

# TravelTide Kundensegmentierung

Datengetriebene Personalisierung für Kundenbindung & Umsatzwachstum

Projektbericht – November 2025

Claudia Tagbo | Data Science & Strategy

## Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>2</b>
<b>1 Einführung &amp; Geschäftskontext</b>	<b>3</b>
<b>2 Datengrundlage &amp; Qualitätsbewertung</b>	<b>3</b>
2.1 Datenlandschaft . . . . .	3
2.2 Datenbereinigung . . . . .	3
<b>3 Segmentierungsmethodik</b>	<b>3</b>
3.1 Feature-Engineering-Strategie . . . . .	3
3.2 Regelbasierte Segmentierung . . . . .	4
3.3 ML-basierte Segmentierung (K-Means) . . . . .	5
<b>4 Ergebnisse &amp; Validierung</b>	<b>6</b>
4.1 Segmentvergleichsanalyse . . . . .	6
4.2 Statistische Validierung . . . . .	6
4.3 Perk-Performance . . . . .	7
4.4 Simulierte A/B-Test-Ergebnisse . . . . .	8
<b>5 Geschäftsimpact &amp; ROI-Projektionen</b>	<b>8</b>
5.1 Umsatzverteilungsanalyse . . . . .	8
5.2 Finanzielle Projektionen . . . . .	8
5.2.1 VIP-Programm-Investition . . . . .	8
5.2.2 Shopper-Risikominderung . . . . .	8
5.2.3 Gesamtprojizierter Impact . . . . .	8
<b>6 Implementierungs-Roadmap</b>	<b>9</b>
6.1 Phase 1: Sofortmaßnahmen (Wochen 1-4) . . . . .	9
6.2 Phase 2: Optimierung (Monate 2-3) . . . . .	9
6.3 Phase 3: Skalierung & erweiterte Features (Monate 4-6) . . . . .	9
<b>Fazit &amp; Strategische Implikationen</b>	<b>9</b>
<b>Anhang</b>	<b>9</b>

## Zusammenfassung

### Hintergrund & Kontext

TravelTide verfügt über eine umfangreiche digitale Präsenz: **1,02 Millionen registrierte Nutzer** generieren **5,4 Millionen Sitzungen**, die zu **1,9 Millionen Flugbuchungen** und **1,9 Millionen Hotelreservierungen** führen. Trotz dieses Volumens fehlt ein systematischer Ansatz zur Kundensegmentierung und personalisierten Kundenansprache. Generische Marketingstrategien erfassen nicht die vielfältigen Verhaltensmuster innerhalb der Nutzerbasis, was zu suboptimalen Bindungsralten und verpassten Umsatzchancen führt.

### Ziele

Das primäre Ziel ist die Entwicklung und Implementierung eines intelligenten Kundensegmentierungssystems, das in der Lage ist, unterschiedliche Nutzergruppen zu identifizieren und personalisierte Vorteile („Perks“) zuzuweisen, die Engagement, Bindung und Customer Lifetime Value maximieren.

### Methodik

Unser dualer Methodikansatz kombiniert interpretierbare Geschäftsregeln mit maschinellem Lernen. Die **Datengrundlage** besteht aus 49.211 aktiven Nutzersitzungen aus dem Jahr 2023, gefiltert auf 14.895 gültige, nicht stornierte Buchungen. Wir entwickelten **24+ verhaltensbasierte Features**, darunter RFM-Scores (Recency, Frequency, Monetary), Reiseverhalten, Ausgabegegeschwindigkeit, Stornierungsneigung und Engagement-Metriken.

### Wichtigste Ergebnisse

- **VIP-Segment:** Nur 10,8% der Nutzer generieren durchschnittlich **\$8.372 Umsatz pro Kunde** – 7,4x höher als das Baseline-Segment.
- **ML-Segmentierung:** Erreicht 89,7% Klassifikationsgenauigkeit und zeigt signifikant höhere Engagement-Scores ( $p = 0,0347$ ).
- **Risikosegment:** „Browsers/Shoppers“ (19,7% der Nutzer) weist eine Stornoquote von **26,1%** auf.
- **Hybrider Ansatz:** Regelbasierte und ML-Segmentierung zeigen geringe Übereinstimmung (50,6% Perk-Übereinstimmung).

