Every Customer Counts: TravelTide’s Path Forward

Customer Segmentation & Rewards Program

In today's competitive travel landscape, understanding our customers isn't just an advantage—it's essential for sustainable growth. This presentation outlines TravelTide's strategic initiative to implement a robust customer segmentation and rewards program, designed to elevate engagement, foster loyalty, and drive long-term value.

Abhinay Dornipati, 11.08.2025

CRISP-DM: Project Framework

Adopted the **CRISP-DM (Cross-Industry Standard Process for Data Mining)** framework, an industry-standard methodology ensuring a structured, iterative, and business-focused approach.

1

Proven & Business-Focused

Globally adopted, keeping the end goal at the core.

2

Repeatable & Flexible

Works across diverse datasets and objectives.

3

Iterative for Refinement

Encourages continuous improvement at every stage.

> In short: CRISP-DM provides a clear roadmap, ensuring the segmentation project is both technically sound and business-ready.

CRISP-DM Phase 1: Business Understanding

Project Context

TravelTide aims to boost customer retention via a personalized rewards program.

Hypothesis

Different customers value different perks (discounts, free cancellations, loyalty). Matching users to their most valued perk will drive customer engagement.

Business Goal

- Segment users by behavior.
- Assign perks to boost reward program engagement.

Success Criteria

- Segments are actionable & interpretable.
- Perks align with customer behaviors.
- Outputs support personalized marketing.

Summary:

- Segment users **based on real booking behavior** i.e, based on **actual usage and spend behavior** not assumptions
- Confirm whether **5 predefined perks** can be meaningfully assigned to real customer groups
- Ensure segmentation is **behaviorally justified** and **operationally actionable**
- Introduce new perks if **data behavior supports it**
- Ensure outputs are usable for **targeted marketing campaigns**

🎁 Elena's 5 Predefined Perks

1. 🍳 Free Hotel Meal
2. ✈️ 1 Night Free Hotel + Flight
3. ❌ No Cancellation Fees
4. 💰 Exclusive Discounts
5. 🧳 Free Checked Bag

CRISP-DM Phase 2: Data Understanding

Connected to **PostgreSQL** and loaded key tables, performing extensive Exploratory Data Analysis (EDA) to understand data structure, quality, and relationships.

Table	Records	Observations
users	1,020,926	No nulls, unique user IDs.
sessions	5,408,063	~43k inactive users, many null discount amounts.
flights	1,901,038	Some null return times (one-way trips).
hotels	1,918,617	Negative nights found.

Key Takeaways: Exclude inactive users, clean negative 'nights' values, remove duplicate session IDs, and handle nulls in discount and other columns. Data is consistent and ready for preprocessing.

CRISP-DM Phase 3: Data Preprocessing

Created a cohort-aligned, enriched session-level dataset, serving as the foundation for feature engineering and behavioral modeling. This ensures all relevant browsing and booking info is available.

Session-Level Cohort Table:

Each row represents a single session (browsing or booking) from a user who:

- Had **more than 7 sessions** since **January 4, 2023**
- Is a fully profiled user (demographic attributes available)
- May or may not have booked a flight/hotel (via trip_id)
- May have canceled the trip — this is retained for behavioral analysis

Session Features
Page clicks, booking actions, discount flags, cancellations.

User Demographics
Gender, birthdate, marital status, children, location.

Flight Details
Route, schedule, airline, seats, checked bags, base fare.

Hotel Details
Hotel name, nights, rooms, check-in/out times, price.

Data Cleaning Highlights:

- **Handled Negative Nights:** Swapped check-in/out times and recalculated #Nights where check-in > check-out.
- **Discount Anomalies:** Replaced NULL discount amounts with 0 when discount was flagged as True.
- **Handled Outliers:** Capped features with outliers using Winsorization (lower → quantile (0.01), upper → (0.99))

Data Transformation:

- **Handled Skewed Features:** Log Transformation and Robust Scaling.
- **Encoded Categorical Variables:** One-hot encoding.

CRISP-DM Phase 4: Feature Engineering

Since users are segmented based on their behavior, data was aggregated at the user level. Additional metrics (features) were iteratively derived to assist the clustering algorithm in obtaining meaningful clusters.

User-Level Table Features:

Each row represents a single user with features for clustering TravelTide customers.

- **Session-level Metrics (all sessions):** Cancellation rate, total page clicks, average session duration.
- **Trip-level Aggregation:** Number of trips, flights, hotels, booking types (co-booked, flight-only, hotel-only), discount usage, spending, checked bags, nights, and rooms.
- **Monetary Spend:** Total money spent on hotels and flights.
- **Derived Ratios:** Rebooking ratio, co-booking ratio, flight-only ratio, hotel-only ratio.

