Fırat Üniversitesi Müh. Bil. Dergisi
36(2), 979-991, 2024

Araştırma Makalesi
https://doi.org/10.35234/fumbd.1408941

Customer Segmentation by using Annual E-Invoice Data

Fahrettin Burak DEMİR[1*], Gürkan ÇELİK[2]
[1] Software Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedi Eylul University, Bandirma, Türkiye
[2] F.I.T. Information Technology Systems Services Industry and Com. Inc., Istanbul, Türkiye
[*1] fdemir@bandirma.edu.tr, [2] gurkan.celik@sovos.com

(Geliş/Received: 25/12/2023; Kabul/Accepted: 27/09/2024)

Abstract: As businesses navigate the digital landscape, the proliferation of electronic transactions has led to an abundance of valuable data that can be harnessed for strategic decision-making. This study explores the application of CRM and RFM analysis for customer profiling and segmentation, utilizing e-invoice data as a rich source of information. By leveraging these advanced statistical techniques, the research aims to uncover hidden patterns within electronic transaction records, allowing for the identification of distinct customer segments based on their purchasing behavior. The methodology involved collecting and pre-processing one year of e-invoice data from Fit IT Company, followed by applying statistical models to uncover underlying structures and relationships. Furthermore, the research examines the implications of customer segmentation on marketing strategies, customer relationship management, and personalized service offerings. CRM and RFM analyses were performed on the annual sales data obtained as a result of e-invoice usage service to customers. When the results of the analysis were analyzed, the number of transactions belonging to the sender, recipient, and parties in the top 10 every month were extracted. It has been demonstrated that customer segmentation can be conducted more comprehensively by using CRM and RFM analyses together. While CRM analysis focuses on transaction volume and customer relationships, RFM analysis provides a more detailed perspective on customer behavior by evaluating purchase frequency, recency, and monetary value. In the study, by analyzing e-invoice data through these two methods, the most valuable customer groups were identified, and how strategic marketing approaches can be developed for these groups was illustrated. The combined use of CRM and RFM analyses allows for more accurate customer segmentation based on both transaction volume and spending habits. This approach concludes that strategies can be developed to increase customer loyalty, optimize marketing strategies, and improve business performance.

Key words: CRM Analyses, Customer Segmentation, Digital Business, e-Invoice, RFM Analyses.

Yıllık E-Fatura Verilerini Kullanarak Müşteri Segmentasyonu

Öz: İşletmeler, dijital alanda gezinirken, elektronik işlemlerin yaygınlaşması stratejik karar alma için kullanılabilecek çok sayıda değerli veriye yol açmıştır. Bu çalışma, e-fatura verilerini zengin bir bilgi kaynağı olarak kullanarak müşteri profili ve segmentasyonu için CRM ve RFM analizinin uygulanmasını araştırmaktadır. Bu gelişmiş istatistiksel tekniklerden yararlanarak, elektronik işlem kayıtlarındaki gizli kalıpları ortaya çıkarmayı ve satın alma davranışlarına göre farklı müşteri segmentlerinin belirlenmesi amaçlamaktadır. Metodoloji, Fit IT Company'den bir yıllık e-fatura verisinin toplanmasını ve ön işlenmesini, ardından altta yatan yapıları ve ilişkileri ortaya çıkarmak için istatistiksel modeller uygulanmasını içermektedir. Ayrıca çalışma, müşteri segmentasyonunun pazarlama stratejileri, müşteri ilişkileri yönetimi ve kişiselleştirilmiş hizmet teklifleri üzerindeki etkilerini incelemektedir. CRM ve RFM analizleri, müşterilere e-fatura kullanım hizmeti sonucunda elde edilen yıllık satış verileri üzerinde gerçekleştirilmiştir. Analiz sonuçları incelendiğinde, her ay ilk 10'da yer alan gönderici, alıcı ve taraflara ait işlem sayısı çıkarılmıştır. CRM ve RFM analizlerinin birlikte kullanılmasıyla müşteri segmentasyonunun daha kapsamlı bir şekilde yapılabileceği gösterilmiştir. CRM analizi işlem hacmi ve müşteri ilişkilerine odaklanırken, RFM analizi satın alma sıklığı, yakınlık ve parasal değeri değerlendirerek müşteri davranışı hakkında daha detaylı bir bakış açısı sunmaktadır. Çalışmada, e-fatura verilerinin bu iki yöntemle analiz edilmesiyle en değerli müşteri grupları belirlenmiş ve bu gruplara yönelik stratejik pazarlama yaklaşımlarının nasıl geliştirilebileceği gösterilmiştir. CRM ve RFM analizlerinin birlikte kullanılması, hem işlem hacmi hem de harcama alışkanlıklarına göre daha doğru müşteri segmentasyonuna olanak sağlamaktadır. Bu yaklaşım, müşteri sadakatini artırmak, pazarlama stratejilerini optimize etmek ve iş performansını iyileştirmek için stratejiler geliştirilebileceği sonucuna varmaktadır.

Anahtar kelimeler: CRM Analizleri, Müşteri Segmentasyonu, Dijital İşletme, e-Fatura, RFM Analizleri.

1. Introduction

In the era of digital transformation, businesses are not only witnessing a surge in electronic transactions but are also presented with unprecedented opportunities to harness the wealth of data generated through these transactions. One such avenue that holds immense potential for strategic insights is the realm of e-invoice data. As

* Sorumlu yazar: fdemir@bandirma.edu.tr. Yazarların ORCID Numarası: [1] 0000-0001-9095-5166, [2] 0009-0005-5282-8365

companies increasingly transition towards electronic invoicing systems, they find themselves in possession of a vast repository of information that extends far beyond mere financial transactions. Traditionally, customer segmentation has been a cornerstone of effective marketing and business strategies, allowing organizations to tailor their approaches based on the unique characteristics and preferences of different customer groups. The integration of e-invoice data into this segmentation process represents a paradigm shift, as it introduces a granular perspective into customer behavior, transcending conventional demographic or psychographic parameters. By adopting advanced statistical methods and machine learning algorithms, businesses can uncover hidden patterns within the intricate web of electronic transactions, leading to the identification of nuanced customer segments.

