Project Summary: Travel_Tide Project

Introduction / Background:

The **Travel_Tide** project enhances customer retention by analyzing user behavior to develop a personalized rewards program. As an e-booking platform facing retention challenges, segmentation-based insights enable targeted marketing, improving engagement and customer satisfaction.

Objectives:

This analysis aims to segment **Travel_Tide** customers based on behavior to enhance loyalty and business performance. It focuses on:

- Identifying customer personas through behavioral traits.
- Developing targeted perks and recommendations to improve engagement.

Methodology:

The analysis employs K-Means clustering, an unsupervised machine learning algorithm, to group customers based on key behavioral attributes. The following steps were undertaken:

1. Data Collection & Cleaning:

- Merging relevant tables on session-level granularity.
- Filtering cohorts after January 4, 2023 for user sessions count greater than 7.
- Handling missing or inconsistent data entries.

2. Exploratory Data Analysis (EDA):

- Analyzing existing columns for identifying key patterns and trends in user engagement.

3. Feature Engineering:

To enhance the customer segmentation analysis, we engineered key features across multiple aspects of customer behavior ensuring a more refined clustering process.

A. Demographics

- Age: Age of customers.
- Gender: Gender of the user.
- Marital Status: Marital status.
- Has Children: Indicating whether a user has children.

B. User Engagement

- Page Click Rate: Number of page clicks per user session in minutes.
- Days Since Last Booking: Average time since a user's last booking.

C. Discount Responsiveness

- Discount Responsiveness: Average responsiveness to promotional offers.

D. Booking Behavior

- Checked Bags: Average number of checked bags per user.
- Seats Booked: Mean number of seats booked per transaction.
- Number of Flights Booked: Total count of flight bookings per user.
- Number of Hotels Booked: Total count of hotel reservations per user.
- Total Nights Booked: Sum of nights booked across all hotel stays.
- Average Trip Duration: Mean trip length in days.

E. Spending & Loyalty

- Total Spending: Cumulative amount spent by the user.
- Average Spending per Booking: Mean spending per booking transaction.
- Average Loyalty Score: Mean loyalty program score indicating brand engagement.

F. Travel Preferences

- Average Flight Distance: Mean travel distance per flight.
- Total Flight Distance: Sum of all flight distances traveled by the user.

G. Seasonality & Trends

- Most Common Booking Month: Mode of the months in which a user books the most frequently.
- These engineered features provide valuable insights for clustering, allowing Travel_Tide to categorize users effectively and tailor marketing strategies accordingly.

4. Clustering Analysis:

A. Data Scaling:

Standardize the customer behavioral features to ensure all variables are on a similar scale, which is crucial for K-Means clustering to perform optimally.

B. Principal Component Analysis (PCA):

Apply PCA to reduce the dimensionality of the dataset while retaining the most important variance. This simplifies the data without losing critical information, improving the efficiency of the clustering process.

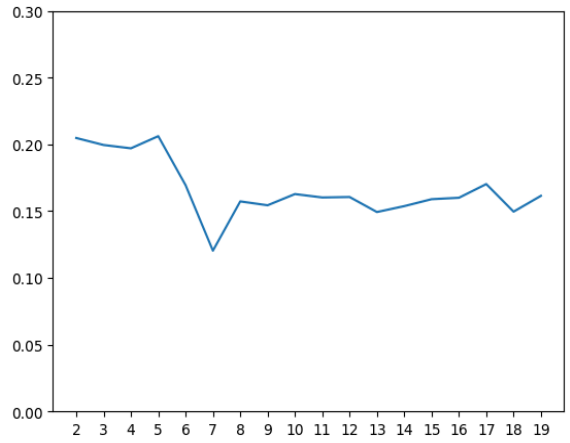
C. Evaluating Cluster Validity:

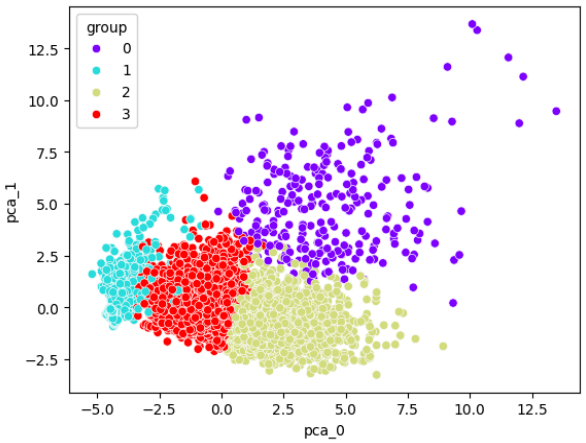
Use the silhouette score to evaluate the effectiveness of the clustering. This metric measures how well-separated and cohesive the customer segments are, guiding the selection of the optimal number of clusters.

