Customer Segmentation and Reward Strategy for TravelTide

1. Introduction

TravelTide has built a strong travel inventory and search engine but struggles with customer retention. To improve this, the company plans to launch a personalized rewards program. This report outlines a data-driven approach to segment customers and tailor rewards based on user behavior and value.

The analysis is based on session and booking data from January 4, 2023, onwards focusing on users with 8 or more sessions. The goal is to classify users based on booking behavior and spending patterns to optimize reward offerings.

2. Data Preparation & Cleaning

Data Merging & Handling Missing Values

To gain a holistic view of customer behavior, data from four tables (hotels, flights, sessions and users) were merged. The only missing values were found in booking-related columns, which is expected:

- **trip_id** was missing when no trip was booked
- **return_time** was missing for one-way flights

These missing values represent natural gaps in user behavior rather than errors. Additionally, canceled trips were removed, including cases where a trip was both booked and later canceled, ensuring data accuracy.

Data Cleaning & Feature Engineering

- **Fixed Negative Nights**: recalculated trip duration using check-in and check-out dates to correct inconsistencies
- **Derived Age**: calculated from birth date
- **Session Duration**: measured in seconds from session start and end times
- **Trip Distance**: computed using latitude and longitude data for both origin and destination airports, utilizing the Haversine formula
- **Booking Rate Calculation**: determined how often a user books per session to enhance segmentation
- **Aggregated User-Level Data**: retained meaningful features, such as total trips booked, session count and total distance flown
- **Age Buckets & Customer Scoring**: segmented users into meaningful age groups and assigned an engagement-based score to better identify high-value customers

3. Customer Segmentation

Customers were categorized based on booking frequency per session:

- **Low Booker** (rare bookings)
- **Occasional Booker** (sporadic bookings)
- **Frequent Booker** (regular bookings)
- **Super Booker** (almost every session leads to a booking)

Customer Scoring

To enhance segmentation a customer value score (1-4) was assigned based on:

- **Spending on hotels and flights**
- **Booking segment** (higher scores for frequent bookers)
- **Adjustment for low bookers** to ensure an accurate ranking
- **Engagement Adjustments**: users with high session duration received a score boost

This scoring system ensures that high-value customers are properly identified, regardless of booking frequency alone.

4. Exploratory Data Analysis (EDA)

Customer Distribution Across Segments

Customer Value Score by Booking Segment

	Frequent Booker	Low Booker	Occasional Booker	Super Booker
1	0	1321	728	0
2	102	248	1760	0
3	456	0	1203	1
4	103	0	40	36

Customer Value Score (y-axis) | Booking Segment (x-axis)

- The majority of users belong to the **Low Booker** and **Occasional Booker** segments, with relatively few **Super Bookers**.
- Customer Value Scores are concentrated in the **1-2 range**, with fewer high-value customers (Score 4).
- Frequent Bookers tend to have higher value scores, but Super Bookers are rare and highly engaged.

Demographic Insights
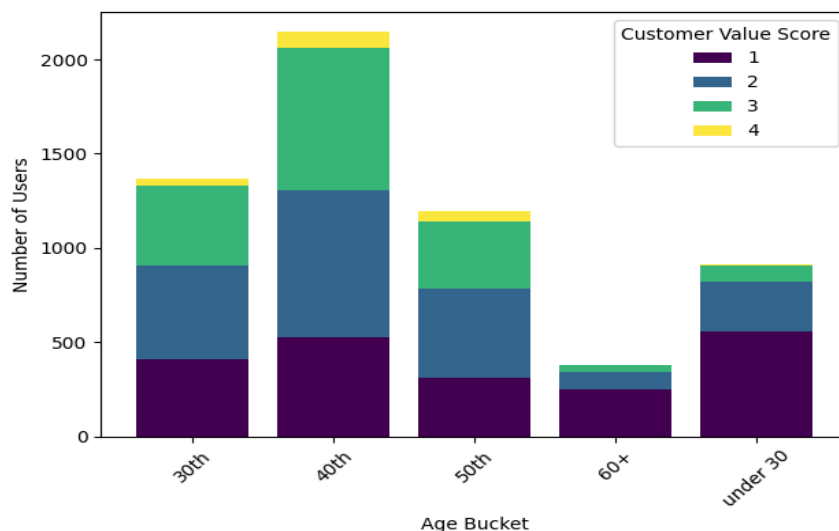
Age and Customer Value

- Users in their **30s and 40s** are the largest groups, while those **under 30 and over 60** are less frequent
- Higher value scores are more common in **30-50-year-olds**, suggesting middle-aged users are more engaged

Behavioral Analysis

- **Session Engagement & Booking Trends:** Users with higher session durations and activity tend to have a higher booking rate
- **Spending Patterns:**
    - Higher scores correlate with higher total spending on flights and hotels
    - Super Bookers and high-value customers contribute disproportionately to revenue

These insights will help tailor the reward strategy to different customer segments.

5. Reward Strategy & Next Steps

Proposed Rewards per Customer Score

Score	Reward	Business Impact
1	10% discount on the next booking (valid for 6 months)	A generous, time-limited incentive designed to trigger immediate action and drive rebooking
2	3% personal discount valid for 1 year	A subtle but lasting benefit that encourages continued usage and loyalty. This extended perk increases the likelihood of repeat purchases and higher total spend over time
3	One-time free breakfast for the entire stay	A cost-efficient but high-perceived-value perk that adds a premium feel to the customer experience, helping improve satisfaction and brand perception without deep discounting.
4	VIP perk – free hotel or flight upgrade	A prestige reward that recognizes high-value customers, boosting long-term loyalty and increasing the chance of word-of-mouth advocacy among top-tier users.

Next Steps & Validation

This segmentation model serves as an initial approach and requires further validation. Future improvements include:

- **A/B Testing**: measure the effectiveness of different reward types
- **Chi-Squared Tests**: assess the relationship between segments and booking behavior
- **User Surveys**: gather direct feedback on preferred rewards
- **Supervised Learning Models**: train a classifier using additional behavioral data for more precise segmentation

6. Conclusion

This analysis provides a data-driven foundation for TravelTide's rewards program. By leveraging customer segmentation and scoring we can tailor incentives to maximize retention and engagement. However, further testing and refinement are necessary to optimize the impact of this strategy.