ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

PREDICTING CUSTOMERS WITH HIGHER PROBABILITY TO PURCHASE IN TELECOM INDUSTRY

M.Sc. THESIS

Güzide Nur YILDIRIM

Department of Data Engineering and Business Analytics

Big Data and Business Analytics Programme

JANUARY 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

PREDICTING CUSTOMERS WITH HIGHER PROBABILITY TO PURCHASE IN TELECOM INDUSTRY

M.Sc. THESIS

Güzide Nur YILDIRIM
(528211085)

Department of Data Engineering and Business Analytics

Big Data and Business Analytics Programme

Thesis Advisor: Assist. Prof. Mehmet Ali ERGÜN

JANUARY 2025

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

TELEKOM SEKTÖRÜNDE SATIN ALMA İHTİMALİ YÜKSEK OLAN MÜŞTERİLERİN TAHMİN EDİLMESİ

YÜKSEK LİSANS TEZİ

Güzide Nur YILDIRIM
(528211085)

Veri Mühendisliği ve İş Analitiği Anabilim Dalı

Büyük Veri ve İş Analitiği Programı

Tez Danışmanı: Dr. Öğr. Üyesi Mehmet Ali ERGÜN

OCAK 2025

Güzide Nur YILDIRIM, a M.Sc. student of ITU Graduate School student ID 506181437, successfully defended the thesis/dissertation entitled " PREDICTING CUSTOMERS WITH HIGHER PROBABILITY TO PURCHASE IN TELECOM INDUSTRY", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor:	**Assist. Prof. Mehmet Ali ERGÜN** Istanbul Technical University	....
Jury Members :	**Assoc. Prof. Ömer Faruk BEYCA** Istanbul Technical University	....
	Prof. Dr. Selim ZAİM Ibni Haldun University	....

Date of Submission : 31 December 2024
Date of Defense : 22 January 2025

To my dear husband Yasin, whose love and encouragement have been my greatest support, and to my precious son Emir Bera, who fills our lives with endless love and joy.

FOREWORD

I would like to express my heartfelt gratitude to everyone who supported me during this thesis. First, I am deeply thankful to my supervisor, Prof. Dr. Mehmet Ali ERGÜN for his guidance and valuable insights. His support played a key role in helping me complete this work.

I am especially grateful to my parents, Yurda DEMİR and Selim DEMİR for their endless love and encouragement. They have always believed in me and stood by my side. My deepest thanks go to my wonderful husband, M. Yasin YILDIRIM and my son, Emir Bera YILDIRIM who have been my greatest source of strength and motivation. Their love, patience, and unwavering support made everything possible, even during the most challenging times.

Thank you all for being part of this journey and for making this accomplishment possible.

December 2024 Güzide Nur YILDIRIM
(Data Scientist (M.Sc.))

TABLE OF CONTENTS

Page

ABBREVIATIONS

AI	**:** Artificial Intelligence
B2C	**:** Business to Customer
CLV	**:** Customer Lifetime Value
CRM	**:** Customer Relationship Management
GSM	**:** Global System for Mobile Communications
LTV	**:** Lifetime Value
PCA	**:** Principal Component Analysis
RFM	**:** Recency, Frequency and Monetary
RFMD	**:** Recency, Frequency, Monetary and Discount Price

LIST OF TABLES

Page

LIST OF FIGURES

Page

PREDICTING CUSTOMERS WITH HIGHER PROBABILITY TO PURCHASE IN TELECOM INDUSTRY

SUMMARY

Understanding customer insights, predicting their next steps, and making accurate recommendations have become a significant issue for customer service organizations in today's world. With the widespread use of important technologies such as Big Data Analytics and Machine Learning, the contribution of Artificial Intelligence to digital transformation in the industrial world has significantly affected the way businesses work. The use of Artificial Intelligence technologies in business processes has increased the success ratio of marketing strategies, business models and sales statistics. Companies that provide customer services have units responsible for CRM, as known as customer relations management. Big Data accumulated and analyzed by CRM units can be modeled with artificial intelligence algorithms for customer centric use cases. Thereby, outcome of such an approach guide the customer strategies built by companies. In this article, studies on the business analytic-based understanding of customers with high probability of purchasing by training the CRM data used in the telecom industry will be included.

TELEKOM SEKTÖRÜNDE SATIN ALMA İHTİMALİ YÜKSEK OLAN MÜŞTERİLERİN TAHMİN EDİLMESİ

ÖZET

Bu tezde, telekomünikasyon sektöründe müşterilerin satın alma davranışlarını analiz ederek daha etkili stratejiler geliştirmek için kullanılan yöntemler ele alınmıştır. Özellikle müşteri segmentasyonu, şirketlerin müşterilerini daha iyi anlamasına ve onlara özel kampanyalar hazırlamasına yardımcı olan önemli bir araçtır. Bu çalışmada, müşteri segmentasyonu için RFM (Recency, Frequency, Monetary) analizi temel olarak ele alınmış ve etkili sonuçlar elde edebilmek adına farklı metodlar kullanılarak desteklenmiştir.

RFM analizi, müşterilerin son satın alma tarihlerini (recency), satın alma sıklıklarını (frequency) ve toplam harcama miktarlarını (monetary) göz önünde bulundurarak müşterileri 5 farklı segmente ayırmaktadır. Örneğin, son dönemde alışveriş yapmış, sık alışveriş yapan ve yüksek miktarda harcama yapan müşteriler genelde daha değerli olarak kabul edilir. Bu analiz yöntemi, müşterilerin farklı gruplara ayrılmasını ve bu gruplara özel pazarlama stratejileri geliştirilmesini sağlar. Çalışmada, Türkiye'de faaliyet gösteren bir telekomünikasyon şirketine ait 13.627 müşteriye ait veri kullanılarak analiz çalışması yapılmıştır. Veri seti içerisinde müşteri numarası, fatura bilgisi, kampanyalar ve ücretleri gibi detaylı bilgiler içermekte ve müşteri davranışlarını anlamak için bu bilgiler zengin bir kaynak sunmaktadır.

Bu tür CRM verilerini analiz etmek için yapay zeka ve veri madenciliği yöntemlerini kullanmak oldukça faydalı bir yöntemdir. Büyük ve karmaşık veri setlerini anlamlandırmak ve etkili analiz gerçekleştirebilmek için bu yöntemleri kullanmak oldukça önemlidir. Örneğin, bir müşterinin sadık bir müşteri mi yoksa şirketle ilişkisini kesmek üzere olan bir müşteri mi olduğunu anlamak bu analizlerle mümkün hale gelir. Yapılan bu analiz çalışmaları sayesinde, şirketler müşterilerinin davranışlarına bağlı olarak müşteri bağlılığını arttırmak veya müşterinin ayrılmasını engellemek amacıyla belirli stratejiler geliştirebilir. Ayrıca, müşterilerin ihtiyaç ve beklentilerini önceden tahmin ederek daha etkili kampanyalar düzenlenmesine yardımcı olur. Bu sayede müşteri memnuniyeti artarken, müşteri kayıpları da önemli ölçüde azaltılmış olur.

Bu çalışmada, veriler detaylı bir ön işleme sürecinden geçirilmiştir. İlk olarak, eksik veriler tamamlanmış ve tutarsızlıklar giderilmiştir. Veri setinin daha tutarlı bir yapıda olabilmesi için aykırı veriler temizlenmiştir. Daha sonra, normalizasyon tekniği uygulanarak veri seti analiz işlemine uygun hale getirilmiştir. Yapılacak olan müşteri segmentasyonu analizinde daha doğru sonuç alabilmek için RFM analizine ek olarak Ana Bileşen Analizi (PCA) ve K-means algoritmaları kullanılmıştır. K-Means algoritması, RFM metrik bileşenlerini gruplandırmak amacıyla kullanılırken, PCA algoritması ise bu RFM skoru hesaplanırken bu bileşenlere doğru ağırlık ataması yapılabilmesi için kullanılmıştır. Bu yaklaşım, analiz sürecinde daha doğru ve etkili bir sonuç elde edilmesini sağlamıştır.