In the contemporary business landscape, where electronic transactions have become ubiquitous, the utilization of e-invoice data for customer segmentation emerges as a cutting-edge strategy. Electronic invoicing not only streamlines financial transactions but also generates a wealth of valuable information that extends beyond monetary exchanges. Customer segmentation, a fundamental aspect of targeted marketing and personalized service delivery, gains a new dimension when augmented by the intricate details embedded within e-invoice data. Unlike traditional segmentation methods reliant on demographic or psychographic factors, leveraging e-invoice data allows for a more nuanced and dynamic approach. Through the application of advanced statistical techniques and machine learning algorithms, businesses can delve into the intricacies of customer behavior, identifying patterns, preferences, and trends that might remain obscured through conventional means. Transaction frequency, monetary values, product categories, and temporal patterns all contribute to the rich tapestry of e-invoice data, enabling the creation of finely tuned customer segments. This approach not only enhances the precision of marketing efforts but also facilitates a more tailored and responsive customer experience. However, the integration of e-invoice data into the segmentation process comes with its set of challenges, including data privacy concerns and the need for robust analytical frameworks. Despite these challenges, the potential benefits are significant—businesses can gain a competitive edge by delivering targeted promotions, personalized services, and strategic communication to specific customer segments, thereby fostering customer loyalty and maximizing the value of each customer relationship. As organizations strive to adapt to the digital age, the exploration of customer segmentation through e-invoice data stands as a pivotal strategy for navigating the complexities of the modern marketplace. This paper delves into the realm of customer segmentation, exploring the innovative application of anonymous e-invoice data as a tool to discern and understand distinct customer profiles.

2. Related Works

There are a significant number of studies on customer segmentation in the literature. Customer segmentation has generally been achieved through approaches such as statistical analysis, clustering and machine learning. It seems possible to make commercial activities more systematic by segmenting customers, promotions, campaigns, discount rate determination, special discounts, etc.

Marcus [1] aims to present an uncomplicated yet potent method for customer segmentation known as the Customer Value Matrix. The strength of this approach is not only in its ability to pinpoint crucial customer segments but also in its capacity to spotlight appropriate marketing strategies and tactics that can be easily conveyed and implemented. Jonker et al. [2] introduce a unified optimization method that tackles two primary concerns: grouping customers into uniform segments, and identifying the most effective strategy for each segment. They apply this integrated optimization framework within the context of a direct-mail scenario for a charitable organization. Kim et al. [3] suggest a structure for assessing customer worth and categorizing customers according to their value. Following the segmentation of customers based on their value, it has been demonstrated the development of strategies tailored to each customer segment through a case study centered around a wireless telecommunication company. It has investigated the contemporary developments in customer lifetime value and models for customer segmentation, pinpointing critical areas for prospective research [4]. The segmentation model based on customer lifetime value, known as CLV-based segmentation, involves categorizing customers into meaningful segments primarily based on their customer lifetime value and, possibly, other contributing factors. The potential of customer segmentation is virtually boundless, serving as a tool to steer companies towards more efficient marketing approaches and the development of novel products [5]. In this conceptual overview, it has been explored the case of Migros Turk, an innovative multinational firm, to analyze the successful formulation of a set of segmentation strategies. Chan [6] introduces an innovative methodology that integrates customer targeting and segmentation for campaign strategies. The study employs a recency, frequency, and monetary (RFM) model to discern customer behavior and utilizes a customer lifetime value (LTV) model to assess the identified segmented customers. The application of a genetic algorithm (GA) to optimize the selection of more suitable customers has been also suggested. Namwar [7] utilizes data mining tools to develop a fresh customer segmentation approach incorporating RFM, demographic, and LTV data. The novel method comprises two phases. Initially, K-means clustering is employed to group customers into distinct segments based on their RFM characteristics.

Subsequently, utilizing demographic data, each cluster undergoes further partitioning into new clusters. Ultimately, a customer profile is crafted based on LTV analysis. Insights gained from customer segmentation empower company leaders to foster strong customer relationships and tailor their marketing strategies to align with customer expectations [8]. To attain the most effective segmentation, it has been devised a soft clustering technique employing a latent mixed-class membership clustering approach. This method categorizes online customers based on their purchasing data across various categories. Soft computing, categorized among data mining methods, has recently found application in the realm of segmentation, showing promise as an influential area for future segmentation research [9]. He scrutinizes the current uses of soft computing techniques in tackling segmentation challenges, emphasizing crucial factors, particularly those linked to segmentation effectiveness, that warrant consideration in every segmentation study. Many customer segmentation methods that rely on customer value often neglect to consider the element of time and the evolving trends in value changes during their analysis. It has been categorized customers according to their value using the RFM model and the K-means clustering method [10]. Subsequently, an evaluation of changes over multiple periods is conducted. It is claimed that its integration of time and the trend of customer value changes enhances the precision of predictions based on customers' past behavior. It explored two distinct data mining methodologies for customer segmentation: clustering and subgroup discovery [11]. The resultant models yielded six market segments and 49 rules, providing enhanced insights into customer preferences within the context of a highly specialized fashion manufacturer/e-tailor in a customized manner. This is where the application of machine learning becomes crucial, utilizing various algorithms to unveil concealed patterns in data for improved decision-making in the future [12]. The somewhat elusive concept of determining which segment to target is clarified through segmentation. Customer segmentation involves grouping customers with similar behaviors into the same segment and distinguishing those with different patterns into separate segments.

This study introduces an innovative approach by simultaneously employing CRM (Customer Relationship Management) [13, 14] and RFM (Recency, Frequency, Monetary Value) [14] analyses, filling a gap in the existing literature. While these two methods are typically applied independently, their combined use in this research offers a more comprehensive understanding of customer behaviors, thereby facilitating the development of more effective customer segmentation and targeting strategies. CRM analysis allows for an in-depth examination of customer relationships, while RFM analysis provides valuable insights based on customers' purchasing behaviors. The distinguishing feature of this work from others is its integration of these two powerful tools, enabling businesses to gain a more detailed understanding of their customer base and tailor their marketing and sales strategies with greater precision. This holistic approach has the potential to enhance customer loyalty, maximize customer lifetime value, and, consequently, increase business revenues. Therefore, this study significantly contributes to the literature on customer segmentation and relationship management, opening new perspectives for both academic research and practical applications.

