

COLUMBIA UNIVERSITY
School of Professional Studies

American Airlines 

Analytical Customer Segmentation

FlyAhead Consultants

SERENA S. IYER

KHUSHI MANDAVIA

YILUN LI

SHARON LEE

RIHAAN SATIA

YAN PING YU

YU WEN HUANG



OUR FOCUS

- Understanding customer preferences is key in the evolving airline industry.
- Our analysis reveals insights to enhance passenger connections.
- These insights enable tailored experiences and growth.
- We aim to transform customer segmentation strategies.





AGENDA



Data Preparation



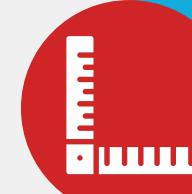
Modeling Strategy & Results



Team Introduction



Feature Construction



Cluster Profiling



Project Objectives

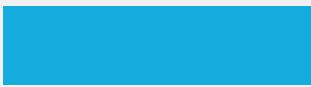
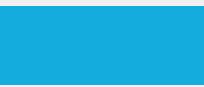


Data Visualization



Strategic Recommendations

TEAM INTRODUCTION

**SERENA S. IYER**

Management Consultant
B.Tech Civil Engg @ BPUT, MBA @
Symbiosis, Ex-McKinsey

**KHUSHI MANDAVIA**

Project Manager/Team Lead
Bachelors in Business,
Accounting & Finance @ DU

**YILUN LI**

Business Analyst
Bachelors in Business, Philosophy
@ Brandeis University

**SHARON LEE**

Business Analyst
BS Accounting @ NYU

**RIHAAN SATIA**

Data Scientist/ML Scientist
B.Tech CS @ VIT University

**YANG PING YU**

Data Scientist/ML Scientist
BA Quant Fin @ National Tsing Hua University

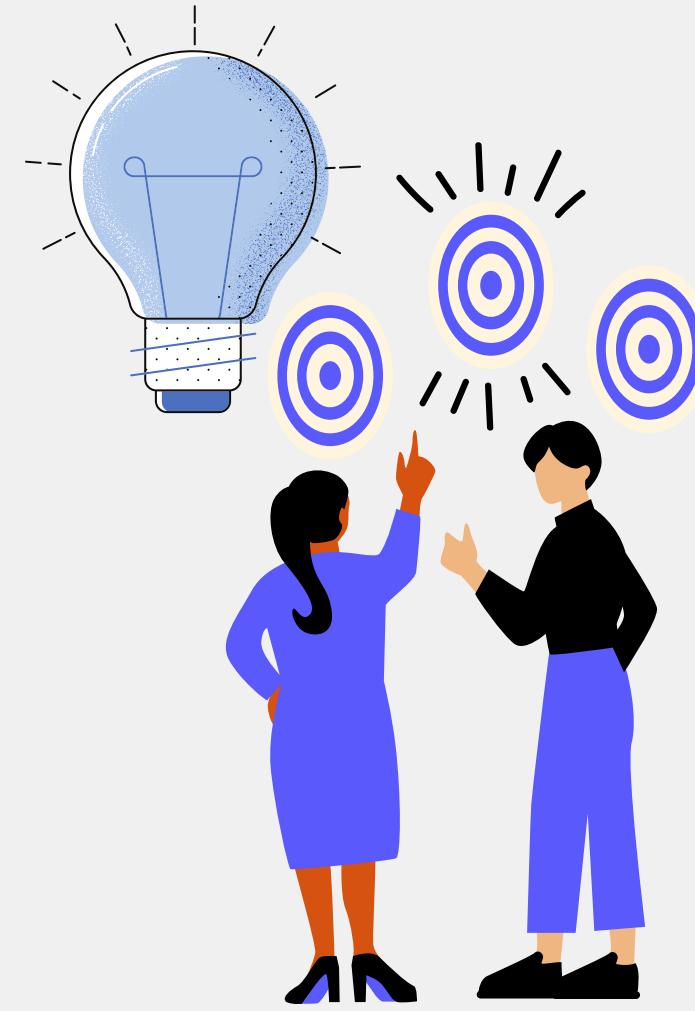
**YU WEN HUANG**

Data Engineer
BS Informatics @ IU Bloomington

PROJECT OBJECTIVES

CHALLENGES

- Complexity in Booking Pattern
- Traditional Segmentation Limitations



GOALS

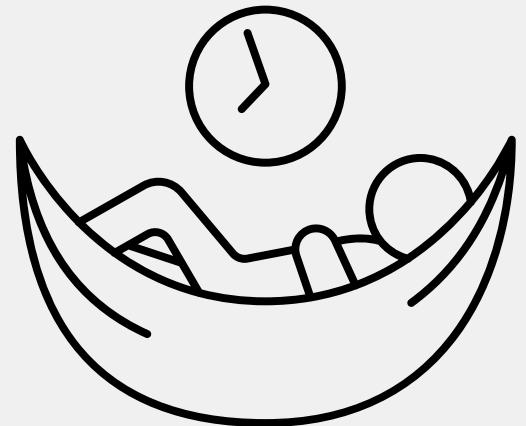
- Improve leisure traveler segmentation to identify distinct groups
- Uncover WTP (Willingness to Pay)

APPROACH

- Advanced Clustering Techniques
- Iterative Model Refinement

DATASET OVERVIEW & FOCUS

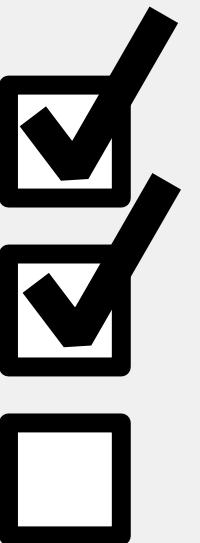
- Variable travel and booking patterns based on booking type



Leisure Travel

%

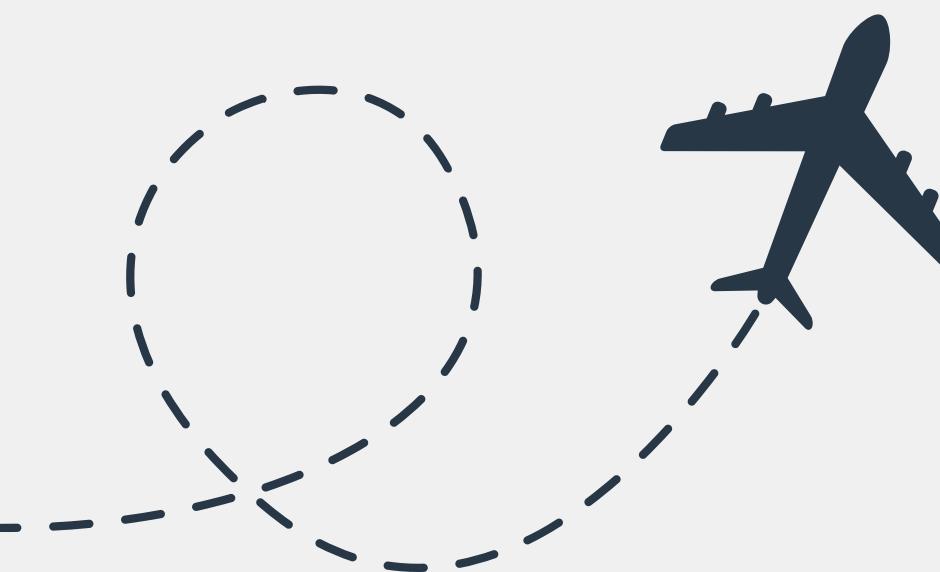
Probability Columns



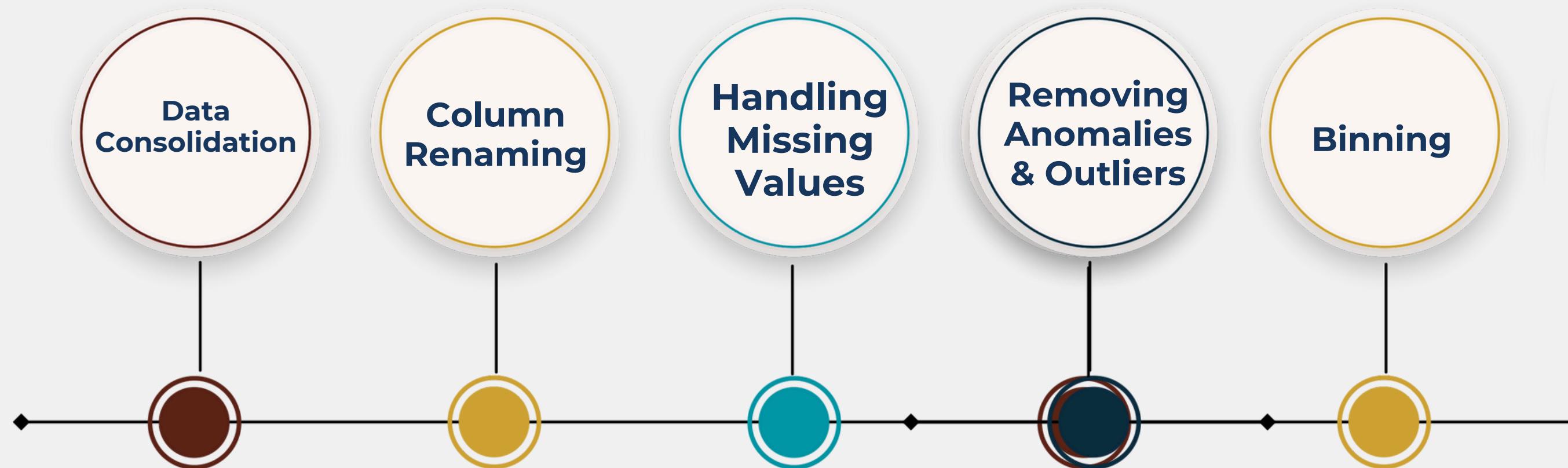
Selection



New Dataset : 2,249,589 entries



DATA PREPARATION



FEATURE CONSTRUCTION



Length of Stay



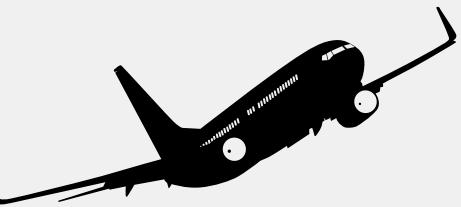
Weekend Departure & Arrival
(PNR_DEPARTURE_DATE
& PNR_RETURN_DATE)