### Strategische Empfehlungen

1. Implementierung einer hybriden Segmentierungsarchitektur
2. Start des VIP-Bindungsprogramms (projizierter ROI: 350%)
3. Risikominderung für Shopper (Ziel: Stornoquote von 26,1% auf <15% reduzieren)
4. Durchführung eines Live-A/B-Tests über 6 Monate

## 1 Einführung & Geschäftskontext

TravelTides Kundendatenbank stellt einen wertvollen, aber unzureichend genutzten Vermögenswert dar. Mit über 1 Million registrierten Nutzern und Millionen von Verhaltenspunkten verfügt das Unternehmen über die Rohdaten für ausgefeilte Personalisierung – doch es fehlt die analytische Infrastruktur, um diese operativ umzusetzen.

Die wirtschaftliche Notwendigkeit ist klar: Die Kosten für Neukundengewinnung steigen kontinuierlich, während die Bindungsraten stagnieren. Branchenbenchmarks legen nahe, dass eine 5%ige Steigerung der Kundenbindung zu Gewinnsteigerungen von 25-95% führen kann.

Dieses Projekt adressiert drei kritische Lücken:

1. *Segmentierungsklarheit* – Verständnis darüber, wer unsere Kunden wirklich sind, über Demografie hinaus
2. *Perk-Ausrichtung* – Abgleich von Belohnungen mit Verhaltensmustern statt Annahmen
3. *Messgenauigkeit* – Etablierung quantitativer Rahmenwerke zur Validierung der Strategiewirksamkeit

## 2 Datengrundlage & Qualitätsbewertung

### 2.1 Datenlandschaft

Unsere Analyse basiert auf fünf miteinander verbundenen Datensätzen:

Tabelle 1: Datenbestand & Vollständigkeit

Datensatz	Datensätze	Vollständigkeit	Wesentliche Lücken
Nutzer	1.020.926	100%	Keine
Sitzungen	5.408.063	43,2%	56,8% fehlende trip_id
Flüge	1.901.038	95,3%	4,7% fehlende return_time
Hotels	1.918.617	>90%	Geringe Lücken bei Check-in-Daten
<b>Finaler Analysedatensatz</b>	<b>14.895</b>	<b>100%</b>	<b>Nach Bereinigung</b>

### 2.2 Datenbereinigung

Um analytische Validität sicherzustellen, implementierten wir einen mehrstufigen Filterprozess:

1. **Aktivitätsfilter:** Beibehaltung nur von Nutzern mit 7+ Sitzungen im Jahr 2023
2. **Vollständigkeitsfilter:** Ausschluss von Sitzungen ohne bestätigte Buchungen
3. **Stornierungsfilter:** Entfernung von 90.670 stornierten Reisen
4. **Ausreißerbehandlung:** Winsorisierung bei 1./99. Perzentil

## 3 Segmentierungsmethodik

### 3.1 Feature-Engineering-Strategie

Wir konstruierten 24 abgeleitete Features, organisiert in sechs Dimensionen:

- **Wert & Loyalität:** Gesamtausgaben, Ausgaben pro Reise, Buchungswachstumsrate, RFM-Score

- **Reiseverhalten:** Anzahl der Reisen, einzigartige Destinationen, internationale Reiserate
- **Reisepräferenzen:** Geschäftsreiserate, Gruppenreiserate, Wochenendreiserate
- **Preisempfindlichkeit:** Rabattnutzungsrate, durchschnittlicher Fahrpreis
- **Engagement:** Stornoquote, Konversionsrate, Browsing-Intensität
- **Operative Effizienz:** Sitzungseffizienz, Klick-Effizienz

