

Von Daten zu Personas: Eine strategische Kundensegmentierung für Travel Tide

Eine datengestützte Kundensegmentierung für Travel Tide



Mastery Projekt | Masterschool

[Interaktive Analyse und Implementierung \(Google Colab\)](#)

von Sinem Ara-Yücel | September 2025

Management Summary

Ziel dieses Projekts war es, eine datengestützte Grundlage für personalisiertes Marketing bei Travel Tide zu schaffen, indem der Kundenstamm in handhabbare Gruppen segmentiert wird. Mittels unüberwachten maschinellen Lernen wurden erfolgreich fünf distinkte Kunden-Personas identifiziert, die von selten buchenden „Effizienten Städtereisenden“ bis hin zu hochprofitablen „Premium-Familienplanern“ reichen.

Jede Persona weist einzigartige Verhaltensweisen und Bedürfnisse auf, was einen maßgeschneiderten Marketingansatz für das neue Prämienprogramm ermöglicht. Die zentrale Handlungsempfehlung ist, diese Erkenntnisse für gezielte E-Mail-Kampagnen zu nutzen, indem für jedes Segment die jeweils passendsten Vorteile beworben werden. Diese personalisierte Strategie zielt darauf ab, das Kundenengagement, die Loyalität und letztlich den Lifetime Value unserer Kunden signifikant zu steigern.

Business Case & Projektziele

Business Case

Der Online-Reisemarkt ist hart umkämpft. Die Gewinnung von Neukunden ist kostspielig, weshalb die Steigerung der Kundenbindung und die Maximierung des Lifetime Values bestehender Kunden von strategischer Bedeutung für Travel Tide sind.

Um ein effektives, personalisiertes Prämienprogramm zu entwickeln, das die Kundenloyalität nachhaltig erhöht, müssen wir zunächst ein tiefes, datengestütztes Verständnis für unsere Kunden entwickeln. Die zentrale geschäftliche Herausforderung besteht darin, die Frage zu beantworten: "Welche unterschiedlichen Kundentypen nutzen unsere Plattform und was charakterisiert sie?" Dieses Projekt wird die notwendigen Erkenntnisse liefern, um diese Frage zu beantworten und damit die Grundlage für gezielte, personalisierte Marketingmaßnahmen zu schaffen.

Projektziel: Entwicklung einer datengestützten Kundensegmentierung als strategische Grundlage für personalisiertes Marketing

Um unsere Kunden effektiver und persönlicher ansprechen zu können, segmentieren wir in diesem Projekt unseren gesamten Kundenstamm in klar definierte Gruppen. Ziel ist es, von einer undifferenzierten Massenansprache wegzukommen und stattdessen maßgeschneiderte Angebote zu schaffen, die exakt auf die Bedürfnisse der jeweiligen Kundengruppe zugeschnitten sind.

Als konkretes Ergebnis wird ein praktischer Leitfaden entwickelt, der 3 bis 5 zentrale Kunden-Personas vorstellt. Für jede dieser Personas werden die folgenden drei Kernfragen beantwortet:

- Wer ist der Kunde? (Ein prägnanter Name und eine griffige Kurzbeschreibung, z.B. "Der effiziente Städtereisende").
- Wie verhält er sich? (Die wichtigsten Merkmale zu Demografie, typischem Buchungsverhalten, Budget und bevorzugten Reisearten).
- Wie relevant ist die Gruppe? (Ihr prozentualer Anteil an unserer gesamten Kundenbasis zur Einschätzung des Marktpotenzials).

Diese Ausarbeitung liefert dem Marketing-Team eine sofort einsetzbare Grundlage, um die Kommunikation, insbesondere für das neue Prämienprogramm, zielgerichtet zu gestalten und die Konversionsraten signifikant zu erhöhen.

Datengrundlage: Schema, Struktur und initiale Qualitätsanalyse

Datenquelle und Aufbereitung

Die Analyse basiert auf internen Produktionsdaten von Travel Tide aus den letzten Jahren. Für die Segmentierung wurde dieser Datensatz gezielt auf eine relevante Kundengruppe gefiltert.

Berücksichtigt wurden ausschließlich aktive Nutzer mit **mehr als 7 Seitenklicks**, deren Interaktionen **nach dem 4. Januar 2023** stattfanden, um eine Analyse des aktuellen Kundenverhaltens zu gewährleisten. Die Datengrundlage bilden die Tabellen: users, sessions, flights und hotels.

