Logo

Description automatically generated with low confidence

T.C.

MARMARA UNIVERSITÄT

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT  
DEUTSCHSPRACHIGE ABTEILUNG FÜR WIRTSCHAFTSINFORMATİK

**KUNDENSEGMENTIERUNG MIT RFM-ANALYSE**

LIZENZARBEIT

HASENE CEREN YIKILMAZ

ISTANBUL, 2022

T.C.

MARMARA UNIVERSITÄT

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT  
DEUTSCHSPRACHIGE ABTEILUNG FÜR WIRTSCHAFTSINFORMATİK

**KUNDENSEGMENTIERUNG MIT RFM-ANALYSE**

LIZENZARBEIT

HASENE CEREN YIKILMAZ

BERATERIN: ARŞ. GÖR. DR. İLKİM ECEM EMRE

ISTANBUL, 2022

**ÖZET**

**RFM ANALİZİ İLE MÜŞTERİ SEGMENTASYONU**

Bu çalışma, müşteri segmentasyonunda kullanılan RFM yönteminden elde edilecek sonuçları incelemek amacıyla hazırlanmıştır. E-ticaretin son derece önemli olduğu bu günlerde, müşteri kayıp analizi bu sektörün olmazsa olmazlarından biridir. Günümüzde müşteriyi en iyi şekilde tanımak ve müşteriye nasıl yaklaşmak gerektiğini belirlemek iş stratejileri arasında son derece öne çıkmıştır. Aynı zamanda sahip olduğu özellikler ile her müşteriye kişisel olarak erişmeyi sağlayan Veri Madenciliği de en önemli araçlardan biri haline gelmiştir.

Bu çalışma kapsamında bir e-ticaret şirketinden elde ettiğimiz veriler ile müşterileri segmente ettik ve sonuçlarını inceledik.

**Anahtar Kelimeler:** RFM, Müşteri Segmentasyonu, Veri Madenciliği

**ABSTRAKT**

**KUNDENSEGMENTIERUNG MIT RFM-ANALYSE**

Diese Studie wurde erstellt, um die Ergebnisse zu untersuchen, die aus der bei der Kundensegmentierung verwendeten RFM-Methode zu erhalten sind. In der heutigen Zeit, in der E-Commerce wichtig ist, ist die Kundenverlustanalyse eine der unabdingbaren Voraussetzungen dieser Branche. Heutzutage ist es bei den Geschäftsstrategien in den Vordergrund gerückt, den Kunden am besten zu kennen und zu bestimmen, wie man auf den Kunden zugeht. Gleichzeitig ist Data Mining, das mit seinen Features jedem Kunden einen persönlichen Zugang ermöglicht, zu einem der wichtigsten Tools geworden.

Im Rahmen dieser Studie haben wir Kunden mit den Daten eines E-Commerce-Unternehmens segmentiert und die Ergebnisse analysiert.

**Schlüsselwörter:** RFM, Kundensegmentierung, Datenverwaltung

**VORWORT**

Bu çalışmanın hayata geçirilmesinde bana destek veren, vakit ayıran, bilgi ve tecrübelerinden faydalandığım danışmanım Arş. Gör. Dr. İlkim Ecem Emre’ye, zorlandığım noktalarda hem benimle bilgisini paylaşan hem de manevi desteğini esirgemeyen Onur Yurteri’ye ve hayatımın her aşamasında olduğu gibi bana destek olup güvenini hiç kaybetmeyen annem Selma Avcı’ya sonsuz teşekkürlerimi sunarım.

HASENE CEREN YIKILMAZ

**INHALTSVERZEICHNIS**

[TABELLENVERZEICHNIS v](#_Toc96652685)

[ABBILDUNGSVERZEICHNIS vi](#_Toc96652686)

[ABKÜRZUNGSVERZEICHNIS vii](#_Toc96652687)

[1. EINLEITUNG 1](#_Toc96652688)

[2. THEORETISCHER TEIL 2](#_Toc96652689)

[2.1. E-COMMERCE 2](#_Toc96652690)

[2.2. KUNDENBEZIEHUNGSMANAGEMENT 2](#_Toc96652691)

[2.3. SEGMENTIERUNG 3](#_Toc96652692)

[2.4. KUNDENSEGMENTIERUNG 3](#_Toc96652693)

[2.5. CLUSTERING-TECHNIKEN IN DER KUNDENSEGMENTIERUNG 5](#_Toc96652694)

[2.5.1. K-Means Clustering 5](#_Toc96652695)

[2.5.2. Hierarchical Agglomerative Clustering (HAC) 5](#_Toc96652696)

[2.5.3. Density Based Spatial Clustering of Applications with Noise (DBSCAN) 5](#_Toc96652697)

[2.5.4. RFM Analyse 6](#_Toc96652698)

[2.5.5. Verwandte Arbeiten zur RFM-Analyse 6](#_Toc96652699)

[3. MATERIAL UND METHODEN 8](#_Toc96652700)

[3.1. VERWENDETE PYTHON BIBLIOTHEKEN UND MODULE 8](#_Toc96652701)

[3.2. ÜBER DATENSÄTZE 9](#_Toc96652702)

[3.2.1. Datensatz mit Verkaufsdaten 9](#_Toc96652703)

[3.2.2. Datensatz mit Kundeninformationen 15](#_Toc96652704)

[4. ERGEBNISSE 18](#_Toc96652705)

[4.1. RETENTIONSRATE 18](#_Toc96652706)

[4.1.1. Datenaufarbeitung 18](#_Toc96652707)

[4.1.2. Retentionsrate 19](#_Toc96652708)

[4.2. RECENCY, FREQUENCY AND MONETARY (RFM) ANALYSE 20](#_Toc96652709)

[4.2.1. Finden von RFM-Werten 20](#_Toc96652710)

[4.2.2. RFM (1-4) Analyse 22](#_Toc96652711)

[4.2.2. RFM (1-5) Analyse 26](#_Toc96652712)

[5. FAZIT 30](#_Toc96652713)

[5.1. KURZE BESCHREIBUNGEN DER SEGMENTE 30](#_Toc96652714)

[5.2. VERGLEICH DER ERGEBNISSE UND EMPFEHLUNGEN 31](#_Toc96652715)

[LITERATURVERZEICHNIS 33](#_Toc96652716)

# TABELLENVERZEICHNIS

[Tabelle 1: Am häufigsten und am wenigsten gekaufte Produkte 10](#_Toc96398656)

[Tabelle 2: Meist und am wenigsten gekaufte Produkte nach Anzahl 11](#_Toc96398657)

[Tabelle 3: Bestellungen mit dem größten Warenkorbwert. 11](#_Toc96398658)

[Tabelle 4: die teuersten Produkte 12](#_Toc96398659)

[Tabelle 5: Segmentierungskarte 22](#_Toc96398660)

[Tabelle 6: Segmentstatistiken 23](#_Toc96398661)

[Tabelle 7: Segmentierungskarte (1-5) 26](#_Toc96398662)

[Tabelle 8: Segmentstatistiken (1-5) 27](#_Toc96398663)

[Tabelle 9: Vergleichen 30](#_Toc96398664)

[Tabelle 10: wichtige Segmente 31](#_Toc96398665)

# ABBILDUNGSVERZEICHNIS

[Abbildung 1: Kundenbeziehungsmanagement Dimensionen 2](#_Toc96393307)

[Abbildung 2: Rohdatensatz 8](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393308)

[Abbildung 3: Datensatz, erste Iteration 9](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393309)

[Abbildung 4: beschreibende Statistik 10](#_Toc96393310)

[Abbildung 5: meistverkaufte Produkte 12](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393311)

[Abbildung 6: Verteilung der Werte 12](#_Toc96393312)

[Abbildung 7: beschreibende Grafik 13](#_Toc96393313)

[Abbildung 8: monatliche Verteilung der Einkäufe 13](#_Toc96393314)

[Abbildung 9: Datensatz mit Kundeninformationen 14](#_Toc96393315)

[Abbildung 10: endgültige Version des Datensatzes 15](#_Toc96393316)

[Abbildung 11: Beschreibende Statistik 15](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393317)

[Abbildung 12: Geschlechterverteilung 15](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393318)

[Abbildung 13: Geschlecht und Altersverteilung 15](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393319)

[Abbildung 14: Altersverteilung 16](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393320)

[Abbildung 15: Datensatz für das Konto der Retention Rate 17](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393321)

[Abbildung 16: Retentionsrate 18](#_Toc96393322)

[Abbildung 17: Recency 19](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393323)

[Abbildung 18: Finden Sie das letzte Einkaufsdatum 19](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393324)

[Abbildung 19: Frequency 20](#_Toc96393325)