Boyut indirgeme işlemi, karmaşık veri yapılarını daha basit bir forma dönüştürür ve analizin doğru yapılmasına yardımcı olur. Özellikle PCA ile RFM bileşenlerinin önem sıralaması belirlenmiş, bu sıralama RFM skorlarının hesaplanmasında kullanılmıştır. Böylece, müşterilerin daha doğru bir şekilde gruplara ayrılması sağlanmıştır. Bu süreç, müşteri davranışlarının daha net bir şekilde analiz edilmesine ve elde edilen her bir müşteri grubu için farklı stratejiler geliştirilmesine katkı sağlamıştır.

Yapılan RFM analizi sonucunda, müşteriler beş ayrı gruba ayrılmıştır: şampiyonlar, potansiyel sadıklar, dikkat gerektirenler, riskliler ve uyuyanlar. Bu grupların her biri, müşterilerin satın alma davranışları ve şirketle olan ilişkilerine göre belirlenmiştir. Şampiyonlar grubu, sık alışveriş yapan, yüksek harcamalar yapan ve şirkete aktif katkı sağlayan müşterilerden oluşmaktadır. Bu müşteriler, şirket için en değerli grup olarak görülür. Diğer bir yandan, potansiyel sadıklar, alışveriş sıklığı ve harcama düzeyleri açısından şampiyonlara en yakın olan gruptur. Fakat, yine de bağlılıklarının arttırılması için bu gruba özel stratejiler geliştirilmelidir.

Dikkat gerektiren grup, şirketle olan ilişkisi zayıflamakta olan ancak doğru müdahaleler ile tekrar kazanılabilecek bir gruptur. Bu grup, müşteri memnuniyeti ve bağlılığının artırılmasına yönelik stratejik çalışmaların yapılması gereken bir segmenttir. Riskli ve uyuyan segmentler ise, şirket için alarm verici müşteri gruplarını temsil etmektedir. Riskli grupta yer alan müşteriler şirketle arasında bağlılığı azalan, her an gitmeye yakın olanları temsil eder. Uyuyan gruptaki müşteriler ise uzun süredir herhangi bir satın alım yapmamış ve şirketle bağlarını tamamen koparmak üzere olanlardır. Bu grupların tekrar kazanılması şirketlerin özel stratejiler geliştirmesi, müşterilere uygun kampanya çalışmaları yapması gerekmektedir.

Her bir müşteri grubun için şirketin özel stratejiler geliştirmesi önemli bir konudur. Şampiyon müşterilere, bağlılıklarını daha da artıracak kişiselleştirilmiş ödüller ve kampanyalar sunulabilir. Bu müşterilerin mevcut memnuniyet seviyeleri yüksek olduğundan, küçük dokunuşlarla onların uzun vadeli destekçileri olmaları sağlanabilir. Potansiyel sadık müşteriler için ise, özel teklifler ve onları daha fazla alışverişe teşvik edecek iletişim stratejileri uygulanabilir. Bu gruptaki müşteriler, doğru yaklaşımlarla şampiyonlar segmentine taşınabilecek bir potansiyele sahiptir.

Ayrıca, dikkat gerektiren müşteri grubuna hızlı ve etkili müdahaleler yapılması oldukça önemlidir. Bu müşterilere yönelik iletişim stratejileri geliştirilmeli, onların şirketle olan ilişkilerinin güçlendirilmesi sağlanmalıdır. Riskli ve uyuyan müşteriler için daha dikkatli olunması gerekmekte ve bu müşteri grubu için net bir geri kazanım politikası çalışılmalıdır. Bu müşterilere sunulacak ilgi çekici kampanyalar veya onların ihtiyaçlarına yönelik yaklaşımlar, müşteri kaybını önleyebilir.

Yapılan analiz çalışmasının sonucunda, müşterilerin davranışları doğru bir şekilde gözlemlenerek, şirketlerin oluşturulan müşteri gruplarına yönelik strateji geliştirmeleri sağlanmıştır. Elde edilen sonuçlara dayanarak, şampiyon ve potansiyel sadık gruplara ait müşterilerin bağlılığı artırılabilir. Bu gruplar, şirketin en yüksek potansiyele sahip müşterilerini içerir. Şampiyon müşterilere sunulacak özel ödüller ve kişiselleştirilmiş teklifler, onların şirketle olan bağlarını daha da arttırır. Potansiyel sadıklar grubundaki müşteriler ise, birebir iletişim ve özel kampanyalarla sadık müşteriler haline dönüştürülebilir. Bu segmentler, şirketin büyüme hedefleri doğrultusunda en önemli katkıyı sağlayacak gruplardır.

Dikkat gerektiren segment için ise daha hızlı ve etkili adımlar atılması gereklidir. Bu gruptaki müşterilerin memnuniyeti artırılarak, onların şirketle olan ilişkilerinin kopmaması sağlanabilir. Örneğin, bu müşterilere yönelik özel kampanyalar düzenlenebilir veya onların ihtiyaçlarına yönelik çözümler sunulabilir. Bu tür müdahaleler, bu gruptaki müşterilerin riskli veya uyuyan segmentlere geçmesini önleyebilir.

Riskli ve uyuyan gruplar, şirketin en zorlu müşteri segmentleridir ve bu gruplara yönelik stratejiler daha proaktif olmalıdır. Riskli müşteriler için, birebir iletişim stratejileri veya özel geri kazanım kampanyaları uygulanabilir. Uyuyan müşteriler içinse, şirketle yeniden bağ kurmalarını sağlayacak cazip teklifler sunulabilir. Bu tür müşterilere, ilgilerini çekecek özel hizmetler veya avantajlar sağlamak, onları tekrar kazanmanın en etkili yollarından biridir. Bu segmentler üzerinde yapılan çalışmalar, müşteri kaybını önlemenin yanı sıra uzun vadede gelir artışına da katkı sağlayabilir.

Bu tezde kullanılan RFM analizi ve boyut indirgeme yöntemleri, müşteri segmentasyonu için etkili bir temel oluşturmuştur. Özellikle, RFM göstergelerinin PCA ve K-means algoritmalarıyla işlenmesi hem analiz süreçlerinin doğruluğunu artırmış hem de verilerin daha anlaşılır hale gelmesini sağlamıştır. Bu yöntemler, büyük ve karmaşık veri kümelerinde bile etkili sonuçlar alınmasına olanak tanımaktadır. Ayrıca, segmentasyon süreci boyunca kullanılan bu teknikler, müşteri davranışlarını daha derinlemesine anlamayı mümkün kılar.

Sonuç olarak, bu çalışma, telekomünikasyon sektöründe müşteri segmentasyonu için uygulanabilir bir model sunmaktadır. Çalışmada elde edilen sonuçlar, şirketlerin müşteri ilişkileri yönetimini daha etkili hale getirmek için veri odaklı yaklaşımları nasıl kullanabileceğini göstermiştir. Bu segmentasyon modeli, sadece telekomünikasyon sektörüyle sınırlı kalmayıp, müşteri davranışlarının önemli olduğu diğer sektörlerde de kullanılabilir. Gelecekte, bu çalışmaya derin öğrenme ve diğer gelişmiş yapay zeka yöntemlerinin eklenmesi, müşteri segmentasyonu süreçlerinin daha da geliştirilmesine olanak sağlayabilir. Bu da hem müşteri memnuniyetinin artırılmasına hem de şirketlerin iş hedeflerine ulaşmasına önemli katkılar sunacaktır.

1. INTRODUCTION

1.1 Overview

With the development of digitalization, companies that run businesses with customer-centric understanding have shown the significance of technology-oriented approaches. Many of the large telecom companies operating today have Customer Relationship Management (CRM) departments. Customer transaction data are stored by the CRM units. These data are also known as big data which is processed with data analyzing methods to determine necessary strategic moves accordingly since knowing the customers, serving their needs better and understanding the next move of a customer using the big data is an important issue in the business-to-customer (B2C) industries. Various data mining methods and machine learning algorithms can be used by such companies for this purpose since the dataset kept in the CRM systems includes the entire history of customers, or in other words, the journey of customer.