Zeilenanzahl je Datensatz in den Rohrtabellen:

- User: 1.020.926
- hotels: 1.918.617
- flights: 1.901.038
- sessions: 5.408.063

Überprüfung der Nullwerte, um diese zu berücksichtigen.

- Users: keine NULL-Werte
- hotels: keine NULL-Werte
- flights:
 - **return_time = 88.734**
- sessions:
 - **tripd_id = 3.072.218**
 - **flight_discount_amount = 4.522.267**
 - **hotel_discount_amount = 4.716.683**

Datensatz

Nachfolgend eine Beschreibung der einzelnen Tabellen und ihrer Spalten:

Benutzer : demografische Informationen zum Benutzer

- user_id : eindeutige Benutzer-ID (Schlüssel, int)
- Geburtsdatum : Geburtsdatum des Benutzers (Datum/Uhrzeit)
- Geschlecht : Geschlecht des Benutzers (nominal)
- verheiratet : Familienstand des Benutzers (binär)
- has_children : ob der Benutzer Kinder hat oder nicht (binär)
- home_country : Wohnsitzland des Benutzers (nominal)
- home_city : Wohnort des Benutzers (nominal)
- home_airport : bevorzugter Heimatflughafen des Benutzers (nominal)
- home_airport_lat : geografische Nord-Süd-Position des Heimatflughafens (dezimal)
- home_airport_lon : geografische Ost-West-Position des Heimatflughafens (dezimal)
- sign_up_date : Datum der TravelTide-Kontoerstellung (Datum/Uhrzeit)

Sitzungen : Informationen zu einzelnen Browsersitzungen (Hinweis: Es werden nur Sitzungen mit mindestens 2 Klicks berücksichtigt)

- session_id : eindeutige Browser-Sitzungs-ID (Schlüssel, Zeichenfolge)
- user_id : die Benutzer-ID (Fremdschlüssel, int)
- trip_id : ID, die Flug- und Hotelbuchungen zugeordnet ist (Fremdschlüssel, Zeichenfolge)
- session_start : Zeitpunkt des Beginns der Browsersitzung (Zeitstempel)
- session_end : Zeitpunkt des Endes der Browsersitzung (Zeitstempel)
- flight_discount : ob ein Flugrabatt angeboten wurde oder nicht (binär)

- hotel_discount : ob ein Hotelrabatt angeboten wurde oder nicht (binär)
- flight_discount_amount : Prozentsatz vom Basistarif (Dezimalzahl)
- hotel_discount_amount : Prozentsatz vom Basisnachtpreis (Dezimal)
- flight_booked : ob der Flug gebucht wurde oder nicht (binär)
- hotel_booked : ob das Hotel gebucht wurde oder nicht (binär)
- page_clicks : Anzahl der Seitenklicks während der Browsersitzung (int)
- Stornierung : ob der Zweck der Sitzung darin bestand, eine Reise zu stornieren (binär)