[Abbildung 20: RFM-Werte 20](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393326)

[Abbildung 21: Monetary 20](file:////Users/cerenyikilmaz/Desktop/x.docx#_Toc96393327)

[Abbildung 22: Viertel 21](#_Toc96393328)

[Abbildung 23: Standardisierung RFM-werte 21](#_Toc96393329)

[Abbildung 24: RFM Score 21](#_Toc96393330)

[Abbildung 25: Segmentierung 22](#_Toc96393331)

[Abbildung 26: Segmentstatistiken 23](#_Toc96393332)

[Abbildung 27: Segmentierung Tree Map 24](#_Toc96393333)

[Abbildung 28: RFM-werte (1-5) 25](#_Toc96393334)

[Abbildung 29: RFM Score (1-5) 25](#_Toc96393335)

[Abbildung 30: Segmentierung (1-5) 26](#_Toc96393336)

[Abbildung 31: Segmentstatistiken (1-5) 27](#_Toc96393337)

[Abbildung 32: Segmentierung Tree Map (1-5) 28](#_Toc96393338)

# ABKÜRZUNGSVERZEICHNIS

**RFM** Recency Frequency Monetary

**CRM** Customer Relationships Management

**OECD** Organisation for Economic Co-operation and Development

**WTO** World Trade Organization

**ITO** İstanbul Ticaret Odası

**HAC** Hierarchical Agglomerative Clustering

**DBSCAN** Density Based Spatial Clustering of Applications with Noise

# EINLEITUNG

Der Kunde ist das wichtigste Kapital im E-Commerce. In diesem Zusammenhang versetzt die Kundensegmentierung E-Commerce-Unternehmen in die Lage, ihre Kunden in verschiedene und homogene Gruppen mit ähnlichen Merkmalen zu unterteilen und mit jedem Kundensegment separat zu interagieren, um die Merkmale und Bedürfnisse ihrer Kunden besser zu verstehen. Darüber hinaus ist die Kundensegmentierung ein entscheidender Erfolgsfaktor, um das Verhalten verschiedener Kundengruppen zu verstehen und deren Geschäftswert zu bewerten. Mit der richtigen Segmentierung können Unternehmen die richtigen Produkte, Dienstleistungen und Assets verwalten, um interessierte Kunden anzusprechen und eine angenehme Beziehung zu ihnen aufzubauen.

Um ein Unternehmen zu erhalten und wachsen zu lassen, müssen viele Aspekte berücksichtigt werden, um sicherzustellen, dass die Erträge maximiert werden. Ein Aspekt, den es zu berücksichtigen gilt, ist die Beobachtung des Kaufverhaltens der Kunden und die Bewertung, wie man hochwertige Kunden entdeckt. Durch die Anpassung der Werbung und die Fokussierung auf hochwertige Kunden können Kundenbeziehungen wachsen und höhere Gewinne erzielen. Die Analyse von RFM-Werten zusammen mit Clustering-Algorithmen zur Beobachtung eines solchen Verhaltens kann aussagekräftige Informationen liefern.

Die RFM-Analyse wird verwendet, um Kundenwerte zu analysieren. Es umfasst drei Aspekte. Der erste Aspekt ist die Aktualität, bei der es darum geht, wie kürzlich der Kunde einen Kauf getätigt hat. Die nächste Überlegung ist die Häufigkeit, wie oft Kunden kaufen. Der letzte Aspekt ist der monetäre Aspekt, der die Summe der Ausgaben des Kunden annimmt. Anhand der empfangenen Werte können dem Kunden Punkte zugeordnet werden. Jedes Segment kann Informationen bereitstellen, die von den wertvollsten bis zu den am wenigsten wertvollen Kunden reichen.

In der Literatur werden mehrere Techniken vorgeschlagen, um eine Kundensegmentierung durchzuführen, und Clustering ist die an der weitesten verbreiteten Methode. Das Clustering kann auf Kundenprofilen oder einer RFM-Analyse basieren.

Während das traditionelle RFM-Modell bei der Kundensegmentierung gute Leistungen erbringt, ignoriert es häufig gekaufte Produkte und Kategorien, die wichtige Einblicke in das Ausgabeverhalten der Kunden liefern. Genauer gesagt berücksichtigt das RFM-Modell nur das Kaufpotenzial, was bedeutet, dass Kunden mit ähnlichen Profilen ein völlig unterschiedliches Ausgabeverhalten haben können.

# THEORETISCHER TEIL

In diesem Abschnitt werden die Themen, die Gegenstand der Bewerbung sind, theoretisch erläutert. E-Commerce, Kundensegmentierung, Analysemethoden der Kundensegmentierung und RFM-Analyse werden theoretisch behandelt.

## E-COMMERCE

Nach Angaben der Welthandelsorganisation (WTO); *„Elektronischer Handel ist die Produktion, Werbung, der Verkauf und der Vertrieb von Waren und Dienstleistungen über Telekommunikationsnetze.“*

Nach einer Definition der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD); *„Elektronischer Handel ist im Allgemeinen die Durchführung von Geschäftsaktivitäten durch Einzelpersonen und Organisationen auf der Grundlage der Übertragung digitaler Daten, einschließlich Text, Ton und visuellen Bildern.“* (ITO 1998; 77) *[[1]](#endnote-1)*

Mobiltelefon, Internet, Fernsehen, Computer, elektronische Zahlungssysteme können zu den grundlegenden Werkzeugen des elektronischen Handels gezählt werden.

## 2.2. KUNDENBEZIEHUNGSMANAGEMENT

CRM ist definiert als die Kombination aus Prozess und Technologie, um den Kunden in Bezug auf die Produkte und Dienstleistungen zu verstehen, die den Kunden angeboten werden. Durch den Einsatz von CRM-Techniken auf Basis der Kundendaten können dem Kunden zielgerichtete Produkte und qualitativ hochwertige Dienstleistungen angeboten werden.

In Bezug auf Parvatiyar & Sheth (2001)[[2]](#endnote-2), und Ngai et al. (2009)[[3]](#endnote-3) , CRM besteht aus vier Dimensionen.

Diagram

Description automatically generated

Abbildung 1: Kundenbeziehungsmanagement Dimensionen

## 2.3. SEGMENTIERUNG

Die Segmentierung ist eine Methode, die darauf abzielt, den Markt, mit dem Marken oder Unternehmen verbunden sind, in Segmente zu unterteilen und eine Strategie für die kleinen Märkte zu entwickeln, die sie aus diesen Segmenten auswählen. Die Segmentierung wird verwendet, um Marketingstrategien für die Bedürfnisse, Wünsche und Wünsche der Zielgruppe des ausgewählten Teilmarkts zu entwickeln.

Marken und Unternehmen identifizieren Gruppen, die ihre Produkte und Dienstleistungen eher kaufen. Auf diese Weise steigt die Verkaufswahrscheinlichkeit des Unternehmens und gleichzeitig sparen sie Geld und Zeit, indem sie keine Zeit mit Kunden verbringen, die das Produkt oder die Dienstleistung nicht kaufen werden.

## 2.4. KUNDENSEGMENTIERUNG

Die Marktsegmentierung steuert die Marketingstrategie eines Unternehmens und die Zuweisung von Ressourcen zu Produkten und Märkten. In der Marketingpraxis ist die Idee, Produkte und Dienstleistungen gezielt auf Untergruppen von Verbrauchern auszurichten, bereits um 1820 dokumentiert.

Deutsche und britische Buchhändler zielten auf Marktsegmente für Ort basierend auf Preis, Geografie, Demografie und Psychografie. (Fullerton 1988) [[4]](#endnote-4)

So scheiterten etwa 85 % der etwa 30.000 neuen Produkteinführungen an unzureichender Marktsegmentierung. (Christensen, Cook, Hall 2005)[[5]](#endnote-5)

Um erfolgreich auf ein Marktsegment abzuzielen, muss das Segment drei Kriterien erfüllen: Die Mitglieder des Segments müssen identische Merkmale aufweisen; das Segment muss zugänglich und für das Targeting verfügbar sein; und das Segment muss groß genug sein, um gemessen werden zu können. (De Sáez, 2002)[[6]](#endnote-6)

In Bezug auf Wedel und Kamakura (2012), *[[7]](#endnote-7)* *“Es gibt vier Möglichkeiten, einen Markt zu segmentieren: nach Geographie, nach demographischen Eigenschaften, nach psychographischen Eigenschaften und nach Verhalten.”*

Der Kunde ist die grundlegendste Grundlage eines jeden Unternehmens. Ohne Kunden kann ein Unternehmen seine Ziele nicht erreichen. Es ist wichtig, den Kunden während des gesamten von uns angebotenen Dienstes zufrieden zu stellen. Wir müssen erkennen, was ein Kunde braucht. Um diese Schwierigkeit zu überwinden, müssen wir die Daten der Kunden analysieren.