Advanced_Purchase



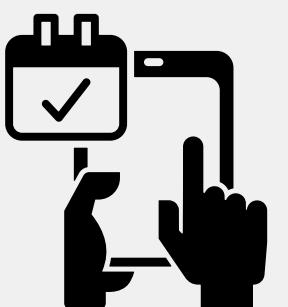
Round Trip Flights
(True/False)



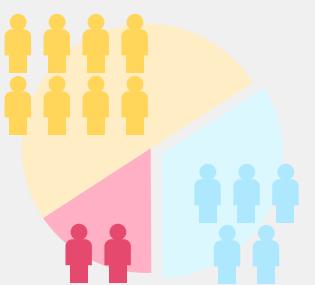
Non-Stop Flights
(True/False)



Travel Distance
(Origin & Destination)



Cumulative Bookings
until PNR Creation



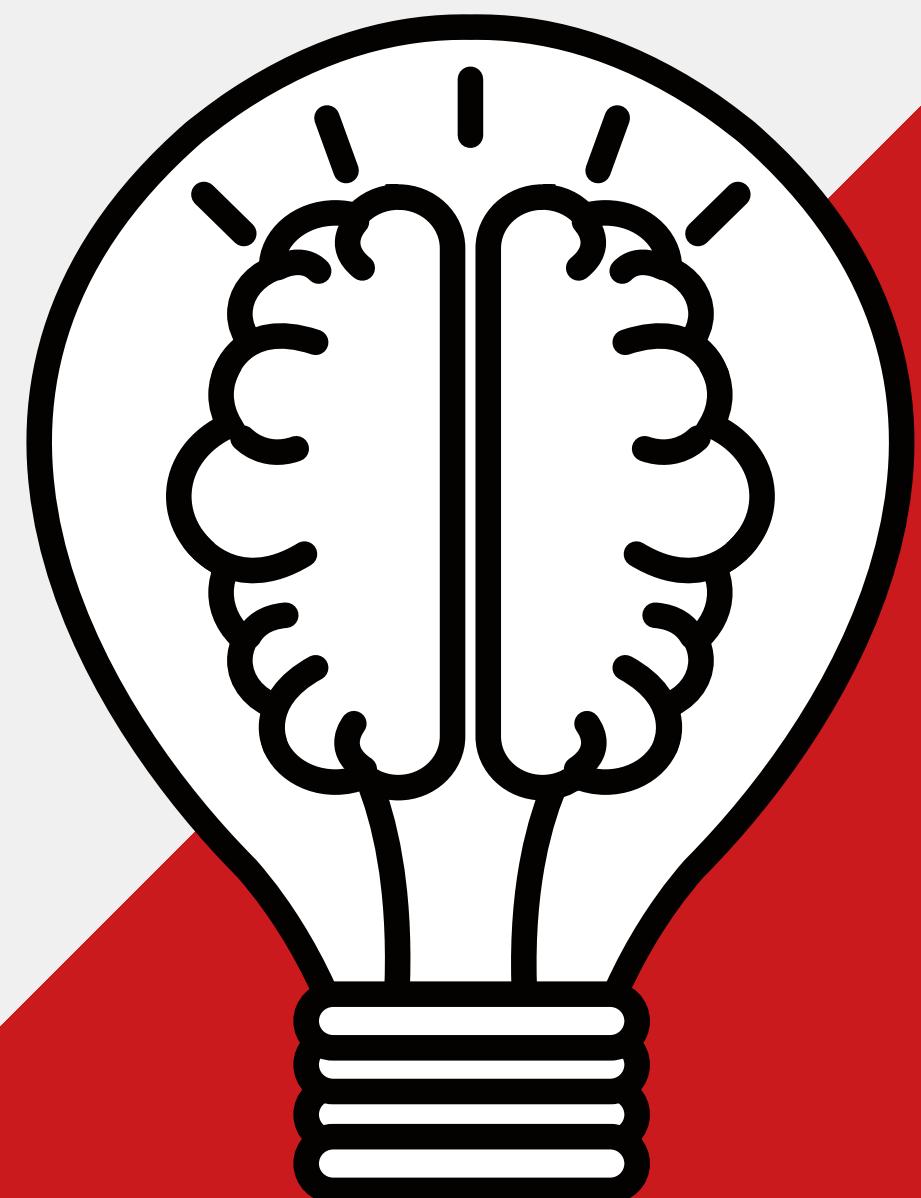
Segment AP days
(1-7, 8-14, 15-
21, 21+)



