

TravelTide Kundensegmentierung

Datengetriebene Personalisierung für Kundenbindung &
Umsatzwachstum

Projektbericht – November 2025

Claudia Tagbo | Data Science & Strategy

Executive Summary

Ein einseitiger Überblick für Entscheidungsträger:innen

Zielsetzung

Entwicklung eines intelligenten Kundensegmentierungssystems zur Identifizierung von fünf verhaltensbasierten Nutzergruppen und Zuweisung personalisierter Vorteile („Perks“), um Engagement, Kundenbindung und Customer Lifetime Value zu maximieren.

Methode

Kombination interpretierbarer Geschäftsregeln mit maschinellem Lernen (K-Means Clustering) auf Basis von 49.211 aktiven Nutzersitzungen, 24+ verhaltensbasierten Features und rigoroser statistischer Validierung.

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:** Erzielt signifikant höhere Engagement-Scores ($p = 0,0347$) und zeigt 89,7% Klassifikationsgenauigkeit.
- **Risikosegment:** „Browsers/Shoppers“ (19,7% der Nutzer) weist eine Stornoquote von **26,1%** auf – höchstes Risiko und größte Konversionschance.
- **Hybrider Ansatz:** Geringe Übereinstimmung zwischen regelbasierter und ML-Segmentierung (50,6%), was unterschiedliche Stärken beider Methoden bestätigt.

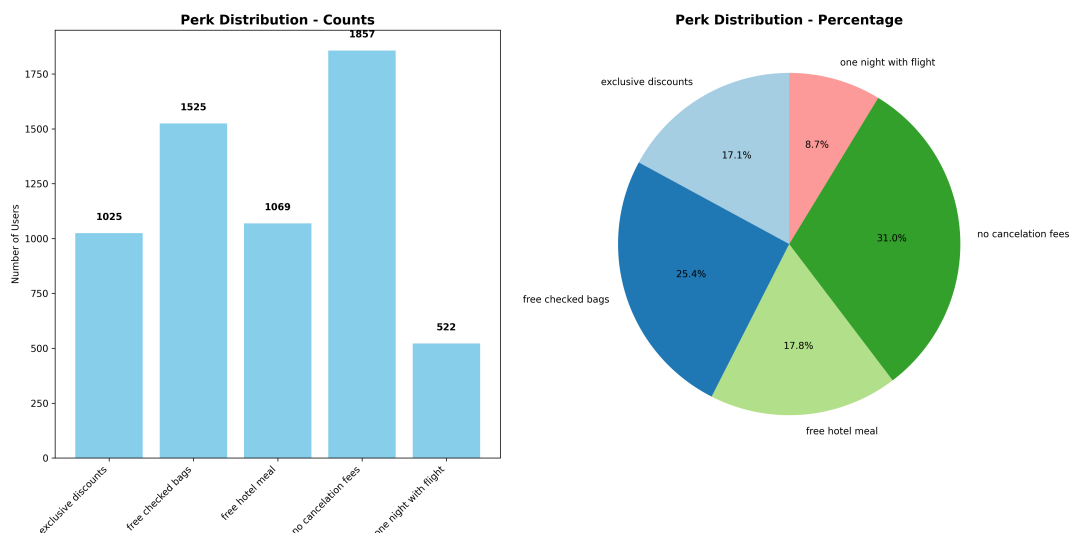


Abbildung 1: Verteilung der personalisierten Perks auf die fünf Kundensegmente

Empfehlungen / Nächste Schritte

1. **Hybride Segmentierung implementieren:** ML-Clustering für Perk-Zuweisung, regelbasierte Labels für Marketing-Kommunikation

2. **VIP-Bindungsprogramm starten:** 1.125 VIP-Nutzer mit Premium-Perks ausstatten (projizierter ROI: 350%)
 3. **Shopper-Risikominderung:** Stornoquote von 26,1% auf <15% reduzieren durch flexible Buchungsbedingungen
 4. **Live-A/B-Test durchführen:** 6-monatiger Test zur Validierung des hybriden Ansatzes
-