D. Applying K-Means Clustering:

Perform K-Means clustering on the transformed data, segmenting customers based on their behavioral attributes. Experiment with different cluster counts to identify the most meaningful customer segments.

E. Visualization & Interpretation:





Key findings:

Characteristics of Group 0: Family Travelers

- **Moderate Travelers with Low Engagement**: This group books fewer flights and hotels, spends modestly, and has a moderate loyalty score of 0.45, indicating low engagement with the platform and a moderate response to discounts.

- **Family-Oriented**: Significant portion (87.8%) females with 43% married and 27% having children, many users likely travel with family. Offering **family-focused perks** would appeal to this segment.

- **Perk Assigned**:

  A **"Family Discount"** that provides a discount for bookings involving multiple **travelers (e.g., 10-15% off for bookings of 2 or more people)**.

Characteristics of Group 1: Loyal Deal Seekers Travelers

- **Moderate Engagement & Booking Frequency:** Users interact moderately with the platform (page click rate of 7.93) but book infrequently, averaging 3.88 flights and 13.17 nights with a trip duration of 4.26 days.

- **Diverse Travel Distances:** Users travel a wide range of distances, from short domestic trips to long-haul international flights, with an average flight distance of 2,141 km.

- **Moderate Spending:** Total spending averages 4,068.59 USD, with each booking typically costing 65.66 USD, indicating a moderate spending pattern.

- **Strong Loyalty:** Despite booking less frequently, users maintain a high loyalty score of 0.94, suggesting strong platform loyalty.

- **Perk Assigned:**

  "Loyalty Upgrade" that offers complimentary upgrades (e.g., seat or room upgrades) for their next booking to encourage higher engagement.

Characteristics of Group 2: Smart Saver Travelers

- **Moderate Engagement & Booking Patterns**: Average page click rate of 8.00 and 362 days since the last booking, with low interest in checked bags and seat selection. They book an average of 3.92 hotels, including some longer stays.

- **Discount & Spending**: Only 10.4% are responsive to discounts, indicating they value more than just promotions. Average total spending is $774.53, with $71.86 per booking, reflecting a budget-conscious but regular traveler.

- **Loyalty & Travel Preferences**: Moderate loyalty score of 0.07, preferring shorter trips with an average flight distance of 12.34 miles, suggesting regional travel with infrequent but cost-sensitive bookings.

- **Assigned Perk**:

  "**Budget-Friendly Weekend Getaways**" or "**Discount on Short Trips**"

Characteristics of Group 3: Luxury Travelers

- **User Engagement**: Group 3 has a moderate engagement with a page click rate of 7.35, but their bookings are less frequent (average days since last booking: 421 days), likely for longer, planned trips.

- **Spending & Booking Patterns**: They spend an average of $9,880.46 per user, with $244.55 per booking, showing moderate spending. They book more checked bags and seats (indicating family/group travel), and often book multi-night stays (average of 2.91 hotels).

- **Travel Preferences & Loyalty**: Group 3 travels long distances (average flight distance: 7,334 miles) for international or special event trips. They have a moderate loyalty score of 0.58, suggesting some level of brand loyalty but less engagement than more frequent users.

- **Assigned Perk:**

  "VIP Travel Experience" or **"Premium Travel Upgrades"**

Recommendations:

- **Recommendation 1:** Optimize platform UX/UI for mid-tier users to encourage more frequent bookings and spending. Enhancing search filters and booking processes can drive conversions.

- **Recommendation 2:** Continuously monitor customer behavior and refine segmentation models for improved personalization. Utilizing real-time data can improve marketing strategies.

- **Recommendation 3:** A/B test different engagement strategies to determine the most effective approach for each segment.

Conclusion:

This analysis successfully segmented Travel_Tide customers based on behavioral attributes, identifying distinct personas to guide targeted engagement strategies. By implementing personalized rewards, re-engagement campaigns, and loyalty-driven initiatives, Travel_Tide can enhance customer retention and maximize revenue. Future work can explore **real-time clustering models** and **predictive analytics** to dynamically adapt marketing efforts to evolving customer behaviors.