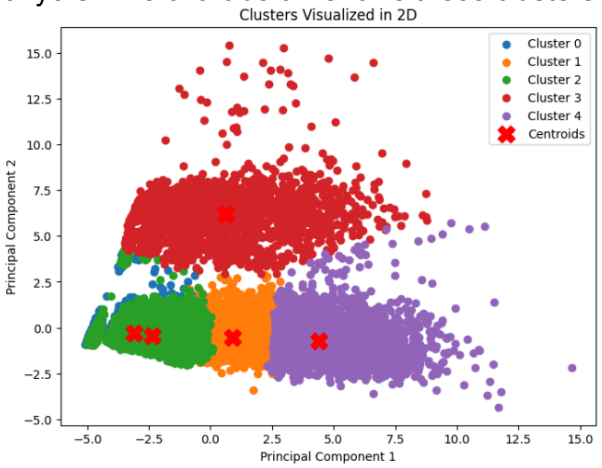
EXECUTIVE SUMMARY

Introduction: TravelTide, a leading e-booking startup, is developing a rewards program to attract and retain customers. The analysis focused on active users after 01/04/2023 with over seven sessions, targeting engaged and recent customers for loyalty initiatives. This work was developed by Diana Torres, Data Analyst/Scientist.

Objectives: The analysis aimed to identify customer segments among recent, high-activity users that align with TravelTide's user base and recommend personalized perks for each segment to maximize engagement and drive rewards program sign-ups.

Methodology: The analysis began with Exploratory Data Analysis (EDA) in SQL and Python to understand customer behavior, followed by feature engineering to create metrics like booking frequency, spending patterns, and demographic insights. Principal Component Analysis (PCA) was used to enhance the performance of the K-means clustering algorithm, which identified five distinct customer segments based on demographics, activity, and spending. Each cluster was then characterized to align customer preferences and behaviors with tailored perks.

Key Findings: Five distinct clusters were identified, ranging from budget-conscious individuals to frequent flyers. The chart below shows these clusters.



Recommendations: Implement tailored perks for each segment, develop a loyalty program, design targeted communication strategies, monitor and iterate.

Cluster 0 *(Budget-Conscious Singles):* **Free Hotel Meal**. Enhances perceived value for low spenders.

Cluster 1 *(Frequent Flyers):* **No Cancellation Fees**. Supports their dynamic and high-frequency travel habits.

Cluster 2 (*Family Travelers)*: **1 Night Free Hotel with Flight**. Appeals to families requiring extended stays.

Cluster 3 *(Adventure Seekers)*: **Exclusive Discounts**. Encourages loyalty among mid-range spenders.

Cluster 4 (*Occasional Travelers*): **Free Checked Bag**. Reduces barriers for cost-conscious infrequent travelers.

DETAILED REPORT

This report details the methodology, findings, and recommendations resulting from a thorough segmentation analysis. Annex 1 contains the schema of the analyzed data. Annex 2 contains the detailed characterization of the clusters by the subdivision found by the EDA after the clustering by K-Means.

Methodology

Data Preparation and Feature Engineering

- *Data Filters:* Only customers active after 01/04/2023 and with more than seven sessions were included to focus on recent, engaged users.
- *Metrics Created:* A total of 42 features were engineered, capturing demographics, activity metrics, reservations, and spending behaviors.

Exploratory Data Analysis (EDA)

- The analyses were conducted in Python and SQL using univariate methods, such as histograms, and bivariate methods, such as bar plots and correlation matrices. Exploratory data analysis (EDA) was performed on the initial data set (see Annex 1) and on the created metrics (42 new columns after feature engineering).
- Reduced features to 31 based on statistical relevance and correlation to ensure meaningful insights.

Segmentation

- *PCA Analysis:* Reduced to 12 components preserving 91.5% variance.

- *K-Means Clustering:* Optimal clustering determined using the Elbow Method and Silhouette Score. Despite slightly better metrics for 2 clusters, 5 clusters were selected for richer interpretability (WCSS = 309278.6, Silhouette Score = 0.16).