Der Prozess der Analyse der Daten und Gruppierung nach Ähnlichkeiten wird als Kundensegmentierung bezeichnet. In einfachen Worten können wir sagen, dass Kundensegmentierung bedeutet, dass es sich um die Identifizierung der Kunden handelt. Die Kundensegmentierung basiert hauptsächlich auf dem Customer-Relationship-Management (CRM).

Wir können die Kundensegmentierung als den primären Schritt bei der Identifizierung von CRM betrachten. Es ist wichtig, die Beziehung zwischen dem Kunden und den Unternehmen zu identifizieren, um das Ziel für die Erzielung ihres Gewinns festzulegen. Die Analyse von Kundendaten hilft dem Unternehmen, sich selbst zu improvisieren, indem es den Kunden gute Dienstleistungen bietet. Es hilft auch, das Produkt individuell anzupassen und hilft, den Service zu verbessern.

Die Kundensegmentierung hilft auch bei der Erhöhung der Kundenbindung, der Loyalität des Kunden und der Ermittlung des Wertes jedes Kunden. Dies hilft dem Unternehmen, seine Marketingstrategie mit Bedacht aufrechtzuerhalten und hilft, auf dem Gebiet wettbewerbsfähig zu bleiben.

Meistens basiert die Segmentierung auf dem Verhalten, psychografischen, geografischen und demografischen Informationen der Kunden. Es gibt mehrere Algorithmen, die für die Segmentierung verwendet werden, nämlich Assoziationsalgorithmen, Clustering-Algorithmen, Klassifizierungsalgorithmen und Regressionsalgorithmen.

Die Kundensegmentierung kann unter diesen Hauptüberschriften untersucht werden: (Ziafat, H. Shakeri, M. 2014)[[8]](#endnote-8)

#### Demographische Segmentierung

Bei dieser Segmentierung wird davon ausgegangen, dass die Gewohnheiten und Ansprüche von Konsumenten mit gleichen demografischen Merkmalen auch in die gleiche Richtung gehen. Potenzielle Kunden werden nach demografischen Merkmalen wie Alter, Bildung, Geschlecht, Familie, ethnischer Zugehörigkeit und sozialer Stellung segmentiert.

#### Geographische Segmentierung

Es ist die Aufteilung der Verbrauchermärkte nach geografischen Kriterien. Die geografische Segmentierung ermöglicht es, spezielle Marketingstrategien nach Ländern oder Regionen zu entwickeln.

#### Psychographische Segmentierung

Bei der Methode der psychografischen Segmentierung werden die Konsumgewohnheiten von Individuen beleuchtet. Sie trennen und segmentieren Kunden nach Lebensstil und Persönlichkeitsmerkmalen.

#### Verhaltenssegmentierung

Es ist die am häufigsten verwendete Segmentierungsmethode. Es zielt darauf ab, Kunden nach ihren aktuellen und vergangenen Nutzungsgewohnheiten zu klassifizieren. Z.B; Untersucht werden Faktoren wie beispielsweise welches Produkt der Kunde nutzt, Häufigkeit der Nutzung der Produkte, Markentreue.

#### Wertbasierte Segmentation

Bei der wertorientierten Segmentierung werden die Kunden nach dem Wert segmentiert. Dies hilft, den wertvollsten Kunden und die Werte jedes Kunden und die Änderungen der Werte durch den zeitlichen Wandel zu identifizieren.

Aktueller Wert = (Durchschnittlicher Zahlungsbetrag für einen Kunden – Kumulierter rückständiger Betrag für den Kunden/Gesamtnutzungsdauer).

## 2.5. CLUSTERING-TECHNIKEN IN DER KUNDENSEGMENTIERUNG

### *2.5.1. K-Means Clustering*

K-Means einen harten Partitional-Clustering-Algorithmus und ist einer der einfachsten und beliebtesten Ansätze. Es wurde in vielen verschiedenen Bereichen eingesetzt, und in letzter Zeit gab es Anwendungen in den Bereichen Finanzen, Signalverarbeitung, Textverarbeitung, Bildverarbeitung, Mobilfunknetze, Windkraftvorhersage, chemische Modellierung und Biologie.

Die K-Means-Clustering-Methode ist ein 1967 von MacQueen entwickelter nicht-hierarchischer Segmentierungs-Clustering-Algorithmus, der die Daten mit Hilfe der Zentren der Cluster charakterisiert. (MacQueen, 1967). [[9]](#endnote-9) Bei der K-Means-Clustering-Methode können die Daten nur zu einem Cluster gehören, daher handelt es sich um einen scharfen Clustering-Algorithmus. K-Means-Clustering kann nur auf kontinuierliche Daten angewendet werden.

Als Vorteile des k-Means-Clustering-Verfahrens lassen sich die einfache Durchführung und die geringere Komplexität gegenüber anderen Verfahren nennen. Als häufige Probleme von Partitionier-Clustering-Methoden kann die Schwierigkeit, die Anzahl der Cluster zu bestimmen, die in der ersten Stufe in der k-means-Clustering-Methode erstellt werden sollen, und zu Beginn der Studie nicht zu wissen, in welcher Phase der angegebene Zyklus enden wird, erstellt werden Probleme. (Han, 2012) [[10]](#endnote-10)

### *2.5.2. Hierarchical Agglomerative Clustering (HAC)*

HAC ist die häufigste Art hierarchischer Clustering-Modelle, die zum Gruppieren von Objekten in Clustern basierend auf ihrer Ähnlichkeit verwendet werden. Das Modell behandelt zunächst jedes Objekt als Singleton-Cluster und berechnet dann den Abstand zwischen verschiedenen Clustern. Je kleiner der Abstand, desto höher die Ähnlichkeit. Als nächstes werden Paare ähnlicher Cluster erfolgreich zusammengeführt, bis alle Cluster zu einem großen Cluster zusammengeführt wurden, der alle Objekte enthält. Das Ergebnis ist eine baumbasierte Darstellung der Objekte, ein sogenanntes Dendrogramm. (L. Liu, H. Mosavat-Jahromi, 2021) [[11]](#endnote-11)

### *2.5.3. Density Based Spatial Clustering of Applications with Noise (DBSCAN)*

DBSCAN (Density Based Spatial Clustering of Applications with Noise) ist ein Dichte-basierter Clustering-Algorithmus. Anders als beim Partitionieren und hierarchischen Clustern gruppiert es den paarweisen Abstand zwischen Eingabedatenmatrix oder Beobachtungswerten, definiert eine Klasse als die maximale Menge von Punkten, die durch Dichte verbunden sind, es kann den Bereich mit ausreichend hoher Dichte in Klassen unterteilen und jede Klassenform erzeugen in verrauschten räumlichen Daten. Der DBSCAN-Algorithmus muss die Anzahl der zu bildenden Klassen nicht im Voraus kennen. Es kann Cluster jeder Form finden, unabhängig von der Reihenfolge der Dateneingabe, und kann mit anormalen Daten (Rauschen) umgehen. (C. Ye, X. Zhao, 2020)[[12]](#endnote-12)

### *2.5.4. RFM Analyse*

RFM steht für Recency, Frequency und Monetary. Die RFM-Analyse ist eine Marketingtechnik, die verwendet wird, um das Kundenverhalten zu analysieren, z. B. wie lange ein Kunde kauft (Aktualität), wie oft ein Kunde kauft (Frequenz) und wie viel ein Kunde ausgibt (Monetär). Es ist eine nützliche Möglichkeit, die Kundensegmentierung zu verbessern, indem Kunden für zukünftige Personalisierungsdienste in verschiedene Gruppen eingeteilt werden, und um Kunden zu identifizieren, die mit größerer Wahrscheinlichkeit auf Werbeaktionen reagieren.

Die aus den RFM-Variablen ermittelten Klassifizierungsregeln liefern Managern nützliche Informationen, um das zukünftige Kundenverhalten vorherzusagen, z. B. wie bald der Kunde voraussichtlich kaufen wird, wie oft der Kunde kaufen wird und was sein Kauf wert sein wird.

Das Konzept des RFM wurde von Bult und Wansbeek (1995) eingeführt und hat sich bei der Anwendung auf Marketingdatenbanken als sehr effektiv erwiesen. (Blattberg 2008). [[13]](#endnote-13) Die RFM-Analyse stützt sich auf die Messwerte Aktualität (R), Häufigkeit (F) und Geldwert (M), drei wichtige kaufbezogene Variablen, die die Wahrscheinlichkeit zukünftiger Käufe der Kunden beeinflussen.