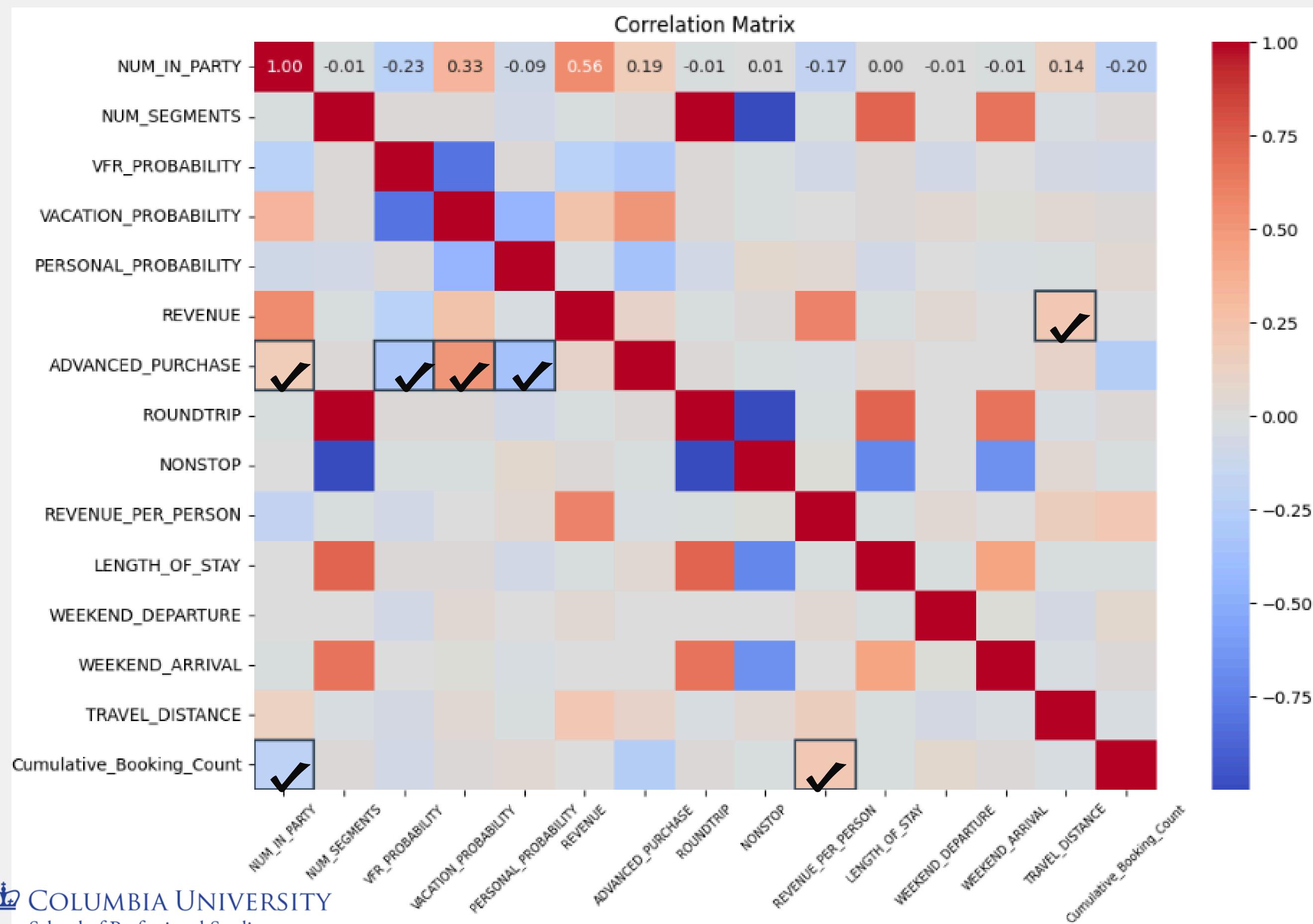
Revenue per person



Holiday Travel

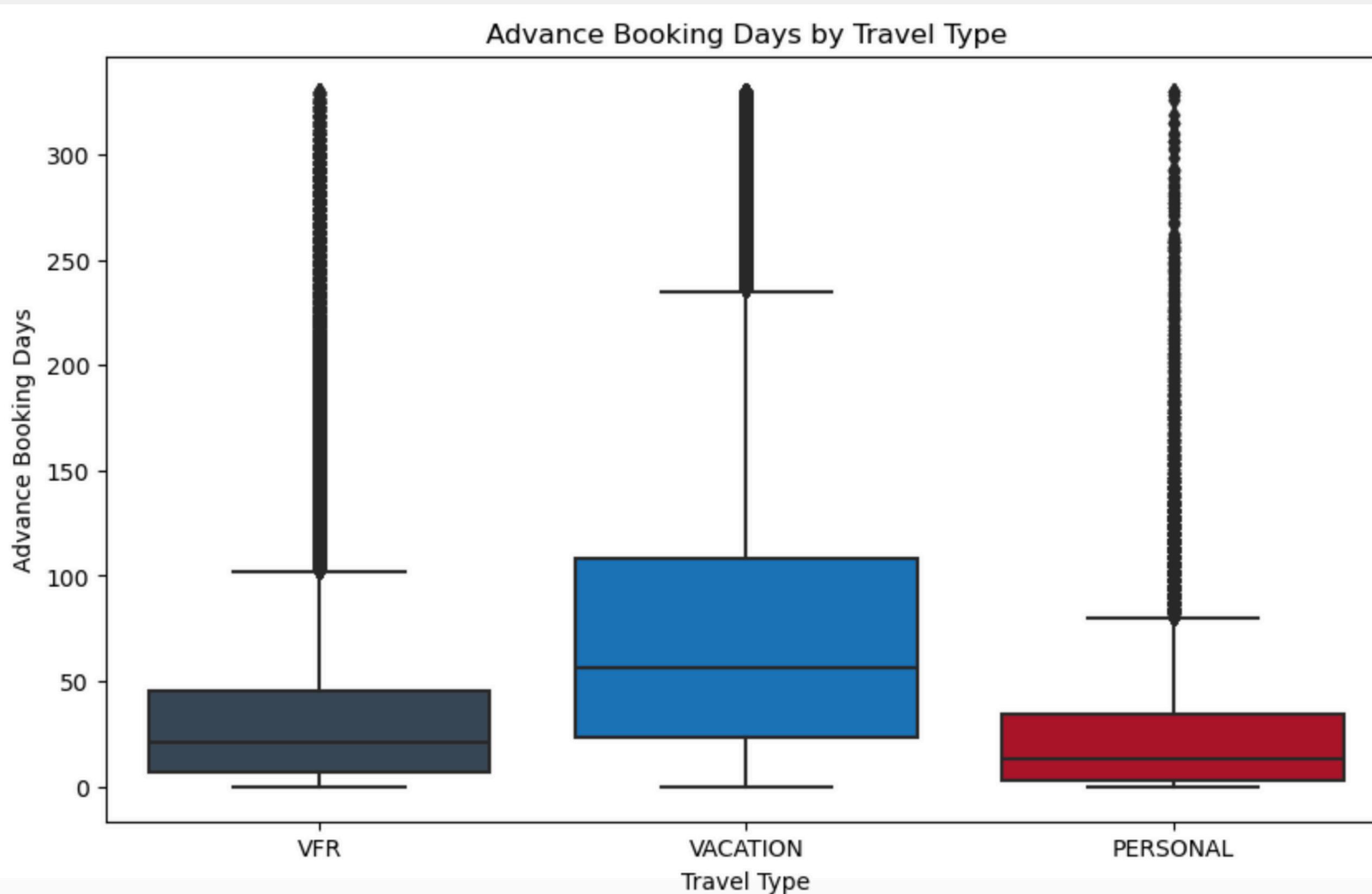


CORRELATION MATRIX



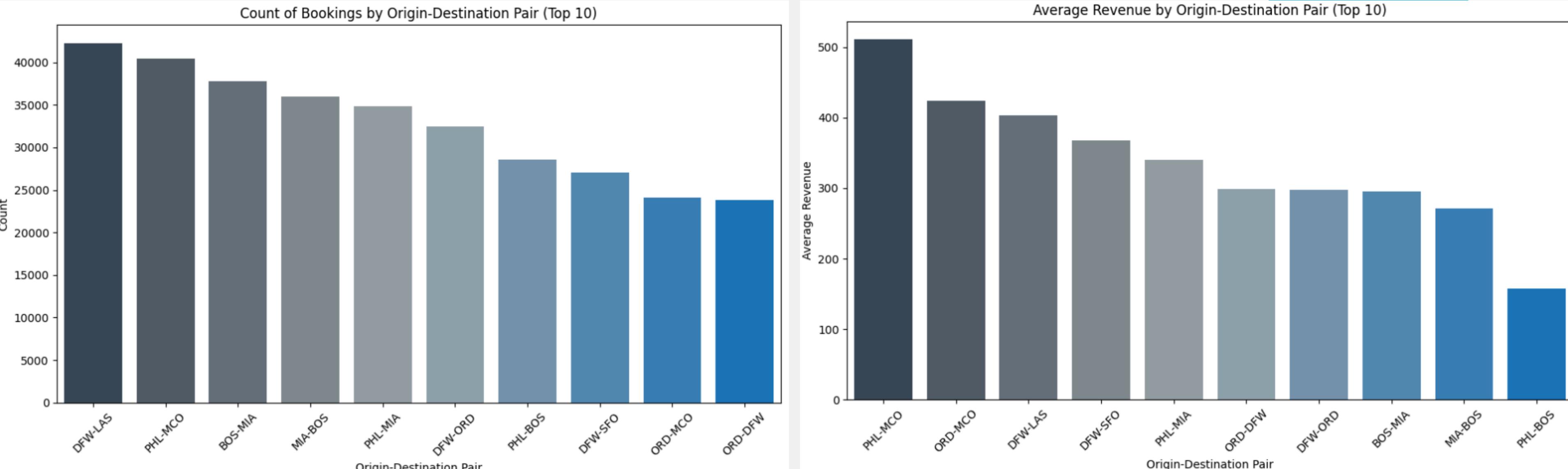
- AP & travel type probabilities
- Number in Party and Cumulative Booking Count
- Revenue & Travel Distance

ADVANCED BOOKING DAYS DISTRIBUTION



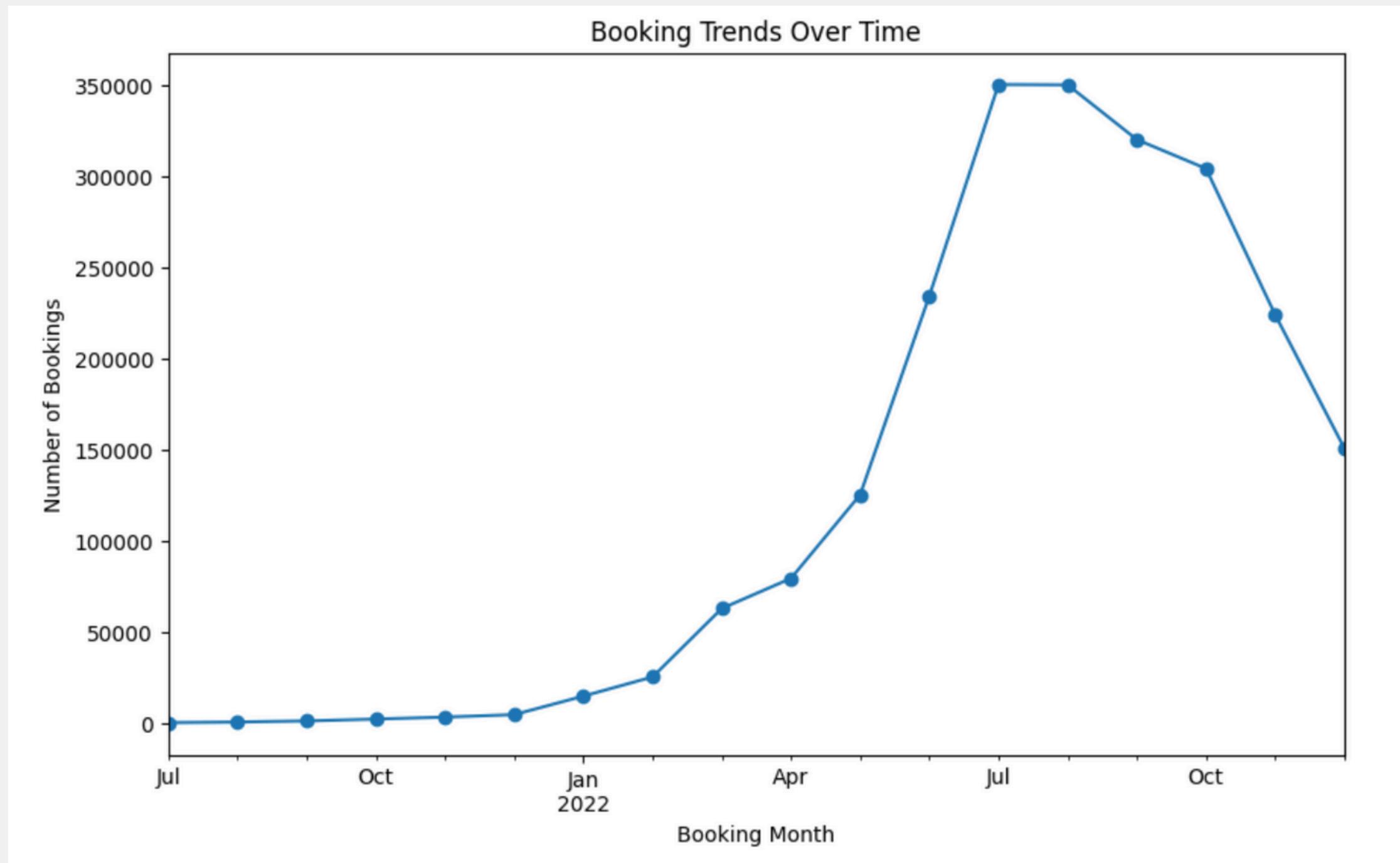
1. **Vacation**
2. **VFR**
3. **Personal**