Post-Clustering Analysis

The EDA was revisited by grouping the data into clusters and focusing on the following: demographics and family status, activity metrics, flight reservations, hotel reservations, and spending behavior.

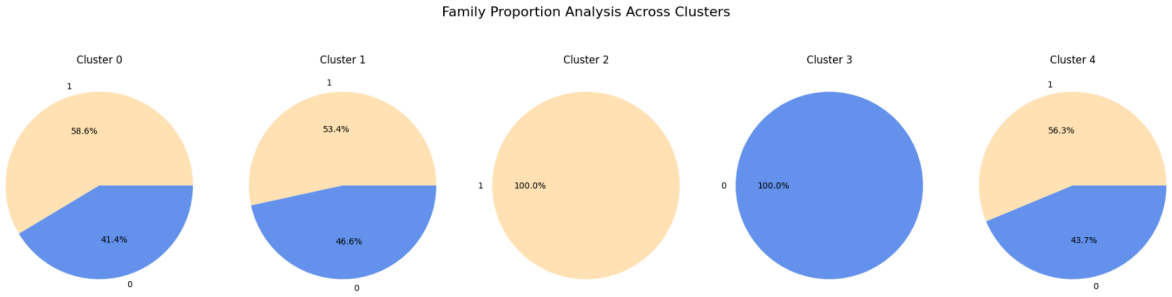
Key Findings

Demographics and Family Status

- **Gender:** Predominantly female across all clusters (73%-80%).

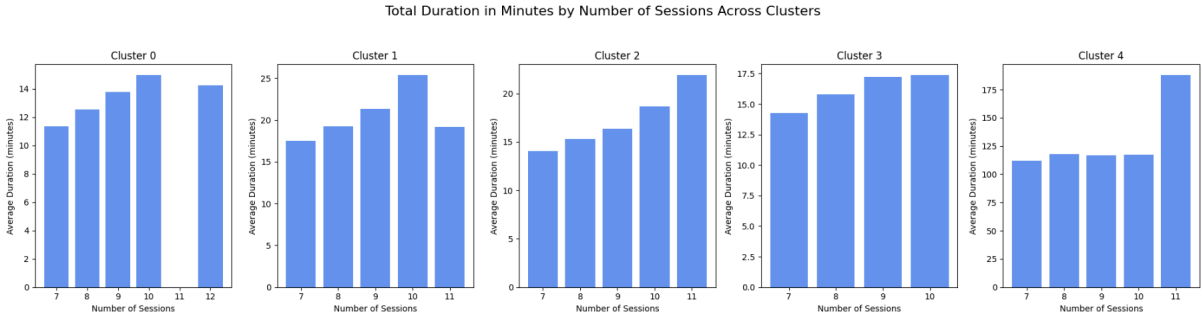
- **Family Status:** Clusters 2 and 4 show higher family-related indicators, while Cluster 3 comprises individuals without families.

- **Age Groups:** Clusters 1, 2, and 4 feature an older demographic (36-55 years), while Clusters 0 and 3 skew younger.



Activity Metrics

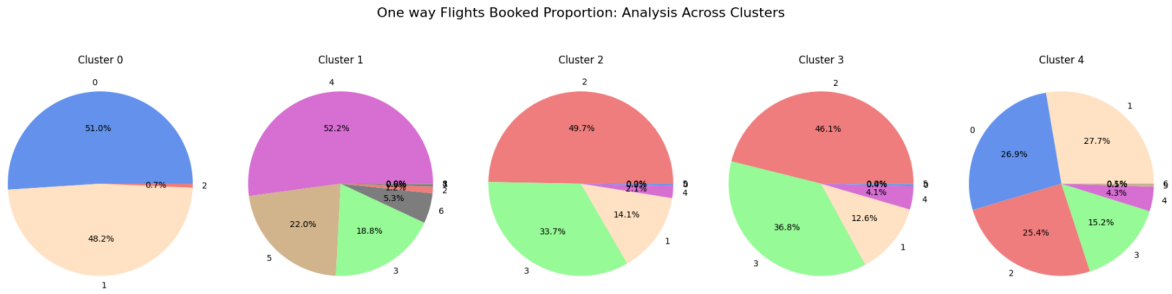
- Clusters 1 and 2 are the most active, with multiple trips and sessions, while Clusters 0 and 4 are less frequent travelers. High session-to-trip conversion rates were observed in Clusters 1 and 2.



