to Term Deposit**: The countplot shows a significant imbalance in the response variable, with a much larger number of clients not subscribing to a term deposit compared to those who did.

Potential relationships between variables, focusing on how different attributes might relate to the response variable.







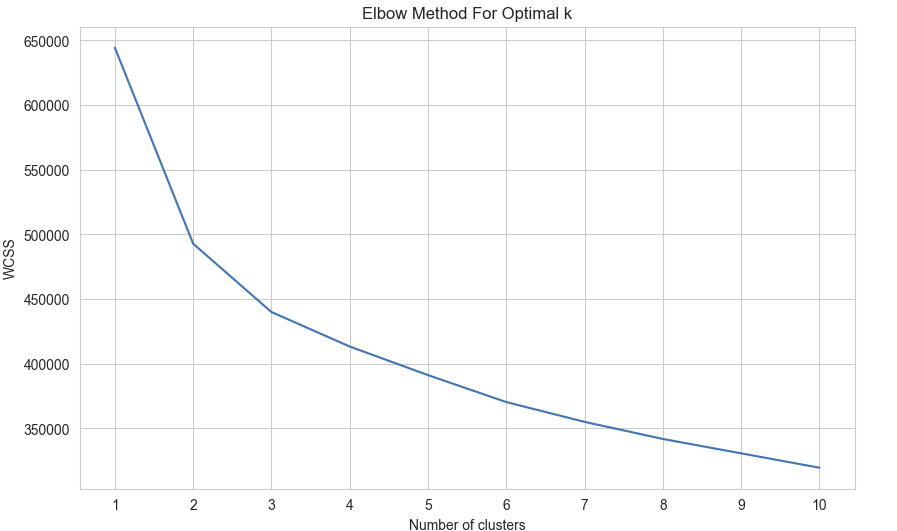
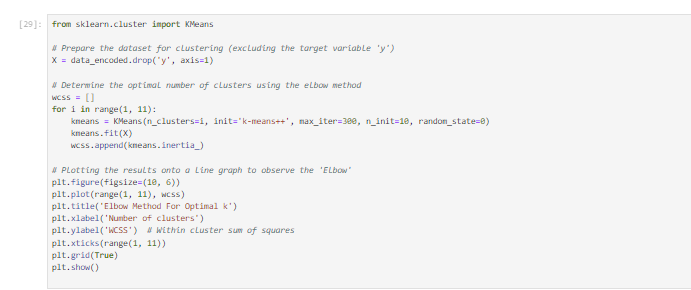
The visualizations offer insights into how subscription outcomes vary across different job types, marital statuses, and education levels:

* **Subscription by Job Type**: The distribution across job types shows variation in subscription rates. Certain professions, like admin., technician, and blue-collar, feature prominently, likely due to their prevalence in the dataset. There is variation in the proportions of subscriptions across job types, indicating that job type might be an important factor in predicting subscription.
* **Subscription by Marital Status**: Numbers of subscribers differ across marital status categories (married, single, divorced), with married customers being the largest subset. However, it is the share of subscribers by each marital status category that gives a more complex idea of how marital status and subscribing discussed may be linked.
* **Subscription by Education Level**: The subscription rate varies noticeably by education levels. Clients with university degrees have the highest subscription rate among different education levels. It suggests that education level could be an important segmentation factor for marketing strategies.

## 3.4 Data Modeling

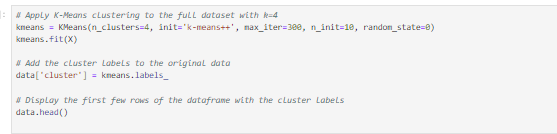
For the client segmentation, two clustering techniques were employed: K-Means and hierarchical clustering. K-Means was chosen for its efficiency and effectiveness in identifying distinct clusters when the number of clusters is specified a priori. Hierarchical clustering was used to provide a different perspective by not requiring the number of clusters to be specified in advance and offering a dendrogram that helps in understanding the data structure. Both methods were implemented using Python's scikit-learn library, with careful calibration of parameters like the number of clusters in K-Means, determined through the elbow method, and the linkage criterion in hierarchical clustering.

