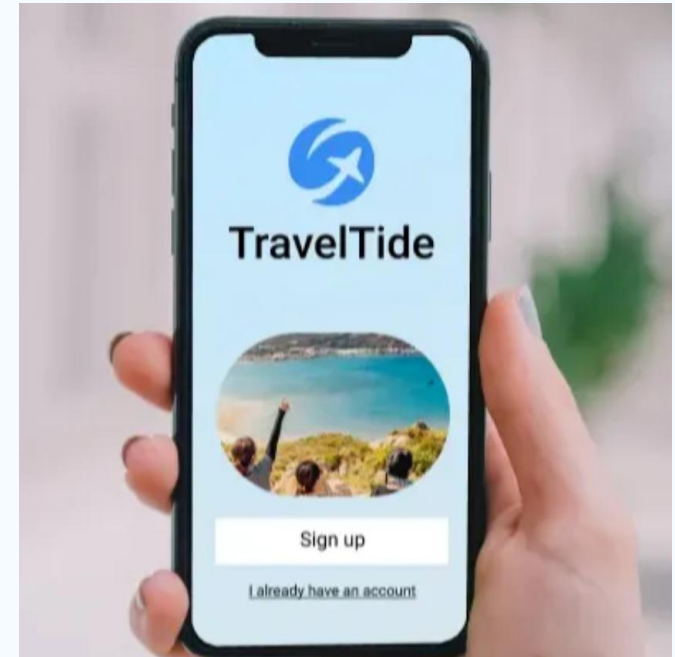


# Travel Tide Customer Segmentation & Personalized Perks

**Jawad Mofleh**

**06.02.2026**



# Introduction



**Comprehensive customer segmentation**

**Using behavioral, transactional, and demographic data**

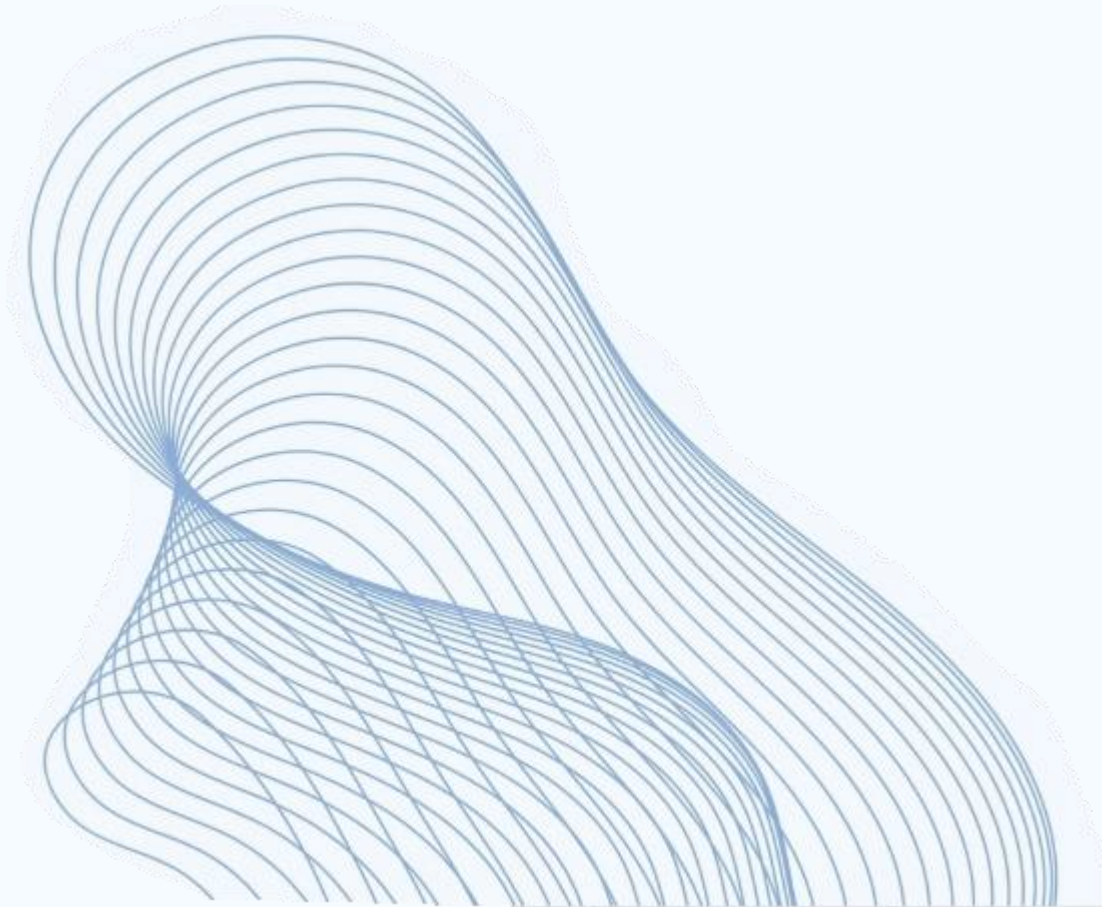
**Goal is to enable personalized perk assignment**

