TravelTide_ExecutiveSummary

Data driven insight

Md Erfan - 5. October 2025

🧭 Executive Summary

Project: TravelTide — Customer Segmentation for an Online Travel e-Booking Platform

The **TravelTide** project aims to identify distinct customer segments and design a **personalized rewards program** to enhance user retention, engagement, and lifetime value. Leveraging real user-session and booking data, the analysis integrates both **Rule-Based** and **Machine Learning (ML)** segmentation approaches to provide data-driven recommendations for targeted marketing and perk allocation.

🔍 Project Workflow

The end-to-end workflow followed the full **data analytics lifecycle**:

1. **Data Understanding & Cleaning:**
   The dataset (5,998 users) was cleaned, normalized, and aggregated at the user level through SQL and Python. Missing values were treated systematically, and key behavioral variables such as *conversion rate, session frequency, discount usage,* and *travel patterns* were engineered.

2. **Feature Engineering & RFM/CLTV Modeling:**
   RFM (Recency, Frequency, Monetary) and CLTV (Customer Lifetime Value) scores were computed, producing loyalty and value tiers. This formed the foundation for persona logic and ML clustering.

🧩 Rule-Based Segmentation

Using RFM, CLTV, and behavioral thresholds, customers were categorized into **10 personas** such as *Loyal High*, *Family High Value*, *Discount Hunter*, *Business Flyer*, *Steady Explorer*, and *Casual / Inactive*.

Validation confirmed distinct behavioral and value differences across personas:

- *Loyal High* and *Family High Value* showed the highest CLTV and engagement.
- *Discount Hunters* displayed strong promo responsiveness, suitable for targeted discounts.
- *Business Flyers* had shorter trips and moderate CLTV, ideal for service upgrades.
- *Casual / Inactive* users were identified as re-engagement opportunities.

The personas were then mapped to **personalized perk offers** (e.g., premium upgrades, loyalty bonuses, special discounts), and validated through post-segmentation EDA and geographic breakdowns.

🤖 Machine Learning Segmentation

A parallel ML pipeline was developed to validate and enhance the segmentation:

- **PCA** reduced 29 behavioral and transactional features to key components explaining 90% of data variance.
- **K-Means clustering** (optimal $k=3$, based on silhouette analysis) revealed three distinct user clusters.
- Cluster profiling highlighted clear behavioral contrasts, enabling interpretive naming such as *Loyal High-Value Traveler*, *Discount Hunter*, and *Family Vacationer*.

Visual analyses (PCA scatterplots, heatmaps, and persona summaries) reinforced the robustness and interpretability of the ML-derived segments.

💡 Business Impact & Recommendations

Both segmentation frameworks converge on clear strategic insights:

- **Retention Focus:** Prioritize “Loyal High” and “Family High Value” segments with exclusive perks and loyalty upgrades.
- **Re-engagement Strategy:** Offer time-limited discounts or personalized content to “Casual / Inactive” users.
- **Discount Optimization:** Target “Discount Hunters” with high-ROI campaigns to increase conversion without excessive cost.
- **Regional Targeting:** Concentrate efforts on high-CLTV geographies (e.g., USA, Canada) and popular destinations.

📦 Deliverables

- Final user-level dataset: `TravelTide_rulebased_persona_final.csv`
- ML cluster summary and visualization notebooks
- Persona-perk mapping table for marketing deployment
- Executive dashboard visuals summarizing segmentation outcomes

Conclusion

The project successfully demonstrates how combining **data-driven machine learning** and **business-aligned rule-based segmentation** can create actionable customer personas. These insights empower TravelTide to deploy **targeted marketing**, **personalized perks**, and **optimized loyalty strategies**, ultimately driving **higher retention and revenue growth**.