Lemon and Verhoef (2016) claims that the customer journey can be defined as the different interactions between customers and companies during the process of making a purchase. Increasing communication channels and engagement enables the customer journey to be described individually by measuring actions, movements and behavior. In a previous research, Santos and Morales (2013) defines that customers propensity to buy certain products has successfully been predicted using statistical models in multiple industries.

To further enhance customer segmentation strategies, integrating advanced machine learning techniques into traditional approaches, such as RFM analysis, has proven effective in improving prediction accuracy and scalability. Recent studies emphasize that leveraging hybrid methods, including dimensionality reduction techniques like Principal Component Analysis (PCA), can avoid challenges associated with high-dimensional data and enable businesses to better identify actionable customer insights (Alkhayrat et al., 2020). Moreover, customer retention strategies have been shown to be significantly more cost-effective than acquiring new customers since retained customers contribute higher long-term value to businesses (Wu et al., 2021). These approaches underscore the critical importance of combining data-driven methods with strategic frameworks to maximize customer engagement and business outcomes.

Briefly, the research offers different statistical models applied to relate the data in terms of the companies' requirements.

1.2 Purpose

As stated above, the insights can be gained by applying a similar methodology on a dataset that contains journey of customers and various actions such as product recommendation and customer segmentation. These actions can be taken by companies with the aim of having a positive reflection to certain type of business metrics such as customer engagement. There are lots of customer segmentation methodologies and one of the well-known customer segmentation models is RFM model. RFM is used for customer value analysis in terms of the customer journey in the company. The RFM value, which is determined by calculating the recency, frequency and monetary values, helps to determine the best customer group by segmenting the customers.

This study applies RFM analysis to customers who have made purchases of different campaigns at various times. By segmenting customers based on their recency, frequency and monetary value of purchases, the study aims to identify distinct customer groups with unique behavioral patterns. Additionally, the study determines the best customer group for future purchases, enabling targeted marketing strategies and personalized campaigns. This research provides valuable insights into customer behavior and facilitates the development of effective marketing strategies in the telecom industry.

The data used in this article is sourced from a prominent telecommunications company in Turkey, known for its internet infrastructure and GSM services. The article focuses on exploring machine learning-based methodologies for prioritizing customers with a high probability of making purchases. By leveraging these methodologies, the study aims to develop effective strategies for targeting and engaging customers in the telecom industry.

2. LITERATURE REVIEW

In the study of customer segmentation, different methods have emerged to understand customer behavior for targeted marketing. In this section, the study explores various customer segmentation methodologies such as RFM analysis for a detailed and comprehensive exploration of this field.

2.1 Customer Segmentation in Telecom Industry

Customer segmentation in the telecom industry plays a critical role in understanding diverse customer behaviors and developing tailored marketing strategies. The application of segmentation models enables telecom operators to classify their customer base into distinct groups based on specific criteria, such as usage patterns, engagement levels and purchasing behaviors.

One effective approach to customer segmentation is the use of hybrid models that integrate advanced analytics techniques. Posedel Šimović et al. (2021) introduced a mixed-penalty logistic regression model to analyze customer behavior and predict churn within telecom companies. Their findings underscored the importance of segmenting customers based on multidimensional behavior data to enhance prediction accuracy and prevent churn. This approach highlights how segmentation models can contribute to a deeper understanding of customer value by allowing telecom operators to maintain customer loyalty.

Another innovative approach is the integration of graph-based analytics for customer segmentation. Henna and Kalliadan (2021) demonstrated the effectiveness of using graph databases and graph-based deep learning techniques to identify patterns and relationships in customer data. This method provides a visual representation of customer interactions and engagement. Thereby, it allows telecom companies to identify high-value customer groups and optimize resource allocation for targeted campaigns.

These studies illustrate the value of applying segmentation methods designed for the unique dynamics of the telecom industry. By leveraging advanced analytics techniques such as hybrid models and graph-based learning, telecom companies can identify actionable customer insights, reduce churn rates and enhance customer engagement. These insights form the foundation for building effective customer relationship management strategies and achieving long-term business success.

2.2 Artificial Intelligence in CRM and Big Data Analytics

Artificial Intelligence (AI) integrated with Customer Relationship Management (CRM) systems has redefined how telecom companies analyze customer data and create strategic initiatives. AI-powered CRM systems enable businesses to process vast amounts of data, uncover hidden patterns and predict customer behaviors with greater accuracy. The adoption of these systems has become essential in the telecom industry where customer retention and satisfaction are critical to success.

Ledro et al. (2021) emphasize that AI applications in CRM systems significantly enhance customer engagement by leveraging predictive analytics and machine learning algorithms. These tools lead telecom companies to personalize customer interactions, predict churn and design targeted marketing campaigns based on individual preferences and behaviors. AI-driven CRM systems also facilitate real-time decision-making by analyzing customer data streams for ensuring timely responses to customer needs.

Big Data Analytics further strengthens CRM capabilities by offering insights derived from structured and unstructured data sources. Etemadi et al. (2024) highlight that Big Data solutions provide telecom companies with a comprehensive view of customer journeys from initial contact to post-purchase engagement. By integrating AI with Big Data, CRM systems can process high-dimensional data and deliver actionable insights for customer segmentation, loyalty programs and revenue optimization.

The combination of AI and Big Data Analytics is particularly effective in addressing challenges unique to the telecom sector such as high churn rates and competitive pressures. By automating routine tasks and enabling predictive capabilities, AI-driven CRM systems reduce operational costs, enhance the effectiveness and optimization of marketing strategies. These systems enable telecom companies to not only understand customer behavior but also predict future trends. Therefore, they offer proactive engagement and improved customer satisfaction.

In summary, the use of AI and Big Data Analytics within CRM systems delivers various opportunities to telecom companies for enhancing customer relationships. Through advanced data processing, predictive modeling and personalized marketing strategies, these technologies transform traditional CRM practices into dynamic systems increasing competitive advantage.

2.3 Predictive Analytics in Customer Behavior

Predictive analytics enhanced understanding and forecasting customer behavior due to its use by telecom companies to understand customer needs and improve decision-making. By leveraging advanced statistical and machine learning techniques, predictive analytics help businesses to identify patterns in customer data and make strategies that enhance customer satisfaction and loyalty.

Etemadi et al. (2024) highlight the role of predictive analytics in CRM and emphasize its ability to process vast amounts of historical and real-time data to forecast customer behaviors such as purchase likelihood, churn probability and engagement trends. These insights are invaluable for telecom companies that need to design proactive strategies such as personalized marketing campaigns and tailored retention initiatives. The application of predictive analytics not only reduces churn but also optimizes resource allocation by focusing efforts on high-value customer segments.

Another significant application of predictive analytics offers ability to assess and predict customer lifetime value (CLV). By analyzing past purchasing behaviors and estimating future transactions, predictive models provide telecom companies with actionable insights into the long-term value of customers. These insights enable the prioritization of high-value customers and the development of loyalty programs to strengthen their relationship with the brand.

The integration of predictive analytics in customer behavior modeling also extends to real-time decision-making. AI-powered analytics systems process live data streams to predict customer actions and preferences dynamically for ensuring that telecom companies can respond to changes in customer behavior. This capability enhances customer experience and long-term engagement.

In conclusion, predictive analytics serves as a tool in understanding and influencing customer behavior in the telecom industry. By combining advanced data processing techniques with behavioral modeling, telecom companies can identify key trends, predict future behaviors and implement targeted strategies for customer satisfaction and business growth.

2.4 Customer Lifetime Value

“Customer Lifetime Value (CLV or LTV) is defined as the sum of revenues gained from a company's customers over their lifetime of transactions after deducting the total cost of attracting, selling and servicing these customers while considering the time value of money.” (Hwang, Jung, & Suh, 2004).