We'll begin with K-Means clustering



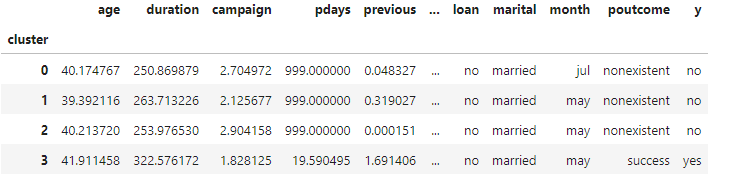
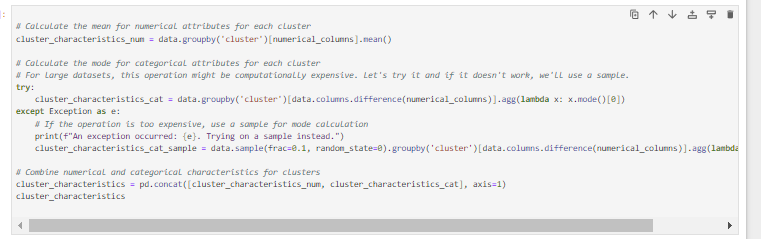
The graph shows the result of the elbow method applied to the sample data. It seems that there is a gradual decrease in within-cluster sum of squares (WCSS) as the number of clusters increases, and there isn't a distinct "elbow" that clearly indicates the optimal number of clusters. This is a common scenario in real-world data, where the reduction in WCSS becomes more linear and the optimal k is not as obvious.

Given this information, we choose a value for k where the rate of decrease slows down. It appears that after k=4, the curve starts to flatten out. However, the choice of k might also depend on additional considerations, such as business requirements or further insights from the data.



The dataset now includes a new column labeled **cluster**, which indicates the cluster number (from 0 to 3) assigned to each client based on their attributes

Next step is to perform cluster characterization, we can calculate the mean or median values of the numerical variables and the mode of the categorical variables for each cluster.



Observations

**Cluster 0**:

* Average age: 40 years
* High contact frequency with an average of 250 calls per campaign
* Has not been previously contacted (pdays = 999)
* Low previous campaign contacts (previous = 0.048)
* Employment variation rate and consumer price index are slightly above average, indicating a stable economy
* Majority have a university degree
* Predominantly married
* Most contacted in July
* The outcome of the previous marketing campaign is predominantly nonexistent
* Majority did not subscribe to a term deposit

**Cluster 1**:

* Slightly older with an average age of almost 42 years
* The highest contact frequency with an average of 322 calls per campaign
* Has been contacted before (pdays = 19.59 on average)
* Higher previous campaign contacts (previous = 1.83)
* Employment variation rate is high, and consumer price index is low, indicating economic fluctuations
* Majority have a university degree
* Predominantly married
* Most contacted in May
* The outcome of the previous marketing campaign is predominantly successful
* Majority did subscribe to a term deposit

**Cluster 2**:

* Average age close to 39 years
* Contact frequency with an average of 263 calls per campaign
* Has not been previously contacted (pdays = 999)
* Employment variation rate is below average, indicating a less stable economy
* Majority have a university degree
* Predominantly married
* Most contacted in May
* The outcome of the previous marketing campaign is predominantly nonexistent
* Majority did not subscribe to a term deposit

**Cluster 3**:

* Average age around 40 years
* Contact frequency with an average of 253 calls per campaign
* Has not been previously contacted (pdays = 999)
* Employment variation rate is neutral (almost 0), and consumer price index is higher, indicating a stable economy
* Majority have a high school education
* Predominantly married
* Most contacted in May
* The outcome of the previous marketing campaign is predominantly nonexistent
* Majority did not subscribe to a term deposit