Flight and Hotel Reservations

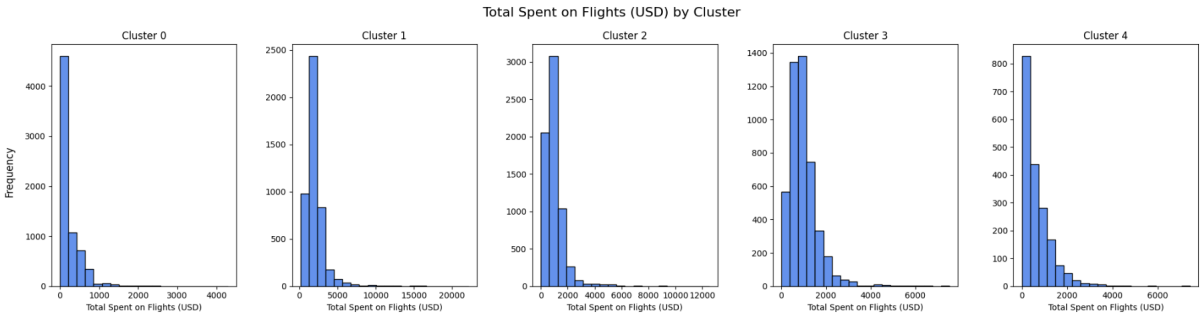
- **Flights:** Cluster 1 leads in flight bookings (4-6 trips annually), while Clusters 0 and 4 book fewer flights. Checked bag preferences also align with these trends.

- **Hotels:** Cluster 1 books the most hotels and nights, followed by Clusters 2 and 3. Cluster 4's booking behavior is minimal but includes notable discounts.



Spending Behavior

- Clusters 1 and 2 exhibit higher spending on flights and hotels (1000+ USD), while Clusters 0 and 4 are more budget-conscious.



Recommendations

The following table provides a characterisation of each group, along with the suggested perk and its justification.

CLUSTERS	NAME	PERK	CHARACTERISTICS
Cluster 0	*Budget-Conscious Singles*	**Free Hotel Meal**	Majority are low spenders (flights: 0-200 USD; hotels: 0-83 USD). Offering a free meal adds value without raising costs significantly and matches their low budget.
Cluster 1	*Frequent Flyers*	**No Cancellation Fees**	High trip frequency (4-6 trips/year) and significant spending (flights: 1000-3000 USD). Removing cancellation fees supports their dynamic travel needs.
Cluster 2	*Family Travelers*	**1 Night Free Hotel with Flight**	High booking frequency for families (2-3 hotels with 7-11 nights on average). This perk aligns with their need for accommodation flexibility.
Cluster 3	*Adventure Seekers*	**Exclusive Discounts**	High activity rate (2-3 flights, 2-3 hotels) with mid-range spending (flights: 400-1200 USD). Exclusive discounts on packages encourage loyalty and repeat bookings.
Cluster 4	*Occasional Travelers*	**Free Checked Bag**	Moderate activity (1-2 flights/hotels) and spending primarily in lower ranges (flights: 0-400 USD, hotels: 0-100 USD). A checked bag reduces their cost concerns.

Conclusion

This segmentation study provides actionable insights into TravelTide's customer base, aligning the rewards program with user preferences to boost engagement. By leveraging these data-driven recommendations, TravelTide can solidify its market position and foster long-term customer loyalty.

ANNEXES

Annex 1: Schema

Below you can find the schema of the analyzed data.

