

TravelTide Customer Segmentation and Perk Recommendation Project Report



1- Introduction

TravelTide, a dynamic e-booking platform renowned for its comprehensive travel data and extensive inventory, initiated this project with the goal of strengthening customer loyalty through a targeted rewards program. This analysis aimed to categorize TravelTide's customers into distinct segments based on their different characteristics. The primary objectives were

- To validate the presence of customer groups receptive to specific perks.
- To recommend a potentially appealing perk for each customer segment.

This data-driven approach is designed to inform a more personalized and effective marketing strategy, ultimately enhancing customer retention and fostering business growth.

2- Data Acquisition and Preparation

The foundation of this analysis involved the retrieval and preparation of data from the TravelTide database. Information was extracted from four key tables: **users** (demographic details), **sessions** (user browsing activity), **flights** (flight booking information), and **hotels** (hotel booking details). The initial dataset comprised over 5.4 million user sessions and around 1.0 million unique users, enriched with demographic and booking information.

To focus the analysis on active and recent users, the dataset was filtered.

- Only records after January 4, 2023, were retained, and
- users with fewer than seven browsing sessions were excluded.

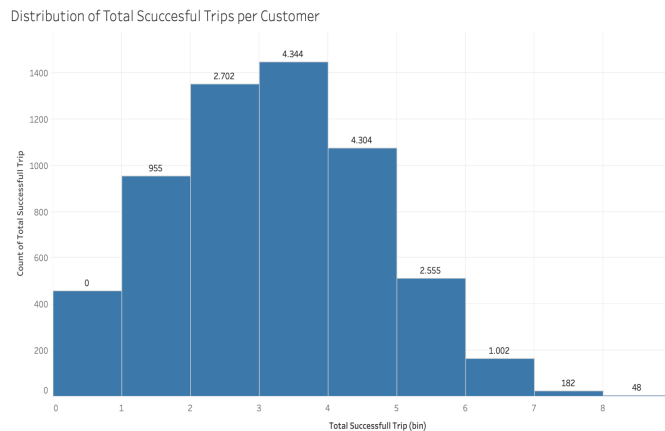
This resulted in a refined cohort of engaged users with sufficient interaction data for meaningful analysis. The subsequent data cleaning and transformation phase involved adjusting data types for consistency, and strategically addressing missing values.

3- Exploratory Data Analysis

A thorough exploratory data analysis (EDA) was conducted to gain initial insights into the characteristics and behaviors of the selected customer cohort. This involved examining the distribution of demographic variables, analyzing booking patterns, including destination preferences and airline choices. For instance it revealed

- A higher number of female users compared to male users,
- A strong concentration of users residing in the USA, and
- NewYork emerging as the most popular travel destination.

Furthermore, anomalies in the recorded hotel stay duration were identified and rectified by calculating the actual duration based on check-in and check-out times or flight return dates. This step ensured the accuracy of subsequent analyses involving stay duration.



4- Feature Engineering

Several new variables were engineered to increase understanding of user behavior/spending patterns, enrich the dataset and to derive more insightful features for segmentation, For example:

- **Session Duration:** The length of user browsing sessions.
- **Trip Duration:** The total duration of a travel booking.
- **User Since:** The duration a user has been registered on the platform.
- **Distance Flown:** The estimated distance of flight bookings.
- **Total Flights Booked:** The total number of flight bookings made by a user.
- **Hotel Expenditure:** The total spending on hotel bookings.
- **Flight Expenditure:** The total spending on flight bookings.

5 User-Based Table Creation and Further Feature Engineering

The session-level data was then aggregated to create a user-centric table, summarizing each user's overall engagement and booking history. This involved calculating metrics such as the total number of successful trips, total trip duration, unique destinations visited, total spending, and the total number of hotel and flight bookings.

Building upon this user-based table, further feature engineering was conducted to derive rates, ratios, and averages that better represent user behavior. Such as

- **Clicks Per Session:** The average number of clicks within a session.
- **Trip Cancellation Rate:** The proportion of booked trips that were canceled.
- **Average Spend:** The average amount spent per trip.
- **Flight Trip Ratio:** The proportion of trips involving flight bookings.
- **Average Hotel Stay:** The average duration of hotel stays.
- **Average Discounted Flight/Hotel:** The average discount amount received on flights and hotels.