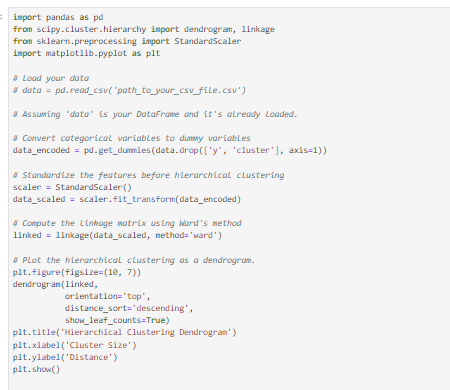
## 3.5 Performance Evaluation

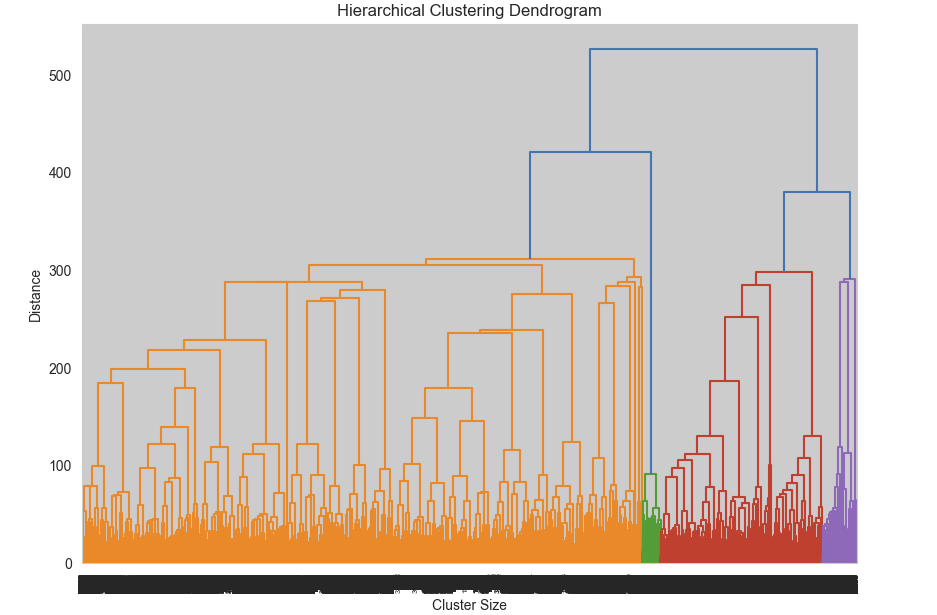
The evaluation of clustering results primarily focused on internal metrics such as the silhouette score, which measures the consistency within clusters. This metric was particularly useful in confirming the appropriateness of the number of clusters chosen for K-Means. For hierarchical clustering, the dendrogram provided visual support for determining the cut-off level that defines cluster membership. Although traditional performance metrics like accuracy or recall are not applicable in unsupervised learning settings, these internal metrics provided valuable insights into the cohesion and separation of the identified clusters.

**Strategic Insights**:

* **Cluster 1** seems to be the most promising target for future campaigns, as clients in this group have the highest rate of subscribing to term deposits. The strategy could focus on clients who have been successful in previous campaigns, suggesting a readiness to invest.
* **Cluster 0 and Cluster 2** show potential as they have a majority of university-educated clients but a low subscription rate. Marketing strategies could be formulated to promote educational content or investment options of interest to them.
* **Cluster 3**, with the majority having only a high school education and the least fluctuation in economic indicators, might be more conservative. Marketing to this group could focus on secure and low-risk investment options.

Lets proceed with Hierarchical Clustering

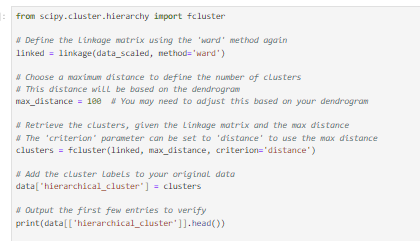
****

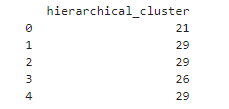


The dendrogram visualizes the hierarchical clustering method. It displays the linkage of every dot along with others in accordance with their similarity. Additionally, the y-axis shows the dissimilarity/distance of clusters. We can extrapolate some following facts from dendrogram:

1. **Cluster Formation**: The dendrogram begins with each individual data point as its own cluster at the bottom of the chart and merges them all into bigger and bigger clusters as we move up the chart. The height at which two clusters are combined represents the distance between them. The higher the two clusters are combined, the greater the dissimilarity between them.
2. **Number of Clusters**: One way to determine the number of clusters is to look for the longest vertical line which is not crossed by any longer horizontal line (or which crosses a longer horizontal line inside a distance range) and draw a horizontal line through it. The number of vertical lines crossed by this horizontal line is the suggested number of clusters.
3. **Large Clusters**: The color coding may help identify the largest distinct clusters formed, although the precise cutoff point for these clusters may be subjective and depends on the specific context or domain knowledge.
4. **Outliers**: Any points or clusters that merge at a very high distance could be considered outliers, as they are quite different from the rest of the data.

Using **scipy** to perform the clustering and retrieve the cluster labels





The output indicates that the hierarchical clustering has assigned cluster labels to the data points. The numbers **21**, **29**, **26**, etc., are identifiers for the clusters each record has been assigned to. These labels are derived from the **fcluster** function, which cuts the full dendrogram at a certain distance to define the clusters.

