

TravelTide Rewards: Detailed Customer Segmentation Report

1. Project Background and Methodology

1.1 Data Preparation and Cleaning

The segmentation project began with **careful** data preparation and feature engineering, which was the most time-intensive phase, executed in the (combined) traveltide-segmentation-analysis (3).ipynb notebook. The goal was to transform raw transactional data into a set of normalized, behavioral metrics that would be directly applicable to segmentation models.

- **Initial Data Filtering & Curation:** The analysis started with a highly selective SQL query executed on the sessions table. The dataset was filtered to include only sessions that started **after 2023-01-04** and retained only users with **more than 7 sessions** in the period, resulting in **5,998 active users**. This filtered sessions base was then joined with users, flights, and hotels tables to create a comprehensive session_base dataset for feature extraction. Missing values were observed but intentionally left for non-booking records, where trip_id, flight, and hotel details were null.
- **Deep Feature Engineering & User Aggregation:** This step was key to converting raw data into actionable insights, culminating in the user_agg_features DataFrame:
 - **Core Behavioral Metrics:** Calculated foundational features such as total sessions, **Total Flights/Hotels Booked** (excluding cancellations), and **Total Cancellations** at the user level.
 - **Engagement Metrics:** Calculated total_page_clicks and avg_page_clicks to quantify user activity.
 - **Spend Features:** Calculated avg_flight_fare_usd, avg_hotel_price_usd, and avg_hotel_total_spend_usd. The **Spend Profile** categorical feature was derived from these values (Budget, Mid-Range, Premium).
 - **Travel Party Features:** The binary feature **has_children** was carried over from the demographics table.
 - **Final Data Output:** This process resulted in a single, finalized feature set for the **5,998 active users**, which was saved as traveltide_final_features.csv, ensuring all subsequent analysis was performed on a clean, consistent, and feature-rich dataset.

1.2 Exploratory Data Analysis (EDA)

The EDA phase provided foundational insights, directly validating the final hierarchical rules:

- **User Demographics:**
 - The user base is predominantly **Female (88.2%)** with 5,998 total users.
 - The **35–49 age bracket** represents the largest single group (**2,951 users**).

- **Parental Status:** 32.6% of the user base is classified as **Family** (has children), while 67.4% are Solo Travelers.
- **Engagement and Conversion:**
 - **Conversion Rate:** The overall booking conversion rate is stable (Flights: 28.7%; Hotels: 29.9% in the USA, and slightly higher in Canada).
 - **Engagement Value:** The median number of page clicks for a session that resulted in a booking was **22.0**, significantly higher than sessions that did not result in a booking (**8.0**). This confirms that high page click volume is the strongest signal of purchase intent and conversion likelihood.
 - **Engagement by Age:** The **35-49 age group** exhibits the highest average sessions (**4.02**) and average page clicks (**18.40**), confirming them as the **Core Age Group**.
- **Spend and Behavior Profiling (Validating Segmentation Splits):**
 - **Spend by Age:** The **65+** group showed the highest average flight fare (**\$921.87**) and total hotel spend (**\$1,326.65**), while the **35-49** group had the lowest average spend in both categories, indicating high-value transactions are often driven by older, less frequent travelers.
 - **Correlation:** There is a high correlation between `total_page_clicks` and `total_cancellations` (**0.744**). This suggests that users who are highly active on the site—spending time clicking through many options—are often engaged in **comparison shopping, seeking the best deal, or building complex itineraries that are prone to last-minute failure**. This high click/high cancellation behavior is an important finding, confirming the need for perks to mitigate purchase anxiety among volatile segments.
 - **Family Market Distinction:** As hypothesized, the Family segment shows a distinct profile, though surprisingly, the *average* number of bookings for Family travelers (Flights: 2.09, Hotels: 2.22) is slightly *lower* than Leisure travelers (Flights: 2.35, Hotels: 2.43), confirming their value lies in **high transaction size, not necessarily high frequency**.

1.3 Segmentation Model Selection

Initial attempts to segment the data using traditional unsupervised Machine Learning techniques (K-Means, documented in `ML Approach.ipynb`) failed to produce clear, non-overlapping segments.