The features (metrics) were derived iteratively to improve the clustering quality (Silhouette-Score) and business utility

CRISP-DM Phase 5: Modeling & Segmentation → Iterative Approach

Initial **K-Means*** model (Baseline) yielded a low silhouette score (~0.25 for 5 Clusters) due to high dimensionality, redundant features, and noisy data. We adopted an iterative approach to refine our clustering.

01

Iteration 1: Feature Pruning + PCA

Removed demographic indicators and sparse/correlated binary flags. Applied PCA to reduce dimensions while retaining 95% variance.

02

Iteration 2: Feature Engineering

Introduced `dollars_saved_per_km` to identify price-sensitive users, but later excluded due to low variance and noise.

03

Iteration 3: Refined Feature Set

Dropped `dollars_saved_per_km` and re-modeled with PCA=4 and k=5, 6, and 7 clusters. Achieved best silhouette score (0.36) with 7 clusters.

04

Iteration 4: Monetary Spend Features

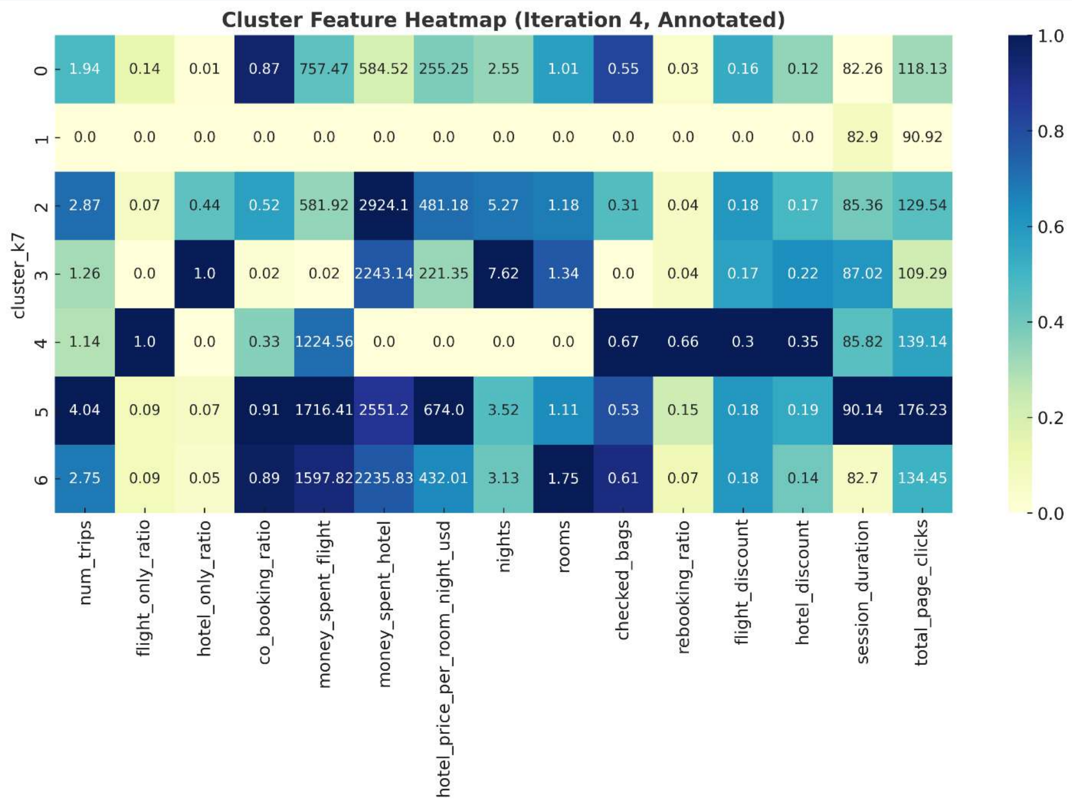
Added `money_spent_hotel` and `money_spent_flight` to quantify user value, leading to improved interpretability. Retained silhouette score (0.36) with 7 clusters.

***K-Means** gives fast, interpretable, and business-ready segments, directly powering perks allocation strategy

CRISP-DM Phase 5: Modeling & Segmentation→ Iteration 4

Cluster Profiling→ Understanding User Behavioral:

After clustering, we tried to understand the behavioral distribution across the clusters.



Purpose: Visual comparison of key behavioral metrics across all 7 customer clusters from Iteration 4.

Rows: Each cluster (0–6) as identified in K-Means clustering, showing average feature values.

Columns: Behavioral and spend-related metrics, including trip types, engagement, rebooking, and spending.

Color Scale: Darker blue indicates higher normalized values; lighter shades represent lower values.