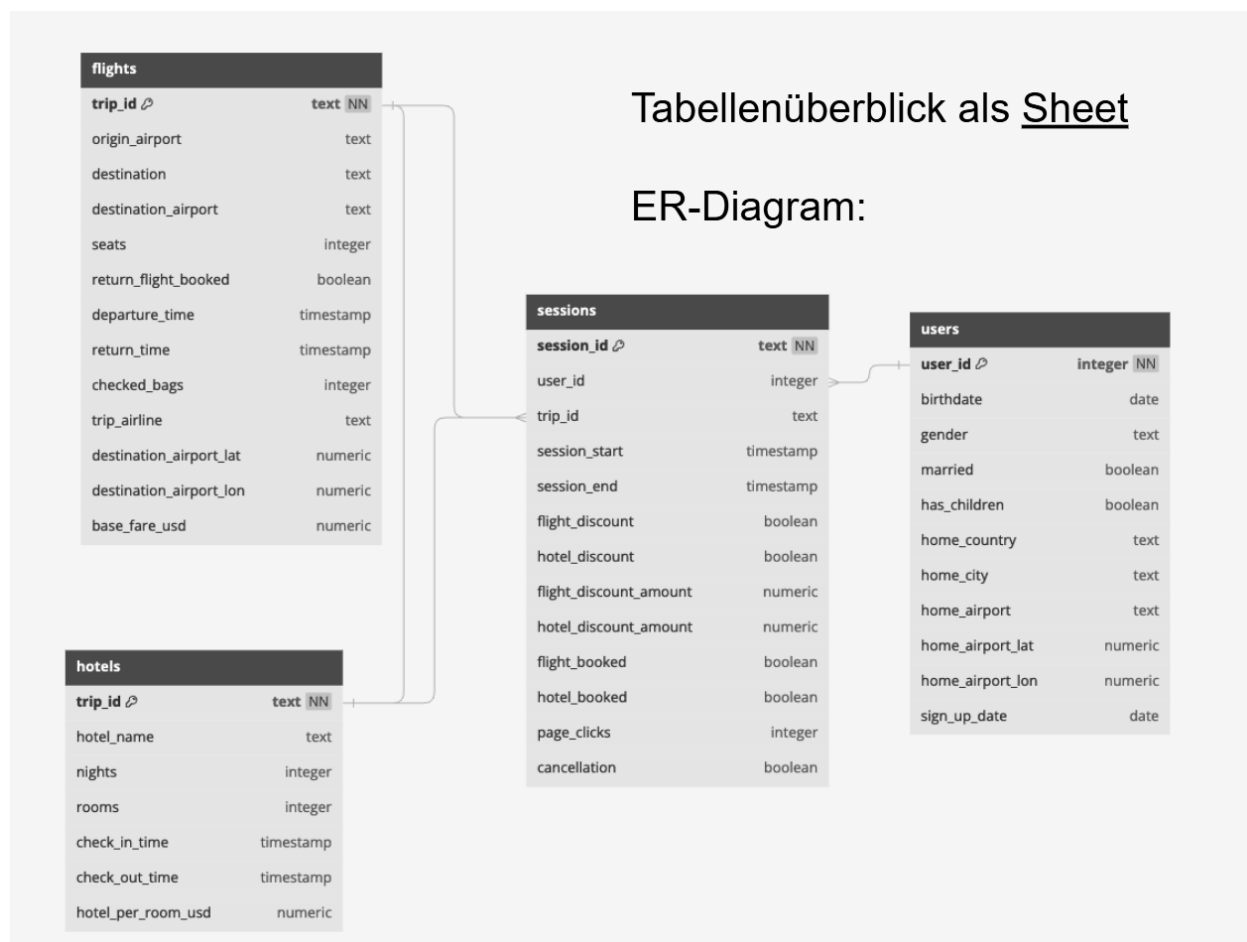
Flüge : Informationen zu gekauften Flügen

- trip_id : eindeutige Reise-ID (Schlüssel, Zeichenfolge)
- origin_airport : Heimatflughafen des Benutzers (nominal)
- Zielort : Zielort (nominal)
- destination_airport : Flughafen in der Zielstadt (nominal)
- Sitzplätze : Anzahl der gebuchten Sitzplätze (int)
- return_flight_booked : ob ein Rückflug gebucht wurde oder nicht (binär)
- departure_time : Abflugzeit vom Abflughafen (Zeitstempel)
- return_time : Zeitpunkt der Rückkehr zum Ausgangsflughafen (Zeitstempel)
- checked_bags : Anzahl der aufgegebenen Gepäckstücke (int)
- trip_airline : Fluggesellschaft, die den Benutzer vom Abflug- zum Zielort bringt (nominell)
- destination_airport_lat : geografische Nord-Süd-Position des Zielflughafens (dezimal)
- destination_airport_lon : geografische Ost-West-Position des Zielflughafens (dezimal)
- base_fare_usd : Preis des Flugpreises vor Rabatt (Dezimalzahl)

Hotels : Informationen zu gekauften Hotelaufenthalten

- trip_id : eindeutige Reise-ID (Schlüssel, Zeichenfolge)
- hotel_name : Hotelmarkenname (nominal)
- Nächte: Anzahl der im Hotel verbrachten Nächte (int)
- Zimmer : Anzahl der beim Hotel gebuchten Zimmer (int)
- check_in_time : Beginn des Hotelaufenthalts des Benutzers (Zeitstempel)
- check_out_time : Uhrzeit, zu der der Hotelaufenthalt des Benutzers endet (Zeitstempel)
- hotel_per_room_usd : Preis für den Hotelaufenthalt pro Zimmer und Nacht vor Rabatt (Dezimal)

Datenbankschema



Vorgehensweise und Methodik

Angewandte Methoden

Um verborgene Muster und natürliche Gruppierungen im Nutzerverhalten ohne vordefinierte Annahmen zu entdecken, wird als primäre Methode das unüberwachte maschinelle Lernen eingesetzt.