BOOKINGS & REVENUE BY OD



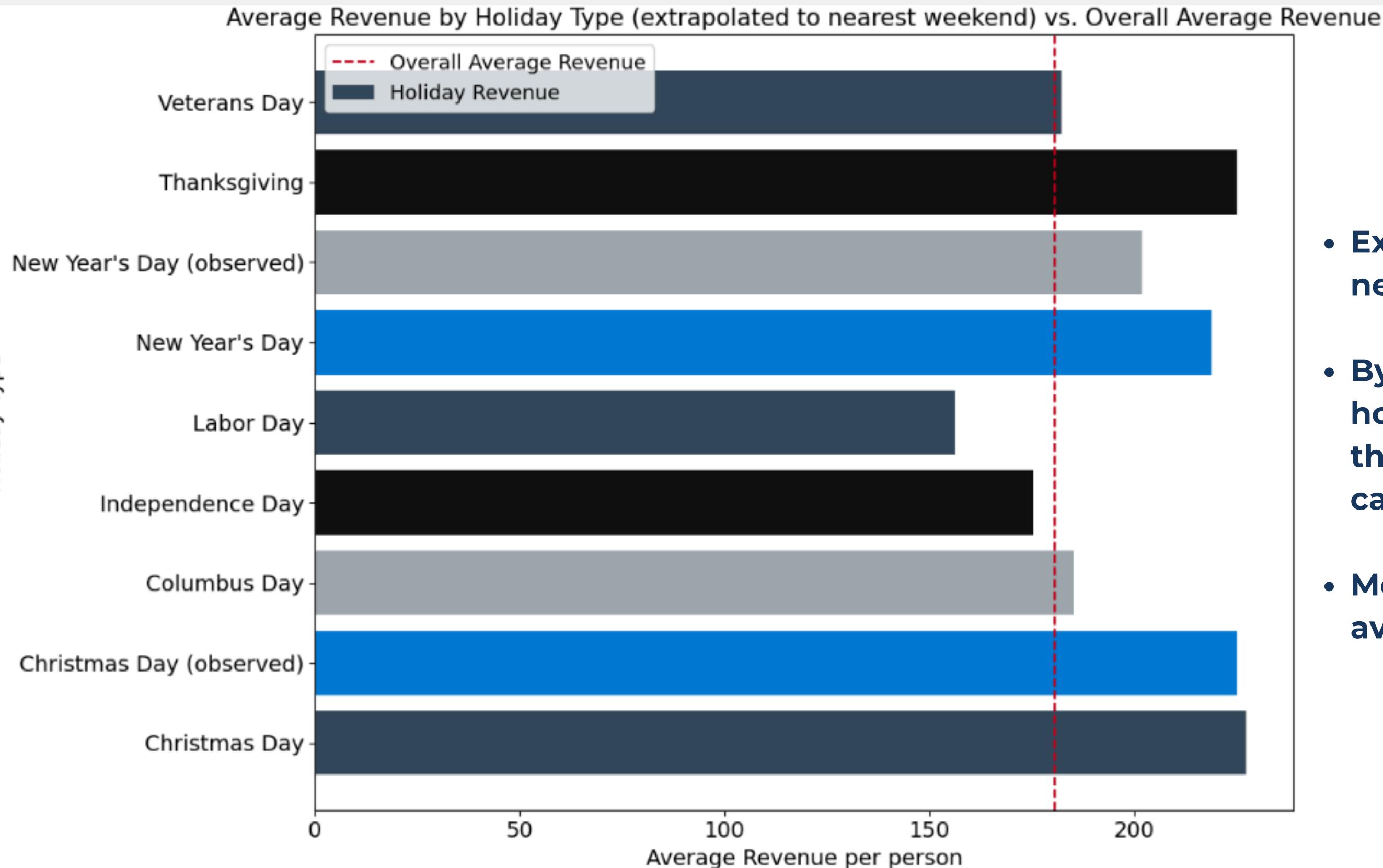
- **DFW-LAS and PHL-MCO are critical for both volume and revenue**
- **PHL-MCO and DFW-LAS are more efficient in generating revenue per booking**

BOOKING DISTRIBUTION BY MONTH



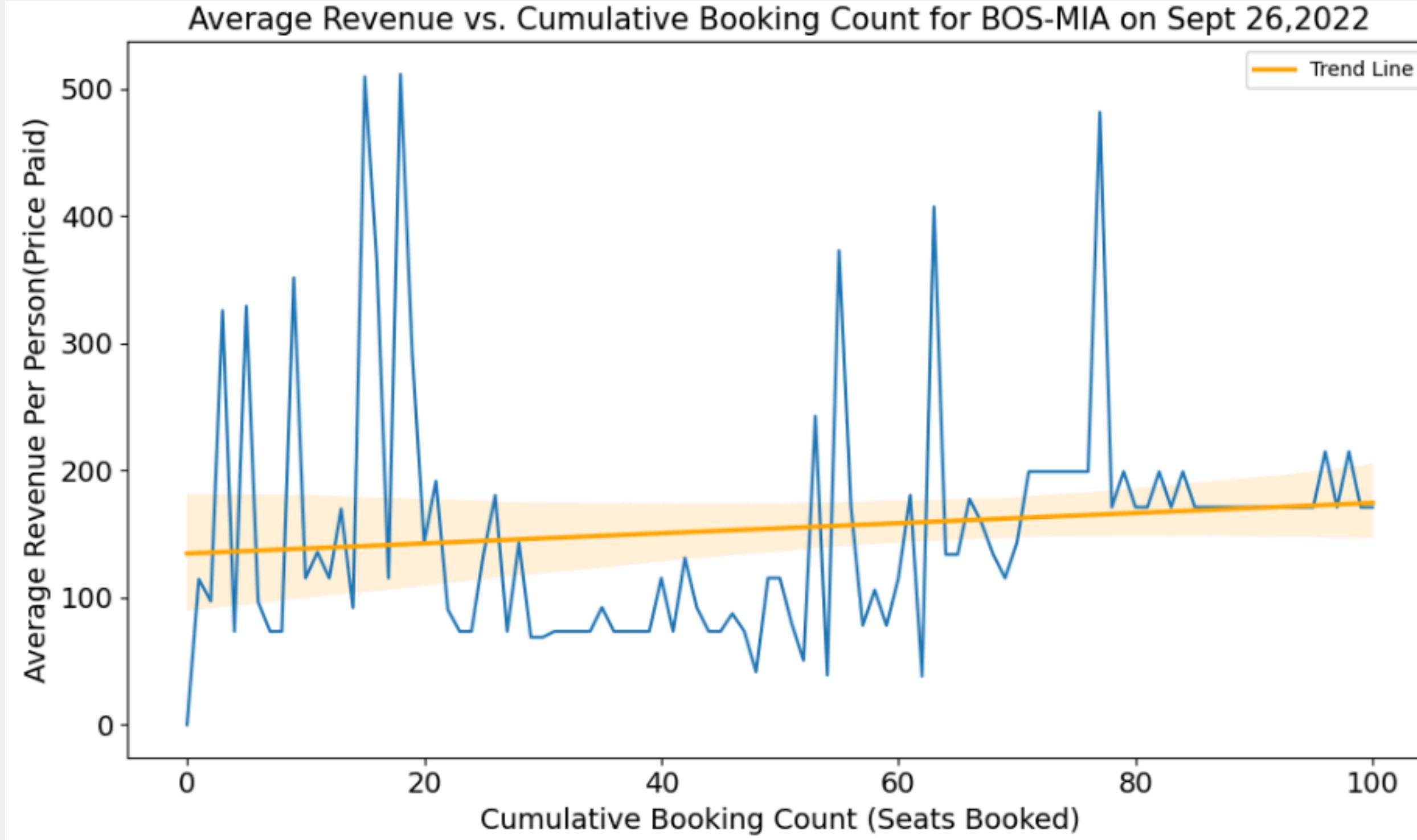
- **Number of bookings peaks in July and August, indicating the busiest months for non-business travelers**

EXTENDED HOLIDAYS



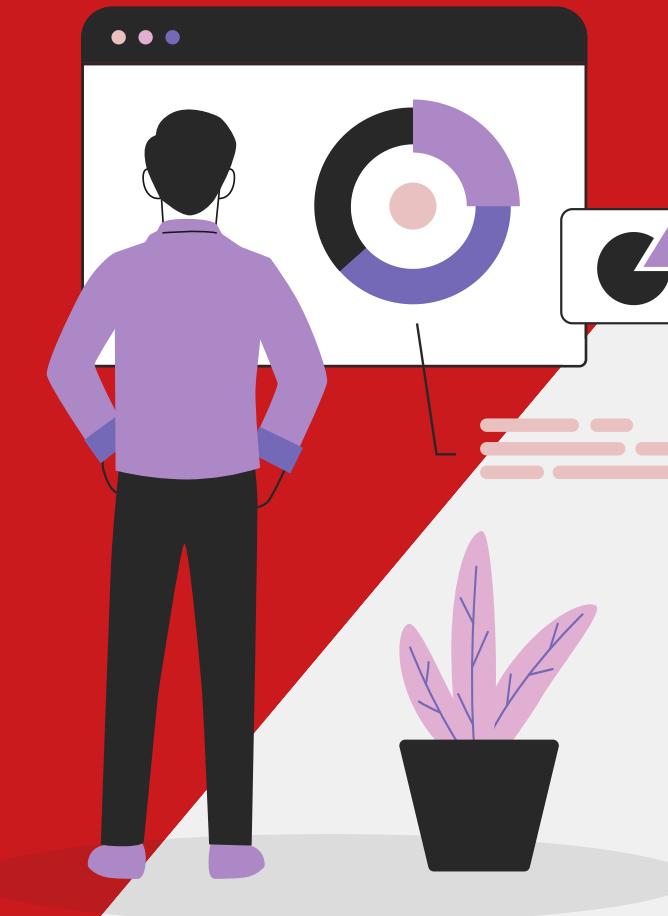
- **Extrapolated the holiday to the nearest weekend - Assumption**
- **By including the days around holidays in the holiday analysis, the real effect on revenue is captured**
- **Most holidays have higher than average revenue**