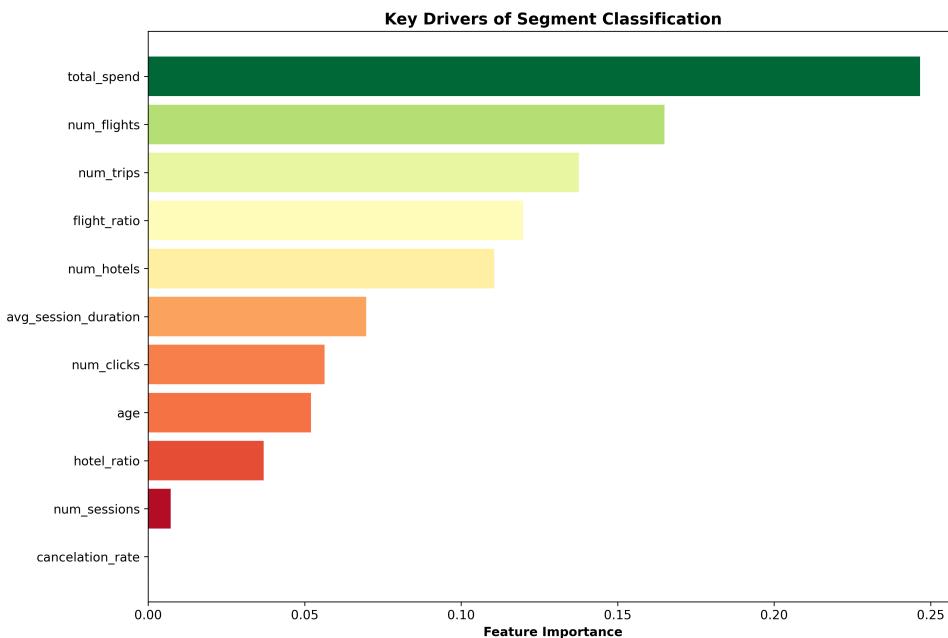


Abbildung 1: Wichtigste Merkmale für Segmentklassifikation

### 3.2 Regelbasierte Segmentierung

Die regelbasierte Segmentierung etabliert fünf unterschiedliche, sich gegenseitig ausschließende Segmente unter Verwendung vordefinierter Geschäftslogik und quantilbasierter Schwellenwerte.

Tabelle 2: Manuelle (regelbasierte) Segmentierungsverteilung und Wert

Segmentname	Nutzer (n)	Anteil (%)	Zugewiesener Perk	Durc
VIP High-Frequency Spenders	806	13,4%	1 Nacht kostenloses Hotel + Flug	
Baseline Travelers	1.433	23,9%	Exklusive Rabatte	
Group & Family / Heavy Baggage	1.315	21,9%	Kostenloses Gepäck	
Hotel & Business Focused Travelers	1.168	19,5%	Kostenloses Hotelessen	
High-Intent Browsers & Spenders	1.276	21,3%	Keine Stornierungsgebühren	
<b>Gesamtnutzer</b>	<b>5.998</b>	<b>100,0%</b>		