- **Explorative Datenanalyse (EDA):** Das Projekt begann mit einer umfassenden EDA, um Datenverteilungen zu visualisieren, Korrelationen zu identifizieren und ein grundlegendes Verständnis der Rohdatensätze zu gewinnen.
- **Feature Engineering & Skalierung:** Bestehende Daten werden zu aussagekräftigen Merkmalen kombiniert. Da Clustering-Algorithmen auf Distanzmessungen basieren, werden alle numerischen Features auf eine einheitliche Skala gebracht (Standardisierung), um Verzerrungen zu vermeiden.
- **Hashing Encoding:** Textbasierte kategoriale Merkmale mit vielen einzigartigen Werten (z. B. Heimatstadt) wurden in eine feste Anzahl numerischer Merkmale umgewandelt. Dieser Prozess macht wertvolle, nicht-numerische Informationen für den Clustering-Algorithmus nutzbar, ohne eine unüberschaubare Anzahl neuer Spalten zu erzeugen.
- **Hauptkomponentenanalyse (PCA):** Um den hochdimensionalen Datensatz für das Clustering zu optimieren, wurde die PCA verwendet, um die Anzahl der Merkmale zu reduzieren und gleichzeitig ein Maximum an Informationen (Varianz) zu erhalten. Dieser Schritt verdichtet komplexe Daten zu ihren wesentlichen Komponenten, was zu einem stabileren und effizienten Clustering führt.

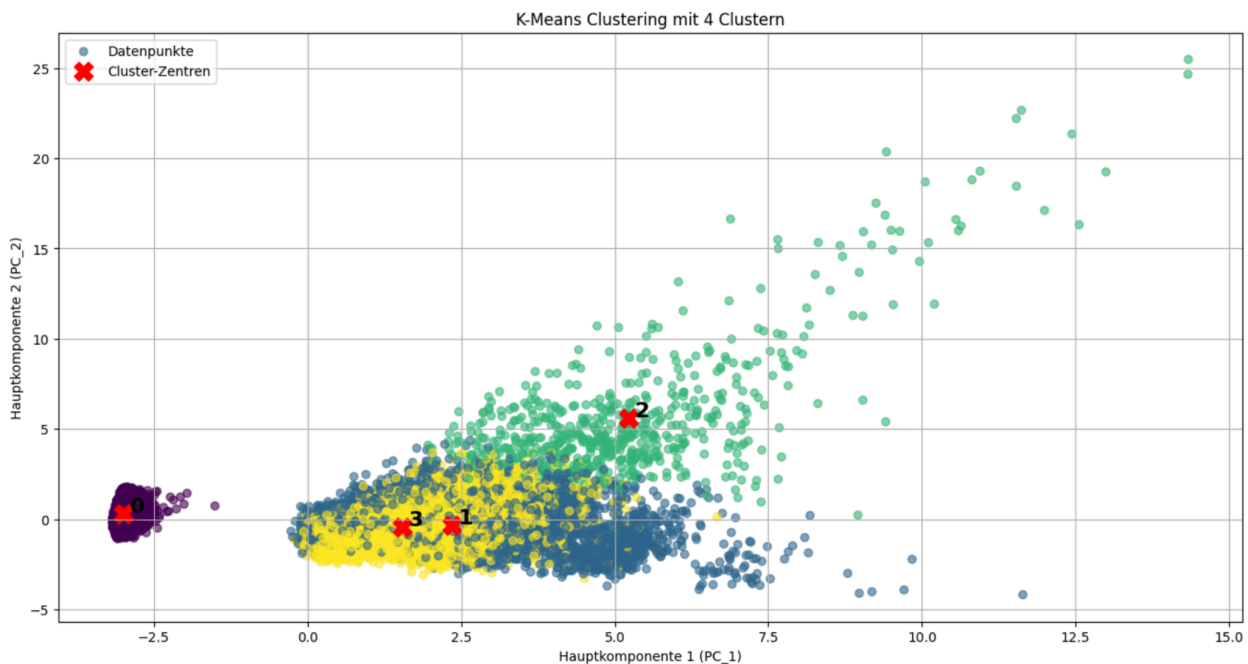
- K-Means-Clustering: Im Kern wird der K-Means-Algorithmus verwendet, um die Nutzerdaten in eine bestimmte Anzahl von Clustern (Segmenten) zu unterteilen. Die Mitglieder eines Clusters ähneln sich in ihrem Verhalten und ihren Merkmalen stark, während sie sich von den Mitgliedern anderer Cluster deutlich unterscheiden