**Data obtained from PostgreSQL( sessions, users, flights, and hotels )**

**Criteria: user had session January 4, 2023 and users with more than seven sessions**

# Objectives

**Assigning perks to eligible customers**



**One Night Free Hotel**

**With Flight Free Hotel Meal**

**Free Cancellation**

**Free Checked Bag**

**Exclusive Discounts**



# Process after data cleaning

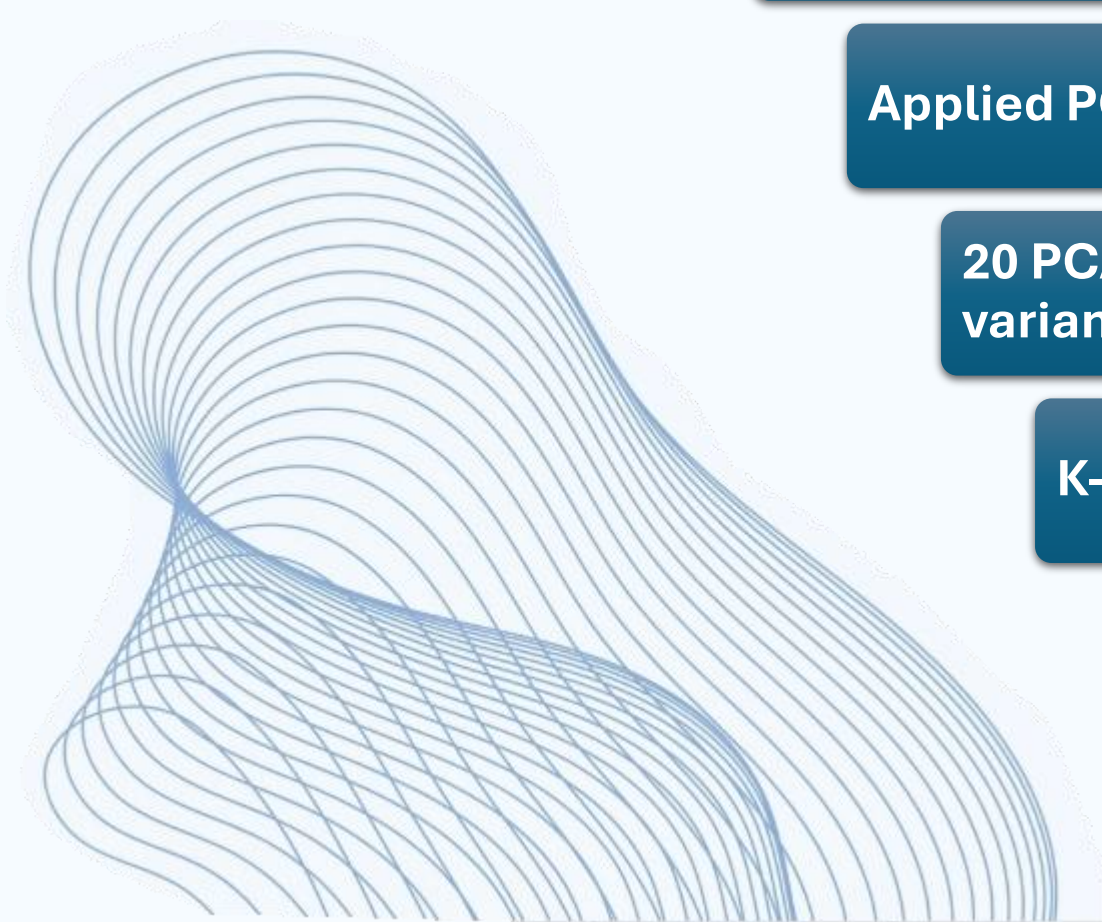
Thirty-seven user-level features engineered