Die Aktualität bezeichnet den Zeitraum zwischen dem letzten Konsumverhalten und dem heutigen Zeitpunkt. Viele Direktvermarkter glauben, dass diejenigen, die zuletzt gekauft haben, eher wieder kaufen werden.

Häufigkeit ist die Anzahl der Transaktionen, die ein Kunde in einem bestimmten Zeitraum tätigt. Diese Metrik wird unter der Annahme verwendet, dass Kunden, die mehr Käufe tätigen, eher Produkte kaufen als Kunden, die weniger Käufe tätigen.

Monetär bezieht sich auf die kumulierte Summe der Geldausgaben eines bestimmten Kunden.

Es ist möglich, Beispiele für die Klassifizierung mit einem 4- oder 5-System in RFM zu sehen.

### *2.5.5. Verwandte Arbeiten zur RFM-Analyse*

Schijns & Schröder (1996) [[14]](#endnote-14)wiesen darauf hin, dass das RFM-Modell eine Methode ist, die hilft, die Stärke der Kundenbeziehung aus Sicht des Konsumverhaltens zu messen.

Kaymak (2001) [[15]](#endnote-15)ist der Ansicht, dass das RFM-Modell eine der bekanntesten Methoden der Kundenwertanalyse ist, da es hilft, kritische Eigenschaften von Kunden durch die Verwendung weniger Kriterien zu bestimmen, was die Komplexität der Kundenwertanalyse reduziert.

Mccarty & Hastak (2007) [[16]](#endnote-16)verwendeten die Kombination aus RFM-Modell und anderen Modellen wie CHAID und logistischer Regression als analytische Methoden für die Direktmarketing-Segmentierung mit zwei verschiedenen Datensätzen.

Chan (2008) [[17]](#endnote-17)führte eine Kombination aus Kunden-Targeting und Kundensegmentierung für Kampagnenstrategien ein, die das Kundenverhalten mit dem RFM-Modell identifizierte und segmentierte Kunden mit dem LTV-Modell bewertete.

Ein weiteres Beispiel ist eine Studie von Chen, Sain & Guo (2012), [[18]](#endnote-18)die eine 29-Fallstudie zur Verwendung von RFM-Modellen mit Data-Mining-Techniken wie K-Means-Clustering und Entscheidungsbauminduktion vorschlug, um die Kunden des Einzelhandels zu segmentieren.

# MATERIAL UND METHODEN

Bei unserer Analyse haben wir die Verkaufsdaten einer E-Commerce-Website für 2021 verwendet.

Wir haben unsere Analyse auf Jupyter Notebook unter Verwendung der Python-Sprache durchgeführt. Wir haben die RFM-Methode verwendet, um Kunden zu segmentieren.

## 3.1. VERWENDETE PYTHON BIBLIOTHEKEN UND MODULE

Wir haben einige Bibliotheken in Python verwendet, um den Datensatz zu laden, die Analyse durchzuführen und die Daten zu visualisieren. Ihre Namen und Kurzdefinitionen sind wie folgt;

* import pandas

Mit dieser Bibliothek werden Operationen wie das Lesen von Daten, Vorverarbeitung und Reinigung durchgeführt.

* import numpy as np

Die Pandas-Bibliothek, die es uns ermöglicht, mit mehrdimensionalen Arrays und Matrizen zu arbeiten, wird auch verwendet, um mathematische Operationen durchzuführen.

* import datetime as dt

Es ist ein Modul, das uns verschiedene Funktionen zur Verfügung stellt, um mit Zeit, Zeit und Datum zu arbeiten.

* import glob

Es ist ein Modul, mit dem wir bestimmte Dateien in einem Ordner in Python auflisten und verwenden können. Wir haben dies verwendet, um 12 separate CSV-Dateien, die für 12 Monate erstellt wurden, als einen einzigen Datenrahmen zu importieren.

* import os

Es ist ein Modul, mit dem wir problemlos mit Dateien und Verzeichnissen arbeiten können.

* from pathlib import Path

Es ist eine Bibliothek, die verwendet wird, um Textdateien zu manipulieren.

* import matplotlib.pyplot as plt

Es ist eine Bibliothek, die es uns ermöglicht, Daten durch 2- oder 3-dimensionale Zeichnungen zu visualisieren.

* import seaborn as sns

Es ist eine Bibliothek, die hauptsächlich für statistische Visualisierungen verwendet wird.

* import squarify

Es ist die Bibliothek, mit der wir Baumkarten erstellen können.

* import hashlib

Es ist die Bibliothek, die verwendet wird, um verschiedene Verschlüsselungsalgorithmen auszuführen.

## 3.2. ÜBER DATENSÄTZE

### *3.2.1. Datensatz mit Verkaufsdaten*

Unser Datensatz enthält Verkaufsdaten einer E-Commerce-Site für 2021.

Unser Datensatz enthält insgesamt 6 Spalten. A picture containing graphical user interface

Description automatically generated

Abbildung 2: Rohdatensatz

*customer\_id* 🡪 Identitätsinformationen, die für jeden Kunden, der einen Kauf tätigt, eindeutig sind

*item\_id* 🡪 Identitätsinformationen, die für jedes gekaufte Produkt eindeutig sind

*item\_name* 🡪 Namen gekaufter Produkte

*unit\_price* 🡪 Einzelpreise der Produkte

*item\_count* 🡪 Anzahl der gekauften Produkte

*purchase\_time* 🡪 Wenn der Kunde den Kauf tätigt

Im Datensatz wurde definiert als die Parameter customer\_id und unit\_price *float*,

item\_id und item\_count Parameter *integer*,

item\_name und purchase\_time Parameter *object*.

Da wir hier keine mathematischen Operationen mit den Parametern customer\_id und item\_id durchführen müssen, werden sie als Objekt verwendet; Da wir auch den Parameter „purchase\_time“ als Datum benötigen, haben wir ihn in „datetime“ geändert. Da der Parameter customer\_id als Ganzzahl kommt, haben wir außerdem die am Ende auftretenden „.0“-Werte entfernt.

Text

Description automatically generated with medium confidenceDas endgültige Erscheinungsbild des Datensatzes in der ersten Stufe ist wie folgt;

Abbildung 3: Datensatz, erste Iteration

Unser Datensatz hat insgesamt 16.105.820 Zeilen, 3.706.002 eindeutige item\_id und 2.847.919 eindeutige customer\_id.

#### Daten Löschen

In unserem Datensatz; Es gibt 406.530 Datensätze mit leerem Parameter “customer\_id” und 2 Datensätze mit leerem Parameter “item\_name”.

Da wir für jeden Kunden in unserer Anwendung eine eindeutige customer\_id benötigen, haben wir diese Datensätze aus dem Datensatz entfernt.

Wir haben auch die Datensätze mit dem unit\_price-Wert kleiner als 0 gelöscht, um Inkonsistenzen zu vermeiden.

Bei der Untersuchung des Datensatzes haben wir festgestellt, dass es zu viele Produkte mit dem Namen “Uçak Bileti” gibt, aber wir haben festgestellt, dass diese Produkte unlogische Werte in Bezug auf den Einheitspreis enthalten. Aus diesem Grund haben wir die Produkte “Uçak Bileti” aus dem Datensatz gelöscht, um Dateninkonsistenzen zu vermeiden.

Nach der Datenbereinigung hat unser Datensatz insgesamt 15.568.066 Zeilen, 3.635.505 eindeutige item\_id und 2.838.497 eindeutige customer\_id.

#### Datenaufarbeitung

Ich benötige eine eindeutige Bestellnummer für jede Bestellung, die in denselben Warenkorb gelegt wird, um sie in beschreibenden Statistiken zu verwenden. Es gibt keine Bestellnummer in unserem Datensatz.

Aus diesem Grund haben wir eine Bestellnummer erstellt, die für jeden Kunden und jede Bestellung einzigartig sein wird.

Dazu haben wir die Variablen customer\_id und purchase\_time verwendet. Wenn es Werte für eine customer\_id mit genau derselben Purchase\_time gibt, ist die dafür erstellte Bestellnummer dieselbe, da sie im selben Warenkorb gekauft wurden.

Wir haben den Hash-Algorithmus in Python verwendet, um dies zu erstellen.

df = df.copy()

key\_combination = ['customer\_id', 'purchase\_time']

df['invoice\_number'] = list(map(lambda x: hashlib.sha1('-'.join([str(col\_value) for col\_value in x])

.encode('utf-8')).hexdigest(), df[key\_combination].values))

Ich brauche auch das Gesamtgeld, das jeder Kunde für jede Bestellung ausgibt, während ich die Analyse durchführe. Aus diesem Grund haben wir den Parameter total\_price mithilfe der Variablen item\_count und unit\_price erhalten.