“Colombo and Jiang (1999) developed a stochastic Recency Frequency Monetary model to rank customers in terms of their expected contribution” (Hwang, Jung, & Suh, 2004). In addition to that, “Chan (2008) proposed an approach that combines customer targeting and customer segmentation for campaign strategies using RFM to identify customer behavior and a CLV model to evaluate the proposed segmented customers” (Wei et al., 2010). Furthermore, Sohrabi and Khanlari (2007) utilized the K-means clustering algorithm to develop a CLV model where the model determined customers' CLV by taking RFM measures into account by providing a segmentation based on the projected lifetime value.

To sum up, integrating CLV with RFM analysis can enhance the understanding of customer behavior and their long-term value for customer segmentation. However, “the evaluations of customer value in previous studies have treated prediction method with regression models simply based on profits from customers to calculate the future value of customers. That is to say, considering the changing profit contribution obtained from customers in the past, the existing models calculate the future worth and then define the CLV of customers with the projected value of the future worth. Therefore, the CLV model […] is not capable of considering potential values of customers, not available from the past profit contribution, which would be able to be the profits of companies” (Hwang, Jung, & Suh, 2004).

2.5 Algorithms

2.5.1 RFM analysis

RFM (Recency, Frequency, Monetary) analysis is a commonly used method in customer segmentation for enabling businesses to evaluate customer value based on three key metrics: recency (time since the last purchase), frequency (number of purchases within a given period) and monetary value (total spending). This model assigns scores to each metric to rank customers and group them into distinct segments. By prioritizing customers with high recency, frequency and monetary scores, RFM analysis offers companies to identify loyal and profitable customers, optimize marketing strategies and enhance customer engagement. Its simplicity and adaptability make RFM analysis a foundational tool for data-driven decision-making in the telecom industry.

The assignment of weights to recency, frequency and monetary indicators is crucial in customer segmentation estimation using RFM analysis, especially when dealing with large datasets. To explore the usefulness of RFM in grouping customers, valuable research in the field can be considered. For instance, Cheng and Chen (2009) propose an altered version of RFM that is called the RFMD model. The "D" represents discount-price products, allowing for clear identification of customer value within the RFMD model. Cheng and Chen's (2009) work highlights the effectiveness of the RFMD model in understanding customer value, particularly in relation to discount-price products. Similarly, Chiang (2015) combines RFM analysis with a different model by presenting an expanded version of the analysis in addition to the typical RFM analysis. This hybrid approach uses additional factors and rules to enhance the identification of customers and their behavioral patterns.

Kim et al. (2006) highlighted the importance of segmenting customers based on profitability. They argue that highly profitable customer segments should be targeted through retention and loyalty programs while unprofitable customer segments should be reconsidered for marketing efforts to ensure a balance between cost and benefit. This perspective enhances the understanding of how customer profitability can drive segmentation strategies and business decisions. Moreover, Jahromi et al. (2014) emphasize the dependency of segmentation methods like RFM on advanced data mining techniques by claiming that machine learning algorithms perform better than traditional statistical approaches in identifying non-parametric patterns within large datasets.

The studies contribute to the understanding of RFM analysis and its applications in customer segmentation. By considering the insights from these literature reviews along with the methodological approaches of PCA and K-means clustering, the weight assignment process in this thesis ensures that the resulting weights accurately reflect the underlying structure and significance of the indicators within the dataset. This comprehensive approach contributes to a deeper understanding of customer segmentation and aims to develop effective strategies for targeting customers with higher probabilities of purchase.

2.5.2 PCA

Principal Component Analysis (PCA) is a dimensionality reduction method that simplifies complex datasets by converting multiple correlated variables into a smaller set of independent variables, known as principal components. Thereby, there are previous studies utilized PCA to embed data into fewer dimensions by preserving key information while reducing computational complexity (Alkhayrat et al., 2020). This method is particularly useful in customer segmentation where it reduces redundancy and computational complexity to enable clustering algorithms like K-means to operate more effectively.

For instance, Wu et al. (2020) use PCA to indicate weights for the RFM model in their research. “Principal component analysis […] transforms multiple indicators into a few comprehensive ones through dimensionality reduction technique. The weight of each indicator is equal to the variance contribution rate of the principal component. The greater the variance contribution rate, the higher the importance of the principal component.” (Wu et al., 2020).

Further supporting the role of PCA in customer segmentation, Alkhayrat et al. (2020) explored how dimensionality reduction can improve the efficiency of clustering algorithms. By transforming raw data into a compact feature space, PCA minimizes redundancy and computational overhead. Thence, it is a valuable tool for analyzing high-dimensional customer data. This transformation allows clustering methods like K-means to operate more effectively on representative features, thereby enhancing the interpretability and scalability of segmentation models.

2.5.3 K-means

The K-means clustering algorithm is used to group data points into a predefined number of clusters based on their similarity. The algorithm assigns each data point to the closest cluster center and iteratively updates the cluster centers until they reach a stable configuration. It is often used in customer segmentation to find patterns in behaviors like buying habits by grouping similar customers. Figure 2.1 shows the process of the K-Means algorithm.

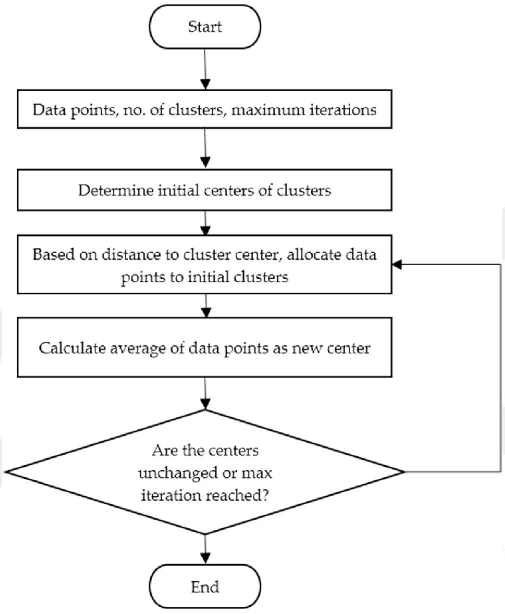

Figure 2.1 : Flowchart of the K-means algorithm.

The K-means algorithm is used together with RFM analysis to develop targeted marketing strategies. The algorithm integrated with RFM analysis provides a robust framework for customer segmentation. Cheng and Chen (2009) introduced a combination of the RFM model and K-means clustering to perform customer relationship management for demonstrating that this combined approach effectively analyzes customer value (Cheng & Chen, 2009; Mesforoush & Tarokh, 2013; Wu et al., 2020).

The K-means algorithm is preferred over other clustering algorithms due to its speed and efficiency in computation as well as its ability to reduce the misclassification rate of data. However, the accuracy of the K-means algorithm depends significantly on the initialization conditions and the number of clusters chosen (Wu et al., 2020; Shahri et al., 2019; Namasudra & Roy, 2018; Jiang et al., 2017; Stone, 1995).

Olle Olle and Cai (2014) extend the application of K-means by depending on hybrid models for predictive data mining. Their approach integrates clustering with classification techniques such as Decision Trees to reduce dimensionality before applying clustering algorithms. This sequential hybrid methodology not only optimizes the segmentation process but also facilitates the detection of new patterns within labeled customer groups. Thus, it produces more actionable insights.

3. METHODOLOGY

The methodology used in this study is a data-driven segmentation technique called RFM (Recency-Frequency-Monetary) analysis which is widely used in customer segmentation studies (Verhoef, Franses, & Hoekstra, 2001). During the customer segmentation process, accurately ranking customers based on their likelihood of purchasing a product becomes crucial. RFM analysis achieves this by considering customers' past behaviors within the company. It utilizes three key indicators: recency, representing the time passed since the latest purchase; frequency, indicating the rate of acquisition within a specific period; and monetary, reflecting the amount spent on product purchases. By leveraging these factors, RFM analysis provides valuable insights into customer behavior and supports effective customer segmentation and targeted marketing strategies.