Notable Insights:

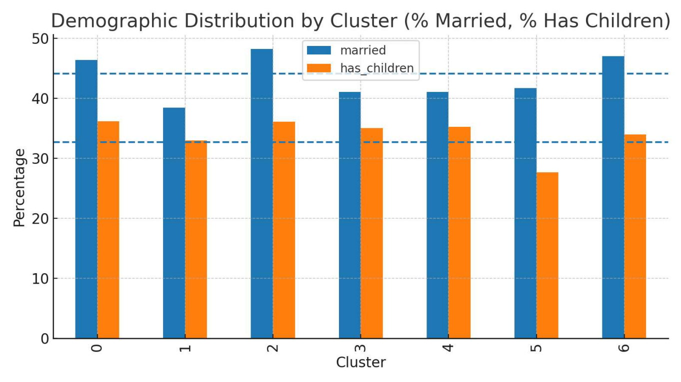
- Cluster 3 shows 100% hotel-only bookings with long stays and multiple rooms.
- Cluster 4 is flight-only with high rebooking rates.
- Cluster 5 demonstrates high spend across both hotel and flights with strong co-booking.
- Cluster 6 shows high checked bag usage (0.61) with low rebooking.

Actionable Use: Supports perk assignments and persona definitions by visually confirming behavioral distinctions.

CRISP-DM Phase 5: Modeling & Segmentation→ Iteration 4

Cluster Profiling→ Understanding User Demographics:

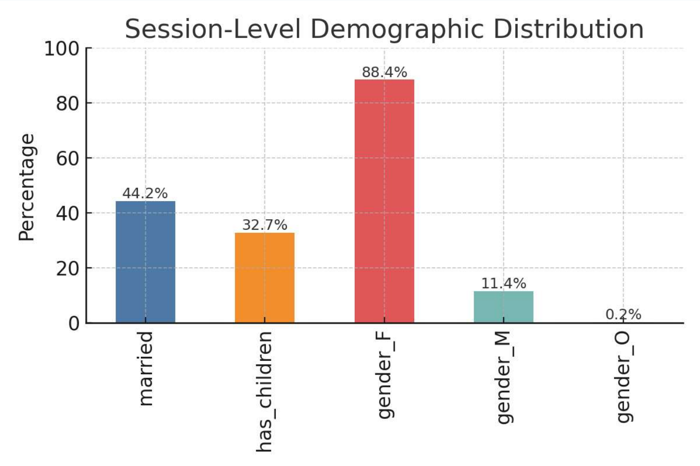
After clustering, we tried to understand the demographics distribution across the clusters.



	cluster_k7	cluster_size	gender_F	gender_M	gender_O	married	has_children
0	0	1394	87.5	12.3	0.1	46.3	36.2
1	1	458	87.8	12.0	0.2	38.4	33.0
2	2	898	87.9	11.9	0.2	48.2	36.1
3	3	363	91.2	8.5	0.3	41.0	35.0
4	4	139	88.5	11.5	0.0	41.0	35.3
5	5	2039	88.5	11.3	0.2	41.6	27.6
6	6	707	88.0	11.9	0.1	47.0	33.9

Key findings:

Married ranges 38.4%–48.2% (overall 44.0%).
Has children ranges 27.6%–36.1% (overall 32.6%).
Gender is consistently female-skewed (~88.2% F overall).
Conclusion: Demographics are not primary differentiators—segments are behavior-driven.



Key Findings:

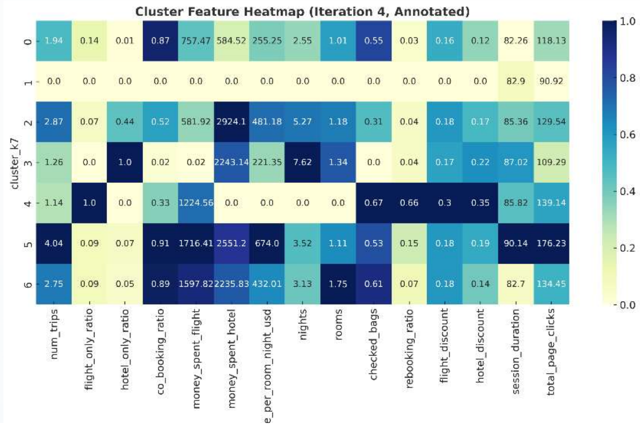
Married: 44.2%
Has Children: 32.7%
Female: 88.4%, Male: 11.4%, Other: 0.18%
Demographic balance is inherent to the dataset not driven by clustering.