3. Materials and Methods

The dataset used in the study was provided by Sovos and was obtained for the period from January 2022 to December 2022. The dataset consists of 30 attributes in total, counting the number of features in each row. The SENDER and RECIPIENT fields in each row represent the sender and recipient companies, and the STATUS and STATUS_DESCRIPTION fields contain the status of each invoice transaction and the text describing this status. The STATUS_DATE field indicates the date each transaction occurred, while the SCENARIO field indicates the transaction scenario. The INVOICE_TYPE field indicates the type of invoices issued, while the DOCUMENT_FX_RATE field contains the exchange rate at which the invoice document was issued. ITEM_NUMBER indicates the number of goods/service items included in each invoice transaction, while MAL_HIZMET_TOTAL_AMOUNT represents the total amount of goods/services for the invoice transaction. The SEND_DATE field indicates when each transaction was sent, while the VAT fields (VAT_0_Amount, VAT_0_MATRAH, VAT_1_Amount, VAT_1_MATRAH, VAT_8_Amount, VAT_8_MATRAH, VAT_18_Amount, VAT_18_MATRAH) contain the VAT amounts and bases broken down according to the different VAT rates in the invoice transaction. Finally, the field GONDERICI_SEHIR indicates the city where the sender is located. The dataset contains detailed information on e-invoice transactions and can be used to understand financial movements in business activities. Table 1 presents a simple example of the dataset, showing the attributes in the dataset.

Table 1. Sample Dataset.

STATUS	STATUS ANNOTATION	STATUS DATE	STATUS DATE	SCENARIO	EDIT DATE	EDIT TIME	INVOIC E TYPE
100	Invoice forwarded to buyer	3.16.2022	BASIC INVOICE	3.15.2022	00:01:00+ 0200	SATIS	TRY
100	Invoice forwarded to buyer	3.16.2022	BASIC INVOICE	3.15.2022	00:01:00+ 0200	SATIS	TRY
100	Invoice forwarded to buyer	3.16.2022	BASIC INVOICE	3.15.2022	00:01:00+ 0200	SATIS	TRY
100	Invoice forwarded to buyer	3.16.2022	BASIC INVOICE	3.15.2022	00:01:00+ 0200	SATIS	TRY
100	Invoice forwarded to buyer	3.16.2022	BASIC INVOICE	3.15.2022	00:01:00+ 0200	SATIS	TRY

CURR	TAXES EXCLUSIVE TOTAL AMOUNT	CURR	TAXES TOTAL AMOUNT	CURR	TOTAL DISCOU NT	AMOUNT	SUBMIS SION DATE
TRY	36658,8	TRY	39591,5	TRY	14256,2	39591,5	59:59,4
TRY	97961,06	TRY	105797,94	TRY	62581,03	105797,94	59:59,0
TRY	294049,28	TRY	317573,22	TRY	196747,4 3	317573,22	59:58,9
TRY	255825,61	TRY	276291,66	TRY	168771,5	276291,66	59:58,7
TRY	238222,46	TRY	257280,26	TRY	164104,5 6	257280,26	59:58,5

VAT 1%	VAT 1% MATRAH	VAT 8%	VAT 8% MATRAH	VAT 18%	VAT 18% MATRA H	SHIPPER_CITY
0	0	2932,7	36658,8	0	0	İSTANBUL
0	0	7836,88	97961,06	0	0	İSTANBUL
0	0	23523,94	294049,28	0	0	İSTANBUL
0	0	20466,05	255825,61	0	0	İSTANBUL
0	0	19057,8	238222,46	0	0	İSTANBUL

3.1. Data Preprocessing

Before performing RFM analysis on the dataset used in the study, data preprocessing stages were performed. Data preprocessing is a critical stage used in data analytics and machine learning projects. It involves cleaning, organizing and preparing the raw dataset. Data preprocessing steps make the dataset suitable for machine learning algorithms by removing potential errors, correcting inconsistencies and organizing the data in a meaningful way. In the study, missing and outlier data were detected as part of the preprocessing step, but no missing or outlier data were found in the dataset.

3.1.1. CRM Analysis

CRM (Customer Relationship Management) is a set of strategies and technologies used to effectively manage a business' customer relationships, increase customer satisfaction and optimize sales performance [15]. CRM includes the processes of collecting and analyzing customer data, managing customer interactions and improving customer relationships. This system helps a business to communicate with its customers in a more interactive and personalized way. CRM aims to improve the customer experience by creating customer-specific campaigns, responding quickly to customer requests, tracking sales opportunities and analyzing customer feedback. It also makes strategic use of customer data to increase customer loyalty, make sales processes more efficient and improve business performance. CRM is usually implemented through specialized software and enables businesses to better understand customer communications, develop customer-centric strategies and sustainably grow their customer base. Therefore, CRM is a key element in today's business world that emphasizes the importance of customer centricity and is a key element towards achieving competitive advantage. For the dataset used in the experiments, CRM analysis was performed by considering the following article [16].

- **Customer Segmentation:**

By segmenting your customers according to the SENDER and RECIPIENT fields, the characteristics of certain customer groups (for example, those who frequently send invoices to a specific recipient) were evaluated.

3.1.2. RFM Analysis

RFM analysis is an effective marketing analysis method used to study customer behavior. This method evaluates a customer's shopping behavior based on three basic criteria: Recency, Frequency and Monetary Value. How recently a customer has shopped, how often they shop and the amount they spend on these purchases are analyzed by measuring these criteria. Recency determines the last time a customer made a purchase and is a value that measures this. Frequency refers to the number of times a customer has shopped in a given time period. Monetary represents the total monetary value spent by the customer. RFM analysis is usually performed by normalizing these criteria. Normalization involves compressing the data into a specific range. This allows it to be transformed into a smaller and often useful range, usually [-1,1] or [0.0, 1.0]. This normalization process is performed by assigning equal weights to different attributes in the data set. This type of normalization is commonly used, especially in classification algorithms. In conclusion, RFM analysis is a powerful tool to help businesses optimize their customer segmentation and marketing strategies. Through this analysis, customers' shopping behavior can be better understood and customized strategies can be developed [15], [16].