Applied PCA to reduce dimensionality

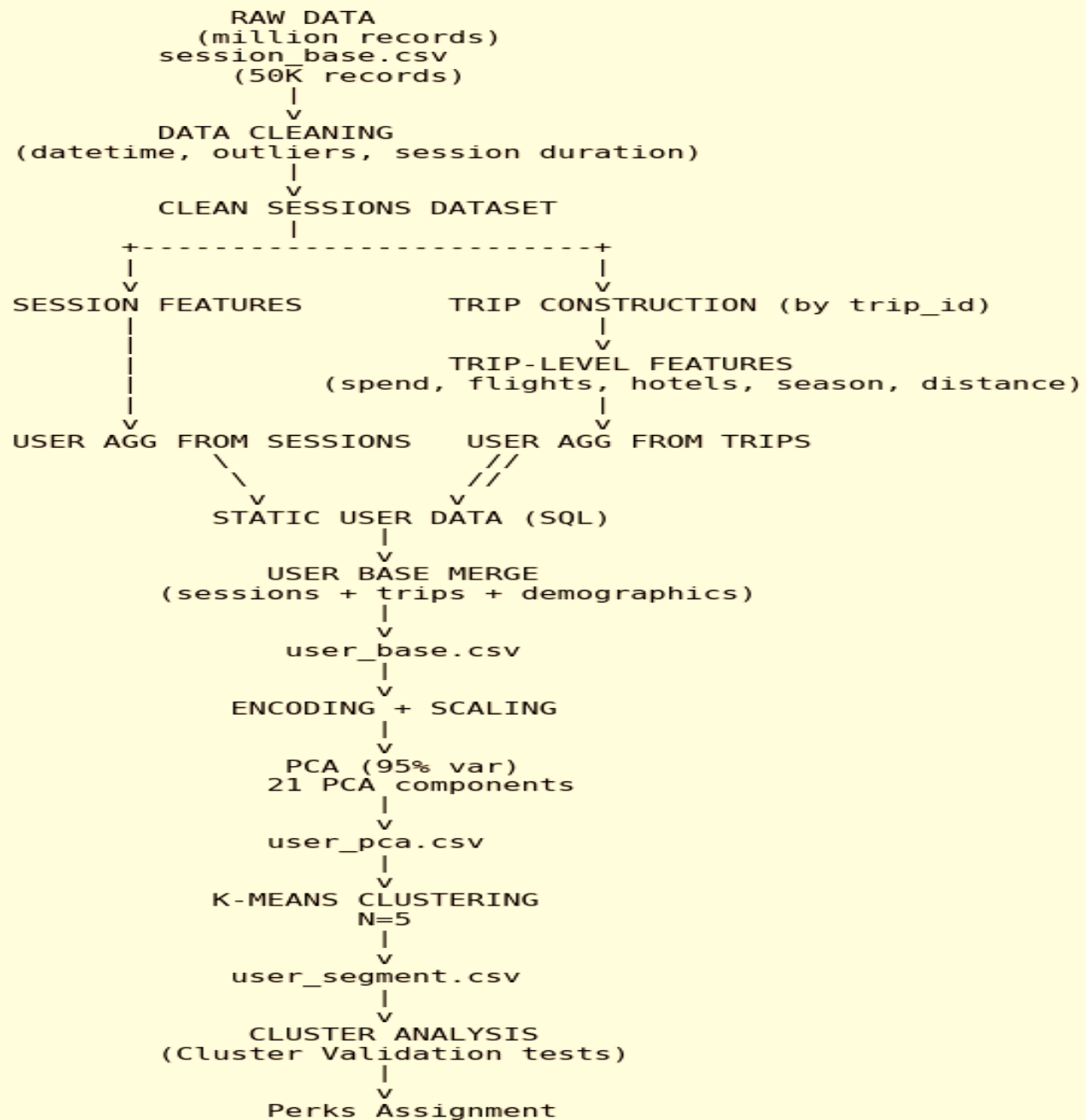
20 PCA components retained 95% of variance