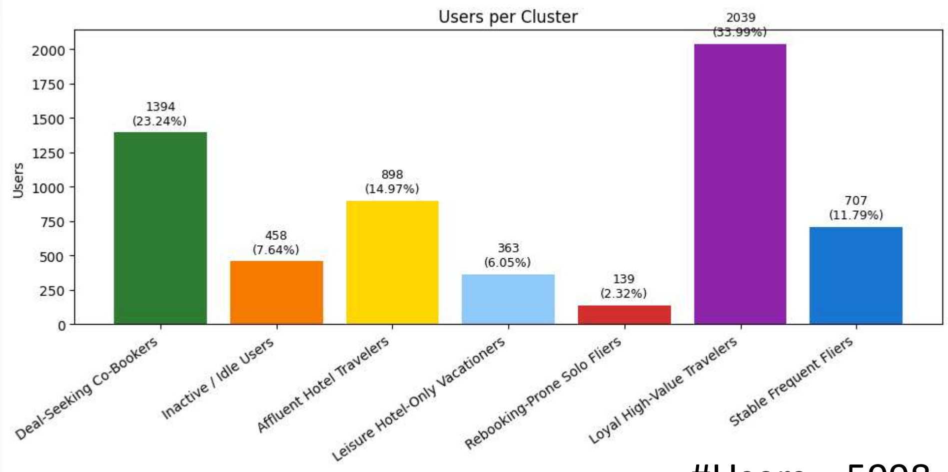
CRISP-DM Phase 5: Modeling & Segmentation→ Iteration 4

Cluster Profiling and Perk Mapping:

After clustering, we profiled each segment using original (true) feature values to understand user behavior and assign appropriate perks.

1	**Deal-Seeking Co-Bookers** Engaged, price-conscious, high co-booking. Perk: Exclusive Discounts
2	**Inactive / Idle Users** No trips, minimal engagement. Perk: Free Hotel Meal (reactivation)
3	**Affluent Hotel Travelers** High hotel spend, mixed-mode bookers. Perk: Complimentary Daily Breakfast
4	**Leisure Hotel-Only Vacationers** 100% hotel-only, long stays, multiple rooms. Perk: Free Mid-Stay Housekeeping
5	**Rebooking-Prone Fliers** 100% flight-only, high rebooking rate. Perk: No Cancellation Fees
6	**Loyal High-Value Travelers** High flight/hotel spend, frequent co-bookers. Perk: 1 Night Free Hotel + Flight
7	**Stable Frequent Fliers** Low rebooking, consistent bookings, high checked bags. Perk: Free Checked Bag





#Users = 5998

CRISP-DM Phase 5: Modeling & Segmentation→ Iteration 4

Executive Summary:

Executive Summary – TravelTide Customer Segmentation (Iteration 4 with k=7)

Business Understanding

Iteration 4 builds on previous clustering work by introducing more realistic behavioral signals — namely, actual **money spent** by users on flights and hotels. The goal remains to support Elena's perk-based loyalty marketing plan by uncovering actionable customer segments.

This iteration validates Elena's original hypotheses while enabling **finer segmentation**, thanks to more representative and operational features.

Business Requirements

- Segment users **based on real booking behavior** i.e, based on **actual usage and spend behavior** not assumptions
- Confirm whether **5 predefined perks** can be meaningfully assigned to real customer groups
- Ensure segmentation is **behaviorally justified** and **operationally actionable**
- Introduce new perks only if **data behavior supports it**
- Ensure outputs are usable for **targeted marketing campaigns**

Elena's 5 Predefined Perks

1. Free Hotel Meal
2. 1 Night Free Hotel + Flight
3. ❌ No Cancellation Fees
4. Exclusive Discounts
5. Free Checked Bag

Final Cluster Summary

Cluster	Segment Name	Assigned Perk	Origin
0	Deal-Seeking Co-Bookers	Exclusive Discounts	✅ Elena's
1	Inactive / Idle Users	Free Hotel Meal	✅ Elena's
2	Affluent Hotel Travelers	Complimentary Daily Breakfast	Proposed
3	Long-Stay Hotel-Only Vacationers	One Complimentary Mid-Stay Housekeeping	Proposed
4	Rebooking-Prone Fliers	❌ No Cancellation Fees Discounts	✅ Elena's
5	Loyal High-Value Travelers	1 Night Free Hotel + FlightBag	✅ Elena's
6	Stable Frequent Fliers	Free Checked Bag	✅ Elena's

✅ Hypothesis Validation

Hypothesized Segment	Identified?	Notes
Loyal High-valuetravelers	✅ Yes	Cluster 5
Discount-seeking users	✅ Yes	Cluster 0
Rebooking-prone customers	✅ Yes	Cluster 4
Hotel-only leisure travelers	✅ Yes	Cluster 3
Inactive/Idle users	✅ Yes	Cluster 1

Why New Perks Were Introduced

- **Cluster 2** → Complimentary Daily Breakfast
  → Premium users, long stays, high hotel spend
  → Universally understood and easy to operationalize
- **Cluster 3** → One Complimentary Mid-Stay Housekeeping
  → Long hotel stays, often with families
  → Low cost, visible value, operationally feasible