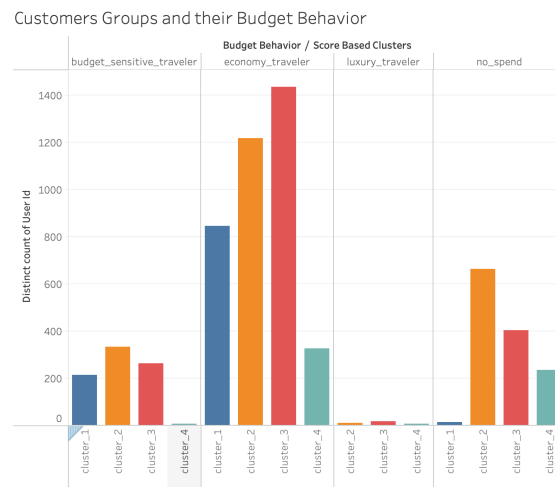
6- Customer Segmentation

The core of the project involved customer segmentation, which is achieved through a combination of rule-based segmentation and value scoring, based on predefined criteria related to their:

- **Travel Behavior:** (e.g., Infrequent, Frequent Travelers) based on the total number of successful trips.
- **Website Engagement:** (e.g., High, Medium) based on average session duration and clicks per session.
- **Budgetary Behavior:** (e.g., Budget Sensitive, Luxury) based on average spend and the presence of discounts.
- **Flight Booking Behavior:** (e.g., Never Flown, Primarily Flights) based on the ratio of flight bookings to total trips.
- **Demographic Trip Type:** (e.g., Family Trip, Work Trip) inferred from marital status, presence of children, and number of seats booked.
- **Hotel Stay Behavior:** (e.g., Short Stay, Long Term Guests) based on the average duration of hotel stays.
- **Hotel Booking Behavior:** (e.g., Never Stayed, Primarily Hotels) based on the ratio of hotel bookings to total trips.

7- Value Scoring and Clustering:

Following the rule-based segmentation, a value scoring system was implemented to quantify user behavior across these categories. Numerical scores were assigned to different levels within each behavior category, and these scores were weighted to reflect the relative importance of each behavior. The aggregated value score for each user provided a quantitative measure of their engagement and spending patterns.



K-means clustering, an unsupervised machine learning algorithm, Following PCA (dimensionality reduction), was applied and resulted into four distinct clusters with a silhouette score of 0.53. A further improvement is required and suggested, one way can be doing with feature selection methods.

8- Cluster Characteristics and Perk Recommendations

The analysis of the four identified clusters revealed distinct behavioral patterns and preferences, allowing for the formulation of tailored perk recommendations:

- **Cluster1: (young_traveller/free bird) Independent, mostly without children, frequent travelers prioritizing flights and often needing hotel stays.**

Suggested Perk:

- "Flexi-Flyer" Pass: Discounted one-way flights and last-minute return options with minimal change fees.
- "Urban Explorer": Hotel Credits (discounts on hotels in major city destinations).

Reason: This caters to their frequent travel and potential need for flexible flight arrangements and hotel needs.

- **Cluster 2: (Dreamers) Infrequent travelers with no strong spending patterns.**

Suggested Perk:

- "Welcome Aboard" Introductory Offers, such as a percentage discount on their next booking.
- "Bring a Friend" Rewards could also incentivize them to introduce new customers.

Reason: As infrequent travelers with no strong spending patterns, the initial focus is to encourage them to travel with attractive entry-level offers and build confidence in booking, making it more appealing by sharing the experience.

- **Cluster 3: (Family_travellers) have single traveller but Individuals traveling with children that travel regularly.**

Suggested Perk:

- "Family saver" packages: offering discounts on multiple bookings or bundled deals for families.
- Discounts on family-friendly attractions (e.g., theme parks, museums with children's areas) to offer discounted tickets

Reason: This directly addresses the needs of individuals traveling with children, making these trips more convenient and affordable. Also it enhances trip value by making family-oriented activities more accessible.

- **Cluster 4: (Explorer) primarily flight users, mainly unmarried, but active travelers likely staying 1-2 nights.**

Suggested Perk:

- Destination-based offers: Highlighting attractions or tours, at their likely travel destinations to incentivize flight bookings.
- "Extended Stay Saver": Discounts for hotel stays of 3 nights or more, potentially including value-added benefits like free breakfast.

Reason: These perks aim to make flight and hotel combinations more attractive and cost-effective by acknowledging their broader travel habits, flight booking and preference for short-length hotel stays.

9- Further Recommendations

Building upon this comprehensive analysis, the following recommendations are proposed to further enhance TravelTide's customer retention strategy:

- **Implement dynamic allocation** of suggested perks based on customer clusters.
- **Conduct A/B testing** of different perk variations within each cluster.
- **Chi-Squared** tests to assess the relationship between segments and booking behavior.
- **Personalize marketing** communications like, Tailored email campaigns to highlight relevant perks, and website adaption to showcase appealing perks for relevant groups.
- **Update perk by** direct customer feedback, recommendations based on evolving customer preferences and doing surveys.

10- Conclusion:

Distinct customer segments at TravelTide have been successfully identified through comprehensive analysis. Personalized perks, tailored to each segment's behavior and preferences, offer a strategic approach to enhance customer retention and loyalty, ultimately driving improved engagement and sustained business growth.

11- Links

- [Collab](#)
- [Presentation](#)
- [Presentation Slides](#)
- [Report](#)
- [Tableau dashboard](#)

Submitted by
Nuzhat Jabeen Amna