K-Means clustering was applied ( $n=5$ )

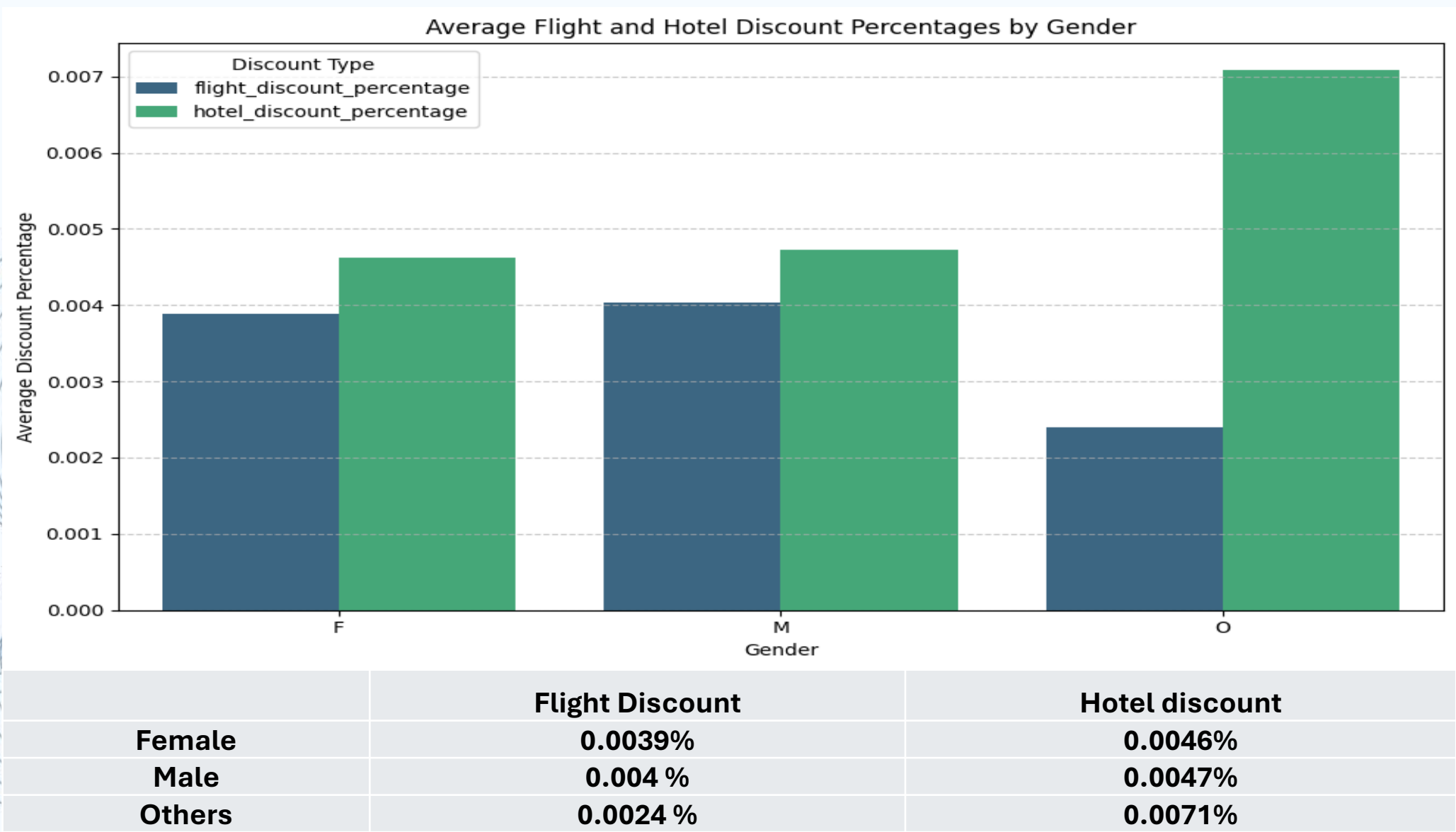
Cluster validation tests ( statistical validation of clusters) applied to ensure validity of clusters.