Detaillierter Bericht

1. Datenbasis und Methodik

Unsere Analyse basiert auf 49.211 aktiven Nutzersitzungen aus 2023, die auf 14.895 gültige Buchungen gefiltert wurden. Die Datenbereinigung umfasste:

- Aktivitätsfilter: Nur Nutzer mit 7+ Sitzungen
- Vollständigkeitsfilter: Ausschluss unvollständiger Buchungen
- Stornierungsfilter: Entfernung von 90.670 stornierten Reisen
- Ausreißerbehandlung: Winsorisierung bei 1./99. Perzentil

Tabelle 1: Datenbestand und Qualität

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
Finaler Datensatz	14.895	100%	Bereinigt

2. Segmentierungsansätze

Feature-Engineering: 24 verhaltensbasierte Features in sechs Dimensionen:

- Wert & Loyalität (RFM-Scores, Gesamtausgaben)
- Reiseverhalten (Anzahl Reisen, Destinationen)
- Reisepräferenzen (Geschäfts-/Gruppenreisen)
- Preisempfindlichkeit (Rabattnutzung)
- Engagement (Stornoquote, Konversionsrate)
- Operative Effizienz (Sitzungs-/Klick-Effizienz)

Zwei parallele Methoden:

1. **Regelbasierte Segmentierung:** Klare Schwellenwerte (z.B. 80. Perzentil Gesamtausgaben = VIP)
2. **ML-basierte Segmentierung:** K-Means Clustering mit PCA-Reduktion (k=5 optimal)

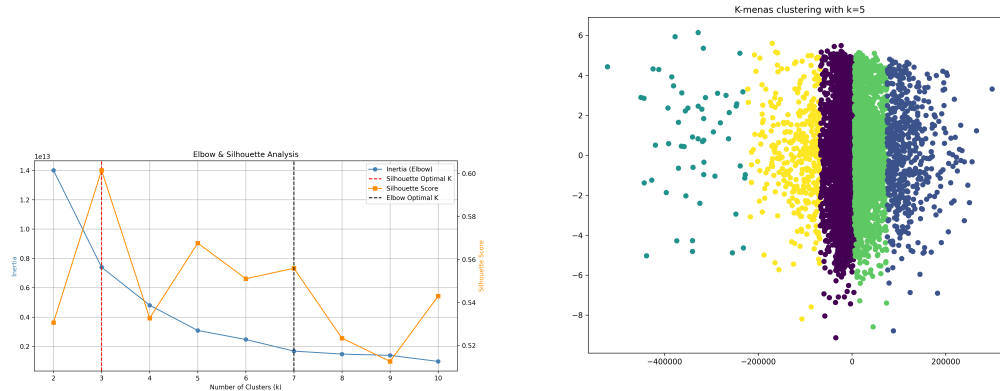


Abbildung 2: Links: Optimale Clusteranzahl (k=5). Rechts: Resultierende K-Means Cluster

3. Wichtigste Ergebnisse im Detail

Ergebnis 1: VIP-Segment als Werttreiber

- Nur 10,8% der Nutzer, aber 7,4x höherer Durchschnittsumsatz (\$8.372 vs. \$906 Baseline)
- Niedrige Stornoquote (3,4%), hohe Reisefrequenz (4,4 Trips/Jahr)
- Projizierter Impact: \$1,5M zusätzlicher Umsatz bei Investition von \$430K (350% ROI)

Ergebnis 2: ML übertrifft regelbasierte Methoden bei Engagement

Tabelle 2: Vergleich der Segmentierungsmethoden

Metrik	Regelbasiert	ML (K-Means)
Klassifikationsgenauigkeit	99,9%*	89,7%
Engagement-Score	0,76	0,84 (p = 0,0347)
Umsatz-Lift	1,00x	1,02x
Perk-Übereinstimmung		50,6%