*df["total\_price"] = df["item\_count"] \* df["unit\_price"]*

#### Beschreibende Statistik

Der in der Analyse zu verwendende Datensatz besteht aus 15.568.066 Zeilen und 6 Spalten.

In meinem Datensatz befinden sich 2.838.497 eindeutige Kunden und 3.635.505 eindeutige Produkte.

Ältestes Datum im Datensatz: 01.01.2021 00:00:11

Neuestes Datum im Datensatz: 2021-12-31 23:59:03

Insgesamt wurden 10.958.403 eindeutige Transaktionen durchgeführt.

Table

Description automatically generated with medium confidence

Abbildung 4: beschreibende Statistik

Hier sind die 5 am häufigsten und am wenigsten gekauften Artikel, basierend darauf, wie oft die Artikel gekauft wurden, unabhängig davon, wie viele der Artikel gekauft wurden:

|  |  |
| --- | --- |
| **Häufigsten 5** | |
| Xiaomi Mi TV Stick Android TV Medya Oynatıcı | 14351 |
| Enjoy Cat Food Tavuklu Yetişkin Kedi Maması 15 KG | 13147 |
| Apple iPhone 11 64 GB (Apple Türkiye Garantili) | 11543 |
| Xiaomi Mi Wifi Pro Türkiye Versiyon 300 Mbps 2.4 Ghz Sinyal Güçlendirici | 10606 |
| Artdeco Akrilik Boya 140 ml Renk Seçmeli Canlı ve Pastel | 9500 |
| **Wenigsten 5** | |
| RED APPLE SULU BOYA FIRÇASI NO:5 | 1 |
| BRONS SULU BOYA FIRÇA NO 00 | 1 |
| Metal Çeki Demiri 1/10 | 1 |
| Mikro Renkli Simli Yapıştırıcı 6 lı Gg-1242 | 1 |
| Keskin Color A5 Haftalık Ajanda - Pond Lily | 1 |

Tabelle 1: Am häufigsten und am wenigsten gekaufte Produkte

Hier sind die 5 meist- und am wenigsten verkauften Artikel, basierend darauf, wie viele der Artikel gekauft wurden:

|  |  |
| --- | --- |
| **Häufigsten 5** | |
| Damlama Sulama Kurtağızlı Tüm Ek Parçalar 16mm-20mm Damlama Set | 51011 |
| Demaks Mat Siyah Lüks Mobilya Dolap Çekmece Kulpları -Tüm Ölçüler | 29109 |
| DAMLAMA SULAMA YEŞİL DAMLATICI DEBİ AYARLI MEME | 21625 |
| DAMLAMA SULAMA KIRMIZI DAMLATICI DEBİ AYARLI MEME | 21069 |
| DAMLAMA SULAMA EK PARÇALAR (16mm.-20mm.) | 19747 |
| **Wenigsten 5** | |
| Kahramanların Görevi/Morgan Rice | 1 |
| Kahramanlar Çelik Yağmur Orta Boy Çelik Çaydanlık Kırmızı | 1 |
| Kahramanlar Çelik Yağmur Mega Boy Çaydanlık Takımı | 1 |
| Kahramanlar Çelik Sırma 20x6 cm Kısa Tencere | 1 |
| Metal Sibop Kapağı // Anahtarlık Hediyeli // Logolu | 1 |

Tabelle 2: Meist und am wenigsten gekaufte Produkte nach Anzahl

Die 10 Bestellungen mit dem höchsten Gesamtbetrag (TL) der in einem Warenkorb platzierten Bestellungen:

|  |  |
| --- | --- |
| **Top 10** | |
| 2cc11788d9b777380f47f6a835a01670eddb9b37 | 11800915.55 |
| aa1def46326573f96c8c8771daa1d9cc06b48d7f | 8249375.00 |
| b158901b0a6d46e2734a8d4b0c5877fa6b545f2e | 7457741.22 |
| f62ab70ad78f747c69db12146dea5aa885df9aca | 7130800.00 |
| b887b2031b7922e17ea3f43404d36bc49e2b7c9a | 5661000.00 |
| 6a91dbe867ee81e7bb70d47b4b7a6260bdf49955 | 5282900.00 |
| 7282895d04f4dbbee3afeee32d93518875a6c5e3 | 4190441.04 |
| c6f98f193325b477dfa0040c7bd233d64bffce16 | 4181338.00 |
| 64e169073617371f0e2c219f29eaa1e3042dc5b5 | 3968000.00 |
| fbcd1a2073eaf88e30da698add778011f1b9a745 | 3796400.00 |

Tabelle 3: Bestellungen mit dem größten Warenkorbwert.

Die teuersten gekauften Produkte:

|  |  |
| --- | --- |
| **Top 5** | |
| NVIDIA Quadro RTX 8000 VCQRTX8000-PB 48 GB 384 Bit GDDR6 Ekran Kartı | 496000.00 |
| 24 Ayar 100 Gram Külçe Altın | 465604.56 |
| 24 Ayar 50 Gram Külçe Altın | 272550.48 |
| 50 Gram Külçe Altın 24 Ayar | 260749.10 |
| 100 Gram Külçe Altın 24 Ayar | 260491.35 |

Tabelle 4: die teuersten Produkte

Meistverkaufte Produkte:

Chart, bar chart

Description automatically generated

Abbildung 5: meistverkaufte Produkte

Verteilung von Stückpreis, Stückzahl und Gesamtpreiswerten:

Diagram, box and whisker chart

Description automatically generated

Abbildung 6: Verteilung der Werte

Chart, funnel chart

Description automatically generated

Abbildung 7: beschreibende Grafik

Chart, bar chart

Description automatically generated

Abbildung 8: monatliche Verteilung der Einkäufe

### 3.2.2. Datensatz mit Kundeninformationen

Wir haben auch einen Datensatz, der das Kundenprofil dieser E-Commerce-Site enthält. Hier gibt es 4 verschiedene Spalten.

Table

Description automatically generated

Abbildung 9: Datensatz mit Kundeninformationen

*customer\_id* 🡪 Identität, die für jeden Kunden einzigartig ist

*gender* 🡪 0: Männlich, 1: Weiblich, 2: Nicht spezifiziert

*birth\_date* 🡪 Geburtsdaten der Kunden

*platform* 🡪 Gerät, auf dem Kunden die mobile App heruntergeladen haben

Im Datensatz sind die Parameter customer\_id und gender integer,

Geburtsdatum und Plattformparameter sind als Objekt definiert.

Hier haben wir die Variable birth\_date auf date und die Variable customer\_id auf object geändert.

Es gibt keine Nullwerte im Datensatz. In seiner endgültigen Form besteht es aus 2.042.094 Zeilen und 5 Spalten.

Da nicht jeder Kunde Informationen zu Geschlecht und Geburtsdatum hat, führen wir eine Überprüfung nur bei Kunden durch, die über diese Informationen verfügen.

#### Datenaufarbeitung

Es gibt eine Änderung, die wir in meinem Datensatz vorgenommen haben, der Kundeninformationen enthält;

Um die Altersverteilung der Kunden einfacher darzustellen, haben wir das Alter der Kunden berechnet, indem wir die Variable „birth\_date“ vom heutigen Datum subtrahiert und eine neue Spalte mit dem Namen „age“ erstellt haben.

Table

Description automatically generated

Abbildung 10: endgültige Version des Datensatzes

#### Beschreibende Statistik

Abbildung 11: Beschreibende Statistik

Chart, box and whisker chart

Description automatically generatedChart, bar chart

Description automatically generatedA picture containing diagram

Description automatically generated

Abbildung 12: Geschlechterverteilung

Abbildung 13: Geschlecht und Altersverteilung

Chart, bar chart, histogram

Description automatically generated

Abbildung 14: Altersverteilung

# 4. ERGEBNISSE

## 4.1. RETENTIONSRATE

### *4.1.1. Datenaufarbeitung*

Bevor ich eine Berechnung der Bindungsrate durchführte, entfernte ich zuerst den Zeitstempel aus dem Parameter „purchase\_time“ in meinem Datensatz und erstellte eine neue Spalte namens „purchase\_date“.

Ich habe eine Funktion verwendet, die das Jahr und den Monat aus dem angegebenen Datum nimmt und den Tag als 1 definiert:

def get\_month(x): return dt.datetime(x.year, x.month, 1)

Mit dieser Funktion habe ich die Tage in purchase\_date auf 1 geändert und eine Spalte namens invoice\_month erstellt, die Spalte, die uns hilft, die Monatsinformationen dieser Bestellung zu verwenden.

Ich habe auch eine Variable mit dem Namen cohort\_month erstellt, indem ich die Variablen invoice\_month und customer\_id gruppiert habe, dies gibt das erste Kaufdatum für jeden Kunden an.