Annex 2: Segmentation Deep Analysis Detailed

Below you can find the detailed facts and characterization of the clusters by the subdivision found by the EDA after the clustering by K-Means.

Demographics and Family Status

CLUSTERS	FACTS
Cluster 0	• 80,4% are female • 58,6% are families • 65,2% do not have children

	● 60,7% are married ● **age group:** ○ 22,5% are 18-25 ○ 21,1% are 26-35 ○ 19,1% are 36-45
Cluster 1	● 79,4% are female ● 53,4% are families ● 73,7% do not have children ● 59,7% are not married ● **age group:** ○ 47,2% are 36-45 ○ 37,7% are 46-55
Cluster 2	● 80,8% are female ● 100% are families ○ 77.5% are married ○ 56,5% has children ● **age group:** ○ 40,4% are 46-55 ○ 35,8% are 36-45
Cluster 3	● 79,9% are female ● 100% are not families ● **age group:** ○ 39,6% are 36-45 ○ 26,2% are 26-35 ○ 21,5% are 46-55
Cluster 4	● 73,1% are female ● 56,3% are families ● 68,3% do not have children ● 59,1% are not married ● **age group:** ○ 30,9% are 36-45 ○ 25,6% are 46-55 ○ 18,8% are 26-35

- The top cities for the 5 clusters are New York, Los Angeles, Chicago, Toronto and Houston.

Activity Metrics

CLUSTERS	**FACTS**

Cluster 0	● 58,1% booked 1 trip ● 29% did not book trips ● 12,8% booked 2 trips ● 99,3% did not canceled
Cluster 1	● 50,1% booked 4 trips ● 33,3% booked 5 trips ● 9,6% booked 6 trips ● 99,2% did not canceled
Cluster 2	● 44,6% booked 3 trips ● 42,6% booked 2 trips ● 10,9% booked 4 trips ● 99,9% did not canceled
Cluster 3	● 45% booked 3 trips ● 39,5% booked 2 trips ● 13,7% booked 4 trips ● 99,9% did not canceled
Cluster 4	● 27,2% booked 2 trips ● 26,5% booked 3 trips ● 20,3% booked 4 trips ● 18% booked 1 trip ● 98,1% canceled 1 trip

- At least 70% in each cluster have a max 7 number of sessions, At least 18% in each cluster have a max 8 number of sessions.

Flight Reservations

CLUSTERS	**FACTS**
Cluster 0	● 51% did not book flight ● 48,2% booked 1 flight ○ and median of 1 seat(1 per flight) ○ median of 0 checked bags ● 86,8% has not **discount** on flights
Cluster 1	● 52,2% booked 4 flights ○ 4 seats(1 per flight) ○ 2 checked bags ● 22% booked 5 flights ○ 6 seats ○ 3 checked bags ● 18,8% booked 3 flights ○ 4 seats ○ 2 checked bags ● 48,4% has not **discount** on flights ● 38% has 1 discount on flights

Cluster 2	● 49,7% booked 2 flights ○ 2 seats (1 per flight) ○ 1 checked bag ● 33,7% booked 3 flights ○ 3 seats (1 per flight) ○ 2 checked bags ● 14,1% booked 1 flight ○ 1 seat (1 per flight) ○ 1 checked bag ● 66,2% has not discount on flights ● 29,5% has 1 discount on flights
Cluster 3	● 46,1% booked 2 flights ○ 2 seats (1 per flight) ○ 1 checked bag ● 36,8% booked 3 flights ○ 3 seats (1 per flight) ○ 2 checked bags ● 12,6% booked 1 flight ○ 1 seats (1 per flight) ○ 1 checked bag ● 64% has not discount on flights ● 30,9% has 1 discount on flights
Cluster 4	● 27,6% booked 1 flight ○ 1 seat (1 per flight) ○ 0 checked bag ● 26,9% did not book a flight ● 25,4% booked 2 flight ○ 2 seats (1 per flight) ○ 1 checked bag ● 73,9% has 2 discount on flights ● 21,3% has 3 discount on flights