# Data Pipeline Overview



# Average flight and Hotel discount %



# Customer Bookings Decisions



**75% had flight and hotel booking**

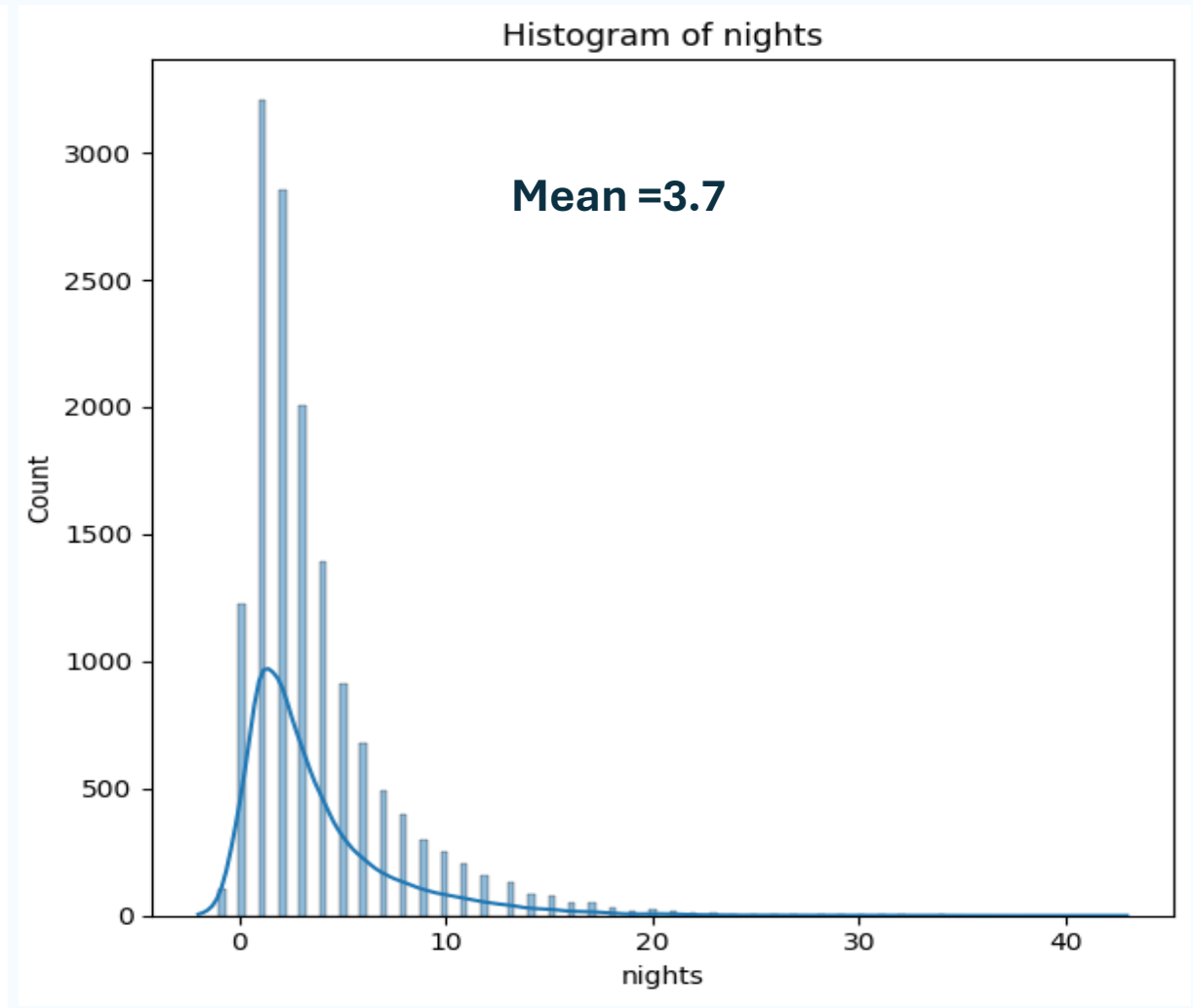
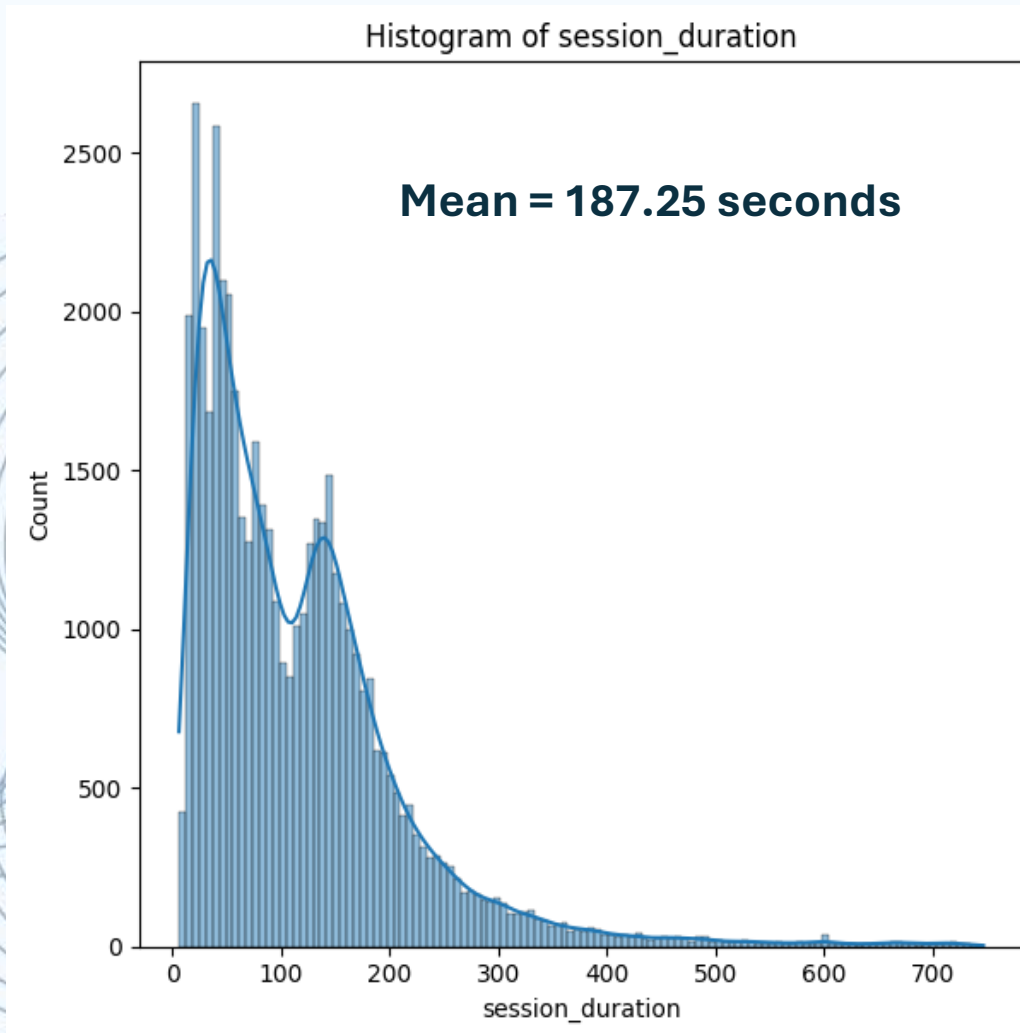
**14.8% only Hotel booking**