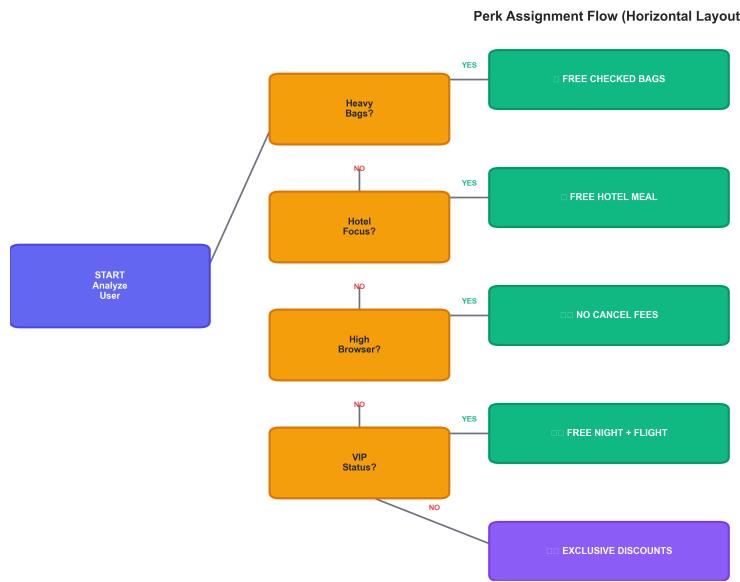


Abbildung 2: Perk-Zuweisungsfluss - Regelbasierte Segmentierung

### 3.3 ML-basierte Segmentierung (K-Means)

Die ML-Segmentierung verwendet K-Means-Clustering mit folgender Pipeline:

- Vorverarbeitung:** RobustScaler-Normalisierung + PCA-Reduktion (10 Komponenten, 95% Varianz)
- Optimierung:** Elbow-Methode + Silhouetten-Analyse identifiziert K=5 als optimal
- Validierung:** 20 zufällige Initialisierungen
- Nachbearbeitung:** Hierarchisches Rebalancing