EFFECT OF SEAT AVAILABILITY



- **Average price increases steadily with increased seat bookings**
- **Can be combined with AP to capture the overall effect of the time of booking**
- **Effect of seat availability is dependent on route and time of year**

MODELING STRATEGY



- Enhance WTP Features

(Revenue per Mile, Advance Purchase Ratios, Distance-Adjusted Revenue)

- Clustering Algorithms

(K-Means, Weighted K-Means)

- Feature Standardization

(Normalization for clustering consistency)

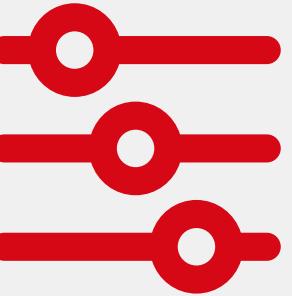
- Cluster Refinement

(Use variable importance to improve segmentation)

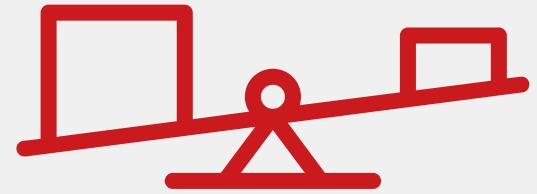
FEATURE MODIFICATIONS TO CAPTURE WILLINGNESS TO PAY (WTP)



REVENUE PER
MILE (RPM)



DISTANCE ADJUSTED
REVENUE



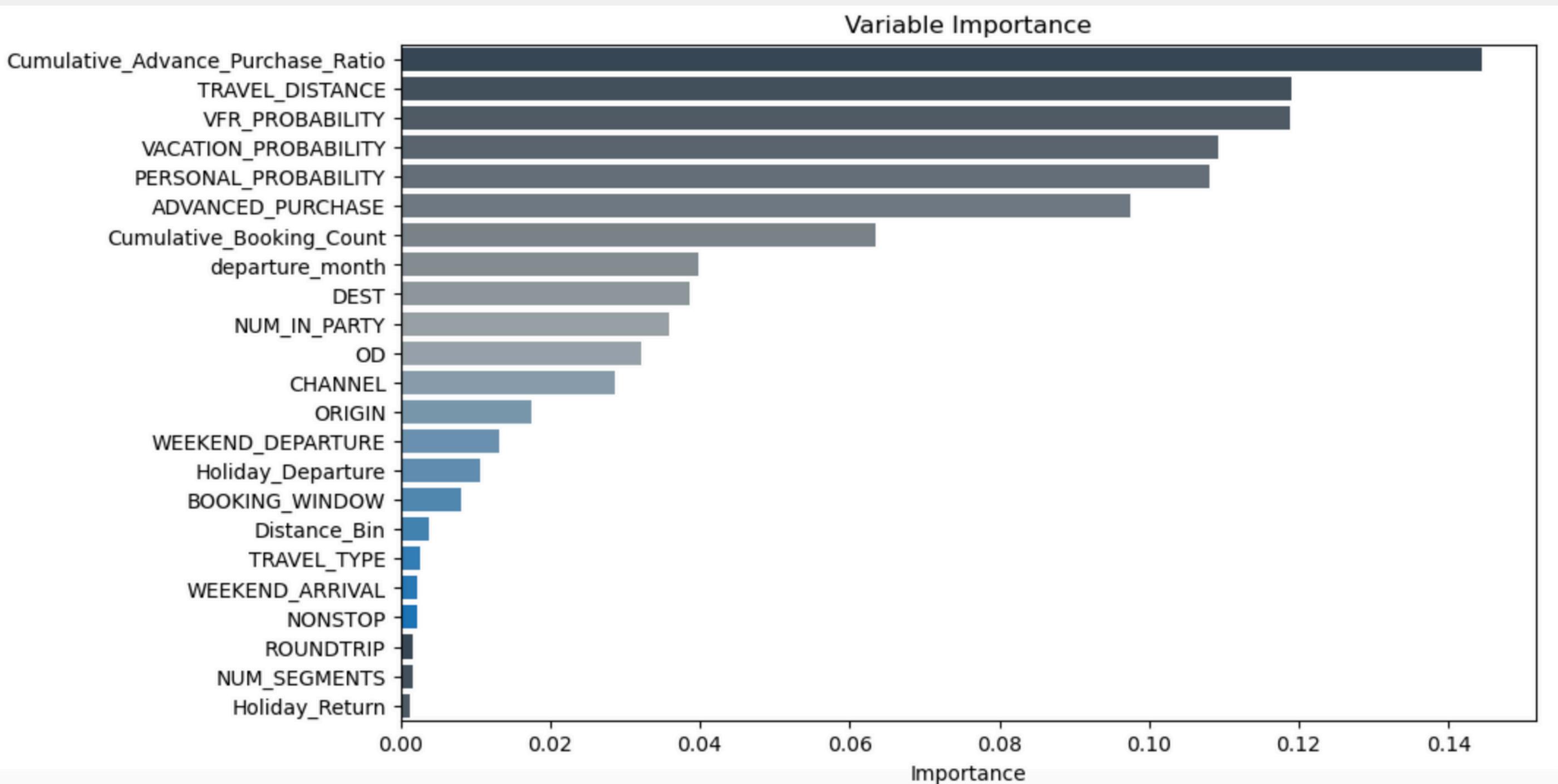
REVENUE AP
RATIO



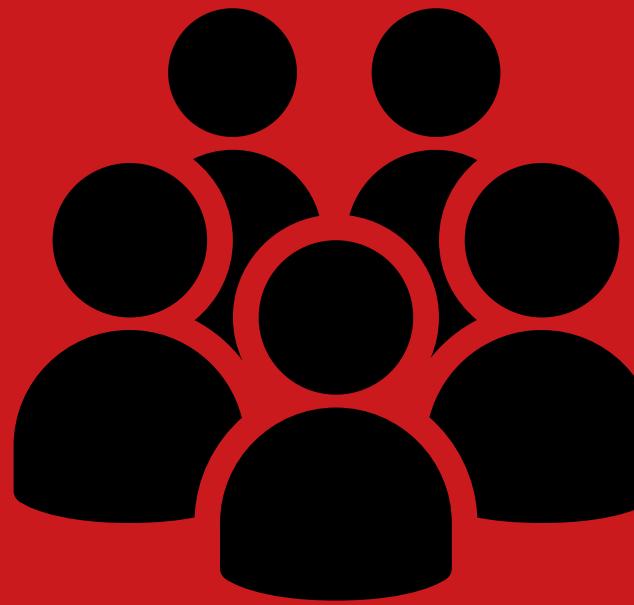
CUMULATIVE
BOOKING - AP
RATIO



VARIABLE IMPORTANCE



- **Decision Tree - Random Forest Regressor(numeric data)**
- **Target: revenue per person**
- **Feature predicting power**