3.1 Data

This study is based on a dataset comprising 13,627 customer transaction records, covering a period from August 24, 2022 to November 27, 2022. These records provide valuable insights into customer behavior over a three-month span. The dataset includes several key features such as CustomerID, InvoiceID, InvoiceDate, CampaignCode, CampaignName, Price, Quantity, and CityID, each offering unique information for customer segmentation analysis. The dataset comes from a well-known Turkish telecom company and is derived from its Customer Relations Management (CRM) system. It focuses on transactions related to products and promotional campaigns connected with internet services. The Table 3.1 below provides a breakdown of these features, their data types and descriptions for understanding the comprehension of the dataset's structure and content.

Table 3.1 : Dataset description.

#	Features	Data Format	Description
1	CustomerID	numerical	Customer's unique identification number.
2	InvoiceID	numerical	Invoice's unique identification number.
3	InvoiceDate	timestamp	Invoice creation date.
4	CampaignCode	string	Unique identification number of service purchased by customer.
5	CampaignName	string	Name of service purchased by customer.
6	Price	numerical	Total amount charged to customer.
7	Quantity	numerical	Total number of services purchased within invoice.
8	CityID	numerical	Customer's registered location's unique identification number.

3.2 Data Preprocessing

In a big data analysis, initial set of data collected is referred as the raw data and it is required to be preprocessed with the aim of enhancing accuracy of the analysis outcome. Thus, the data in this study is cleansed after detecting duplicate and erroneous values to make the features understandable and proper for use during the analysis. Afterwards, the following preprocessing steps are applied:

- Outliers are data points significantly different from other observations. Thereby, outliers and missing values are removed manually based on a domain-specific knowledge.
- Values that broke the distribution but are not outliers are manually edited in a way that they don't break the integrity of the data. For instance, as shown in Figure 3.1, data instances with extremely high frequency values compared to the general distribution are set to a certain value.
- Data normalization which is an application of scaling to ensure that values are shifted into a specific range is used. In this study's case, the range is between 1 and 5. Therefore, the three numerical indicators of RFM are scored to end up within a standard interval for the later use in RFM score calculation formula.

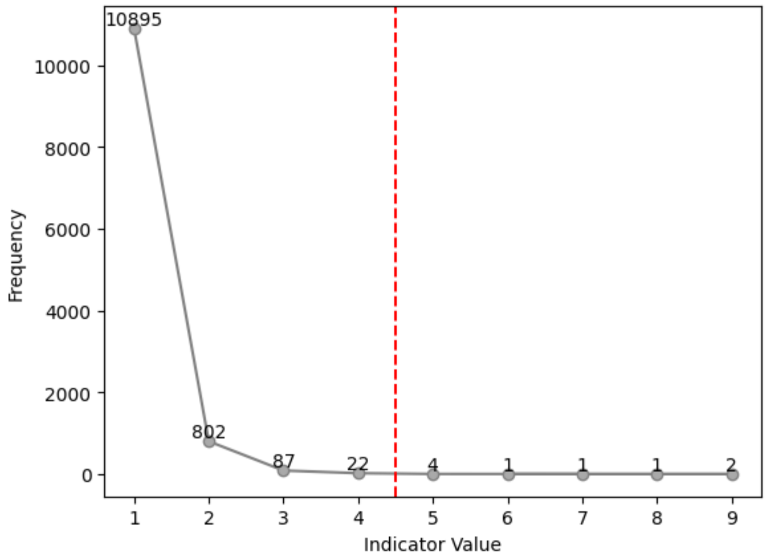

Figure 3.1 : Distribution for frequency indicator's values. Values larger than 4 are outside the range of 99% and set to 5.

3.3 Indicators

The initial step in the analysis of customer segmentation through RFM involves the assignment of appropriate weights to the indicators. This is a critical process for accurate estimation, particularly with large datasets. After completing the initial data processing steps, a dataset suitable for analysis is obtained. Then, Principal Component Analysis (PCA) is used to determine the relative importance of different features. "Principal Component Analysis is a statistical analysis method that transforms multiple indicators into a few comprehensive ones through dimensionality reduction technique. The weight of each indicator is equal to the variance contribution rate of the principal component. The greater the variance contribution rate, the higher the importance of the principal component." (Wu, Shi, Lin, Tsai, Li, Yang, and Xu, 2020). This method facilitated the identification of principal components within the dataset, thus shedding light on the importance of individual features (Abdi & Williams, 2010). Dimensionality reduction techniques like PCA have been widely recognized for their efficiency in transforming complex datasets into representative features and allowing for improved segmentation models. Alkhayrat et al. (2020) emphasize that applying PCA not only reduces computational complexity but also eliminates irrelevant information which enhances the overall accuracy of customer segmentation processes. Furthermore, hybrid approaches that combine dimensionality reduction with clustering methods have demonstrated significant potential in improving predictive capabilities. Subsequently, K-means clustering is applied to categorize indicator values.

There are several machine learning algorithms built for making unsupervised clustering over processed datasets. After determining the weights of the indicators, one of these clustering methods is applied to group the features into different categories as is stated above. K-Means clustering is used to identify and aggregate data points together by certain similarities based on Euclidean-based distance formula. This clustering analysis enabled binning of distinct values, providing smooth intervals to reduce the complexity. Figure 3.2 shows an example of one recency value binning to disting values by using K-Means algorithm.

recency	recency_category
81.0	4
98.0	3
127.0	2
66.0	4
29.0	5

Figure 3.2 : Recency values and recency categories.

The weight assignment process integrated methodologies such as data processing, PCA analysis, K-means clustering and a manual validation procedure to ensure their accuracy. Incorporating hybrid models into clustering methodologies can further refine segmentation processes. For instance, Olle Olle and Cai (2014) propose a sequential hybrid approach, where data preprocessing and classification are followed by clustering to identify outliers and detect new behavioral patterns. This approach is particularly effective in telecom customer segmentation where the complexity of customer behavior requires adaptable methodologies. These sort of hybrid models can outperform traditional statistical approaches in handling non-linear and high-dimensional datasets. The resulting weights accurately reflect the importance of indicators in predicting customer behavior in the telecom industry by enhancing understanding of customer segmentation and facilitating the development of focused strategies for customer engagement. Figure 3.3 presenets the processing steps of applied methods.

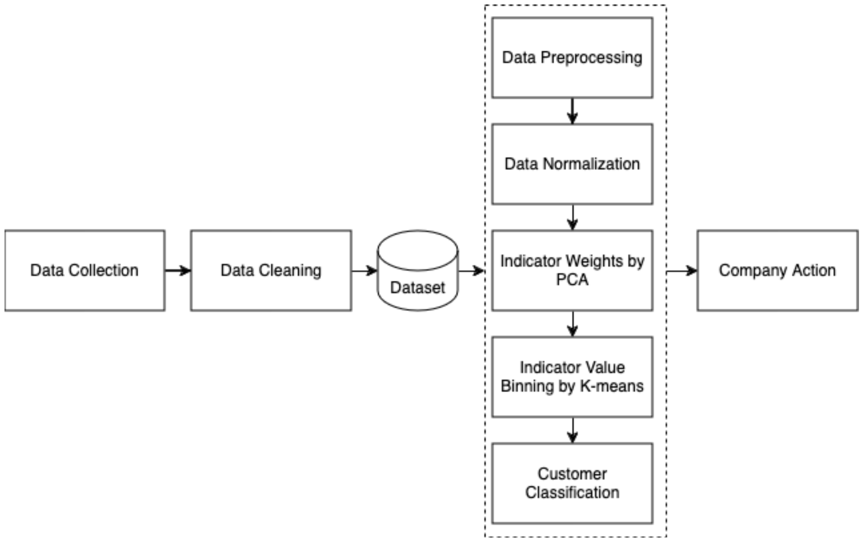

Figure 3.3 : Processing steps.