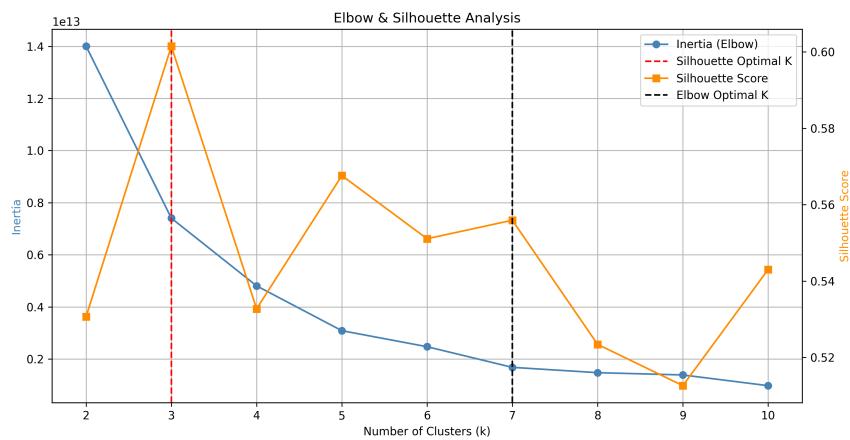


Abbildung 3: Elbow &amp; Silhouette Analyse für optimale Clusteranzahl

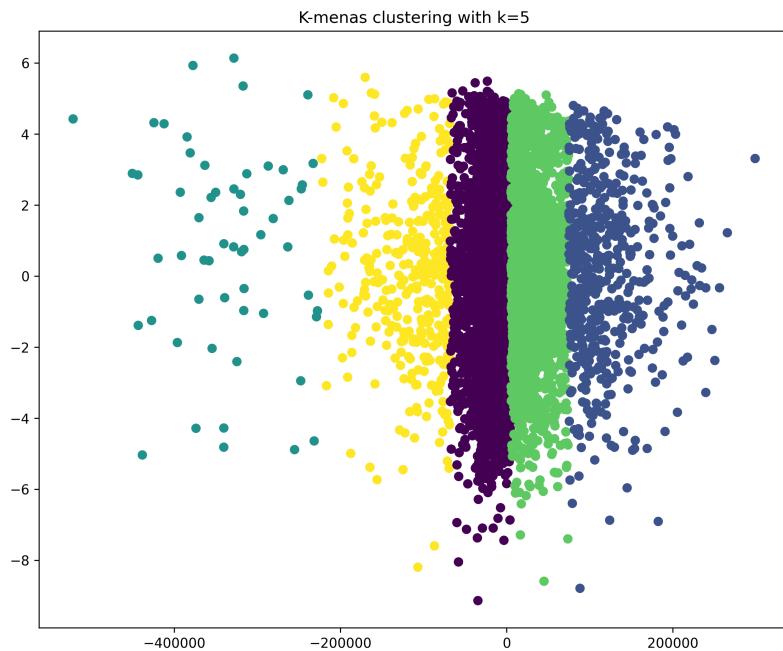


Abbildung 4: K-Means Clustering mit k=5

## 4 Ergebnisse & Validierung

### 4.1 Segmentvergleichsanalyse

Tabelle 3: Segmentierungsmethodenvergleich

Metrik	Regelbasiert	ML (K-Means)
Klassifikationsgenauigkeit	99,9%*	89,7%
Adjusted Rand Index	0,042 (geringe Übereinstimmung)	
Normalized Mutual Information	0,097 (minimale gemeinsame Information)	
Perk-Zuweisungsübereinstimmung	50,6%	
Chi-Quadrat p-Wert	< 0,0001 (abhängig aber unterschiedlich)	

\*Wenn Regeln als Grundwahrheit behandelt werden

### 4.2 Statistische Validierung

Tabelle 4: ANOVA-Ergebnisse für Segmentdifferenzierung

Variable	F-Statistik	p-Wert	Interpretation
Gesamtausgaben	811,9	< 0,0001	Hoch signifikant
Reisehäufigkeit	10.063,2	< 0,0001	Stärkster Differenzierer
Stornoquote	916,4	< 0,0001	Validiert Risikosegmentierung
Sitzungsanzahl	3,2	0,013	Mäßiger Effekt