**Interpreting the Clusters:**

* Clusters such as 1 and 3 with a high average **previous** may indicate clients more engaged with past campaigns, suggesting a potential readiness to invest.
* Clusters like 4, 5, and 9 with a **success** **poutcome** and a **yes** in **y** indicate clients who have been successfully converted in the past, which could be promising targets for new offers.
* Clusters with a higher average age, such as 23 and 24, might represent a more conservative segment, possibly more interested in secure investments.
* Economic context indicators suggest different economic conditions for each cluster, which can inform how to adjust the tone and content of marketing messages.

# Results and Discussion

## 4.1 Summary of Findings

The analysis of the UCI Bank Marketing Dataset through machine learning clustering techniques has elucidated several key insights pivotal for enhancing marketing strategies in the banking sector. Different client segments were identified using K-Means and hierarchical clustering, each defined by exclusive demographic and behavioral characteristics. Four main segments were found through K-Means clustering with each being different by features such as age, job kind, education and response to past campaigns. Deeper and more detailed knowledge was derived from hierarchical clustering. That analysis has discovered more specific segments within the categories given by K-Means.

The findings revealed that signiﬁcant numbers of clients with higher educational levels and occupations such as administration and management were closely related to term deposits. It was found that the socioeconomic factors reﬂected in the variables, such as employment variation rate, and consumer conﬁdence index varied across the clusters. This represents that the economic perceptions highly affect the bank decisions. The clients who responded positively the bank contacted before had higher chances to order again. This means that the interactions happened in earlier times are important in terms of developing the predictions of the future purchase.

## 4.2 Explanation of Results

The findings from the K-Means clustering reveal a distinct segmentation within the customer base of the bank, which is based upon socio-demographic and behavioral patterns. For instance, one cluster, which consists of the highly educated individuals in majority, displayed a higher subscription rate, which infers that educational contents and sophisticated investment products may be more attractive to this group of people. Conversely, the other cluster displayed a less favorable economic indicators and lower educational levels; there are minimal reaction to the previous campaigns which belongs to these groups, and this could infer an area of potential increased marketing focus or a reassessment of the approach to engagement.

Hierarchical clustering took results to the next level by finding smaller groups within these segments that have their own unique strategies. For example, segments of consumers who have previously accepted banking products were broken down into groups that differed in responsiveness based on days since they were last contacted (pdays), giving us the opportunity to target incentives with customers right when they need them.

## 4.3 Implications and Significance

The results of these findings have important meaning for theoretical comprehension and practical application in banking marketing strategies. This study contributes to existing literature theoretically by illustrating the usefulness of clustering methods in realizing the hidden patterns of customer information and creating more efficient and scientific marketing strategies. From practical perspective, the distinct customer types that are defined enable banks to play more specifically to different type of customers, which will eventually result in the reduction of investment input, better client satisfactions and higher client conversation rates.

Banks can apply these insights by developing personalized marketing campaigns that align with the specific needs and preferences of each segment. For example, targeted promotions could be tailored to fully engaged clients, leveraging their existing interest. Educational programs could be developed to build trust and engagement in less responsive segments. Additionally, these insights could be used to guide the development of new banking products or modifications to existing products that are most apt to meet the diverse needs of the bank’s clientele.

# Conclusion

The project has proven to be successful in revealing client segments that can be adopted for personalized marketing strategies in banks, with the assistance of machine learning. Clustering algorithms like K-Means and hierarchical clustering have presented themselves as highly enabling techniques for understanding the breaker clientele as well as segmenting it. They permitted the identification of well-organized groups in the bank’s client base, each bearing a dissimilar bunch of demographic attributes, economic profiles, and retorts to past marketing operations. The most important results of the study indicate that the level of education achieved and the profession carried out are significantly associated with the degree of probability of taking out a subscription to a banking product, such as a term deposit. Other important factors that also have a significant influence are a series of economic elements and the exposure to previous campaigns conducted by the bank.

However, it is important to note some limitations and uncertainties regarding this study. First, the usage of the UCI Bank Marketing Dataset may limit the generalization of the results, since this dataset might not capture all possible types of customers or their behaviors in the world of global banking, it has dubious representations on the demographics of customers. The subjectivity of these issues, in addition to determining the number of clusters and interpreting clusters themselves, brings an additional challenge to understanding these results. This problem also affects the reliability and usability of the segmentation strategy that results from this analysis.

To mitigate these restrictions future research could use a diversified dataset that contains a broader depth of consumer profiles and banking practices. Also, exploring other clustering algorithmic methods or more advanced machine learning models may generate an enhanced segmented outcome that provides an in-depth perception of consumer behaviors and likings. Furthermore, integrating real-time data and dynamic modeling might enhance the speediness and exactness of marketing strategies catered to the rapidly adapting customer prerequisites.