Vorgehensweise (Projektschritte)

Das Projekt folgt einem strukturierten Prozess:

1. **Datenbeschaffung & -aufbereitung:** Die relevanten Daten aus den users, sessions, flights und hotels Tabellen werden mittels der entwickelten SQL-Abfrage extrahiert und zu einem einzigen Analyse-Datensatz zusammengeführt.
2. **Explorative Datenanalyse (EDA):** Untersuchung der Daten auf Muster, Anomalien und Zusammenhänge, um Hypothesen für die Segmentierung zu bilden.
3. **Feature Engineering & Skalierung:** Aufbereitung der Daten für das Modelltraining, inklusive der Normalisierung der numerischen Features.
4. **Modelltraining:** Anwendung des K-Means-Algorithmus. Mittels der "Ellenbogenmethode" wird die optimale Anzahl von Clustern für die Segmentierung ermittelt.
5. **Segmentanalyse & Interpretation:** Detaillierte Analyse der resultierenden Cluster. Jedes Segment wird auf seine charakteristischen Merkmale (z.B. Preisbewusstsein, Reisedauer) untersucht, um greifbare "Personas" zu definieren.
6. **Ergebnispräsentation & Handlungsempfehlungen:** Die finalen Kundensegmente werden visualisiert und beschrieben. Darauf aufbauend werden konkrete, datengestützte Empfehlungen für das Design des personalisierten Prämienprogramms formuliert.

Analyseergebnisse: Die 4 Travel Tide-Personas



Cluster 0 = Die Schaufensterbummler

Diese Nutzer bilden mit fast 40% die mit Abstand größte Gruppe in unserer Datenbank. Sie sind auf der Plattform aktiv, klicken sich durch Angebote und zeigen damit klares Interesse an Reisen. Der entscheidende Punkt ist jedoch, dass sie noch nie eine Buchung getätigt haben und ihre Konversionsrate bei null liegt. Hier schlummert das größte, ungenutzte Wachstumspotenzial für TravelTide.

Top 3 Beweise aus den Daten:

- **Größtes Segment:** Bildet mit ca. 39,8% den größten Anteil aller Nutzer.
- **Keine Buchungen:** Der `num_trips`-Wert für diese Gruppe ist exakt 0.
- **Interesse ohne Kauf:** Sie weisen eine moderate Aktivität (`page_clicks`) auf, aber die `conversion_rate` ist 0.

Marketing-Tipp: Ziel ist es, diese riesige, zögernde Gruppe zu aktivieren und die Hürde für die allererste Buchung so weit wie möglich zu senken.

- **Top 👍: Exklusive Rabatte** (als klarer Anreiz) und **Keine Stornogebühren**, um das Risiko der ersten Buchung zu minimieren.
-

Cluster 1 = Die aufstrebenden Stammkunden

Mit über 20% der Nutzer stellen diese Kunden unsere Zukunft dar. Sie haben bereits mehrfach gebucht und zeigen damit erste Anzeichen von Loyalität. Ein entscheidendes Merkmal ist ihr hohes Engagement: Sie verbringen deutlich mehr Zeit auf der Plattform (höhere **page_clicks**) als Einmal-Bucher. Das zeigt, dass sie TravelTide aktiv für ihre Reiseplanung nutzen und ein hohes Potenzial haben, zu treuen Stammkunden zu werden.

Top 3 Beweise aus den Daten:

- **Mehrfach-Bucher:** Der **num_trips**-Wert liegt im Schnitt bei über 1,5.
- **Hohes Engagement:** Die **page_clicks** sind signifikant höher als bei den Städtereisenden.
- **Wachstumspotenzial:** Bilden mit ca. 21,5% eine kritische Masse für zukünftiges, loyales Wachstum.