4. Experimental Evaluations

In this study, experiments on customer segmentation are carried out on the e-invoice dataset. The purpose of customer segmentation is to divide customers into groups to reflect the similarities between the receivers and senders in each group. In addition, in the experiments, it is decided how to establish a relationship with customers in each segment to maximize the value of each buyer and sender customer for the business. For the dataset used in the study, the top 10 sending and receiving companies with the highest number of transactions per month are presented in Table 2-5. The relationships between the sender and receiver are then re-analyzed according to total and average item-number and amount values, and the most significant 10 are presented in Table 6 for each month.

Table 2. Top 10 most valuable buyers and sellers and number of transactions in January, February and March (Winter Period).

January				February				March			
Sender	Number	Receiver	Number	Sende	Number	Receiver	Number	Sende	Number	Receiver	Number
G15	3841	A4269	755	G1561	5747	A1288	1046	G1561	5753	A1288	999
G42	2973	A10561	501	G15	2965	A15102	871	G15	3985	A10561	384
G1460	1244	A1288	442	G1499	1864	A151	638	G42	2640	A32723	366
G1448	928	A6532	319	G21	1826	A4269	578	G2025	1212	A2350	313
G39	872	A15102	282	G44	1692	A345	411	G1460	1093	A151	294
G44	684	A63543	240	G1448	1561	A6532	389	G1627	970	A3051	220
G1505	598	A31578	165	G1460	1499	A16440	278	G21	824	A4476	217
G1499	502	A10445	165	G1751	1137	A384547	246	G3708	764	A3047	207
G49	499	A5745	164	G22	908	A4476	231	G1406	752	A17422	202

Table 2 provides a CRM analysis for the winter period, detailing transactions between top buyers and sellers for January, February, and March. This analysis offers insights into market dynamics, highlighting how transaction volumes and partnerships evolve in the cold season.

In January, G15 stands out with 3841 transactions, marking a significant presence in the market, with A4269 receiving the highest number of transactions at 755. This indicates a robust demand for G15's offerings and possibly a strategic focus on A4269 as a key market segment or partner. The high volume of transactions underscores the importance of understanding market needs and maintaining strong relationships with significant buyers to ensure sustained business growth.

February sees a shift with G1561 leading the transactions at 5747, directed primarily towards A1288 with 1046 transactions. This substantial increase in transactions for G1561, coupled with the focus on A1288, suggests a strategic pivot or a successful marketing initiative capturing significant market attention. G15 maintains a strong presence, although with slightly fewer transactions, indicating ongoing competitiveness and market activity.

March continues the trend with G1561 at the helm, closely mirroring February's numbers, suggesting a sustained strategic focus and possibly cementing G1561's market position. A1288 remains a principal receiver, highlighting its pivotal role in the market and potential strategic value to sellers like G1561. G15's resurgence in

transactions to 3985, primarily towards A10561, demonstrates market fluidity and the strategic shifts businesses may employ to adapt to changing market dynamics and consumer demands.

This winter period analysis underscores the fluid and dynamic nature of market transactions, emphasizing the critical role of strategic agility and adaptability. Businesses that can effectively navigate changing market conditions, identify emerging trends, and maintain strong partnerships can capitalize on opportunities for growth and consolidation. The transaction volumes and the evolving relationships between buyers and sellers not only reflect the operational strengths and market positioning of these entities but also provide a lens through which broader market trends and consumer demands can be discerned. Such insights are invaluable for businesses aiming to refine their strategic approaches, optimize their market positioning, and enhance responsiveness to the evolving marketplace.

Table 3. Top 10 most valuable buyers and sellers and number of transactions in April, May and June (Spring Period).

April				May				June			
Sender	Number	Receiver	Number	Sender	Number	Receiver	Number	Sender	Number	Receiver	Number
G15	3249	A1288	1369	G15	3636	A6532	482	G4479	12488	A1288	1178
G42	2953	A10561	476	G42	1892	A10561	373	G5666	10455	A151	1157
G39	2487	A55036	427	G1460	750	A5745	312	G15	4826	A10561	699
G6	2049	A16440	388	G28	717	A1288	303	G42	3390	A6532	580
G44	1689	A15102	329	G169	477	A3864	87	G21	2265	A5745	309
G3708	1360	A4476	311	G1	462	A31027	67	G39	2213	A16440	262
G1448	1331	A53459	253	G1406	456	A3599	67	G1	1743	A44	258
G49	1201	A58660	236	G44	447	A7327	56	G1460	1445	A2669	250
G1460	1088	A3047	235	G45	412	A6706	50	G1448	1168	A4476	237

Table 3's CRM analysis for the spring period, spanning April, May, and June, presents a detailed overview of the transactional dynamics between the leading buyers and sellers, showcasing how market interactions evolve over these months. April sees G15 leading with 3249 transactions, signifying a strong start to the spring period, with A1288 being the primary receiver. This highlights a robust demand for G15's offerings and possibly a strategic partnership with A1288, underscoring the critical nature of understanding customer needs and preferences to maintain a competitive edge in the market.

In May, G15 continues its dominance with 3636 transactions, further establishing its market presence. The shift towards A6532 as the top receiver with 482 transactions indicates a broadening of G15's market reach or possibly diversification in its product or service offerings. This month's dynamics underscore the importance of agility and adaptation in business strategies to cater to evolving market demands.

June introduces a significant change with G4479 leading the transactions at an astonishing number of 12488, directing a majority to A1288. This dramatic increase suggests a potential market disruption or a successful promotional campaign that significantly enhanced G4479's market share. Concurrently, G5666 emerges with 10455 transactions, primarily to A151, indicating a vibrant market with active competition and strategic maneuvering among key players.

This period exemplifies the fluidity of market dynamics and the significance of strategic business decisions. The continuous presence of G15 as a significant sender across two months and the sudden emergence of G4479 and G5666 in June reflect the competitive nature of the market and the potential for businesses to rapidly ascend in transactional volume through effective marketing strategies and product offerings. The varied receivers across these months, from A1288 to A6532 and A10561, further highlight the diverse consumer base and the necessity for businesses to tailor their offerings to meet distinct customer segments.