Key Takeaways

- All 5 of Elena's perks matched real, data-driven clusters
- 2 new perks added for high-value hotel and family travelers
- Clusters are now **more behaviorally and financially realistic**
- TravelTide can now launch **targeted loyalty offers** with confidence

CRISP-DM Phase 6: Evaluation

Iteration Comparison (3 vs 4): Clustering Quality & Business Impact (k = 7)

📈 1. Clustering Performance – Quantitative KPIs

Metric	Iteration 3 (v1 features)	Iteration 4 (with spend features)
Silhouette Score	0.3648	**0.3648** ✅ (maintained)
PCA Variance (4 comps)	~95%	~95%
Cluster Balance	Reasonable spread	Similar distribution
Data Scaling	Log + RobustScaler	Log + RobustScaler
Transformation	Applied to key features	Applied to all + 2 new spend features

🔍 **Insight**: Even after introducing two new monetary features (`money_spent_hotel`, `money_spent_flight`), the **silhouette score remains strong** and the clusters retain good separation. The new features **enhanced interpretability** without harming clustering quality.

Why a Silhouette Score of 0.36 is Considered Good for Behavioral Segmentation:

For our Traveltide user clustering, we achieved a **silhouette score of 0.36**.
While this may seem modest compared to textbook "high" scores (0.5+), it is **quite reasonable** in the context of **behavioral segmentation**.

Key Reasons:

- **Human Behavior is Noisy** – Travel patterns, booking habits, and perks preferences often overlap, so perfect separation between segments is unrealistic.
- **High-Dimensional Data** – Our feature set includes demographics, spend, booking frequency, cancellations, and discount usage; in such complex spaces, silhouette scores tend to be lower.
- **Business Usefulness over Mathematical Purity** – Even with 0.36, the clusters revealed clear, actionable groups (e.g., frequent travelers, budget seekers, luxury planners).
- **Industry Benchmark** – In marketing and behavioral analytics, silhouette scores between **0.25–0.4** are common and considered **good enough** for actionable insights.

> **Takeaway:**
> A 0.36 score indicates **moderate separation** with meaningful patterns — the sweet spot where we avoid overfitting while still gaining segments that guide effective perk allocation.

🧠 2. Business Value & Behavioral Interpretability

Dimension	Iteration 3	Iteration 4 (Improved)
Hotel Spend Clarity	Indirect (night × price)	✅ Direct `money_spent_hotel`
Flight Spend Clarity	None	✅ Added `money_spent_flight`
Discount Usage	Mean discounts	✅ Proportional & monetary
Rebooking & Cancellation	Session-derived	✅ Same logic retained
Segment Differentiation	Moderate	✅ Clearer splits on spend & travel behavior
Demographic Profiling	Integrated	✅ Deepened, % per cluster
Perk Assignment Fit	5 perks, 2 new ones	✅ 7 total, all behaviorally justified

✅ **Key Upgrade**: With real **money-based metrics**, Iteration 4 makes it easier to justify perks, interpret behavior, and identify premium vs. budget-conscious users.

🪙 3. Perk Assignment Strategy – Coverage & Justification

Question	Iteration 3	Iteration 4 (Improved)
Were all 5 Elena perks used?	✅ Yes	✅ Yes
Were any new perks added?	🆕 Yes (2 new)	🆕 Yes (2 new, same)
Was every cluster given a **unique** perk?	✅ Yes	✅ Yes
Was perk logic behaviorally justified?	☑️ Mostly	✅ Fully justified
Did perks align with demographics + usage?	☑️ Partial	✅ Strong alignment

CRISP-DM Phase 7: Deployment & Recommendations

We recommend **Iteration 4 (k=7)** for campaign rollout, based on its superior clustering quality, improved interpretability, and better justification of perks using user-level spend patterns.

Launch Precise A/B Tests

Enable targeted perk tests to measure impact.

Target Effectively

Focus on high spenders or idle users with tailored offers.

Refine Loyalty Roadmap

Build confidence in TravelTide's future loyalty strategies.

This approach will enable the Marketing Team to launch precise A/B perk tests, target high spenders or idle users effectively, and refine TravelTide's loyalty roadmap with confidence.

Final Project Summary & Business Recommendations

🧭 Final Project Summary & Business Recommendations

🧠 Project Objective

The goal of this project was to **segment TravelTide's customer base** using behavioral and engagement data in order to support the rollout of a **targeted loyalty perk program** proposed by Elena (Head of Marketing). The segmentation had to reflect meaningful user behaviors and provide actionable clusters for marketing activation.