Marketing-Tipp: Belohne ihre bisherige Treue und gib ihnen einen starken Anreiz für die nächste Buchung, um sie fest an die Marke zu binden.

- **Top 👍: 1 Hotelübernachtung gratis** ist ein starker Anreiz. Auch **Exklusive Rabatte** für Mitglieder wirken als Belohnung.

Cluster 2 = Die Premium-Familienplaner

Dies ist mit ca. 2,3% unsere kleinste, aber mit Abstand wertvollste und profitabelste Gruppe. Sie buchen zwar nicht oft, aber wenn, dann geben sie sehr viel Geld für lange Reisen mit der ganzen Familie aus. Der Name "Familienplaner" wird durch die Daten bestätigt, da fast die Hälfte dieser Gruppe mit Kindern reist. Das sind die VIPs, die wir mit hochwertigen Angeboten pflegen und unbedingt halten müssen.

Top 3 Beweise aus den Daten:

- **Extrem hohe Ausgaben:** Der `total_trip_cost`-Wert ist mit über 5.500 € mit Abstand am höchsten.
- **Lange Reisen & viel Gepäck:** Die `nights` (ca. 7) und `checked_bags` (ca. 1.6) sind die höchsten aller Cluster.
- **Oft mit Familie:** Der `has_children`-Wert ist mit ca. 50% am höchsten.

Marketing-Tipp: Behandle diese VIPs mit hochwertigen Angeboten, die ihren Bedürfnissen nach Komfort und einem Rundum-sorglos-Paket entsprechen.

- **Top 👍:** Alle Vorteile sind relevant! Besonders wirksam sind **Kostenfreies Aufgabegepäck** und **1 Hotelübernachtung gratis**, da sie bei langen, teuren Reisen einen hohen monetären Wert darstellen.

Cluster 3 = Die effizienten Städtereisenden

Diese riesige Gruppe mit über 36% bildet das Herzstück und die solide Basis unserer aktiven Kunden. Es handelt sich typischerweise um Einmal-Bucher, die gezielt und unkompliziert kurze Reisen von 2-3 Nächten planen. Ihr Reiseverhalten ist effizient und auf den Punkt gebracht, was sich auch darin zeigt, dass sie fast ausschließlich mit leichtem Gepäck reisen.

Top 3 Beweise aus den Daten:

- **Solide Kundenbasis:** Bildet mit ca. 36,4% den Kern der aktiven Bucher.
- **Kurzreise-Fokus:** Die durchschnittliche Reisedauer (**nights**) liegt bei ca. 2-3 Nächten.
- **Leichtes Gepäck:** Der **checked_bags**-Wert ist extrem niedrig, was auf typische Handgepäckstreisen hindeutet.

Marketing-Tipp: Die Strategie für diese Gruppe sollte auf Flexibilität und Einfachheit abzielen, um den unkomplizierten Charakter ihrer Reisen zu unterstützen.

- **Top 👍: Keine Stornogebühren** ist der perfekte Vorteil, um die Spontaneität und Flexibilität dieser Gruppe zu fördern.

Zusammenfassende Handlungsempfehlungen

Die datengestützte Analyse unseres Kundenstamms hat vier klar voneinander abgrenzbare Kundensegmente aufgedeckt. Jedes dieser Segmente repräsentiert eine einzigartige Persona mit spezifischen Verhaltensmustern und Bedürfnissen, die eine gezielte Marketing Ansprache erfordert.

Im Folgenden werden die Ergebnisse der Clusteranalyse detailliert vorgestellt. Für jede der vier identifizierten Personas wird eine Beschreibung ihres typischen Verhaltens geliefert und daraus die jeweils wirksamste Handlungsempfehlung abgeleitet, um sie erfolgreich für das Travel Tide-Prämienprogramm zu gewinnen.

Fazit und Ausblick

Dieses Projekt hat erfolgreich Rohdaten in ein klares, umsetzbares Segmentierungsmodell überführt, das die strategische Grundlage für personalisiertes Marketing bei Travel Tide bildet. Der nächste logische Schritt ist die Implementierung gezielter A/B-Test-Kampagnen auf Basis der abgeleiteten Persona-Empfehlungen, um die direkten Auswirkungen auf Konversionsraten und Kundenloyalität zu messen und eine kontinuierliche Optimierung unserer Marketingmaßnahmen zu ermöglichen.