Overall, the CRM analysis of Table 3 emphasizes the importance of leveraging transactional data to inform strategic decisions, adapt to market changes, and identify growth opportunities. Businesses must remain attuned to market trends, customer behaviors, and the competitive landscape to sustain growth and profitability in a dynamic market environment. This analysis not only sheds light on the operational capacities and strategic positioning of the entities involved but also offers insights into broader market trends and potential strategies for businesses seeking to enhance their market positioning and responsiveness to consumer demands.

Table 4. Top 10 most valuable buyers and sellers and number of transactions in July, August and September (Summer period).

July				August				September			
Sender	Number	Receiver	Number	Sender	Number	Receiver	Number	Sender	Number	Receiver	Number
G49	3786	A6532	1012	G39	2486	A6532	1292	G15	3493	A1288	1756
G15	3053	A1288	631	G1499	1435	A1288	855	G42	2851	A6532	1046
G42	2511	A5745	559	G3708	1364	A5745	721	G6	2194	A23484	709
G21	1187	A10561	370	G3180	1322	A151	630	G1460	1664	A181888	544
G1460	986	A123961	358	G2025	1124	A15102	594	G1	1531	A5745	473
G39	920	A4269	347	G3194	1066	A16440	537	G7182	1315	A10561	471
G44	898	A151	345	G21	1028	A345	394	G1118	1286	A51444	326
G1448	741	A10445	241	G22	997	A2850	255	G7538	1277	A151	301
G1	659	A126123	195	G49	987	A162039	246	G21	1243	A4476	295

The CRM analysis of the transaction activities during the summer months of July, August, and September meticulously delineates the fluctuations in transaction volumes between prominent buyers and sellers, illustrating the dynamic structure of the market. In July, G49's position as the most active sender with 3786 transactions reflects a significant demand for its offerings, a testament to the company's strategic maneuvers or seasonal influences driving the market at that time. The consistent reception of a large volume of transactions by A6532 underscores its substantial purchasing power and pivotal role within the market dynamics, highlighting its capacity to engage in significant transactions with major sellers like G49.

Transitioning into August, G39 ascends as the leading sender with 2486 transactions, signaling a shift in market interactions and strategic positioning among the market players over time. The sustained prominence of A6532 as a receiver, coupled with the increased transactions involving A1288, points to evolving market relationships, possibly indicative of new product launches or the fulfillment of specific demands of these key market players.

September introduces another layer of dynamics with G15 taking the helm in transaction volumes, underscoring its influential market presence and operational scale. The substantial increase in transactions received by A1288, emerging as the top receiver, could signify strategic realignment or efforts to bolster partnerships with key suppliers such as G15. Additionally, the ongoing transactions to A6532 throughout these months exemplify its crucial role in the supply chain, indicating its strategic value to sellers.

This analysis derived from Table 4 emphasizes the market's fluid nature and the imperative for businesses to exhibit strategic flexibility, adapting to market shifts to cater to changing demands. The detailed transaction volumes between sellers and buyers not only reveal the operational capabilities and market standings of the entities involved but also offer insights into broader market trends and consumer demands, serving as valuable intelligence for firms aiming to refine their strategic orientations, optimize market positioning, and enhance responsiveness to market evolutions. This comprehensive view underscores the importance of leveraging transactional data to inform strategic decisions, highlighting the potential for businesses to gain a competitive edge and expand their market share by aligning more closely with customer needs and preferences.

Table 5. Top 10 most valuable buyers and sellers and number of transactions in October, November and December (Autumn Period).

October				Novermber				December			
Sender	Number	Receiver	Number	Sender	Number	Receiver	Number	Sender	Number	Receiver	Number
G15	3615	A6532	1068	G8013	6828	A1288	993	G8013	5174	A1288	974
G42	3167	A10561	661	G15	2444	A239093	809	G42	3247	A51444	712
G1460	1828	A5745	493	G39	2154	A15102	737	G15	3226	A10561	564
G1448	1005	A1288	457	G1460	1644	A6532	536	G8244	2706	A6532	410
G1	826	A10445	207	G1	1590	A16440	470	G39	2489	A3224	300
G266	824	A46632	180	G1499	1565	A151	449	G1460	2108	A3370	292
G51	736	A126123	175	G1448	1424	A345	425	G1118	1381	A158534	268
G44	639	A57998	152	G42	1350	A53200	414	G51	1081	A5745	240
G45	628	A10544	147	G22	1072	A2669	380	G1448	978	A4476	225

Table 5's CRM analysis meticulously examines customer segmentation and market dynamics based on e-invoice data during the months of October, November, and December. This period showcases significant transactional activities from sellers such as G15 and G8013, reflecting the efficacy of their marketing strategies and customer relationship management. The substantial increase in transactions by G8013 in November and December particularly highlights its dominance in the market during these months, suggesting a successful expansion of its customer base through effective marketing campaigns or loyalty programs. Conversely, the consistently high transaction volumes associated with buyers like A6532 and A1288 emphasize their strategic importance as robust purchasers to the sellers, underscoring the necessity for businesses to cultivate long-term relationships with such clients. The CRM analysis provides pivotal insights into understanding customer behaviors and market trends, enabling businesses to devise customized strategies targeted at specific customer segments and solidify their market positions. These insights are crucial for businesses aiming to gain a competitive edge and expand their market share by enhancing customer satisfaction. This analysis not only underlines the dynamic nature of the market but also highlights the potential for businesses to leverage detailed transactional data to refine their marketing and customer relationship strategies for better alignment with customer needs and preferences.

Table 6. 12-month province-based relationship between sender and receiver.