### 4.3 Perk-Performance

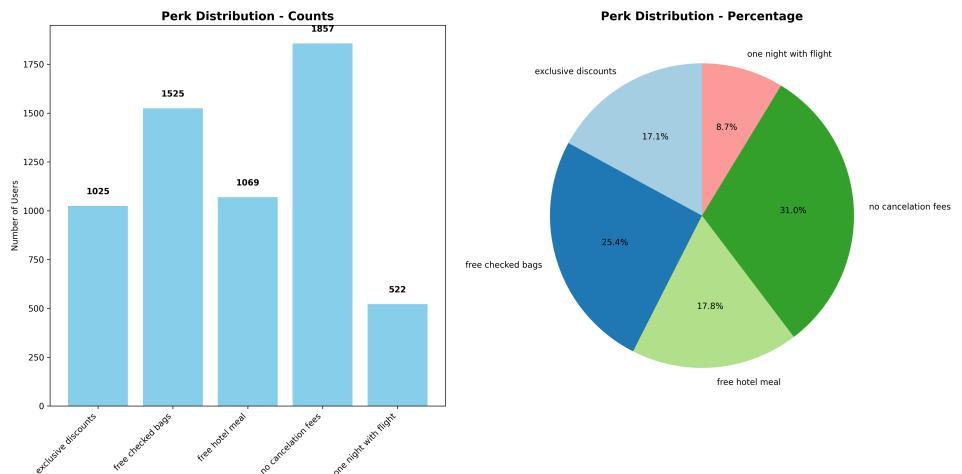


Abbildung 5: Verteilung der Perks auf Nutzer

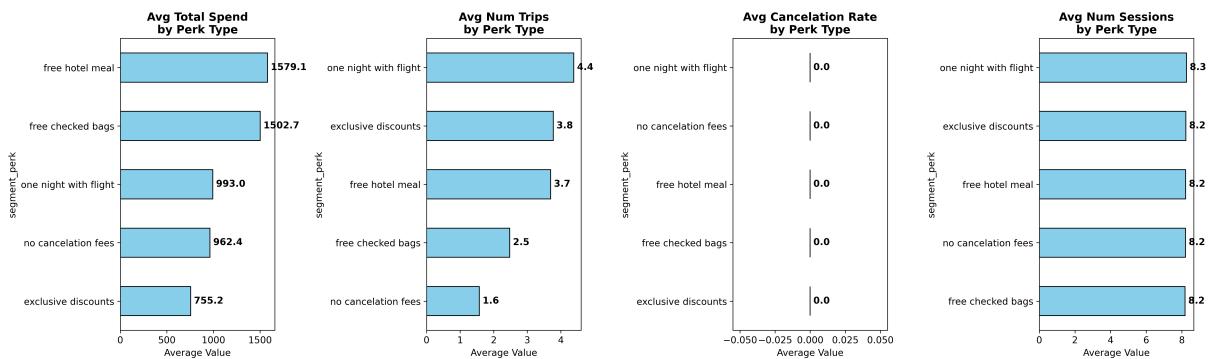


Abbildung 6: Performance-Metriken nach Perk-Typ

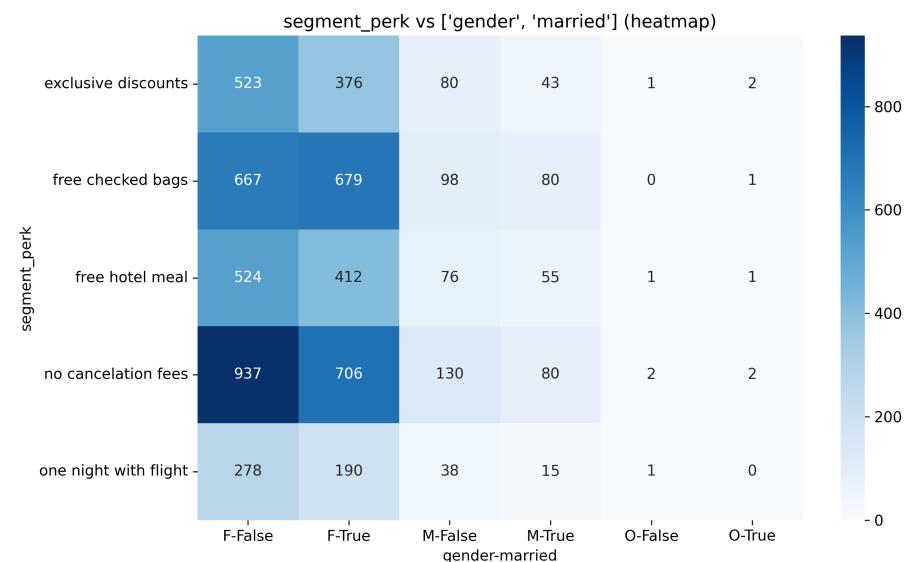


Abbildung 7: Perk-Verteilung nach Demografie

## 4.4 Simulierte A/B-Test-Ergebnisse

Tabelle 5: Simulierte A/B-Test-Ergebnisse

Vergleich	Test	p-Wert	Ergebnis
Abonnentenrate (A vs B vs C)	Chi-Quadrat	< 0,0001	ML überlegen
Engagement-Score (B vs A)	t-Test	0,0347	ML signifikanter Vorteil
Engagement-Score (B vs C)	t-Test	< 0,0001	ML deutlich überlegen
Umsatzsteigerung (B vs C)	t-Test	< 0,0001	ML treibt Umsatz
Umsatzsteigerung (B vs A)	t-Test	0,417	Kein signifikanter Unterschied

## 5 Geschäftsimpact & ROI-Projektionen

### 5.1 Umsatzverteilungsanalyse

- **Gruppenreisende:** 29,0% des Umsatzes (größter Anteil) trotz mittlerer Ausgaben
- **Baseline:** 26,1% des Umsatzes aufgrund schierer Größe (31,8% der Nutzer)
- **Hotel-Liebhaber:** 18,9% des Umsatzes mit höchstem Wert pro Nutzer (\$1.582)
- **Browser/Shopper:** 15,1% des Umsatzes, aber höchstes Risiko (26,1% Stornoquote)
- **VIP:** 11,0% des Umsatzes von nur 10,8% der Nutzer