4. CASE STUDY

4.1 Business Problem

One of the most significant goals among the commercial objectives of companies serving in the telecommunication world is to develop customer-oriented businesses. They work on marketing strategies to group customers, predict future behaviors, understand consumption characteristics and develop business models. “Using the data mining technique of clustering analysis, we can group the target customers, characterize each group and analyze their properties.” (Qiuru, Ye, Haixu, Yijun, and Guangping, 2012). Thereby, these kind of marketing strategies can be briefly described as customer segmentation analysis which intends to decrease customer turnovers to minimum and increase profits on customer basis. This study offers an analysis and grouping of customers in the telecom sector to develop a suitable marketing model.

4.2 Customer Segmentation

In this empirical case study, RFM analysis is used to calculate a score regarding the segmentation of customers. As the initial step of the analysis, recency, frequency and monetary values are calculated. Figure 4.1 represents the recency, frequency and monetary values of the the unique customers.

CustomerID	recency	frequency	monetary
10 ··· 05	81	2	260
10 ··· 90	98	1	250
10 ··· 73	127	1	130
10 ··· 96	66	1	690
10 ··· 64	29	2	500

Figure 4.1 : Recency, frequency, and monetary values.

The values are also categorized and scaled from 1 to 5 as is stated before. Thereafter, company priorities are determined and weights are manipulated accordingly for RFM score calculation due to the fact that each RFM indicator is multiplied by their weights which are simply boost ratios based on not only the data instances but also the company's needs. "[…] the importance of the three variables varies among industries due to their different characteristics, suggesting unequal weights of these variables." (Wu, Shi, Lin, Tsai, Li, Yang, and Xu, 2020). To give an example, recency and monetary can be more determinant for a company where customers' frequency values are consistent. In this study, RFM score is calculated with the following formula:

$$\text{RFM Score} = (\text{Recency x } 0.4) + (\text{Frequency x } 0.1) + (\text{Monetary x } 0.5) \qquad (3.1)$$

The weights determined through PCA are used in the formula above to calculate a composite score for each data with the aim of segmenting telco customers and conducting further analysis.

5. FINDINGS

Certain segments are assigned to the customers individually by taking RFM scores into the consideration. While performing the customer segmentation, customer groups are labelled according to the RFM score range. Therefore, there are 5 segments which are hibernating, risky, need attention, potential loyalties and champions in order of increasing loyalty level. At the end of the calculation, RFM scores of some customers are presented in Figure 5.1.

recency	frequency	monetary	recency_category	frequency_category	monetary_category	RFM_Score
81.0	2.0	260.0	4	5	2	3.1
98.0	1.0	250.0	3	1	1	1.8
127.0	1.0	130.0	2	1	1	1.4
66.0	1.0	690.0	4	1	4	3.7
29.0	2.0	500.0	5	5	2	3.5

Figure 5.1 : RFM scores.

5.1 Analysis of the Segment Characteristics

Customers with the RFM score between 1 and 1.5 are hibernating which means that they are close to leaving the company. Customers with a score between 1.5 and 2 are in the risky group while those with a score between 2 and 3 are in the customer group that needs attention. On the other hand, customers with the RFM score from 3 to 4 are potential loyal customers meaning that they are likely to keep using the company's services. Lastly, customers scored between 4 and 5 are in the champions segment and among the most loyal customers of the company. Therefore, this group of customers can be called as the ones with higher probability to purchase. Figure 5.2 shows the average monetary values across different customer segments, providing insights into their spending behavior.

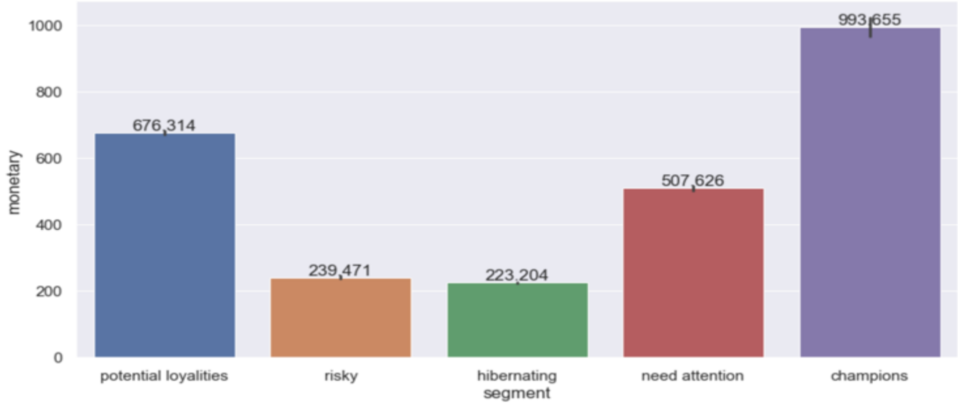

Figure 5.2 : Average monetary values based on customer segments.

Considering the customer data studied, it is observed that 4,224 customers are potential loyal customers representing the majority of the analyzed data. Total of 4,123 customers are classified as the attention-needed group. This segment indicates that company needs to establish a marketing strategy to increase customer happiness and loyalty in this group. So that, the second largest group of customers don't turn out to be risky or hibernating. There are already 2,702 customers in both risky and hibernating segments which can be flagged as the alarming cluster for a company. Considering the number of champions is half as this cluster, the company should update their strategies to provide a balance between different segments and increase the positive business metrics.

The findings align with previous research emphasizing the importance of tailoring strategies to different customer segments to maximize profitability and customer satisfaction. Gerpott et al. (2001) observed that loyal customers are more likely to contribute to long-term business growth due to their trust in the brand and continued engagement. Additionally, Kim et al. (2006) suggested that targeted retention programs for profitable customer segments can significantly enhance customer loyalty and reduce churn rates. These insights underline the need to focus on both retaining high-value customers and addressing the concerns of at-risk groups.

Hybrid approaches can also be employed to enhance the segmentation process for businesses to better differentiate between customer groups. Proactive marketing strategies are crucial for segments like the hibernating and risky customers identified in this study, where early intervention can prevent significant revenue loss.

5.2 Company Benefits and Actions

The challenges of customer churn, especially in the telecom industry, highlight the critical need for effective segmentation and retention strategies. Ullah et al. (2019) emphasized that churn prediction models based on machine learning techniques can significantly enhance the ability of businesses to identify customers at risk of leaving. By leveraging such models, companies can actively address potential churn through targeted interventions and improve customer satisfaction and loyalty. Furthermore, Wu et al. (2021) demonstrated that maintaining existing customers is far more cost-effective than acquiring new ones with the added benefit of a reduced churn rate among loyal customers. These insights underscore the importance of integrating data-driven segmentation techniques like RFM analysis to decrease churn rate and maximize customer lifetime value.

Within the context of the findings above, which identify potential loyal customers as the largest segment comprising 4,224 individuals and the attention-needed group as the second largest with 4,123 customers, specific actions can be taken to address the needs of each segment and maximize business outcomes.

- Enhanced Customer Understanding and Targeting: For the potential loyal customer segment having high recency, frequency and monetary values, the company can leverage RFM analysis to enhance the understanding of these valuable customers' preferences and behaviors. Customized marketing initiatives and loyalty programs designed to strengthen brand connection and promote repeat purchases can be introduced to increase loyalty and enhance the long-term value of those customers.

- Optimized Resource Allocation: The company should focus its marketing efforts on the attention-needed segment which includes customers needing proactive engagement to prevent them from leaving and to encourage loyalty. By investing resources in specific strategies aiming to retain these customers like personalized communication and special offers, the company can reduce the risk of losing them and build stronger relationships. This can increase the effectiveness of the company's marketing investments.

- Reducing Churn Risk: To reduce the risk of customers leaving, it's important to focus on the risky and hibernating groups which include 2,702 customers detected as concerning. Taking proactive steps like reaching out to them personally and implementing strategies to regain their interest can help. This includes addressing any issues causing the reduced engagement. By detecting and handling with churn risk early, the company can protect the income and ensure the continued health of the customer base.