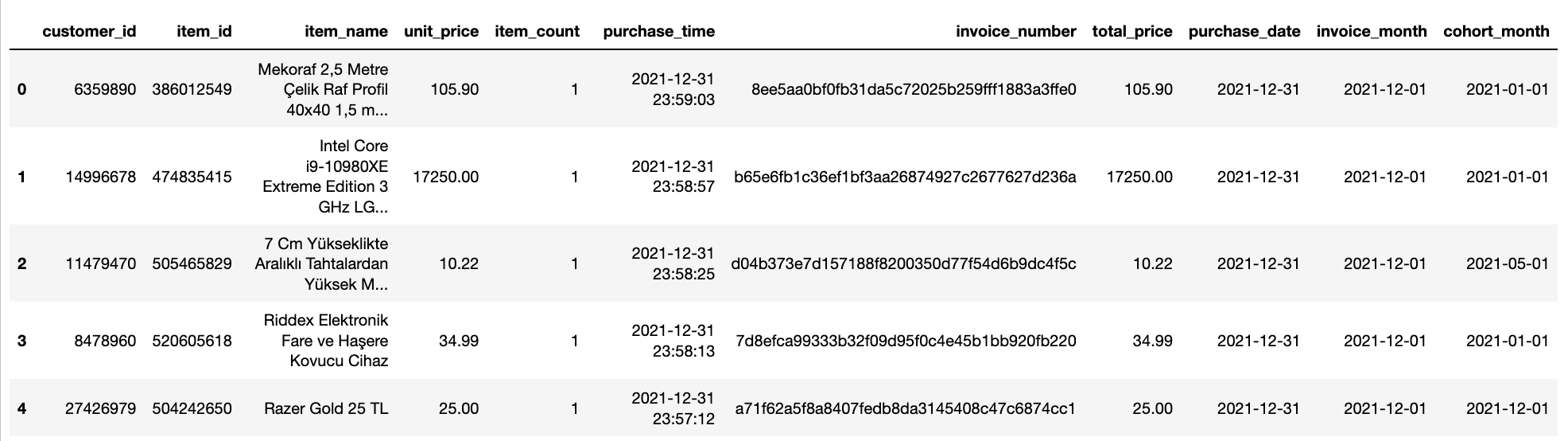
Ich benötige die Tageswerte nicht, da ich meine Analyse monatlich durchführen werde.

Abbildung 15: Datensatz für das Konto der Retention Rate

Mit Hilfe einer Funktion habe ich die Werte für Tag, Monat und Jahr aus den Variablen invoice\_month und cohort\_month gezogen.

def get\_date\_int(df, column):

year = df[column].dt.year

month = df[column].dt.month

day = df[column].dt.day

return year, month, day

Ich habe die Spalte cohort\_index für jeden Benutzer erstellt, um sie zu erstellen, habe ich Folgendes getan:

**Years Difference =** invoice\_year - cohort\_year

**Months Difference =** invoice\_month - cohort\_month

**Cohort Index =** Years Difference \* 12 + Months Difference + 1

### *4.1.2. Retentionsrate*

**Chart

Description automatically generated**

Abbildung 16: Retentionsrate

## 4.2. RECENCY, FREQUENCY AND MONETARY (RFM) ANALYSE

Als ich die Fallstudien untersuchte, stellte ich fest, dass die Bewertung unterschiedlich war. In einigen Studien wurden Aktualität, Häufigkeit und monetäre Variablen mit Werten zwischen 1 und 4 bewertet, während sie in einigen Studien mit Werten zwischen 1 und 5 bewertet wurden. Aus diesem Grund habe ich die Ergebnisse mit beiden Scorings in der Studie verglichen.

### *4.2.1.* *Finden von RFM-Werten*

#### a. Recency: Tage seit dem letzten Kauf

Hier haben wir zuerst den letzten Tag ermittelt, an dem jeder Kunde eingekauft hat, und dann das Datum des letzten Einkaufs des Kunden vom heutigen Datum abgezogen.

Da die Daten für 2021 gelten, haben wir das heutige Datum als 1. Januar 2022 angenommen.

*recency\_df['recency'] = recency\_df['last\_purchase\_date'].apply(lambda x: (now - x).days)*

Recency = Today (01.01.2022) – Last Purchase Date

Auf diese Weise haben wir für jeden Kunden die Variable „Recency“ definiert.Table

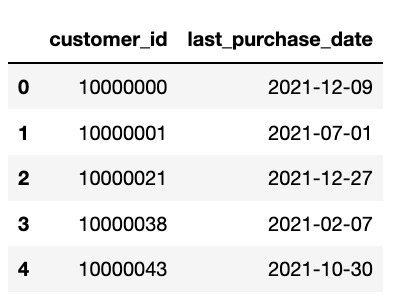
Description automatically generated

Abbildung 17: Recency

Abbildung 18: Finden Sie das letzte Einkaufsdatum

#### b. Frequency: Gesamtzahl der Käufe

Ich möchte die Gesamtzahl der Einkäufe ermitteln, die ein Kunde im Frequency-Konto getätigt hat, also muss ich berechnen, wie viele von jeder customer\_id vorhanden sind. Dadurch erhalte ich die Gesamtzahl der Bestellungen.

*frequency\_df = df\_copy.groupby(by=['customer\_id'], as\_index=False)[ 'frequecy'].count()*

Auf diese Weise haben wir für jeden Kunden die Variable „Häufigkeit“ definiert.

*Table

Description automatically generated*

Abbildung 19: Frequency

#### c. Monetary: Gesamtbetrag, den dieser Kunde ausgegeben hat

Ich musste den Gesamtbetrag finden, der von jedem Kunden im Geldkonto ausgegeben wurde. Dafür habe ich die gefundenen total\_price-Werte für jeden Kunden gesammelt.

*monetary\_df = df.groupby(by='customer\_id',as\_index=False).agg({'total\_price': 'sum'})*

Table

Description automatically generatedAuf diese Weise habe ich für jeden Kunden die Variable „Monetary“ definiert.Table

Description automatically generated

Abbildung 20: RFM-Werte

Abbildung 21: Monetary

### *4.2.2. RFM (1-4) Analyse*

Da ich 1-4 Scoring verwenden werde, habe ich jeden Wert in 4er-Gruppen unterteilt.

Table

Description automatically generated

Abbildung 22: Viertel

Dann habe ich die Recency, Frequency und Monetary für jeden Kunden als 1-4 standardisiert.

Table

Description automatically generated

Abbildung 23: Standardisierung RFM-werte

Nach der Standardisierung habe ich meinen RFM-Score mit diesen drei Werten erstellt, die zusammenkommen.

**Table

Description automatically generated**

Abbildung 24: RFM Score

Wir werden die Aktualitäts- und Häufigkeits-Scores für die Kundensegmentierung verwenden. Wir gehen davon aus, dass ein Kunde, der kürzlich und häufig eingekauft hat, einen hohen RFM-Score haben sollte.

Wir haben eine Segmentierungskarte basierend auf den Werten „Recency“ und „Frequency“ erstellt und unsere numerischen Bewertungen in Zeichenfolgenwerte in der Spalte „RFM\_segment“ konvertiert.

|  |  |  |
| --- | --- | --- |
| **Recency Score** | **Frequency Score** | **Segment** |
| 1 | 1 | Hibernating |
| 1 | 2 | At Risk |
| 1 | 3 | At Risk |
| 1 | 4 | Can't Loose |
| 2 | 1 | About to Sleep |
| 2 | 2 | Need Attention |
| 2 | 3 | Loyal Customers |
| 2 | 4 | Loyal Customers |
| 3 | 1 | Promising |
| 3 | 2 | Potential Loyalists |
| 3 | 3 | Loyal Customers |
| 3 | 4 | Loyal Customers |
| 4 | 1 | New Customers |
| 4 | 2 | Potential Loyalists |
| 4 | 3 | Champions |
| 4 | 4 | Champions |

Tabelle 5: Segmentierungskarte

Table

Description automatically generated

Abbildung 25: Segmentierung

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **recency** | **frequency** | **monetary** |  |
| **RFM\_Segment** | **mean** | **mean** | **mean** | **count** |
| About to Sleep | 189.9 | 1.0 | 475.9 | 262658 |
| At Risk | 297.4 | 2.8 | 762.9 | 277267 |
| Can't Loose | 288.3 | 9.7 | 1842.3 | 48266 |
| Champions | 22.5 | 14.9 | 5749.1 | 495295 |
| Hibernating | 306.1 | 1.0 | 388.5 | 377097 |
| Loyal Customers | 127.9 | 7.8 | 2367.7 | 679479 |
| Need Attention | 188.4 | 2.0 | 724.4 | 142831 |
| New Customers | 25.7 | 1.0 | 781.0 | 129309 |
| Potential Loyalists | 58.8 | 2.0 | 992.3 | 210138 |
| Promising | 82.3 | 1.0 | 544.3 | 216157 |