	CITY	SENDER	RECEIVING	PEN NUMBER	AMOUNT
January	İSTANBUL	G42	A10561	7814	1871006,55
	İSTANBUL	G1	A335462	1634	624271,38
	NİĞDE	G927	A31578	1559	24910400,2
	İSTANBUL	G44	A339090	1475	155335,56
	İSTANBUL	G34	A1288	1390	268774,37
	İSTANBUL	G547	A63543	1315	152174,24
	BURSA	G168	A46632	1273	487803,97
	İSTANBUL	G1499	A15102	1054	427971
	İSTANBUL	G1	A1288	793	130920,6
February	İSTANBUL	G1118	A357158	8852	2169844,83
	İSTANBUL	G1118	A51282	7751	1854046,4
	İSTANBUL	G21	A151	7574	5858144,68
	İSTANBUL	G1118	A43782	6027	1701150,29
	İSTANBUL	G1118	A38968	4934	1738852,15
	İSTANBUL	G44	A6706	4529	97313,7
	İSTANBUL	G42	A10561	4009	857450,81
	ANKARA	G74	A170899	3090	2680321,81
	İZMİR	G1745	A494	2742	4459331,46
March	İSTANBUL	G42	A10561	7341,00	2401341000000,00
	İSTANBUL	G547	A23973	5619,00	153677100000,00
	İSTANBUL	G21	A151	3382,00	2320540000000,00
	İSTANBUL	G2103	A4476	3000,00	5131700000000,00
	İZMİR	G1745	A494	2815,00	8108454000000,00
	BURSA	G169	A3051	1834,00	435023600000,00
	ISTANBUL	G1	A1288	1824,00	442142900000,00
	ANKARA	G1561	A18489	1628,00	48162880000,00
	ANKARA	G3222	A32723	1543,00	3690151000000,00

Table 7. 12-month province-based relationship between sender and receiver (Cont'd).

	CITY	SENDER	RECEIVING	PEN NUMBER	AMOUNT
April	İSTANBUL	G42	A10561	7999,00	2097325,58
	İZMİR	G3708	A61718	3386,00	4563442,68
	İSTANBUL	G3202	A61985	3300,00	73433,19
	İSTANBUL	G21	A151	2627,00	2067914,31
	BURSA	G168	A46632	2181,00	987281,55
	ISTANBUL	G1	A1288	1670,00	347343,42
	İSTANBUL	G1713	A53459	1659,00	45454,10
	İSTANBUL	G828	A54881	1622,00	91692,10
	İSTANBUL	G3127	A63777	1376,00	69368,08
May	İSTANBUL	G1118	A51282	27133,00	5678617,13
	İSTANBUL	G1118	A68515	15471,00	4259040,83
	İSTANBUL	G1118	A43782	14031,00	4648043,96
	İSTANBUL	G1118	A38968	13621,00	4149428,00
	İSTANBUL	G1118	A68537	7536,00	2592643,82
	İSTANBUL	G1118	A51275	7257,00	2218249,06
	İSTANBUL	G1118	A51278	6959,00	1719773,06
	İSTANBUL	G1118	A68538	6696,00	2232065,45
	İSTANBUL	G42	A10561	6485,00	2575899,83
June	İSTANBUL	G1118	A51282	29827,00	5099106,21
	İSTANBUL	G1118	A68515	22898,00	5158290,06
	İSTANBUL	G1118	A43782	16095,00	3657718,71
	İSTANBUL	G1118	A38968	15520,00	4094732,72
	İSTANBUL	G42	A10561	15328,00	5371360,78
	İSTANBUL	G21	A151	15145,00	15314276,42
	İSTANBUL	G1118	A68537	8824,00	2028063,06
	İSTANBUL	G1118	A51275	8656,00	1997136,99
	İSTANBUL	G1118	A68538	7834,00	2388308,20
July	İSTANBUL	G6312	A129326	7501,00	2071997,49
	İSTANBUL	G42	A10561	7335,00	6604315,46
	İSTANBUL	G21	A151	4429,00	6065923,83
	İSTANBUL	G1	A1288	1815,00	463083,57
	İSTANBUL	G50	A6532	1393,00	554760,13
	İSTANBUL	G45	A5745	1288,00	8002590,85
	İSTANBUL	G6312	A129330	1255,00	375934,62
	İSTANBUL	G44	A31027	1069,00	76782,35
	İSTANBUL	G44	A63349	1007,00	57450,05
August	İSTANBUL	G21	A151	9748,00	9309818,65
	İSTANBUL	G21	A44	3168,00	3092323,76
	İSTANBUL	G1499	A15102	2410,00	1125961,35
	GAZİANTEP	G995	A79237	1763,00	192066,32
	İSTANBUL	G1118	A38968	1754,00	837490,02
	İSTANBUL	G2201	A15608	1684,00	20295040,72
	İSTANBUL	G50	A6532	1668,00	652986,63
	İSTANBUL	G45	A5745	1622,00	9203701,90
	İSTANBUL	G76	A41985	1527,00	510502,73
September	ANTALYA	G7538	A193270	254211824,00	845488,92
	ANTALYA	G7538	A193265	248834848,00	254669,00
	ANTALYA	G7538	A126817	240225328,00	1291750,83
	ANTALYA	G7538	A193282	221442272,00	294280,30
	ANTALYA	G7538	A193272	173889568,00	166931,22
	ANTALYA	G7538	A193305	162011808,00	633883,97
	ANTALYA	G7537	A193265	152756416,00	206815,53
	ANTALYA	G7538	A193278	143439072,00	1243042,77
	ANTALYA	G7538	A193280	126773792,00	44756,64
October	İSTANBUL	G42	A10561	13608,00	5722217,82
	İSTANBUL	G2126	A1288	2002,00	320610,25
	BURSA	G168	A46632	1742,00	1572671,58
	ANKARA	G763	A34627	1386,00	564253,42
	İSTANBUL	G45	A5745	1147,00	6781245,20
	İSTANBUL	G52	A6532	1146,00	409998,76
	İZMİR	G266	A158534	1090,00	683462,31
	ANKARA	G56	A6532	1042,00	1028152,50
	BURSA	G168	A49518	830,00	438219,11

Table 8. 12-month province-based relationship between sender and receiver (Cont'd).