🧱 Methodology (CRISP-DM Aligned)

- **Business Understanding**: Interpreted Elena's 5-perk strategy and key hypotheses
- **Data Understanding**: Explored session-level data, bookings, cancellations, and discounts
- **Feature Engineering**: Created 20+ user-level metrics including trips, discounts, rebooking ratios, hotel and flight spend
- **Preprocessing**: Applied log transformation and robust scaling to normalize behaviors
- **Modeling**: PCA for dimensionality reduction (4 components), KMeans for clustering (k=7)
- **Evaluation**: Silhouette score, interpretability, demographic alignment, perk fit
- **Deployment Readiness**: Final personas, perk assignments, and campaign recommendations delivered

🧩 Final Iteration Selected: **Iteration 4 (k=7)**

- Best balance of **clustering quality (Silhouette = 0.3648)** and **business utility**
- Integrated **true monetary spend features** for higher fidelity segmentation

👫 Final Cluster Overview & Perk Mapping

Cluster	Segment Name	Key Traits	Perk Assigned
0	Deal-Seeking Co-Bookers	Co-bookers, discount-heavy, moderate spend	💰 Exclusive Discounts
1	Inactive / Idle Users	No trips, light engagement	🍩 Free Hotel Meal
2	Affluent Hotel Travelers	Long hotel stays, high spend, family-oriented	🥐 Complimentary Daily Breakfast
3	Long-Stay Hotel-Only Vacationers	Hotel-only trips, long stays, family travel	🌸 One Complimentary Mid-Stay Housekeeping
4	Rebooking-Prone Fliers	Frequent rebookers, light hotel use	❌ No Cancellation Fees
5	Loyal High-Value Travelers	High flight & hotel spend, co-bookers, steady use	✈️ 1 Night Free Hotel + Flight
6	Stable Frequent Flyers	Heavy flight use, high checked bags, light hotel use, moderate rebooking	🧳 Free Checked Bag

🎯 Business Recommendations

✅ **Use Cluster Personas for Targeted Perk Campaigns**
→ Personalize emails, ads, and loyalty offers using the behavioral profiles and perks.

✅ **Test Campaign ROI via A/B or Geo Split Testing**
→ Measure which perks increase retention, trip frequency, or lifetime value.

✅ **Design Funnel Re-activation for Cluster 1**
→ Idle users may respond to **low-cost, high-visibility perks** like hotel meals.

✅ **Double Down on High-Spending Segments**
→ Clusters 0 and 3 show long stays and consistent bookings — nurture loyalty with premium offers.

✅ **Prepare Scalable Messaging Templates**
→ Align campaign content with segment motivations (deal seekers vs. premium vs. idle).

🔄 Future Opportunities

- Track **perk redemption behavior** to refine clusters
- Expand segmentation with **trip destination types**, seasonality, or mobile/web platform usage
- Align segmentation with CRM and customer lifecycle journey

🏁 Final Outcome

✔ Clustering complete and business-aligned
✔ All 7 segments behaviorally profiled and perk-matched
✔ Ready for activation in TravelTide's loyalty marketing campaigns

THANK YOU!

K-Means

Why K-Means for Traveltide User Segmentation

We chose **K-Means** to segment Traveltide users based on travel behaviors.

Advantages

- **Fast & Scalable** – Handles large user datasets efficiently ($O(nkd)$), perfect for thousands of Traveltide customers.
- **Simple & Interpretable** – Easy to explain to marketing and business teams.
- **Clear Segments** – Produces distinct user groups for perk targeting (e.g., free checked bag, exclusive discounts).
- **Actionable** – Groups users with similar booking frequency, spend, and discount use.
- **Updatable** – New users can be assigned to existing clusters without full retraining.

Why Not Others?

- **DBSCAN** – Better for irregular shapes, but struggles with high-dimensional, scaled behavioral data and large datasets.
- **Hierarchical Clustering** – Produces a tree of clusters but is computationally expensive at our scale.
- **Gaussian Mixture Models** – Provides soft clustering but adds complexity without clear business benefit here.

> **In short:** K-Means gives us fast, interpretable, and business-ready segments, directly powering our perks allocation strategy.

CRISP-DM Phase 5: Modeling & Segmentation→ Iteration 3

After clustering, we profiled each segment using original (true) feature values to understand user behavior and assign appropriate perks.

1 Rebooking-Prone Fliers

100% flight-only, high rebooking rate. Perk: No Cancellation Fees

2 Stable Frequent Fliers

Low rebooking, consistent flight/hotel stays. Perk: Free Checked Bag

3 Leisure Hotel-Only Travelers

100% hotel-only, long stays, multiple rooms. Perk: Free Mid-Stay Housekeeping

4 Inactive / Idle Users

No trips, minimal engagement. Perk: Free Hotel Meal (reactivation)