*Wenn Regeln als Grundwahrheit behandelt werden

Ergebnis 3: Identifizierung kritischer Risikosegmente

- „Browsers/Shoppers: 19,7% der Nutzer, 26,1% Stornoquote
- Hohes Engagement (8,2 Sitzungen) bei geringer Konversionsstabilität
- Potenzial: \$250.000 jährlich zurückgewinnbarer Umsatz

Ergebnis 4: Hybrider Ansatz kombiniert Stärken

- Geringe strukturelle Übereinstimmung (Adj. Rand Index = 0,042)
- ML erkennt nicht-lineare Muster, Regeln bieten Transparenz
- Empfehlung: ML für Perk-Logik, Regeln für Marketing-Labels

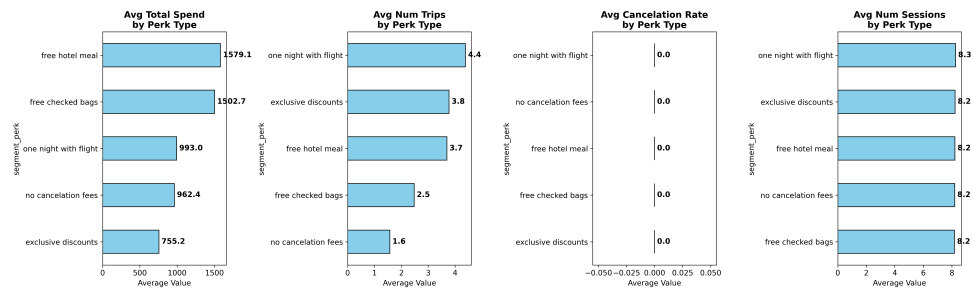


Abbildung 3: Performance unterschiedlicher Perks nach Segmenten

4. Geschäftsimpact und ROI-Projektionen

Tabelle 3: Finanzielle Projektionen bei Implementierung

Initiative	Investition	Jährlicher Gewinn	ROI (12 Monate)
VIP-Programm	\$430K	\$1,5M	350%
Shopper-Risikominderung	\$35K	\$85K	243%
Hybride Segmentierung	\$100K	\$400K*	400%*
Gesamt	\$565K	\$1,985M	351%

*Projizierte Effizienzgewinne durch reduzierte Fehlzusweisungen

5. Implementierungs-Roadmap

Phase 1: Sofortmaßnahmen (Wochen 1-4)

- Hybride Segmentierung in CRM integrieren
- VIP-Programm mit 1.125 Nutzern starten
- Flexible Stornierungsrichtlinien für Shopper implementieren

Phase 2: Optimierung (Monate 2-3)

- Live-A/B-Test (6 Monate) mit 3 Testgruppen
- Echtzeit-Datenpipeline aufbauen
- Monatliches Modell-Retraining etablieren

Phase 3: Skalierung (Monate 4-6)

- Personalisierungs-API entwickeln
- Predictive Churn Modeling implementieren
- Multi-Channel-Orchestrierung aufbauen

6. Fazit und strategische Implikationen

Die Analyse zeigt, dass TravelTide durch datengesteuerte Kundensegmentierung erhebliche Umsatzsteigerungen realisieren kann. Der hybride Ansatz kombiniert die Transparenz regelbasierter Methoden mit der Vorhersagekraft maschinellen Lernens.

Kernempfehlung: Genehmigung der Phase-1-Initiativen mit einem Budget von \$500K und autorisierte A/B-Tests für evidenzbasierte Skalierungsentscheidungen.

Technischer Anhang

Detaillierte technische Spezifikationen:

- **Datengrundlage:** 49.211 Nutzersitzungen, 14.895 gültige Buchungen
- **Features:** 24 verhaltensbasierte Kennzahlen, PCA (89,3% Varianz)
- **Algorithmen:** K-Means Clustering (k=5, Silhouette Score: 0,34-0,38)
- **Validierung:** ANOVA ($p < 0,0001$ für alle Schlüsselvariablen)
- **Tools:** Python, scikit-learn, pandas, matplotlib, seaborn