**11% only Flight booking**



# Histogram of session duration and nights

after outlier clipping





# Key Distributions

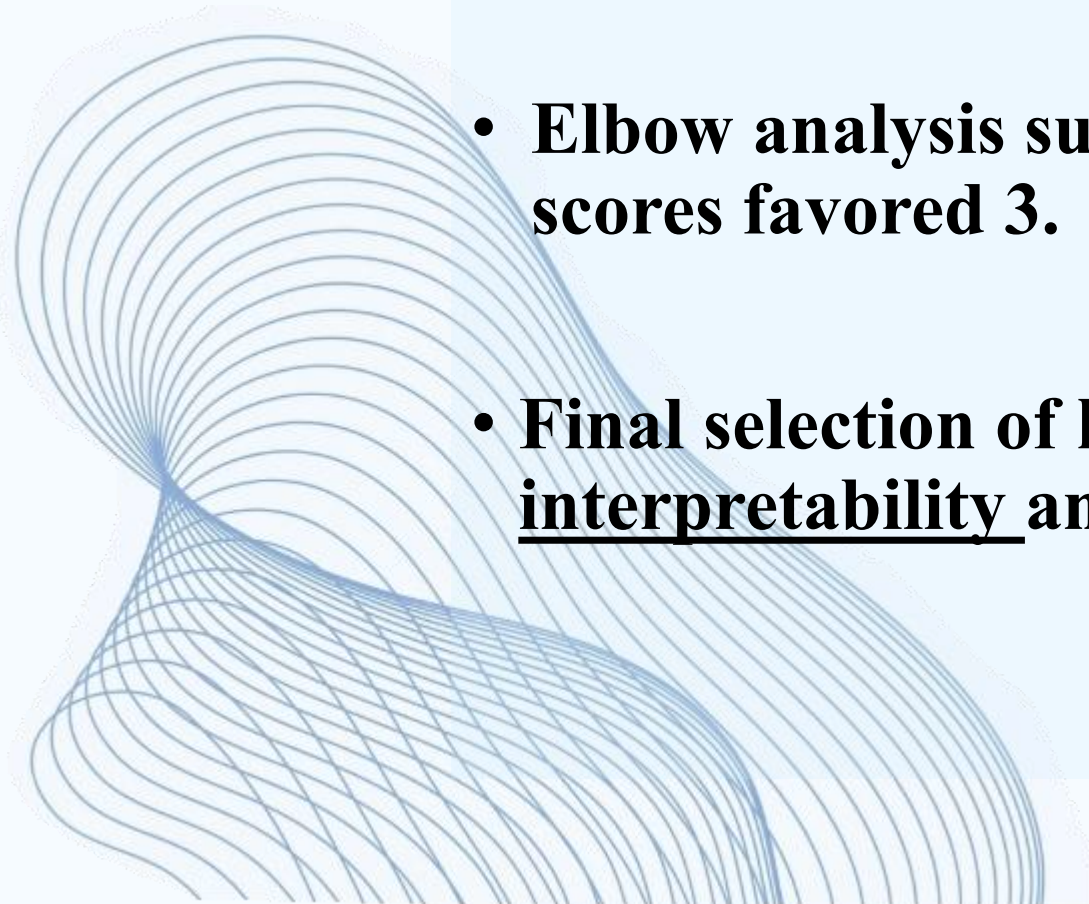
- **Cancelled Trips: 610 (3.9% cancellation rate)**
- **User Age Range: 20-87 years (Mean: 44.3 years)**
- **Page Clicks: Heavily right-skewed (most users < 50 clicks/session)**
- **Session Duration: Bimodal distribution with peaks at 0-5 minutes and 30+ minutes**
- **Booking Patterns: 75% of trips include both flights and hotels**
- **Discount Utilization: Hotel discounts more common than flight discounts**
- **Seasonal Patterns: Clear peaks in winter and spring travel**