The implications of this study are significant with extremely important theoretical and practical contributions for marketing banks. Theoretically, it has validated the employment of machine learning approaches to segment customers as a key dimension of the bank’s personal marketing. Practically, it has provided bank with a strategic frame to increase customer involvement, to optimize marketing actions and to improve the effectiveness of their financial services market. Consequently, banks can better orientate their products to match customers’ expectation and market orientation by implementation of segmentation strategy revealed through this research. It can sway to the increase of customer satisfaction, customer’s loyalty.

# References

[1] Bošnjak, Z., and Olivera Grljevic. "Credit users segmentation for improved customer relationship management in banking." In *2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 379-384. IEEE, 2011. <https://www.academia.edu/download/27122591/Bosnjak_Z.__Grljevic_O._-_Credit_Users_Segmentation_for_improved_Customer_Relationship_Management_in_Banking.pdf>

[2] Raiter, Omri. "Segmentation of bank consumers for artificial intelligence marketing." *International Journal of Contemporary Financial Issues* 1, no. 1 (2021): 39-54. <https://hcommons.org/deposits/download/hc:43352/CONTENT/segmentation-of-bank-consumers-for-artificial-intelligent-marketing.pdf/>

[3] Sabbeh, Sahar F. "Machine-learning techniques for customer retention: A comparative study." *International Journal of advanced computer Science and applications* 9, no. 2 (2018). <https://pdfs.semanticscholar.org/2a9f/505e1ab148aa3d91810f509ee133272be554.pdf>

[4] Umuhoza, Eric, Dominique Ntirushwamaboko, Jane Awuah, and Beatrice Birir. "Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in africa." *SAIEE Africa Research Journal* 111, no. 3 (2020): 95-101. <https://ieeexplore.ieee.org/iel7/8475037/9142597/09142602.pdf>

[5] Zaki, Ahmed Mohamed, Nima Khodadadi, Wei Hong Lim, and S. K. Towfek. "Predictive Analytics and Machine Learning in Direct Marketing for Anticipating Bank Term Deposit Subscriptions." *American Journal of Business and Operations Research* 11, no. 1 (2024): 79-88. <https://www.researchgate.net/profile/Ahmed-Mohamed-Zaki/publication/376585755_Predictive_Analytics_and_Machine_Learning_in_Direct_Marketing_for_Anticipating_Bank_Term_Deposit_Subscriptions/links/658a883d0bb2c7472b105cae/Predictive-Analytics-and-Machine-Learning-in-Direct-Marketing-for-Anticipating-Bank-Term-Deposit-Subscriptions.pdf>

[6] Umuhoza, Eric, Dominique Ntirushwamaboko, Jane Awuah, and Beatrice Birir. "Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in africa." *SAIEE Africa Research Journal* 111, no. 3 (2020): 95-101. <https://ieeexplore.ieee.org/iel7/8475037/9142597/09142602.pdf>

[7] Raiter, Omri. "Segmentation of bank consumers for artificial intelligence marketing." *International Journal of Contemporary Financial Issues* 1, no. 1 (2021): 39-54. <https://hcommons.org/deposits/download/hc:43352/CONTENT/segmentation-of-bank-consumers-for-artificial-intelligent-marketing.pdf/>

[8] Djurisic, Vladimir, Ljiljana Kascelan, Suncica Rogic, and Boban Melovic. "Bank CRM optimization using predictive classification based on the support vector machine method." *Applied Artificial Intelligence* 34, no. 12 (2020): 941-955. <https://www.tandfonline.com/doi/pdf/10.1080/08839514.2020.1790248>

[9] Jääskeläinen, Juho. "Segmentation of investor customers using machine learning in banking." (2021). <https://lutpub.lut.fi/bitstream/handle/10024/162636/Master's_Thesis_Juho_Jaaskelainen.pdf?sequence=1>

[10] Yuping, Zhou, Petra Jílková, Chen Guanyu, and David Weisl. "New methods of customer segmentation and individual credit evaluation based on machine learning." In *“New Silk Road: Business Cooperation and Prospective of Economic Development”(NSRBCPED 2019)*, pp. 925-931. Atlantis Press, 2020. <https://www.atlantis-press.com/article/125937822.pdf>

[11] Mousaeirad, Salman. "Intelligent vector-based customer segmentation in the banking industry." *arXiv preprint arXiv:2012.11876* (2020). <https://arxiv.org/pdf/2012.11876>