Travel Tide Customer Segmentation

Introduction
The goal of this project was to create a customer segmentation model for TravelTide to optimize its email marketing by assigning tailored perks to distinct customer groups. The database provided session-level data, allowing for granular analysis. We focused on a cohort of sessions starting after January 4, 2023 and included users with more than 7 sessions, ensuring we analyzed active customers.

Data Cleaning and Feature Preparation
Extensive data cleaning was performed:

- Missing values were addressed and inconsistencies corrected. For instance, errors in trip-related data (e.g., negative or zero values in the nights column) were fixed based on check-in and check-out information.
- Several key features were engineered:
    - **Trip Length:** Derived from the difference between departure and return times for booked flights or taken from the hotel nights if no flight was booked.
    - **Lead Time:** Calculated as the number of days from when the booking (session end) occurred to the start of the trip.
    - **Distance and Discount per Kilometer:** The distance between airports was computed using the Haversine formula, and discount information was normalized by distance.
- Categorical features were encoded into binary or one-hot representations, and numerical features were scaled using standard techniques to ensure consistent contribution during clustering.

Segmentation Approach

Rule-Based Segmentation
We began by segmenting the customers into three clear, rule-based groups:

1. **High-Spending Customers:** Identified by high overall trip expenses; these customers were assigned VIP-Lounge perks.
2. **Inactive or Canceled Customers:** Those who canceled or did not book any trips received Free Cancellation perks.
3. **Discount-Oriented Customers:** Customers using frequent discounts and operating on lower budgets were targeted with Exclusive Discounts.

Hierarchical Clustering
Next, hierarchical clustering was applied on the remaining customers using a focused set of features (total trip cost, trip length, age, and average trip costs). Although five clusters emerged initially, two clusters with similar characteristics were combined to yield four final segments:

- **Cluster 1:** With longer trips and moderate spending, this group was given the Free Night perk.
- **Cluster 2:** Representing average-spending customers, this cluster was assigned the Free Meal perk.
- **Cluster 3:** Consisting of younger, budget-conscious travelers, these customers were offered Free Breakfast.
- **Cluster 4:** Comprising older customers with higher luggage needs, this cluster received the Free Bag perk.

Insights and Recommendations

The segmentation strategy provides a framework for targeted marketing. However, the clusters derived from hierarchical clustering are only moderately separated. This suggests that while the approach is promising, further steps are needed to validate and refine the strategy. Recommendations for future work include:

- **A/B Testing:** Running controlled experiments to gauge the effectiveness of each perk and adjust the segmentation strategy accordingly.
- **Classification Models:** Using customer response data to build predictive models that can further refine the segmentation.
- **Additional Data Collection:** Gathering more detailed customer information—such as whether a traveler is a business or private traveler—could help tailor the perks more precisely.
- **Revisiting Feature Selection:** Exploring additional or alternative features may enhance cluster separation and yield more distinct customer segments.

Conclusion

The project successfully developed a multi-step segmentation strategy that combines rule-based segmentation with hierarchical clustering. Although the current clusters provide a useful basis for targeting perks via email marketing, further testing and data enrichment will be crucial to optimize the approach. Moving forward, TravelTide can use these insights to design more personalized marketing campaigns, ultimately improving customer engagement and retention.