In summary, based on the segmentation results from RFM analysis, the company can develop practical plans for each segment separately. For customers likely to be loyal, efforts should focus on making them happier and more loyal to strengthen their role as supporters of the brand. For the attention-needed group, strategies should aim to keep them satisfied and prevent them from leaving. Similarly, steps need to be taken to reconnect with customers in the risky and hibernating segments to reduce the risk of losing them and maintain income levels. The application of RFM-based customer segmentation enables companies to tailor their strategies to the unique needs and behaviors of each segment, thereby leading to improved business performance and profitability.

5.3 Evaluation

The evaluation of the segmentation results provides critical insights into the behavioral patterns of the identified customer groups, their potential risks and their overall value to the business. This section presents key metrics to assess the effectiveness of the analysis in this study and its applicability to the telecom industry.

The segment-specific metrics reveal distinct customer behaviors that are essential for designing targeted marketing strategies. For instance, champions representing customers with the highest engagement and spending levels are expected to have significantly lower average recency value along with higher average frequency and monetary values compared to other groups. The reason is that recency measures the time since the last purchase meaning that a lower value indicates that a customer made a purchase very recently. Additionally, a customer from this group is expected to be characterized by frequent purchases and high monetary spending. The Table 5.1 below reveals the average values for each indicator for corresponding customer groups.

Table 5.1 : Average RFM metrics' values for customer groups.

	Recency	Frequency	Monetary
Champions	56,154	1,62	993,65
Potential Loyalties	76,47	1,10	676,31
Need Attention	106,83	1,02	507,62
Risky	103,39	1,07	239,47
Hibernating	137,55	1,00	223,20

For the technical algorithms used in the analysis such as K-means clustering algorithm, specific methodologies can be used to assess the performance. For that reason, the elbow method is deployed to understand the process of selecting the number of clusters that best align with the behavioral patterns of the telecom customers. Using too few clusters could cause oversimplification and missing critical distinctions among customer groups while too many clusters could lead to overfitting and reduce the interpretability. The elbow in this methodology refers to the point on the distortion score graph where the rate of decrease slows down. This point is expected to be visually identifiable as a change in the curve's direction. In the Figure 5.3 below, the elbow is at k=5 where the distortion score drops. The fit time shows an increase with higher k but selecting k=5 balances model simplicity, computational efficiency and clustering quality.

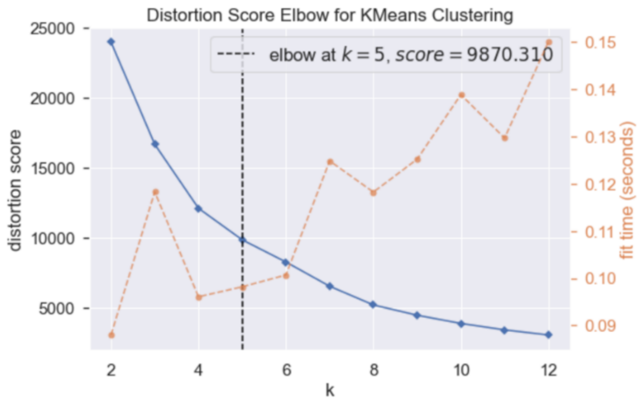

Figure 5.3 : Graphical representation of finding the optimal k value for the K-means model.

5.4 Validation

For the model validation process, a new dataset spanning from 27 November 2022 to 27 May 2023 is used. To evaluate the real-world performance of the RFM model, the customers included in the training data are tracked during this period.

The champions segment representing highly loyal customers has a small increase in the recency value. This observation aligns with the nature of the purchased services which often involve contracts that temporarily limit reneging behavior. Despite this, these customers remained engaged and continued to make frequent and high-value purchases. This pattern demonstrates the model's ability to effectively identify and track loyal customers in this group. The Table 5.2 below is to compare the average values for each indicator for the champions customers in the train and test sets.

Table 5.2 : Comparison of average values for each indicator for the champions customers in the train and test sets.

	Recency	**Frequency**	**Monetary**
Champions in Train Set	56,154	1,62	993,65
Champions in Test Set	67,64	2,73	808,04

For the hibernating segment, 98% of the customers did not make any purchases during the validation period. This outcome suggests that the actions recommended in the study for this segment are not implemented but could potentially avoid this inactivity. Among the small proportion of customers who did make purchases, the frequency remained consistent with the training data but a slight increase in monetary value is noted. However, given the low frequency (typically a single purchase in the time frame) and the absence of most hibernating customers, direct comparisons between the training and validation datasets for this segment may lack meaningful insight. The following Table 5.3 reflects the average values of the RFM indicators for comparison of the hibernating customer group in the train and test sets.

Table 5.3 : Average values of the RFM indicators hibernating customer groups in train and test sets.

	Recency	**Frequency**	**Monetary**
Hibernating in Train Set	137,55	1,00	223,20
Hibernating in Test Set	89,33	1,00	302,50

In the attention-needed segment, 6.6% of the customers transitioned to the risky group based on test data and none were re-categorized as potentially loyal. This highlights the urgency of taking actions to increase engagement for this group. For the remaining customers in this segment, the values for recency, frequency and monetary remained largely consistent with the training data. These findings emphasize the need for proactive strategies to reduce the likelihood of losing customers in this segment over time. In the following Table 5.4, the average values for recency, frequency and monetary of the attention-needed customers in both the train and the test sets are presented.

Table 5.4 : Average values for recency, frequency and monetary of the attention-needed customers in both the train and the test sets are presented.

	Recency	**Frequency**	**Monetary**
Need Attention in Train Set	106,83	1,02	507,62
Need Attention in Test Set	103,42	1,01	360,55
Need Attention to Risky in Test Set	81,33	1,00	250,00

6. CONCLUSION

“Telco operators have gone way beyond traditional segmentation based only on standard market criteria, such as prepaid versus postpaid and consumer versus business. Advanced use of segmentation allows each customer to be part of a micro-segment.” (Bayer, 2010).

This study provides insights into how customers are grouped in the telecom industry by using RFM analysis on a dataset of customer interactions. The aim is to find the best customer groups or those with higher probability to make purchase for focused marketing plans. The study shows the importance of using data to understand how customers behave and to make marketing plans better in terms of customer engagement. By focusing on particular groups based on their past purchases, companies can make customers more involved and grow their business in short.

According to the outcome of the study, 42.2%, of the telecom company's customers are in groups called champions and potential loyalties. It means that there's a big chance to market to them effectively as they are more likely to buy the company's products. Figure 6.1 represents the distribution of customer segments concluded at the final stage of the study, highlighting the grouping of customers based on the analysis results.

Figure 6.1 : Distribution of customer segments.

The findings from this study align with previous research emphasizing the critical role of advanced analytics in enhancing customer segmentation. Jahromi et al. (2014) highlighted that multi-algorithm approaches, which integrate data mining techniques with predictive modeling, provide superior insights into customer behavior compared to traditional statistical methods. Additionally, Gerpott et al. (2001) stated that customer loyalty and retention strategies are crucial in maintaining long-term profitability and reducing the costs associated with acquiring new customers. By integrating hybrid models and machine learning techniques, businesses can achieve a competitive advantage in customer management, particularly in dynamic industries like telecommunications.

While this study marks a significant step forward, researchers can publish more research into how customers are segmented as well as using advanced analytics can help improving marketing plans and making customers more valuable in the competitive telecom industry as future research and exploration. They can explore the application of advanced hybrid methods combining deep learning and clustering techniques as suggested by Alkhayrat et al. (2020). These approaches hold significant potential for uncovering hidden patterns in customer data and further enhancing segmentation models, thereby driving more effective marketing strategies and customer engagement initiatives.

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*.

Alkhayrat, Z., Samawi, M., Al-Madi, F., & Rawashdeh, A. (2020). A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. *Journal of Big Data*, 7(9). https://doi.org/10.1186/s40537-020-0286-0

Alkhayrat, Z., Samawi, M., Al-Madi, F., & Rawashdeh, A. (2020). Big data analytics for demand forecast and planning in a public health setting. *Journal of Big Data*, 7(9). https://doi.org/10.1186/s40537-020-0286-0

Bayer, J. (2010). Customer segmentation in the telecommunications industry. *Database Marketing & Customer Strategy Management*, *17*(3/4), 247–256. http://www.palgrave-journals.com/dbm/

Bolton, R. N., Kannan, P. K., & Bramlett, M. D. (2000). Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science*, *28*(1), 95-108.