	CITY	SENDER	RECEIVING	PEN NUMBER	AMOUNT
November	İSTANBUL	G1118	A51282	17752,00	6613919,24
	İSTANBUL	G1118	A43782	11457,00	4902972,52
	İSTANBUL	G1118	A68515	10840,00	4084359,68
	İSTANBUL	G1118	A38968	9656,00	4792558,32
	İSTANBUL	G6312	A129326	7744,00	3061943,99
	İSTANBUL	G21	A151	6368,00	5020598,01
	İSTANBUL	G42	A10561	6330,00	3426754,47
	İSTANBUL	G1118	A51275	6151,00	2695834,89
	İSTANBUL	G8013	A239093	5573,00	660680,08
December	İSTANBUL	G42	A10561	10244,00	5591535000000,00
	İSTANBUL	G1118	A68515	5016,00	1885258000000,00
	İSTANBUL	G1684	A25376	4466,00	2368756000000,00
	İSTANBUL	G1118	A51275	3669,00	1779095000000,00
	İSTANBUL	G21	A151	2625,00	2748752000000,00
	İSTANBUL	G3202	A61985	1653,00	41227280000,00
	İSTANBUL	G1	A1288	1582,00	6180978000000,00
	İZMİR	G266	A158534	1581,00	1762418000000,00
	İSTANBUL	G23	A6907	1496,00	249631000000000,00

Table 6-8 represents a monthly shipment table with sender, recipient, city, number of items and amount. Each row contains information for one month. A detailed analysis is made according to the senders and recipients of the months. In the data set for January, the highest number of items and total amount was realized in Istanbul. Senders coded G42, G1 and G927 made significant transactions to receivers coded A10561, A335462 and A31578. Other cities, such as Niğde and Bursa, also witnessed heavy sending activity. In February, senders coded G1118 and G42 stood out in shipments from Istanbul. Significant shipment volumes were also recorded from other cities such as Ankara and Izmir. The total number of items and total value indicate a similar intensity to the previous month. In March, there was a general increase in transactions recorded from cities such as Istanbul, Izmir and Niğde. High-value transactions by senders coded G42 and G2103 were particularly noteworthy. The total amount shows a significant increase compared to the previous months. In April, the high volume of transactions from Istanbul and Izmir stood out. Senders coded G3202 and G1713 recorded significant amounts in their transactions with buyers coded A61985 and A53459. The total number of items and the total amount followed a similar course to the previous months. In May, high value transactions of the sender coded G1118 to different recipients were noteworthy. Generally, high amounts were recorded in transactions from Istanbul. This month indicates a significant increase in terms of the total number of items and the total amount. In June, senders coded G1118 and G21 in Istanbul were observed to have realized particularly high value transactions. There was a significant increase in one transaction from Kocaeli. The total amount continues the growth trend of the previous months. In July, senders coded G6312, G42 and G1 stood out in transactions from Istanbul. A significant increase was observed in one transaction towards the end of the day. In terms of total amount, a course parallel to the previous months was observed. In August, transactions from Istanbul generally recorded high amounts. Transactions from Gaziantep and Kocaeli also involved significant amounts of shipments. In addition, a minor typographical error was detected in the field GONDERER_SEHIR. In September, there was a significant increase in transactions from Antalya. This month is generally characterized by very large transactions. Except for a typo in Istanbul, the size of the transactions is noteworthy. In October, transactions from Istanbul and Ankara were dominated by senders coded G42 and G52. There was a significant increase in transactions from sender G52 to recipient A6532. The total amount reflects the general trend of the previous months. In November, a large number of transactions from Istanbul and Ankara stood out. The sender coded G1118 was observed to transact extensively with different recipients. The total amount continues the general upward trend of the previous months. In December, transactions from Istanbul, Ankara and Bursa were noteworthy. In particular, it was observed that the sender coded G8026 made high volume transactions to the recipient coded A253131. In general, this month followed a similar trend to the previous months in terms of total amount.

When Table 6-8 is re-examined, the lowest value is recorded as the number of items 662 and the amount of 40,337.74 TL in the submission made from Eskişehir in January. This is the submission with the lowest number of items in January. In the same month, another shipment from Istanbul had 793 items and an amount of 130,920.60 TL. In February, the number of items was 400 and the amount was recorded as 28,994.51 TRY in a submission from Icel. In March, a shipment from Ankara was recorded as 972 items and the amount as 9,175,680,000.00 TL. In April, in another shipment from Istanbul, the number of items was 995 and the amount was recorded as

765,543.33 TL. In a shipment from Istanbul in May, the number of items was 1,583 and the amount was 339,052.89 TL. In June, a shipment from Kocaeli reported 2,892 items and an amount of 42,702,541.84 TRY. In July, a shipment from Ankara reported 793 items and an amount of TRY 598,157.29. In another shipment from Istanbul in August, the number of items was 806 and the amount was recorded as 23,553.35 TL. In September, the number of items in a shipment from Antalya was quite high and the amount was recorded as 845,488.92 TL. In October, the number of items in a shipment from Istanbul was 555 and the amount was recorded as 13,965,840.57 TL. In November, the number of items in a shipment from Bursa was 830 and the amount was recorded as 438,219.11 TL. In December, the number of items in a shipment from Istanbul was 629 and the amount was 730,721.00 TL.

In this study, following the CRM analysis, RFM analysis was performed and the results are given in Table 9.

Table 9. RFM analysis results.

Master_Id	Recency	Frequency	Monetary	Recency_Score	Frequency_Score	Monetary_Score	Rf_Score
52 (G7538)	19 (A193270)	254211824	845488,9	4	5	3	45
34 (G42)	125 (A10561)	248834848	254669	4	5	2	45
52	5 (A126817)	240225328	1291751	5	5	3	55
52	23 (A193282)	221442272	294280,3	4	5	2	45
52	20 (A193272)	173889568	166931,2	4	5	1	45
44 (G59)	44 (A26089)	434	4607550	4	1	4	41
44	69 (A36769)	414	874649,9	3	1	3	31
1 (G1099)	6 (A1288)	410	293034,1	5	1	2	51
9 (G154)	61 (A334750)	400	28994,51	3	1	1	31

According to the RFM analysis results in Table 9, customer segmentation was done in two different groups. G7538 (Master_Id 52) and A193270 (Recency 19) represent a highly active and valuable customer segment, with a transaction volume of 254211824 units and a monetary value of 845488.9 units. They scored 4, 5, and 3 for Recency, Frequency, and Monetary values, respectively, achieving a total RF score of 45. This indicates they have recently made purchases, are frequent shoppers, and spend above-average amounts.