- **ML Attempt:** After cleaning outliers (removing **2,014 rows** via IQR) and scaling, **PCA** was used, demonstrating 94.39% of variance explained by the first 5 components. The K-Means analysis resulted in a peak Silhouette Score of **0.525 at K=5**. However, the resulting cluster characteristics were overly complex and overlapping (e.g., clusters combining different spend and age groups without a clear defining logic), rendering them impractical for marketing execution.

Therefore, a **Manual Hierarchical Segmentation** model was adopted (detailed in `Customer Segmentation Final (1).ipynb`). This method applies a sequence of business-logic-driven rules to define six mutually exclusive and collectively exhaustive customer segments. The hierarchy

prioritized high-impact behavioral traits:

1. **Spend Profile** (Premium, Mid-Range, Budget) - *Used as the primary split for Non-Family users.*
2. **Travel with Children** (Differentiates Family from others) - *Used to capture the highest-value group early.*
3. **Age Group** (Core 35-49 vs. Other Ages)
4. **Engagement Level** (High vs. Low, defined by \$\geq 22.0\$ total page clicks)

2. Detailed Segment Profiles and Strategy

The segmentation process yielded six highly distinctive customer segments using the assign_final_segment_v4 logic.

Segment Name	% of Base	User Count	Defining Characteristics
1. Family Market	22.17%	1,330	Defined exclusively by Traveller Type = 'Family' (has children). High average transaction value.
2. Premium High Engaged Other	19.67%	1,180	Premium Spend and High Engagement (\$\geq 22\$ Clicks); outside the Core Age (35-49).
3. Mid-Ranged High Engaged Core	18.34%	1,100	Mid-Range Spend, High Engagement, and in the Core Age (35-49) . Highly frequent and loyal.
4. Mid-Ranged High Engaged Other	14.95%	897	Mid-Range Spend and High Engagement; outside the Core Age (35-49).

5. Price Sensitive Buyer	14.49%	869	Budget Spend and High Engagement. Highly transactional but highly price-focused.
6. Premium High Engaged Core	10.37%	622	Premium Spend, High Engagement, and in the Core Age (35-49). The ideal, highest-LTV customer.

Segment 1: The Family Market (22.17%)

Metric	Detail
Profile	Users traveling with children. Their low frequency is offset by high average transaction size (booking larger rooms/multiple flights).
Strategic Focus	Retention & Upsell. Focus on convenience. Offer bulk discounts on activities and accommodations suitable for families.
Perk Alignment	Priority Booking, Free Checked Bags, or Second Room Discount.

Segment 5: Price Sensitive Buyer (14.49%)

Metric	Detail
Profile	Budget-segment users who exhibit high engagement (\$\ge 22\$ clicks) . This group is actively shopping for the best deal and highly risk-averse regarding non-refundable purchases.

Strategic Focus	Conversion & Risk Mitigation. Validate their purchase decision by removing friction points associated with budget travel.
Perk Alignment	A specific risk mitigation perk that addresses purchase anxiety.

Segment 6: Premium High Engaged Core (10.37%)

Metric	Detail
Profile	The highest-LTV group: premium spenders in the core age range who are frequently on the platform.
Strategic Focus	Exclusivity & Advocacy. Do not use discounts. Focus on superior, proactive service and generating referrals.
Perk Alignment	Dedicated Concierge, Early Access to Inventory, or Premium Lounge Access.

3. Conclusion and Recommendations

The segmentation provides a clear framework for resource allocation. The combined **Family Market (22.17%)** and **Premium High Engaged Core (10.37%)** segments represent the primary high-value drivers for TravelTide, while the **Price Sensitive Buyer (14.49%)** offers a key opportunity for conversion using tactical perks.

Overall Strategic Mandates:

- Differentiated Marketing:** All future rewards invitation emails must be segment-specific. Tactical, risk-mitigating incentives should be precisely targeted to the highest-risk segments for maximum conversion impact.
- Loyalty Tier Alignment:** Design the top loyalty tiers (e.g., Platinum/Diamond) to deliver the non-discount perks most valued by Premium segments (Exclusivity) and Family segments (Convenience).
- Engagement Focus:** Leverage the finding that sessions resulting in a booking have a median of 22 clicks. Track high-click behavior in lower-value segments (e.g., Mid-Ranged Other) for immediate, contextual marketing nudges rather than relying on blanket

discounts.