Travel Tide: Customer Analytics & Segmentation Project

Prepared by: Shraddha Ingle
Date: 06-10-2025

Executive Summary

This project analyzes customer behavior in the travel domain using user, session, flight, and hotel data. We built an integrated analytical pipeline to perform feature engineering, customer segmentation, and lifetime value analysis.

Key Objectives:

- Understand customer engagement, loyalty, and spending patterns.
- Build RFM (Recency, Frequency, Monetary) segmentation to categorize customers.
- Estimate Customer Lifetime Value (LTV) for high-value user targeting.
- Apply ML clustering (KMeans + DBSCAN) to uncover distinct customer personas.

Highlights:

- **Feature Engineering: Created 30+ features covering engagement, booking behavior, spending, demographics, and travel patterns.**
- **RFM Segmentation: High vs Low RFM groups identified, with strong differences in recency and monetary value.**
- **LTV Analysis: High-value customers contribute ~70–80% of total revenue.**
- **ML Pipeline: Standardization → PCA → Clustering. Optimal 4 clusters identified.**

- **Personas Identified:**

    - **Frequent Flyers: High spending, low churn, loyal.**
    - **Seasonal Travelers: Book during peak holidays, medium frequency. Discount Seekers: Highly responsive to promotions, low LTV**
    - **Inactive Users: High recency (long inactivity), low engagement.**

- **Business Impact: Enables targeted marketing campaigns and improves retention and revenue by focusing on high-LTV segments.**

Visualization Note: Charts and graphs were created for feature distributions, RFM segmentation, LTV analysis, and clustering results to support insights.

Detailed Report

1. Data Integration

- **Merged four tables (users, sessions, flights, hotels) on `user_id` and `trip_id`.**
- **Applied filtering: sessions after 2023-01-04, users with >7 sessions.**
- **Created final_table with enriched user-level features.**

Visualization Placeholder: A data integration diagram summarizing table joins and filtered data.

---

2. Feature Engineering

Created advanced features across eight domains:

1. **Engagement: Session duration, page click rate, days since last booking**
2. **Booking Behavior: Trip duration, booking gaps, cancellation rates, total bookings**
3. **Travel Patterns: Flight distance (Haversine), route analysis, domestic vs international**
4. **Loyalty & Retention: Loyalty score, high-value flag, churn proxy**
5. **Demographics: Age, age groups, marital & family status**
6. **Hotel Preferences: Average nights/rooms per booking**
7. **Spending Patterns: Total spending, per booking spending**
8. **Seasonality: Booking month, seasonal demand trends**

Visualization Placeholder: Histograms and scatterplots for engagement, spending, and booking behavior features.

---

3. RFM Segmentation

- **Recency: Days since last booking**
- **Frequency: Number of total bookings**
- **Monetary: Total spending per customer**

- **Segmented customers into High RFM (loyal, frequent spenders) and Low RFM (inactive or low-value)**

Visualizations: RFM scatter plots, segment bar charts

---

4. Lifetime Value (LTV) Analysis

- **Estimated LTV: `avg_spending_per_booking × booking_frequency × retention`**
- **Observed Pareto effect (80/20 rule): Top 20% of customers contribute majority revenue**

Visualizations: LTV distribution charts, high vs low LTV boxplots

---

5. Machine Learning Clustering

- **Pipeline: StandardScaler → PCA (2D) → KMeans / DBSCAN**
- **Elbow Method: Optimal K = 4 clusters**

Cluster Personas:

- **Frequent Flyers: High distance traveled, high spending**
- **Seasonal Vacationers: Medium frequency, clustered around holiday months**
- **Discount Seekers: Sensitive to promotions, lower average revenue**
- **Inactive Users: High recency, low engagement**

Visualizations: PCA scatter plots with clusters, Elbow curve, DBSCAN density map

---

6. Business Insights & Recommendations

- **Target High RFM & High LTV customers with loyalty programs**
- **Re-engage Low RFM users with personalized promotions**
- **Optimize discounts for price-sensitive clusters**
- **Use travel seasonality insights to plan peak-season campaigns**

Placeholder for visualizations: *Marketing plan diagram, customer persona chart.*