# PCA Analysis

- **Thirty-seven features created, include Preference indices:**
  - **hotel\_hunter\_index**
  - **flight\_fanatic\_index**
  - **bundle index**
- **Features were standardized and fed into PCA**
- **Twenty principal components retained 95% of total variance**

# CLUSTERING & SEGMENT PROFILES

- **K-Means was tested across multiple k values**
- **Elbow analysis suggested 5–6 clusters, while silhouette scores favored 3.**
- **Final selection of k=5 was chosen for business interpretability and evidence of five behavioral peaks.**





# Cluster Assignment

- **Cluster (0): Younger, hotel-focused**
  - low overall travel engagement, taking fewer but more luxurious or longer hotel-centric trips. Low flight activity.
- **Cluster (1): Highly engaged, frequent travelers**
  - seek and use discounts, often bundling their travel, high session counts, clicks, and usage of flight/hotel discounts
- **Cluster (2): Efficient and active travelers,**
  - less emphasis on discounts and lower hotel spending. Value bundled services,
- **Cluster (3): High-spending, (family or group) , long-haul flights, less interested in hotels as a primary focus.**
- **Cluster (4) Highly engaged users** who spend a lot of time browsing and exploring options, possibly looking for shorter stays, and often interested in bundled deals.



# Perks assigned

- **Cluster (0): Free hotel night with flight**
- **Cluster (1): Exclusive discount**
- **Cluster (2): Free meal**
- **Cluster (3): Free checked bag**
- **Cluster (4) Free cancellation fee**



# Statistical Validation of Clusters

## Segment Validation Methods

ANOVA , Eta square, chi-square tests confirm meaningful differences among segments based on behavior and demographics.

## Segment Profiling

Detailed profiling of segments reveals unique characteristics to guide targeted marketing efforts.





# Limitations:

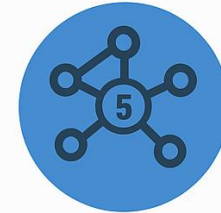
- **Static snapshot (one row per user, probably the same user different trips with different objectives, holiday or business)**
- **Limited psychographic data (no customer satisfaction data and information is available in the dataset.)**



**Static snapshot**  
one row per user



**Limited psychographic data**  
no customer satisfaction data



**K-Means assumptions**  
mathematical assumption

# Suggested Validation Strategies

- **Case-control (A/B testing)** framework is proposed to validate perk effectiveness.
- **Measure lift** in conversion and revenue per cluster: Quantify how much a perk increases bookings and spending for each customer segment compared to users who did not receive the perk.





Marketing  
Perks  
ANNOVA  
validation  
Segmentation  
Chi\_square  
PCA  
Clusters  
Trip\_id  
MIST  
Thanks  
kmeans  
Traveltime  
ABtesting

# Justifications

**Cluster (0): Justification:** While they are hotel-focused, a "free hotel night with flight" could encourage them to combine their hotel stays with flights, potentially increasing their overall engagement and flight activity. This is a compromise from the ideal "Premium Hotel Upgrades" due to the limited options provided.

**Cluster (1) Justification:** This cluster already shows high engagement and actively seeks discounts. Exclusive discounts would directly reward their behavior and encourage continued high activity and loyalty. This aligns perfectly with their discount-seeking nature.

**Cluster (2): Justification:** This cluster is efficient and active. A free meal can add perceived value and convenience to their trips without directly focusing on discounts or large financial incentives. It's a simple, tangible benefit that enhances their travel experience, especially if they are looking for efficiency.

**Cluster (3):** A free checked bag directly addresses a practical need and cost associated with their travel style, providing significant value for group or long-distance trips.

**Cluster (4) Justifications:** This cluster is highly engaged in browsing and exploration, suggesting they might be indecisive or frequently changing plans. A "free cancellation fee" perk offers flexibility and reduces booking friction, encouraging them to book more frequently knowing they have an option to change without penalty, fitting their exploratory nature.