Tabelle 6: Segmentstatistiken

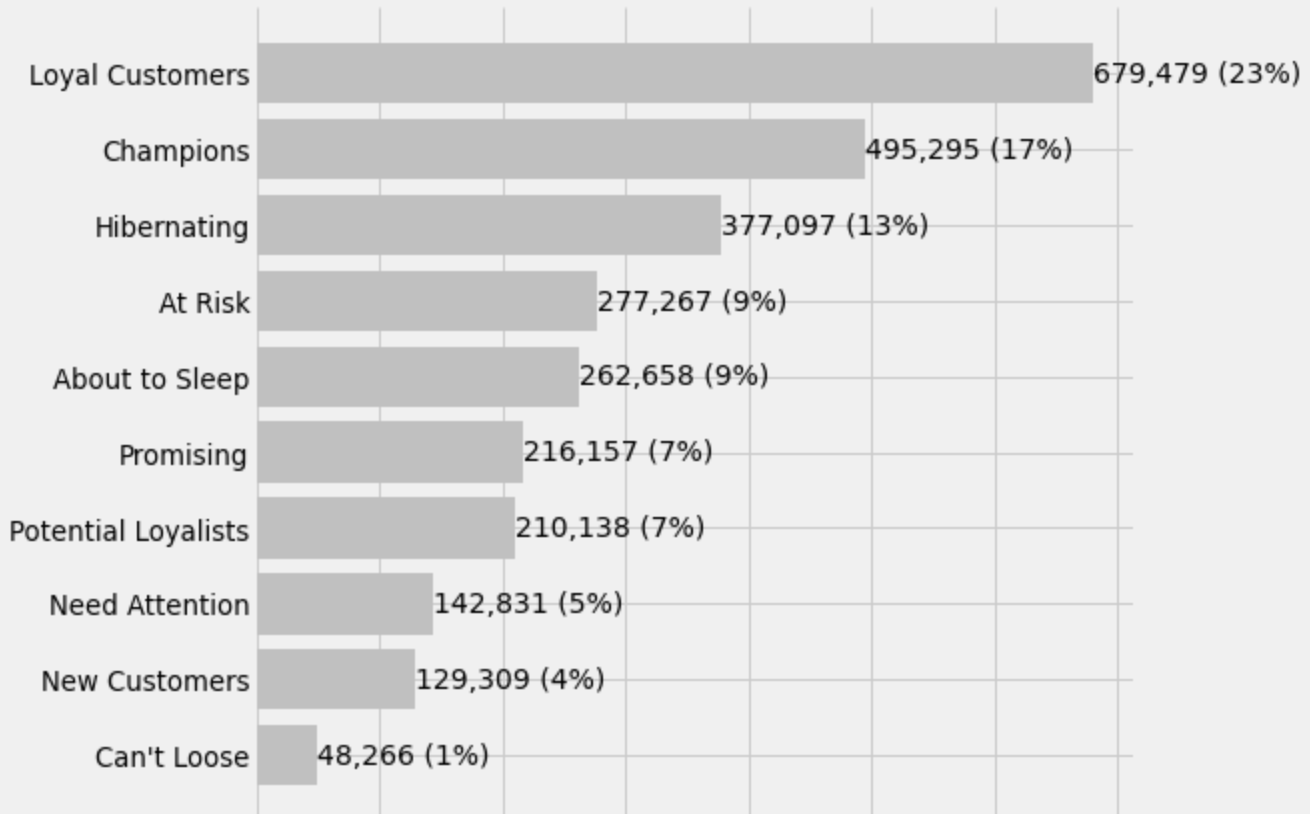
****

Abbildung 26: Segmentstatistiken

**Chart, treemap chart

Description automatically generated**

Abbildung 27: Segmentierung Tree Map

### *4.2.2. RFM (1-5) Analyse*

Dieses Mal habe ich meine Singular-Scores und RFM\_Score erstellt, indem ich die Recency-, Frequency- und Monetary-Werte auf die Scores 1-5 standardisiert habe.

Table

Description automatically generated

Abbildung 28: RFM-Werte (1-5)

**Table

Description automatically generated**

Abbildung 29: RFM Score (1-5)

Anschließend habe ich diese Integer-Werte gemäß der von mir vorbereiteten Segmentierungskarte in String-Werte umgewandelt.

|  |  |  |
| --- | --- | --- |
| **Recency Score** | **Frequency Score** | **Segment** |
| 1 | 1 | Hibernating |
| 1 | 2 | Hibernating |
| 1 | 3 | At Risk |
| 1 | 4 | At Risk |
| 1 | 5 | Can't Loose |
| 2 | 1 | Hibernating |
| 2 | 2 | Hibernating |
| 2 | 3 | At Risk |
| 2 | 4 | At Risk |
| 2 | 5 | Can't Loose |
| 3 | 1 | About to Sleep |
| 3 | 2 | About to Sleep |
| 3 | 3 | Need Attention |
| 3 | 4 | Loyal Customers |
| 3 | 5 | Loyal Customers |
| 4 | 1 | 'Promising |
| 4 | 2 | Potential Loyalists |
| 4 | 3 | Potential Loyalists |
| 4 | 4 | Loyal Customers |
| 4 | 5 | Loyal Customers |
| 5 | 1 | New Customers |
| 5 | 2 | Potential Loyalists |
| 5 | 3 | Potential Loyalists |
| 5 | 4 | Champions |
| 5 | 5 | Champions |

Tabelle 7: Segmentierungskarte (1-5)

**Table

Description automatically generated**

Abbildung 30: Segmentierung (1-5)

**A picture containing table

Description automatically generated**

Abbildung 31: Segmentstatistiken (1-5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **recency** | **frequency** | **monetary** |  |
| **RFM\_Segment** | **mean** | **mean** | **mean** | **count** |
| About to Sleep | 127.5 | 1.1 | 553.1 | 208287 |
| At Risk | 259.7 | 3.3 | 916.0 | 436327 |
| Can't Loose | 240.8 | 12.1 | 2665.9 | 78262 |
| Champions | 17.6 | 17.0 | 6713.0 | 370595 |
| Hibernating | 276.3 | 1.1 | 437.9 | 619815 |
| Loyal Customers | 90.1 | 9.6 | 3023.0 | 502096 |
| Need Attention | 126.4 | 2.4 | 932.7 | 120553 |
| New Customers | 19.8 | 1.0 | 810.7 | 43884 |
| Potential Loyalists | 46.6 | 1.8 | 931.2 | 400475 |
| Promising | 63.2 | 1.0 | 642.7 | 58203 |

Tabelle 8: Segmentstatistiken (1-5)

**Chart, treemap chart

Description automatically generated**

Abbildung 32: Segmentierung Tree Map (1-5)

# 5. FAZIT

## 5.1. KURZE BESCHREIBUNGEN DER SEGMENTE

***Champions:*** Kürzlich gekauft, oft gekauft und am meisten ausgegeben. Belohnen Sie diese Kunden. Sie können Early Adopters für neue Produkte werden und helfen, Ihre Marke zu promoten.

***Loyal Customers:*** Kaufen Sie regelmäßig ein. Reagiert auf Werbeaktionen.

***Potential Loyalist*:** Letzte Kunden mit durchschnittlicher Häufigkeit. Bieten Sie Mitgliedschafts- oder Treueprogramme an oder empfehlen Sie verwandte Produkte, um sie weiter zu verkaufen und ihnen zu helfen, Ihre Loyalisten oder Champions zu werden.

***New Customers:*** Kürzlich gekauft, aber nicht oft. Beginnen Sie mit dem Aufbau von Beziehungen zu diesen Kunden, indem Sie Onboarding-Support und Sonderangebote anbieten, um ihre Besuche zu steigern.

***Promising*:** Neue Käufer, aber nicht viel ausgegeben.

***Need Attention:*** Überdurchschnittliche Aktualität, Häufigkeit und monetäre Werte. Habe es aber vielleicht erst vor kurzem gekauft.

***About to Sleep:*** Unterdurchschnittliche Aktualität und Häufigkeit. Verlieren sie, wenn sie nicht reaktiviert werden.

***At Risk:*** Einige Zeit seit dem Kauf. Ich muss sie zurückbringen! Senden Sie ihnen personalisierte Reaktivierungskampagnen, um die Verbindung wiederherzustellen, und bieten Sie Verlängerungen und hilfreiche Produkte an, um einen weiteren Kauf zu fördern.