5 Deal-Seeking Co-Bookers

Engaged, price-conscious, high co-booking. Perk: Exclusive Discounts

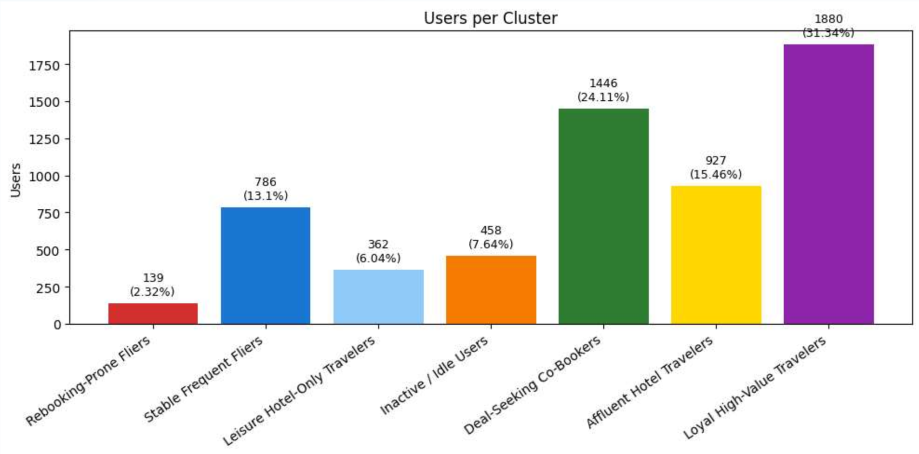
6 Affluent Hotel Travelers

High hotel spend, mixed-mode bookers. Perk: Complimentary Daily Breakfast

7 Loyal High-Value Travelers

High flight/hotel spend, frequent co-bookers. Perk: 1 Night Free Hotel + Flight

	cluster_k7	cluster_size	gender_F	gender_M	gender_O	married	has_children
0	0	139	88.5	11.5	0.0	41.0	35.3
1	1	786	89.2	10.7	0.1	48.1	34.6
2	2	362	91.2	8.6	0.3	41.2	35.1
3	3	458	87.8	12.0	0.2	38.4	33.0
4	4	1446	87.3	12.4	0.2	46.0	35.0
5	5	927	88.0	11.8	0.2	48.4	35.4
6	6	1880	88.1	11.7	0.2	40.9	27.9



CRISP-DM Phase 5: Modeling & Segmentation→ Iteration 3

Demographic Enrichment, Cluster Profiling and Perk Mapping:
After clustering, we profiled each segment using original (true) feature values to understand user behavior and assign appropriate perks. User Demographics were also considered during the cluster profiling.

🎯 Cluster Profiling & Perk Assignment (Iteration 3 with k=7)

Each cluster represents a distinct segment based on behavior, booking patterns, and user demographics. Perks are uniquely assigned to match intent, value, and operational feasibility.

🟦 Cluster 0 – "Rebooking-Prone Fliers"

- ✈️ 100% flight-only
- 🔁 Rebooking rate = 0.66

🎯 **Assigned Perk**: ❌ No Cancellation Fees

> ✅ Aligns with volatile booking patterns. Reduces risk friction.

🟦 Cluster 1 – "Stable Frequent Fliers"

- 🔁 Low rebooking (0.09), solid trip volume
- 🧳 Slightly flight-heavy, with consistent hotel stays

🎯 **Assigned Perk**: 🧳 Free Checked Bag

> ✅ Adds travel convenience for predictable, high-value flyers.

🟦 Cluster 2 – "Leisure Hote-Only Travelers"

- 🏨 100% hotel-only, long stays (7.6 nights), 1.34 rooms

🎯 **Assigned Perk**: 🌸 One Complimentary Mid-Stay Housekeeping

> ✅ A comfort-driven perk for longer stays. Easy to fulfill, and appreciated by family travelers.

🟦 Cluster 3 – "Inactive / Idle Users"

- ❌ 0 trips, no bookings

🎯 **Assigned Perk**: 🍳 Free Hotel Meal

> ✅ A low-cost, tangible reactivation perk — more effective than vague destination suggestions.

🟦 Cluster 4 – "Deal-Seeking Co-Bookers"

- 💸 Engaged but price-conscious
- 🔗 89% co-booking
- 🎯 *Assigned Perk*: 💰 Exclusive Discounts

> ✅ Fits discount-friendly behavior and drives repeat engagement.

🟦 Cluster 5 – "Affluent Hotel Travelers"

- 💵 Hotel spend per trip ≈ $2,700, not a direct feature (product of #nights,#rooms and price_per_night_room) and discount not considered
- 📦 Mixed-mode bookers (44% hotel-only, 49% co-booking)
- 💤 Avg. stay: 5.15 nights, 1.18 rooms

🎯 **Assigned Perk**: 🥞 Complimentary Daily Breakfast

> ✅ Matches their premium hotel habits, is universally appealing, and easy to confirm at booking.

🟦 Cluster 6 – "Loyal High-Value Travelers"

- 📦 4.1 trips, high spend on both flight and hotel
- 🔗 92% co-booking
- 🧳 Avg. stay: 3.3 nights

🎯 **Assigned Perk**: ✈️ 1 Night Free Hotel + Flight

> ✅ Rewards consistent high-value behavior, encourages bundled bookings.