G42 (Master_Id 34) and A10561 (Recency 125) have a transaction volume of 248834848 units and a monetary value of 254669 units. Their scores for Recency, Frequency, and Monetary values are 4, 5, and 2, respectively, resulting in a total RF score of 45. This shows that they have also been active shoppers recently, but despite frequent shopping, their spending is lower compared to the previous group.

Other "52" Master_Id customers: They have conducted transactions with high monetary values with different receiver codes (A126817, A193282, A193272) and have high Recency and Frequency scores. Notably, transactions to A126817 achieved the highest RF score of 55, indicating these customers shop very frequently and have recently made high-volume purchases.

Lower Score Customers (G59 and G154): Their lower Frequency and Monetary scores indicate these customer segments shop less frequently and spend lower amounts. The specified RF score of 51 for G1099 (Master_Id 1) and A1288 suggests this customer, despite infrequent shopping, possesses a high monetary value.

When Tables 6-8 and 9 are re-examined, it is observed that for all months, shipper G42 and seller A10561 are jointly found to be the best performing senders in both RFM and CRM analyses. Thus, it is observed that the results of both CRM and RFM analyses are consistent with each other.

CRM and RFM analyses share similarities, but they emphasize different aspects of customer behavior. CRM analysis focuses on managing customer relationships and tracking customer interactions, while RFM analysis evaluates customer behavior in terms of purchase frequency, recency of purchases, and monetary value. Both analyses contribute to customer segmentation and understanding of customer behavior, but CRM is more focused on transaction volume and customer loyalty, whereas RFM emphasizes shopping habits and spending patterns. In Table 10, the G42 - A10561 pair is identified as a frequent shopper with low monetary value in both analyses, while the G7538 - A193270 pair shows strong performance in both CRM and RFM, with high.

Table 10. Comparison of RFM and CRM analyses.

Customer ID	CRM Analysis (Sender-Receiver Pairs with Highest Transaction Volume)	RFM Analysis (Recency, Frequency, and Monetary Value)	Evaluation
G42 - A10561	One of the sender-receiver pairs with the highest transaction volume	Recency: 4, Frequency: 5, Monetary: 2, RF Score: 45	Among the most active pairs in both CRM and RFM analyses, but the monetary value is relatively low according to the RFM analysis
G7538 - A193270	The pair with the highest transaction volume	Recency: 4, Frequency: 5, Monetary: 3, RF Score: 45	Strong performance in both analyses, showing high frequency and monetary value.
G1099 - A1288	Moderate transaction volume	Recency: 5, Frequency: 1, Monetary: 2, RF Score: 51	Despite infrequent purchases, the high monetary value makes this customer strategically significant.

5. Discussion and Conclusion

In this study, CRM and RFM analyses were performed on the annual sales data obtained as a result of e-invoice usage service to customers. When the results of the analysis were analyzed, the number of transactions belonging to the SENDER, RECIPIENT and parties in the top 10 every month were extracted. Thus, the most active companies in e-invoice service transactions were determined. Although the number of transactions is important in the activity of companies, it should be stated that the number of purchases and material amount values of each transaction are also important values. With the study, valuable customer segmentation based on the number of transactions and number of items was made for the customer profiles of the sender and receiver sides. In addition, transaction volumes by province were analyzed in terms of senders and receivers. With this study, analysis results that will contribute to more systematic pricing, campaign and customer relationship management for the sender and receiver companies (customers) to which e-invoice service is provided were obtained. In the future, it aimed to carry out studies for the predictive analysis of these data to shape future projections with regression-based models and to identify sectoral-oriented customer patterns in trade data.

References

[1] Marcus C. A practical yet meaningful approach to customer segmentation. J of Cons Mark 1998;15(5):494–504.
[2] Jonker JJ, Piersma N, Van Den Poel D. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. Exp Sys w Apps 2004;27(2):159–68.
[3] Kim SY, Jung TS, Suh EH, Hwang HS. Customer segmentation and strategy development based on customer lifetime value: A case study. Exp Sys w Apps 2006;31(1):101–7.
[4] Lemon KN, Mark T. Customer Lifetime Value as the Basis of Customer Segmentation. J of Relation Mark 2006;5(2–3):55–69.
[5] Cooil B, Aksoy L, Keiningham TL. Approaches to Customer Segmentation. J of Relation Mark 2008;6(3–4):9–39.
[6] Chan C. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. Exp Sys w Apps 2008;34(4):2754–62.
[7] Namvar M, Gholamian MR, KhakAbi S. A Two Phase Clustering Method for Intelligent Customer Segmentation. In: Modelling and Simulation 2010 International Conference on Intelligent Systems 2010; p. 215–9.
[8] Wu RS, Chou PH. Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. Elect Com Res and Apps 2011;10(3):331–41.
[9] Hiziroglu A. Soft computing applications in customer segmentation: State-of-art review and critique. Exp Sys w Apps 2013;40(16):6491–507.
[10] Hosseini M, Shabani M. New approach to customer segmentation based on changes in customer value. J Market Anal 2015; 3(3):110–21.
[11] Brito PQ, Soares C, Almeida S, Monte A, Byvoet M. Customer segmentation in a large database of an online customized fashion business. Robo and Comp-Integ Manufact 2015;36:93–100.
[12] Kansal T, Bahuguna S, Singh V, Choudhury T. Customer Segmentation using K-means Clustering. In: 2018 International Conference on Computational Techniques, Elec and Mec Sys (CTEMS) [Internet]. 2018. p. 135–9.
[13] Amnur H. Customer relationship management and machine learning technology for identifying the customer. JOIV: Inter J on Inf Vis 2017; 1(1):12–5.
[14] Al-Fedaghi S, Al-Otaibi M. Service-oriented systems as a thining machine: A case study of customer relationship management. In: 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT) [Internet]. IEEE; 2019; p. 235–42.

[15] Christy AJ, Umamakeswari A, Priyatharsini L, Neyaa A. RFM ranking–An effective approach to customer segmentation. J of King S Uni-Com and Inf Sci 2021; 33(10):1251–7.
[16] Taşabat SE, Esra A. Recyclıng project wıth rfm analysıs ın ındustrıal materıal sector. Sig J of Eng and Nat Sci. 2020; 38(4):1681–92.