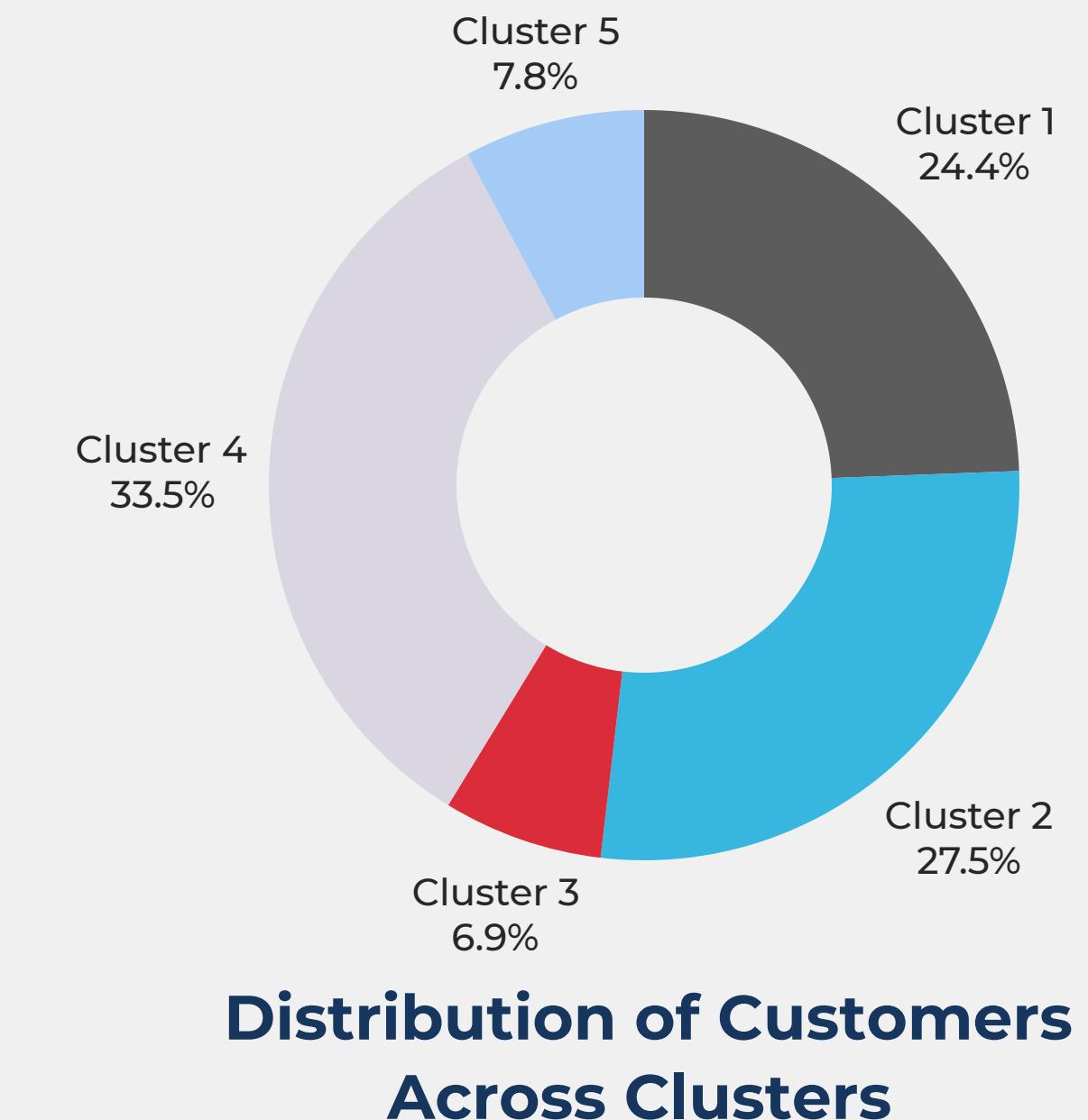
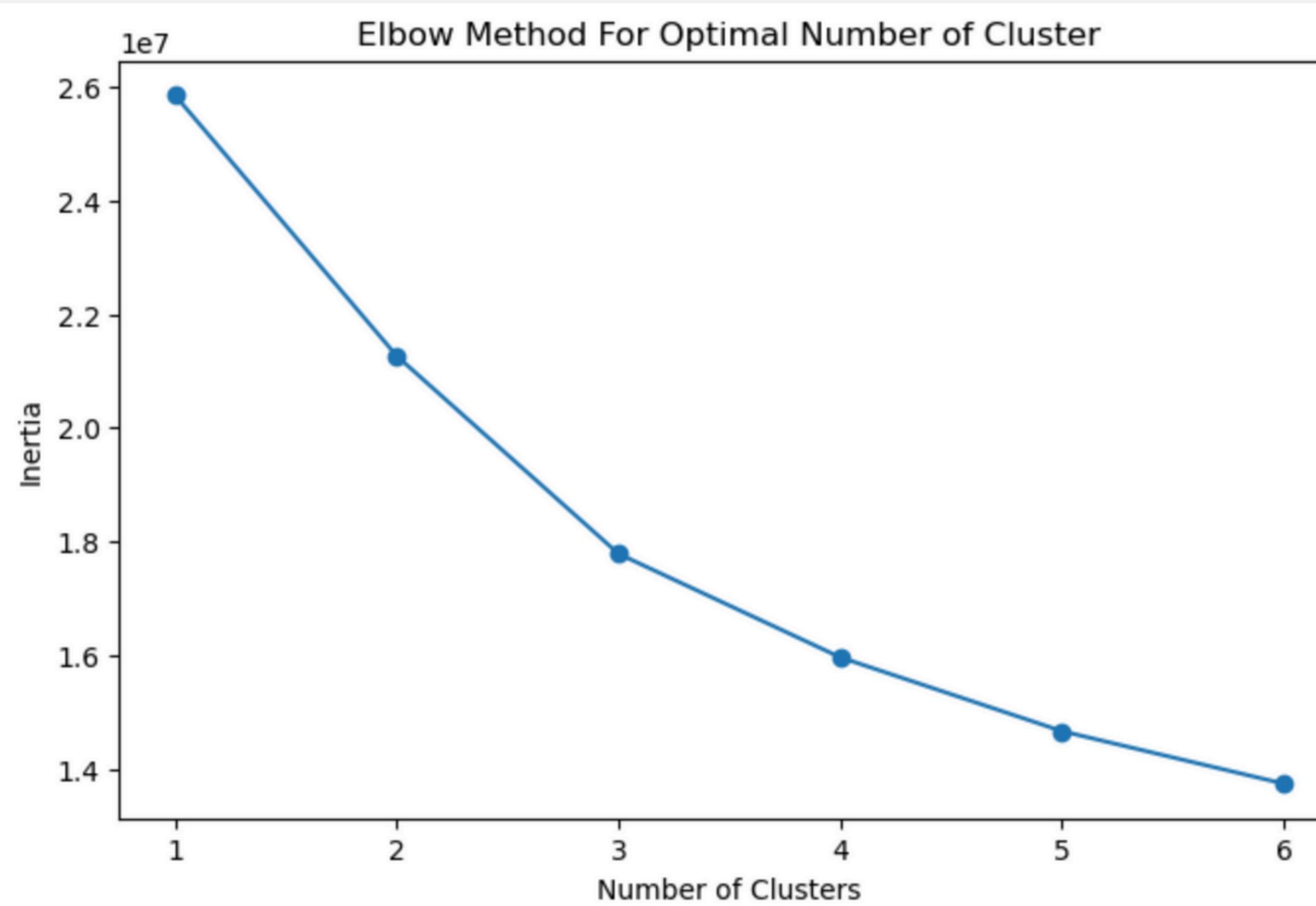
PROPOSED VARIABLE SET

FEATURE SELECTION	WILLINGNESS TO PAY
<ul style="list-style-type: none">• Vacation, VFR, Personal• Advanced Purchase• Cumulative Booking Count• Departure Month• Holiday• Weekend	<ul style="list-style-type: none">• RPM (Revenue Per Mile)• Distance Adjusted Revenue• Revenue AP Ratio• Cumulative Advanced Purchase Ratio

UNWEIGHTED VS. WEIGHTED K-MEANS COMPARISON

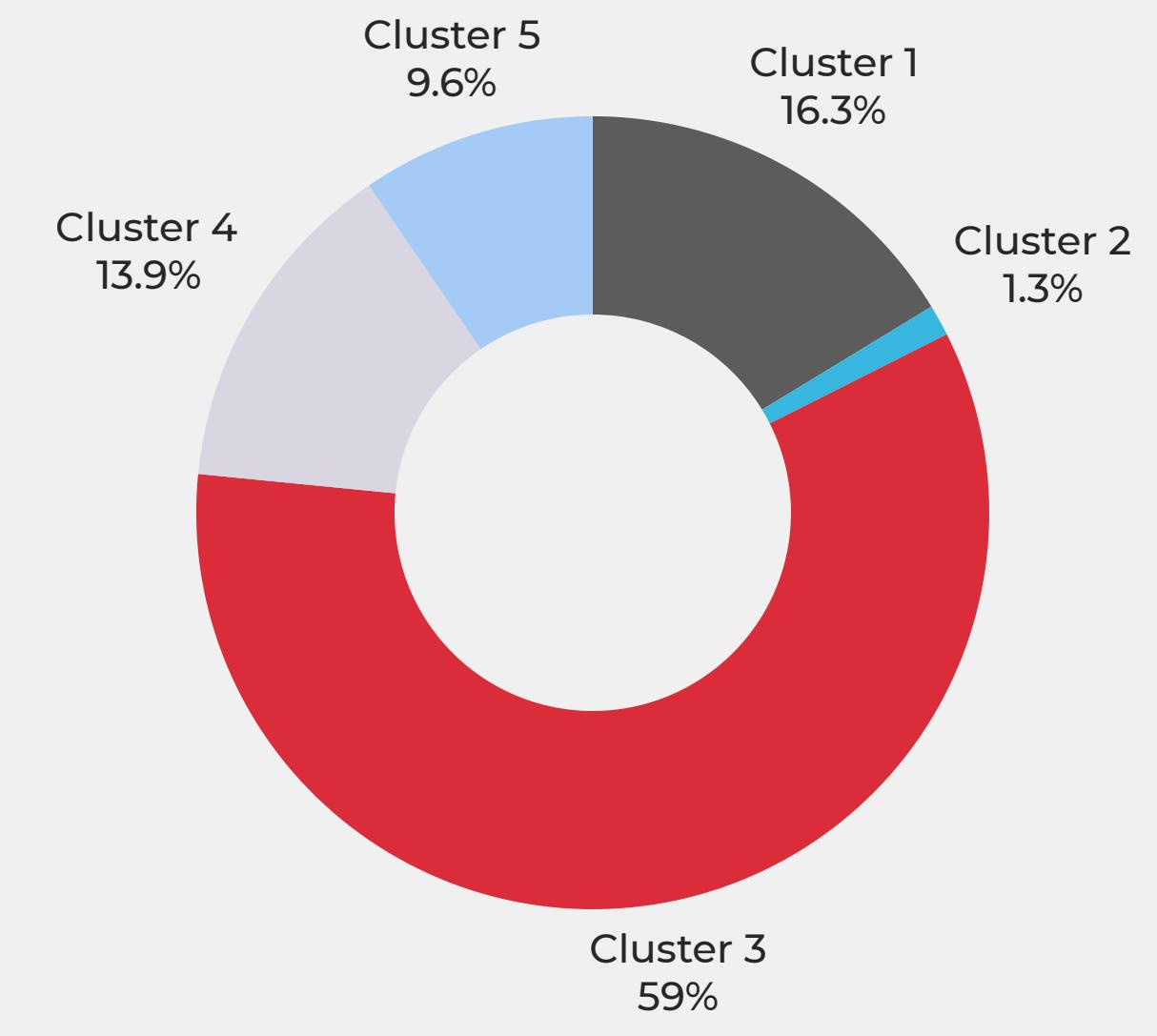
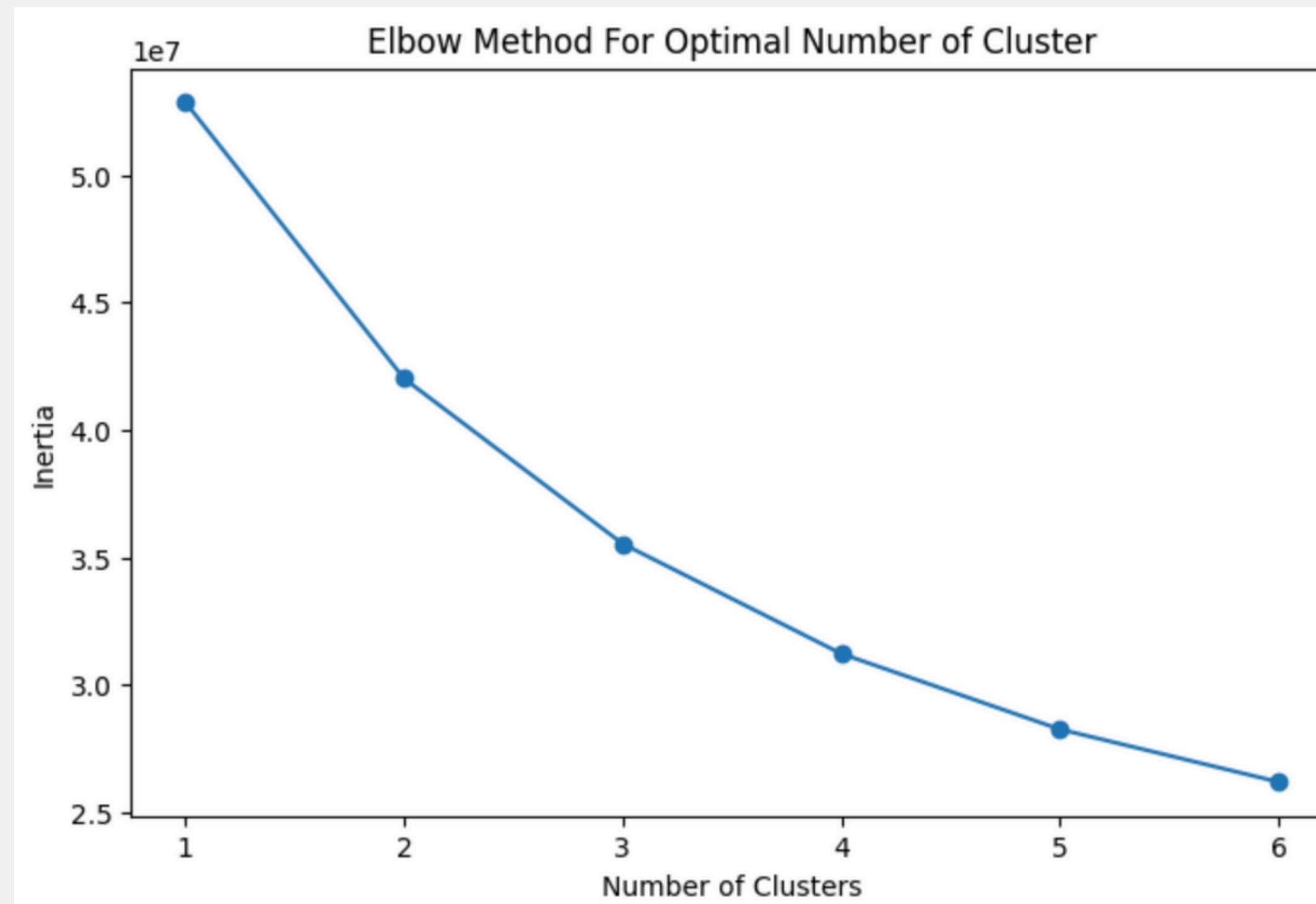


