Detailed Report

Detailed Report (≤ 3 pages)
TravelTide — Customer Segmentation for a Personalized Rewards Program
Author: Athanasios Kardamis (Data Analyst / Data Scientist) | Date: August 2025

1) Context & Business Goal
TravelTide seeks to increase retention by personalizing rewards around real behavior. Rather than predefining tiers, we used unsupervised learning to discover natural segments and then aligned a single most-relevant perk to each segment.

2) Data, Cohort & Feature Engineering
• Tables: users (demographics), sessions (engagement, bookings, discounts, cancellations), flights (itineraries, fares, bags), hotels (brand, rate, rooms, check-in/out).
• Cohort: sessions after 2023-01-04; users with > 7 sessions.
• User-level aggregation with engineered features (21 total), including:
– Engagement: total_sessions, total_clicks, avg_session_duration_sec.
– Outcomes: total_completed_trips (booked & not cancelled), booking_conversion_rate.
– Spend: money_spent_flight, money_spent_hotel (discount-adjusted).
– Travel scope: total_distance_km (Haversine), avg_trip_duration_days.
– Behavior mix: flight_only_rate, hotel_only_rate, both_booked_rate.
– Logistics: total_checked_bags, avg_days_booking_to_trip.
– Loyalty: hotel_loyalty_score = 1 / # distinct hotel brands (completed trips).
– Risk/price sensitivity: discount_usage_rate, cancellation_per_booking_rate.
• Data quality considerations addressed:
– Hotel nights: computed as max(check_out, check_in) – min(check_out, check_in) to prevent negative durations.
– Cancellation sessions: excluded from finalized-behavior metrics; used a “completed_trips” CTE for outcomes.
– Hotel brand proxy: split “hotel_name” on “Brand – City” for loyalty score.

3) Modeling Approach
• Preprocessing: standardization; log transforms for skew (spend, booking lead time).
• Model selection: compared K-Means, GMM, DBSCAN across 300+ variants (feature subsets, k from 4–8).
• Final choice: K-Means, k=6 with silhouette ≈ 0.474. Selected for separation + interpretability.
• Validation: stable segment patterns in PCA projection; coherent behavioral distinctions aligned to business logic.

4) Segment Profiles & Perk Rationale
Cluster 0 — Premium Frequent Travelers (≈26.9%)
• Highest conversion and spend; long distances; package-oriented (both-booked ≈78%).
• Perk: 1 Free Hotel Night with Flight — amplifies package behavior and partner value.
Cluster 1 — Change-Prone Travelers (≈7.1%)
• High cancellations (~35%); long browsing; moderate spend; decent conversion.
• Perk: Free Cancellation — reduces commitment anxiety; encourages earlier booking.
Cluster 2 — Window Shoppers (≈9.6%)
• Very low conversion (~3%); minimal spend and short sessions; many have zero trips.
• Perk: First-Booking Discount — unlocks the very first conversion.
Cluster 3 — Hotel Loyalists (≈6.3%)
• Almost exclusively hotel-only; high brand concentration; long stays.
• Perk: Extended Stay Discounts — deepens loyalty; cost-efficient via hotel partners.
Cluster 4 — Serial Cancellers (≈1.7%)

• Youngest; multiple bookings but 100% cancellations; no realized spend.
• Perk: “Travel or Credit” Offer — if they cancel, convert intent into future revenue.
Cluster 5 — Balanced Frequent Travelers (≈48.6%)
• Large, reliable base; steady bookings and low cancellations; balanced spend across products.
• Perk: Free Checked Bag + Hotel Upgrade — drives goodwill without over-subsidizing.

5) Go-Live Plan & Measurement
• Targeting: assign each user a cluster and send a perk-specific campaign (email/in-app).
• KPIs (per cluster and overall): conversion lift vs. control, cancellation rate delta, repeat-booking rate, ARPU, email CTR/OR, redemption rate, supplier contribution.
• Experiment design: 80/20 hold-out per segment; 2-week read; sequential testing for overlapping offers.
• Guardrails: cap perk cost per booking; monitor misuse and cancellation gaming.

6) Risks & Mitigations
• Data quirks in cancellation sessions → continue fixing flags upstream; rely on completed-trip logic in modeling.
• Segment drift over time → re-train quarterly; add seasonality features.
• Over-incentivizing high-value users → prefer experiential perks (upgrades, nights) over pure cash discounts.

7) Roadmap
• Short term: productionize the cluster assignment pipeline; launch initial perk tests.
• Mid term: add seasonality & recency features; expand to more recent data.
• Long term: train supervised uplift/recommendation model using A/B outcomes to automate perk assignment at scale.