Hotel Reservations

CLUSTERS	**FACTS**
Cluster 0	● 54,1% booked 1 hotel ○ avg 5 nights booked ○ avg 1 room booked ● 35,8% did not book a hotel ● 87,5% have no discounts in hotel ● 11,6% have 1 discount in hotel
Cluster 1	● 47% booked 4 hotels ○ avg 13 nights booked ○ avg 5 rooms booked ● 23,4% booked 5 hotels

	○ avg 17 nights booked ○ avg 6 rooms booked ● 20,5% booked 3 hotels ○ avg 10 nights booked ○ avg 4 rooms booked ● 48,7% have no discounts in hotel ● 37,4% have 1 discount in hotel
Cluster 2	● 46,8% booked 2 hotels ○ avg 7 nights booked ○ avg 2 rooms booked ● 35,4% booked 3 hotels ○ avg 11 nights booked ○ avg 4 rooms booked ● 10,6% booked 1 hotel ○ avg 4 nights booked ○ avg 1 rooms booked ● 67,4% have no discounts in hotel ● 28,8% have 1 discount in hotel
Cluster 3	● 44% booked 2 hotels ○ avg 7 nights booked ○ avg 2 rooms booked ● 37,5% booked 3 hotels ○ avg 11 nights booked ○ avg 4 rooms booked ● 10,4% booked 1 hotel ○ avg 4 nights booked ○ avg 1 rooms booked ● 67,1% have no discount in hotel ● 29,1% have 1 discount in hotel
Cluster 4	● 28,5% booked 1 hotel ○ avg 12 nights booked ○ avg 3 rooms booked ● 28,2% booked 2 hotels ○ avg 15 nights booked ○ avg 4 rooms booked ● 21,3% did not book hotels ● 16,8% booked 3 hotels ○ avg 18 nights booked ○ avg 6 rooms booked ● 74,4% have 2 discounts on hotels ● 21,5% have 3 discounts on hotels

Spending Behavior

CLUSTERS	**FACTS**
Cluster 0	● 66,7% spent between 0-200 USD on flights ● 15,6% spent between 200-400 USD on flights ● 10,3% spent between 400-600 USD on flights

	• 51,9% spent between 0-83 USD on hotels • 19,4% spent between 83-166 USD on hotels • 11,4% spent between 166-249 USD on hotels
Cluster 1	• 53,1% spent between 1000-2000 USD on flights • 21,4% spent between 0-1000 USD on flights • 18,1% spent between 2000-3000 USD on flights • 25,8% spent between 400-600 USD on hotels • 23,3% spent between 600-800 USD on hotels • 15,8% spent between 800-1000 USD on hotels • 15,2% spent between 200-400 USD on hotels
Cluster 2	• 46,5% spent between 667-1334 USD on flights • 31% spent between 0-667 USD on flights • 15,6% spent between 1334-2000 USD on flights • 23,3% spent between 284-426 USD on hotels • 22% spent between 142-284 USD on hotels • 17% spent between 426-568 USD on hotels • 11,7% spent between 0-142 USD on hotels
Cluster 3	• 29,3% spent between 800-1200 USD on flights • 28,6% spent between 400-800 USD on flights • 15,8% spent between 1200-1600 • 12% spent between 0-400 USD on flights • 24,6% spent between 167-334 USD on hotels • 24,5% spent between 334-500 USD on hotels • 17,4% spent between 500-667 USD on hotels • 13,8% spent between 0-167 USD on hotels
Cluster 4	• 43,9% spent between 0-400 USD on flights • 23,2% spent between 400-800 USD on flights • 14,8% spent between 800-1200 USD on flights • 8,9% spent between 1200-1600 USD on flights • 33,1% spent between 0-100 USD on hotels • 14,9% spent between 100-200 USD on hotels • 13,4% spent between 200-300 USD on hotels • 10,8% spent between 300-400 USD on hotels