SELECTING OPTIMAL K AND CLUSTER DISTRIBUTION: NORMAL K-MEANS



- Elbow suggests 3, but we chose 5 for easy comparison

SELECTING OPTIMAL K AND CLUSTER DISTRIBUTION: WEIGHTED K-MEANS



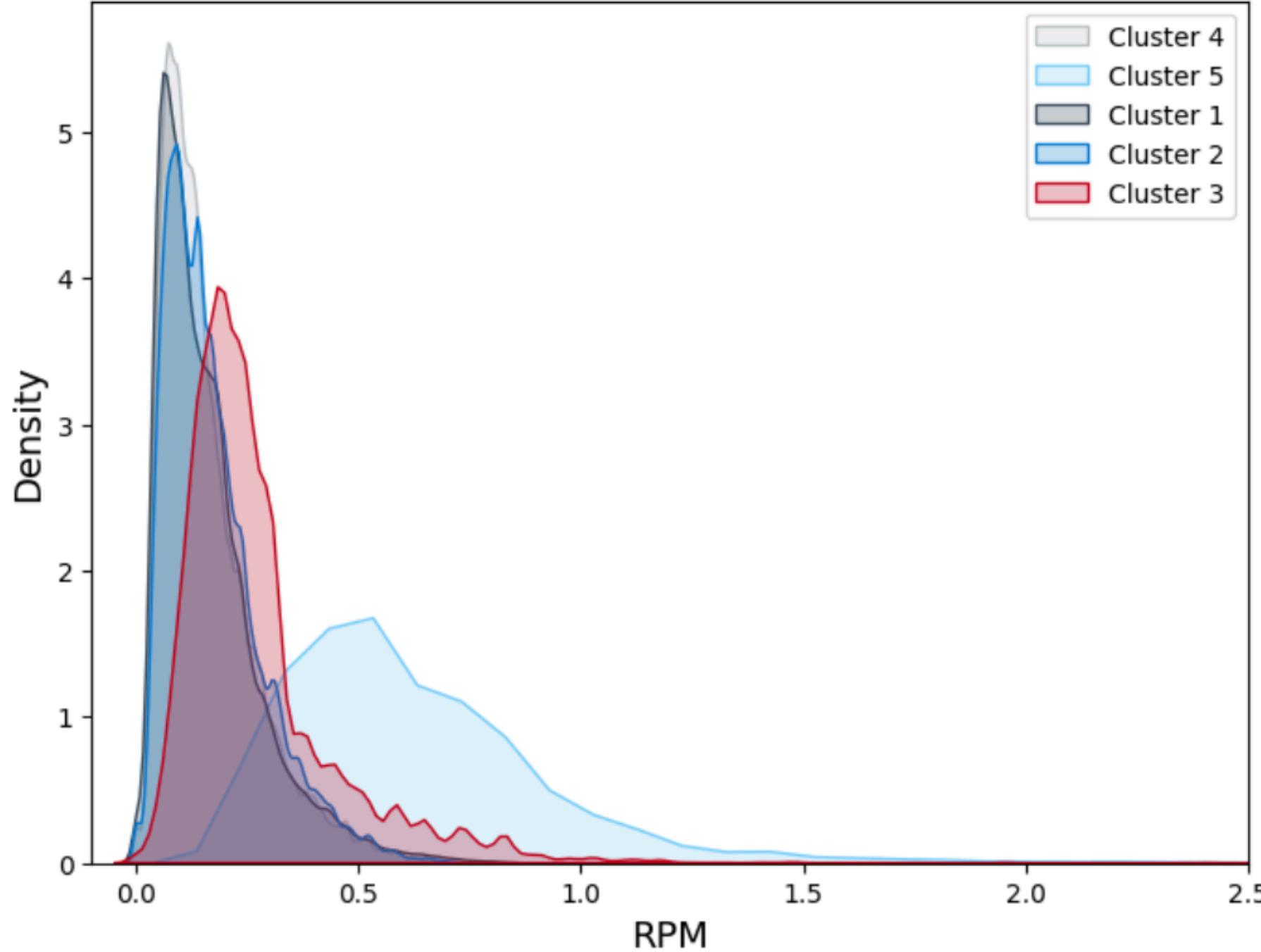
**Distribution of Customers
Across Clusters**