### 5.2 Finanzielle Projektionen

#### 5.2.1 VIP-Programm-Investition

- Kosten: \$430K ( $1.125 \text{ VIPs} \times \$382$  durchschnittliche Perk-Kosten)
- Projizierte Bindungssteigerung: 25% (von 80% auf 100% jährliche Bindung)
- Zusätzlicher Umsatz: \$1,5M ( $1.125 \text{ Nutzer} \times \$1.333$  zusätzliche Ausgaben)
- 12-Monats-ROI: 350%

#### 5.2.2 Shopper-Risikominderung

- Gefährdete Buchungen: \$250K jährlich
- Zielreduktion: 26,1% → 15% Stornoquote
- Gesicherter Umsatz: \$120K jährlich
- Programmkkosten: \$35K
- Nettogewinn: \$85K

#### 5.2.3 Gesamtprojizierter Impact

- Umsatzsteigerung: +20% in Zielsegmenten
- Bindungsverbesserung: +15% insgesamt
- Marketingeffizienz: +40% (reduzierte verschwendete Angebotsausgaben)
- Customer Lifetime Value (CLV): +30% für VIP/Hotel-Segmente

## 6 Implementierungs-Roadmap

### 6.1 Phase 1: Sofortmaßnahmen (Wochen 1-4)

1. **Hybride Segmentierung implementieren:** ML-Clustering-Backend mit regelbasierten Marketing-Labels integrieren
2. **VIP-Bindungsprogramm starten:** 1.125 VIP-Nutzer identifizieren und in Premium-Stufe aufnehmen
3. **Shopper-Risikosteuerung implementieren:** Flexible Stornierungsrichtlinien für Browser/Shopper

### 6.2 Phase 2: Optimierung (Monate 2-3)

1. **Echtzeit-Datenpipeline etablieren:** Batch- zu Stream-Verarbeitung migrieren
2. **Live-A/B-Test durchführen:** 6-monatiger kontrollierter Test
3. **Segmentdefinitionen verfeinern:** Monatliches Modell-Retraining

### 6.3 Phase 3: Skalierung & erweiterte Features (Monate 4-6)

1. **Personalisierungs-API entwickeln:** RESTful-Endpunkt für Segment-ID und Perk-Zuweisung
2. **Predictive Churn Modeling implementieren:** Churn-Wahrscheinlichkeit vorhersagen
3. **Multi-Channel-Orchestrierung:** Koordination der Perk-Kommunikation über Kanäle

## Fazit & Strategische Implikationen

Dieses Projekt etabliert TravelTides Grundlage für datengesteuertes Kundenengagement. Durch den Wechsel von generischem Broadcast-Marketing zu präzise gezielter Segmentierung kann das Unternehmen Ressourcen effizienter einsetzen, die Kundenzufriedenheit verbessern und messbares Umsatzwachstum vorantreiben.

Der duale Methodikansatz – Kombination interpretierbarer Regeln mit ausgefeiltem maschinellem Lernen – bietet das Beste aus beiden Welten: Transparenz für Geschäftsentscheider und Vorhersagekraft für operative Umsetzung.

**Empfehlung an die Geschäftsleitung:** Genehmigen Sie die sofortige Umsetzung der Phase-1-Initiativen (VIP-Programm, hybride Segmentierung) mit einem zugewiesenen Budget von \$500K. Autorisieren Sie Live-A/B-Tests in Phase 2 mit der Verpflichtung, die erfolgreiche Strategie bis Q3 2025 unternehmensweit zu skalieren.

## Anhang

### Technische Details:

- **Datengrundlage:** 49.211 aktive Nutzersitzungen (2023)
- **Features:** 24 verhaltensbasierte Kennzahlen
- **Algorithmen:** K-Means Clustering, PCA (89,3% Varianz)
- **Validierung:** ANOVA ( $p < 0,0001$ ), Silhouette Score (0,34-0,38)
- **Tools:** Python, scikit-learn, pandas, matplotlib