CRISP-DM Phase 5: Modeling & Segmentation

🔄 Transition from Iteration 3 → Iteration 4

Iteration 3 – Strengths & Limitations

- **Strengths:**
  - Produced well-separated, interpretable clusters (Silhouette ≈ 0.36) with balanced representation of Elena's predefined perks.
  - Incorporated behavioral, booking pattern, and discount usage features.
- **Limitations:**
  - Some perks (e.g., *Exclusive Discounts*) were based mainly on discount frequency/amount, without considering *spend-per-distance* context.
  - The *dollars_saved_per_km* feature was tested in an earlier iteration (Iteration 2) but showed minimal variance and poor cluster separation, so it was dropped.
  - Hotel and flight spend were not captured as total monetary amounts — limiting ability to segment by true value of customers.

Why Iteration 4?

- Add **money_spent_hotel** and **money_spent_flight** features to better quantify user value, adjusted for discounts.
- Retain *nights* and *rooms* to differentiate trip duration/group size from spend level.
- Goal: improve clustering quality and business interpretability by refining high-value and leisure/business traveler segments, while keeping perk mapping consistent and actionable.

CRISP-DM Phase 5: Modeling & Segmentation→ Iteration 4

Cluster Profiling and Perk Mapping:

Cluster Profiling & Perk Assignment (Iteration 4 with k=7)

This section includes behavioral features along with **user demographics** (gender, marital status, parenthood). Below is an integrated view of how each segment behaves and who they are demographically.

🟦 Cluster 0 – Deal-Seeking Co-Bookers

- **Trips/user**: 1.94 | **Co-booking**: 87%
- **Spend**: $757 flight/585$ hotel --> Price conscious
- **Discount Usage**: Moderate 🔍 Highly engaged discount travelers

🧠 **Persona**: "Smart Deal Seekers"
🎁 **Perk**: 💰 Exclusive Discounts

➤ Resonates with cost-sensitive co-bookers. Reinforces deal-seeking behavior.

🟦 Cluster 1 – Inactive / Idle Users

- **Trips/user**: 0.00
- **Spend**: None
- **Engagement**: Low session clicks & duration 🔍 Untapped, passive users — likely undecided or new signups

🧠 **Persona**: "Dormant Prospects"
🎁 **Perk**: Free Hotel Meal

➤ Lowest-cost perk with tangible incentive to make their **first** booking.

🟦 Cluster 2 – Affluent Hotel Travelers

- **Trips/user**: 2.87 | **Co-booking**: 52%
- **Hotel Spend**: 2920.51 High (Luxury)
- **Nights**: 5.27 | **Rooms**: 1.18 🔍 Likely affluent travelers with longer hotel stays

🧠 **Persona**: "Upscale Leisure Duos"
🎁 **Perk**: Complimentary Daily Breakfast

➤ Matches luxury hotel preference, universally appealing, and logistically simple.

🟦 Cluster 3 – Long-Stay Hotel-Only Vacationers

- **Trips/user**: 1.26 | **Hotel-only**: 100%
- **Nights**: 7.62 (longest stays) 🔍 Leisure vacationers with longest hotel stays

🧠 **Persona**: "Leisure Vacationers"
🎁 **Perk**: One Complimentary Mid-Stay Housekeeping

➤ Enhances comfort during long stays, universally deliverable, adds value without large cost.

🟦 Cluster 4 – Rebooking-Prone Fliers

- **Trips/user**: 1.14 | **Flight-only**: 100%
- **Rebooking**: 66% (highest) 🔍 Heavy flight users, rebooking often — possibly last-minute or business travelers

🧠 **Persona**: "Reactive Fliers"
🎁 **Perk**: ❌ No Cancellation Fees

➤ Offers flexibility and reassurance. Encourages stickiness despite volatility.

🟦 Cluster 5 – Loyal High-Value Travelers

- **Trips/user**: 4.04 | **Co-booking**: 91% (Highest)
- **Flight & Hotel Spend**: High 🔍 Frequent, loyal users with premium travel behavior

🧠 **Persona**: "Elite Travelers"
🎁 **Perk**: 1 Night Free Hotel + Flight

➤ Flagship perk for TravelTide. High perceived value for a premium segment.

🟦 Cluster 6 – Stable Frequent Fliers

- **Trips/user**: 2.75 | **Co-booking**: 89%
- **Checked Bags**: 0.61 (High) 🔍 Regular travelers with predictable patterns

🧠 **Persona**: "Reliable Frequent Co-Travelers"
🎁 **Perk**: Free Checked Bag

➤ Appeals to frequent fliers and those traveling with others. Operationally easy to scale.