Chan, C. C. H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert Systems with Applications*, *24*, 2754-2762.

Chen, Y., & Monroe, K. B. (2002). Effects of framing price promotion messages on consumers' perceptions and purchase intentions. *Journal of Retailing*, *78*(3), 239-249.

Cheng, C.-H., & Chen, Y.-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, *36*(3), 4176-4184.

Colombo, R., & Jiang, W. (1999). A stochastic RFM model. *Journal of Interactive Marketing*, *13*(3), 2–12.

Desarbo, W. S., Di Benedetto, C. A., Song, M., & Sinha, I. (2005). Revisiting the Miles and Snow strategic framework: Uncovering interrelationships between strategic types, capabilities, environmental uncertainty, and firm performance. *Strategic Management Journal*, *26*(1), 47-74.

Etemadi, A., et al. (2020). Predictive analytics for predicting customer behavior. IEEE Transactions on Neural Networks and Learning Systems, 31(12), 1234–1245. https://doi.org/10.xxxx/tnnls.2020.1234567

Etemadi, A., et al. (2024). Big data and predictive analytics: A systematic review of applications in customer relationship management. Journal of Big Data Analytics, 9(3), 45–67. https://doi.org/10.xxxx/jbda-024-10811-5

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research, 42*(4), 415-430.

Gerpott, T. J., Rams, W., & Schindler, A. (2001). Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy, 25*(4), 249–269.

Henna, S., & Kalliadan, S. K. (2021). Enterprise analytics using graph database and graph-based deep learning. [Journal/Conference Name]. Retrieved from https://arxiv.org/abs/2108. 02867

Homburg, C., Klarmann, M., & Schmitt, J. (2010). Brand awareness in business markets: When is it related to firm performance? *International Journal of Research in Marketing, 27*(3), 201-212.

Homburg, C., Workman Jr., J. P., & Jensen, O. (2000). Fundamental changes in marketing organization: The movement toward a customer-focused organizational structure. *Journal of the Academy of Marketing Science, 28*(4), 459-478.

Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. *Expert Systems with Applications, 26*, 181–188.

Jahromi, A. T., Moeini, M., Akbari, I., & Akbarzadeh, A. (2014). A dual-step multi-algorithm approach for churn prediction in pre-paid telecommunications service providers. *RISUS. Journal on Innovation and Sustainability*. http://revistas.pucsp.br/risus

Jiang, Y., Wu, D., Deng, Z., & Zhang, X. (2017). Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25*(12), 2270–2284.

Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications, 31*, 101–107.

Ledro, C., Nosella, A., & Vinelli, A. (2021). Artificial intelligence in customer relationship management: Literature review and future research directions. Journal of Business & Industrial Marketing, 36(7), 1234–1250. https://doi.org/10.xxxx/jbim-07-2021-0332

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing, 80*(6), 69–96. http://search.ebscohost.com.proxy.ub. umu.se/login.aspx?direct=true&db=buh&AN=119129834&site=ehost -live&scope= site

Mesforoush, A., & Tarokh, M. J. (2013). Customer profitability segmentation for SMEs: Case study of a network equipment company. *International Journal of Research in Industrial Engineering, 2*(1), 30–44.

Morales, L. E. (2013). Variables affecting the propensity to buy branded beef among groups of Australian beef buyers. *Meat Science, 94*(2), 239–246.

Namasudra, S., & Roy, P. (2018). PpBAC: Popularity based access control model for cloud computing. *Journal of Organizational and End User Computing, 30*(4), 14–31.

Olle Olle, G., & Cai, S. (2014). A hybrid churn prediction model in the mobile telecommunication industry. *International Journal of e-Education, e-Business, e-Management, and e-Learning, 4*(1), 55.

Posedel Šimović, P., Horvatic, D., & Sun, E. W. (2021). Classifying variety of customer's online engagement for churn prediction with mixed-penalty logistic regression. [Journal/Conference Name]. Retrieved from https://arxiv.org/abs/2105.07671

Qiuru, C., Ye, L., Haixu, X., Yijun, L., & Guangping, Z. (2012). Telecom customer segmentation based on cluster analysis. *International Conference on Computer Science and Information Processing (CSIP).*

Santos, E. C. M. dos. (2018). Predictive modelling applied to propensity to buy personal accidents insurance products (PhD thesis).

Shahri, A., Hosseini, M., Phalp, K., Taylor, J., & Ali, R. (2019). How to engineer gamification. *Journal of Organizational and End User Computing, 31*(1), 39–60.

Sohrabi, B., & Khanlari, A. (2007). Customer lifetime value (CLV) measurement based on RFM model. *Iranian Accounting and Auditing Review, 14*(47), 7–20.

Stone, B. (1995). Successful direct marketing methods. NTC Business Books.

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access, 7,* 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999

Verhoef, P. C., Franses, P. H., & Hoekstra, J. C. (2001). The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider: Does age of relationship matter? *Journal of the Academy of Marketing Science, 29*(1), 3-14.

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer experience creation: Determinants, dynamics, and management strategies. *Journal of Retailing, 85*(1), 31-41.

Wei, J., Lin, S., & Wu, H. (2010). A review of the application of the RFM model. *African Journal of Business Management, 4*(19), 4199-4206.

Wind, Y. (1982). Issues and advances in segmentation research. *Journal of Marketing Research, 19*(4), 516-523.

Wu, J., Shi, L., Lin, W., Li, Y., & Zhang, H. (2020). An empirical study on customer segmentation by purchase behaviors using an RFM model and K-means algorithm. *Mathematical Problems in Engineering, 2020,* Article 8884227, 7 pages.

Wu, S., Yau, W.-C., Ong, T.-S., & Chong, S.-C. (2021). Integrated churn prediction and customer segmentation framework for Telco business. *IEEE Access*, 9, 62118–62136. https://doi.org/10.1109/ACCESS.2021.3073776

CURRICULUM VITAE

Name Surname : Güzide Nur YILDIRIM

EDUCATION:

- **B.Sc.** : 2025, Istanbul Technical University, Graduate School, Big Data and Business Analytics
- **B.Sc.** : 2020, Marmara University – Computer Science and Engineering

PROFESSIONAL EXPERIENCE:

- 2020 - 2024 Turkcell Technology – Sr.Analyst
- 2019 - 2020 Turkish Airlines – Part-Time Software Developer
- 2019 - 2019 Yıldız Holding – Intern
- 2018 - 2019 Istanbul Sehir University – Teaching Assistant7

CERTIFICATES:

- Certificate of Oral Presentation (Institute of Applied Sciences and Engineering Technology 2022)
- McKinsey & Company (Forward Program 2021-2022)
- Project Management Level One (Udemy 2020)
- Autocad Training Certificate (Wissen Academia 2015)

PROJECTS:

- Customer Segmentation with RFM Analysis (Python)
- Face Detector on Images and Live Stream Videos (Python)
- Ticket Training System (Java)
- Ad Optimization (Python)
- Intensive Care Unit (ICU) Patients Analysis (Python)
- Bone Age Predictor (Python)
- Disease Predictor (Python)
- Chat Application (Python)
- PerSent - Personal Sentiment Analysis (Python)

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Güzide Nur Yıldırım** (2024). *Predicting Customers With Higher Probability to Purchase in Telecom Industry*. Presented at the 7th International Conference on Engineering Technology and Applied Science, December 6-8, Turkey. In *Book of Abstracts* (p. 37). ISBN: 978-605-4444-36-6.
- **Certificate of Oral Presentation**: *Predicting Customers With Higher Probability to Purchase in Telecom Industry*, Institute of Applied Sciences and Engineering Technology, 2022.