***Can’t Lose***: Früher oft gekauft, aber schon lange nicht mehr zurückgegeben. Holen Sie sie mit relevanten Werbeaktionen zurück und führen Sie Umfragen durch, um herauszufinden, was schiefgelaufen ist, und vermeiden Sie, sie an einen Konkurrenten zu verlieren.

***Hibernating:*** Letzter Einkauf liegt lange zurück und geringe Anzahl an Bestellungen. Kann verloren gehen.

## 5.2. VERGLEICH DER ERGEBNISSE UND EMPFEHLUNGEN

Unten sehen wir die Ergebnisse der Segmente als Ergebnis der beiden Anträge, die wir gestellt haben, indem wir sie in 4 Cluster und 5 Cluster unterteilt haben.

Segmente mit einem auffällig hohen Unterschied zwischen ihnen; At Risk, Hibernating, New Customers, Potential Loyalists und Promising.

Die Segmente At Risk und Hibernating beinhalten Kunden, die ich verloren habe oder fast verlieren werde. Während ihre Anzahl niedrig ist, wenn ich 4 Cluster verwende, sehen wir, dass sie ziemlich hoch sind, wenn wir die Ergebnisse eingrenzen und sie in 5 Cluster unterteilen. In diesem Zusammenhang wäre es logischer, die Kunden in 5 Clustern, statt in 4 Clustern zu untersuchen. Auf diese Weise können solidere Schritte unternommen werden, um den Kundenverlust zu minimieren.

|  |  |  |
| --- | --- | --- |
| **RFM\_Segment** | **Count (1-5 Scoring)** | **Count (1-4 Scoring)** |
| About to Sleep | 208287 | 262658 |
| At Risk | 436327 | 277267 |
| Can't Lose | 78262 | 48266 |
| Champions | 370595 | 495295 |
| Hibernating | 619815 | 377097 |
| Loyal Customers | 502096 | 679479 |
| Need Attention | 120553 | 142831 |
| New Customers | 43884 | 129309 |
| Potential Loyalists | 400475 | 210138 |
| Promising | 58203 | 216157 |

Tabelle 9: Vergleichen

Wir sollten versuchen, den Kontakt zu Kunden im Segment *About to Sleep* wieder herzustellen. Wir können sie zurückgewinnen, indem wir Sonderrabatte und beliebte Angebote anbieten und unsere wertvollen Ressourcen mit ihnen teilen.

Um Kunden im *“At Risk”* wieder zu verbinden, versuchen wir möglicherweise, sie zurückzubringen, indem wir ihnen personalisierte E-Mails oder andere personalisierte Nachrichten senden.

Das Kundensegment *Can't Lose* ist eines der Segmente, denen Bedeutung beigemessen werden sollte. Das sind die Kunden, die wir nicht verlieren wollen. Diese Gruppe kauft schon lange nicht mehr ein, aber die Gesamtzahl der Einkäufe ist ziemlich hoch. Wir können Umfragen durchführen und ihre Meinung einholen, um herauszufinden, was in diesem Segment schiefgelaufen ist. Wir sollten auch Kampagnen anbieten, mit denen sich diese Kunden wertgeschätzt fühlen.

Die Konzentration darauf, Kunden in den Segmenten *Champions, Loyal Customers, Potential Loyalists und Promising* zufrieden zu stellen, sollte Priorität haben. Es wird ein Plus sein, sie auf personalisiertere Weise zu erreichen, indem ihre individuellen Vorlieben und Affinitäten stärker analysiert werden. So können beispielsweise persönliche Nachrichten wie Geburtstagswünsche versendet werden. Neue Aktionen können für Produkte frühzeitig gemeldet werden.

Menschen im *Need Attention*-Segment sind die Kundengruppe, die wir tendenziell verlieren, indem wir Durchschnittswerte sowohl zeitlich als auch in Bezug auf die Gesamtzahl der Käufe anzeigen, während es gleichzeitig nicht viel Aufwand erfordert, um zu gewinnen. Wir können in diesem Segment zeitlich begrenzte Angebote machen, die sie dazu bringen, bei uns einzukaufen. Wir können Produkte basierend auf früheren Einkäufen empfehlen. Auf diese Weise können wir sie wieder aktivieren, damit sie einkaufen können.

Die Kundengruppe *New Customers* sind Gruppen, die kürzlich neue Kontakte hatten. Wir müssen die Gesamteinkäufe dieses Segments erhöhen. Wir sollten sie oft zu uns kommen lassen, auch für kleine Summen. Wir können spezielle Aktionen für Neukunden organisieren.

Gemäß dem 5-Clustering befindet sich der größte Teil unseres Kundenstamms in den folgenden Segmenten;

|  |  |
| --- | --- |
| **RFM\_Segment** | **Count (1-5)** |
| Hibernating | 619815 |
| Loyal Customers | 502096 |
| At Risk | 436327 |
| Potential Loyalists | 400475 |

Tabelle 10: wichtige Segmente

In diesem Zusammenhang sollten wir schnell handeln, um Kunden in den Segmenten *Hibernating* und *At Risk* zurückzugewinnen und Kampagnen für diese Gruppen zu planen.

Um unsere Kunden in den Segmenten „*Loyal Customers*“ und „*Potent Loyalists*“ in das Segment „*Champions*“ zu bringen, müssen wir personalisierte Kampagnen für sie organisieren und die Markenloyalität erhöhen.

# LITERATURVERZEICHNIS

1. Ekin, N. (1998). Bilgi Ekonomisinde Elektronik Ticaret, İstanbul Ticaret Odası Yayınları [↑](#endnote-ref-1)
2. Parvatiyar, A., & Sheth, J. N. (2001). Customer Relationship Management: Emerging Practice, Process, and Discipline. Journal of Economic and Social Research 3(2), 1-34. [↑](#endnote-ref-2)
3. Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2), 2592-2602 [↑](#endnote-ref-3)
4. Fullerton, R.A. (1988). How Modern is Modern Marketing? Marketing's Evolution and the Myth of the “Production Era” [↑](#endnote-ref-4)
5. Christensen, C. M., Cook, S., & Hall, T. (2005). Marketing Malpractice. Harvard Business Review [↑](#endnote-ref-5)
6. De Sáez, Eileen Elliott. 2002. Marketing concepts for libraries and information services. London: Facet Publishing. [↑](#endnote-ref-6)
7. Wedel, Michel, and Wagner A. Kamakura. 2012. Market segmentation: Conceptual and methodological foundations. Vol. 8. New York: Springer Science & Business Media. [↑](#endnote-ref-7)
8. H. Ziafat and M. Shakeri, "Using Data Mining Techniques in Customer Segmentation", Int. Journal of Engineering Research and Applications, vol. 4, no. 9, pp. 70-79, 2014 [↑](#endnote-ref-8)
9. MacQueen J (1967) Some Methods for Classification and Analysis. Proc. Symp. Math.  
   Statist. and Probability (5th), 281-297 [↑](#endnote-ref-9)
10. Han J, Kamber M, Pei, J (2012) Data Mining Concepts and Techniques , Third Edition. Morgan Kaufmann. [↑](#endnote-ref-10)
11. L. Liu, H. Mosavat-Jahromi, L. Cai and D. Kidston, "Hierarchical Agglomerative Clustering and LSTM-based Load Prediction for Dynamic Spectrum Allocation," 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), 2021, pp. 1-6, doi: 10.1109/CCNC49032.2021.9369631. [↑](#endnote-ref-11)
12. C. Ye and X. Zhao, "Automated Operational Modal Analysis Based on DBSCAN Clustering," 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2020, pp. 864-869, doi: 10.1109/ICITBS49701.2020.00190. [↑](#endnote-ref-12)
13. Blattberg, R.C.; Kim, B-D. & Neslin, S.A. (2008). Database Marketing: Analyzing and Managing Customers, Chapter 12, pp. 323-337, Springer, ISBN: 978-0387725789, New York, USA. [↑](#endnote-ref-13)
14. Schijns, J. M., & Schröder, G. (1996). Segment selection by relationship strength. Journal of Direct Marketing, 10(3), 69-79 [↑](#endnote-ref-14)
15. Kaymak, U. (2001). Fuzzy target selection using RFM variables. 9th IFSA World Congress and 20th NAFIPS International Conference,2, 1038-1043. [↑](#endnote-ref-15)
16. Mccarty, J., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. Journal of Business Research,60(6), 656-662 [↑](#endnote-ref-16)
17. Chan, C. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. Expert Systems with Applications, 34(4), 2754-2762 [↑](#endnote-ref-17)
18. Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. Journal of Database Marketing & Customer Strategy Management, 19(3), 197-208 [↑](#endnote-ref-18)