- Chose optimal clusters as 5

REVENUE PER MILE ACROSS THE CLUSTERS

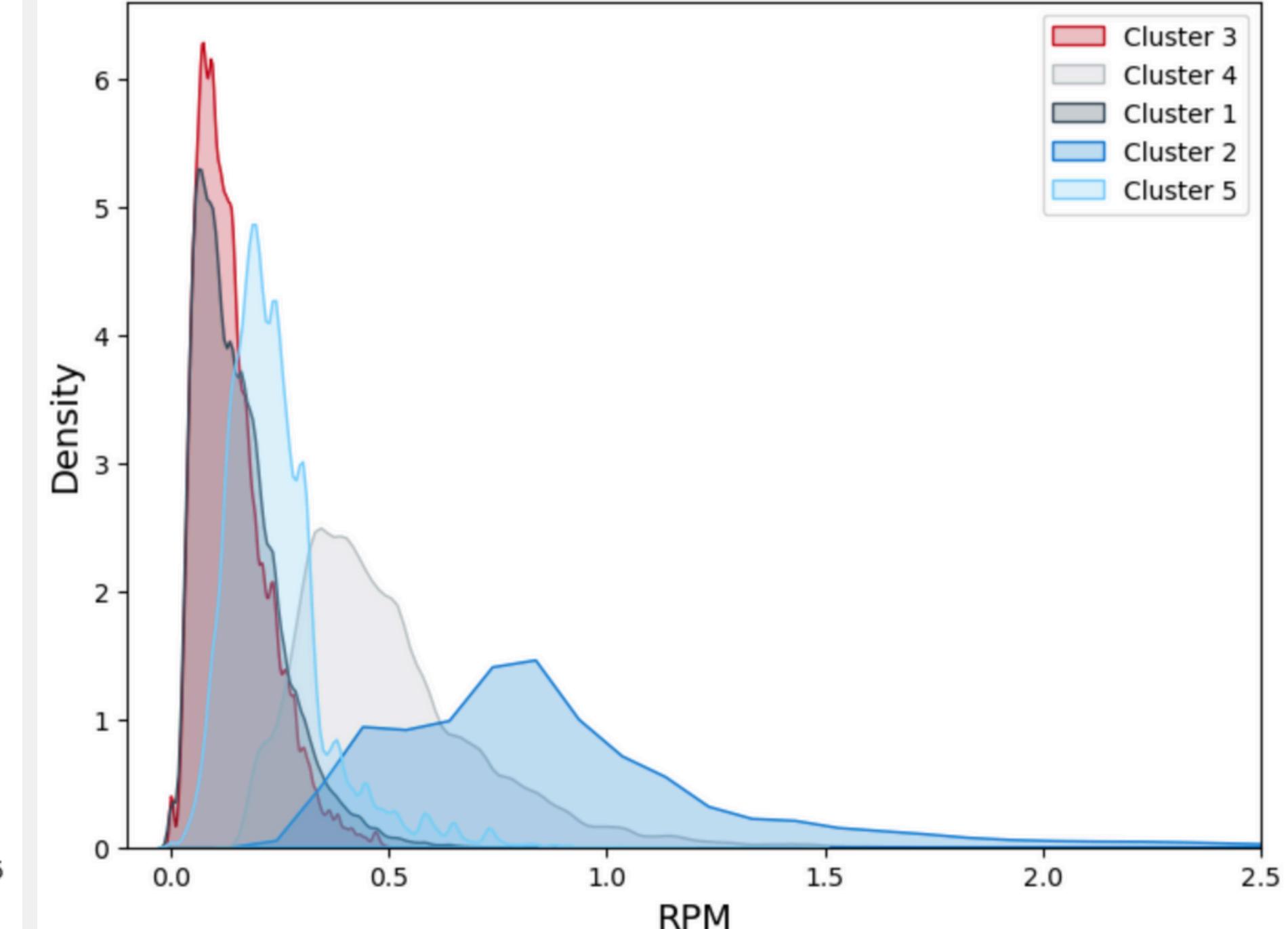


Density Plot of Revenue per Mile by Cluster (Unweighted)



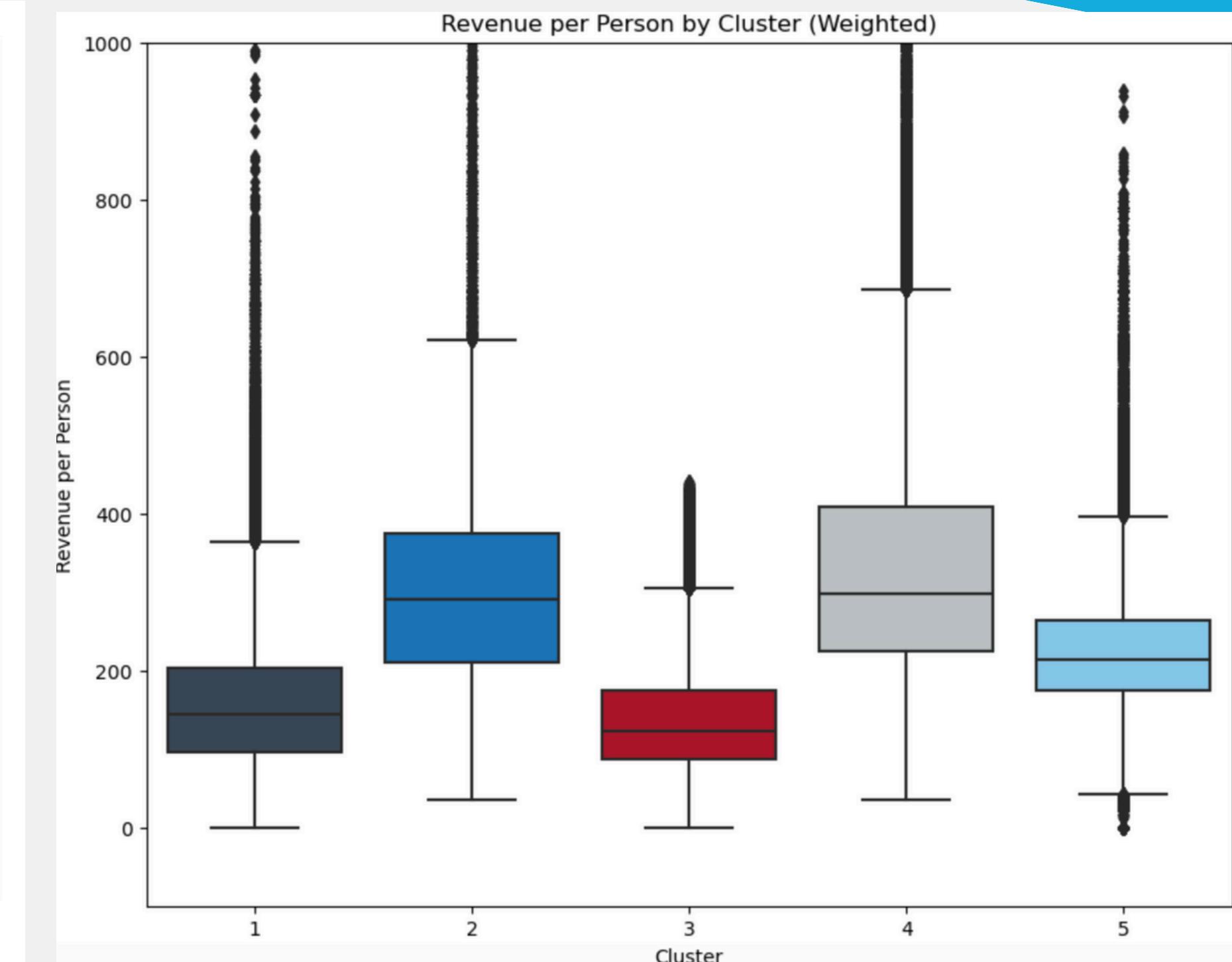
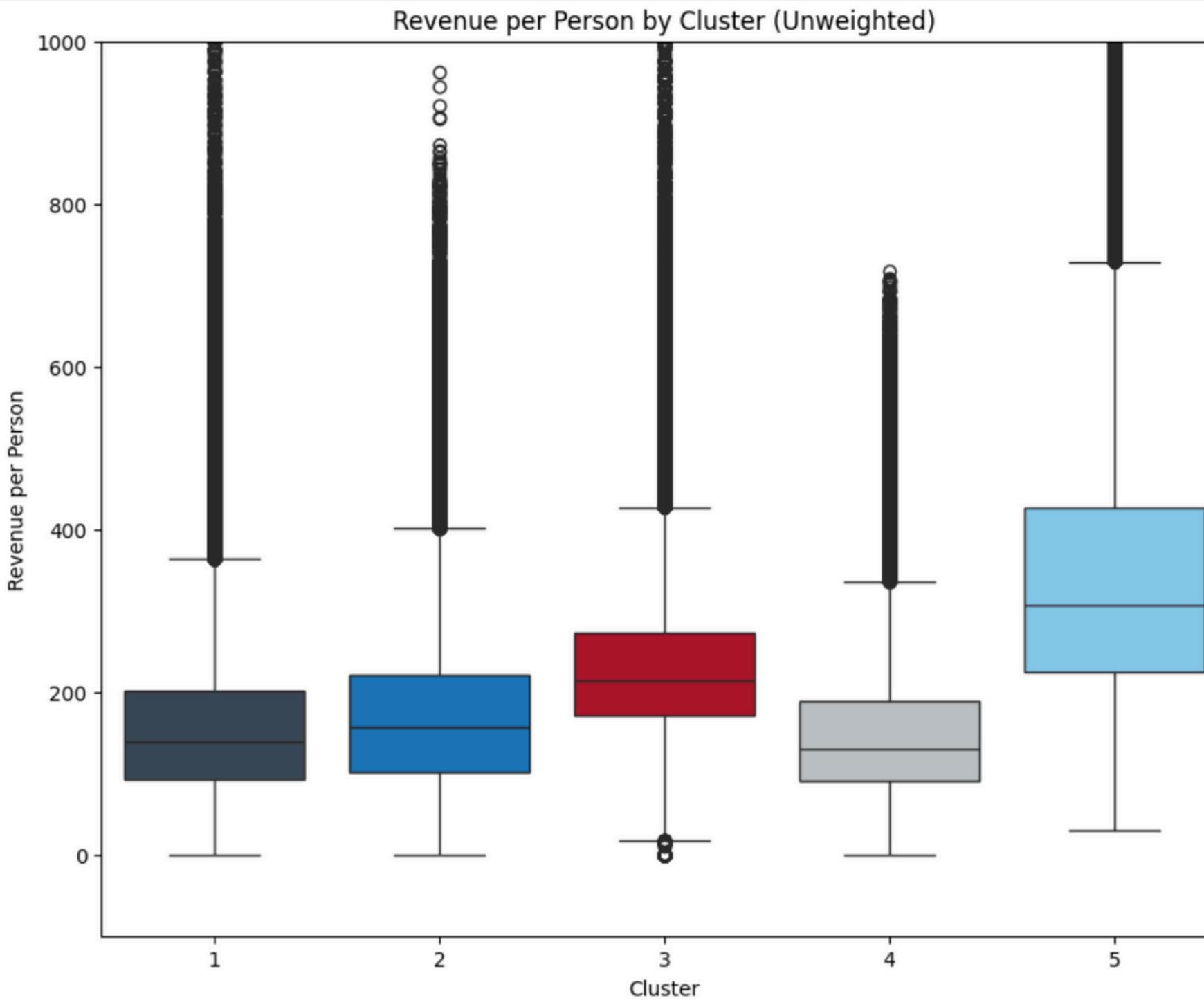
- Narrow peaks, implying tighter revenue distributions around lower RPM values

Density Plot of Revenue per Mile by Cluster (Weighted)



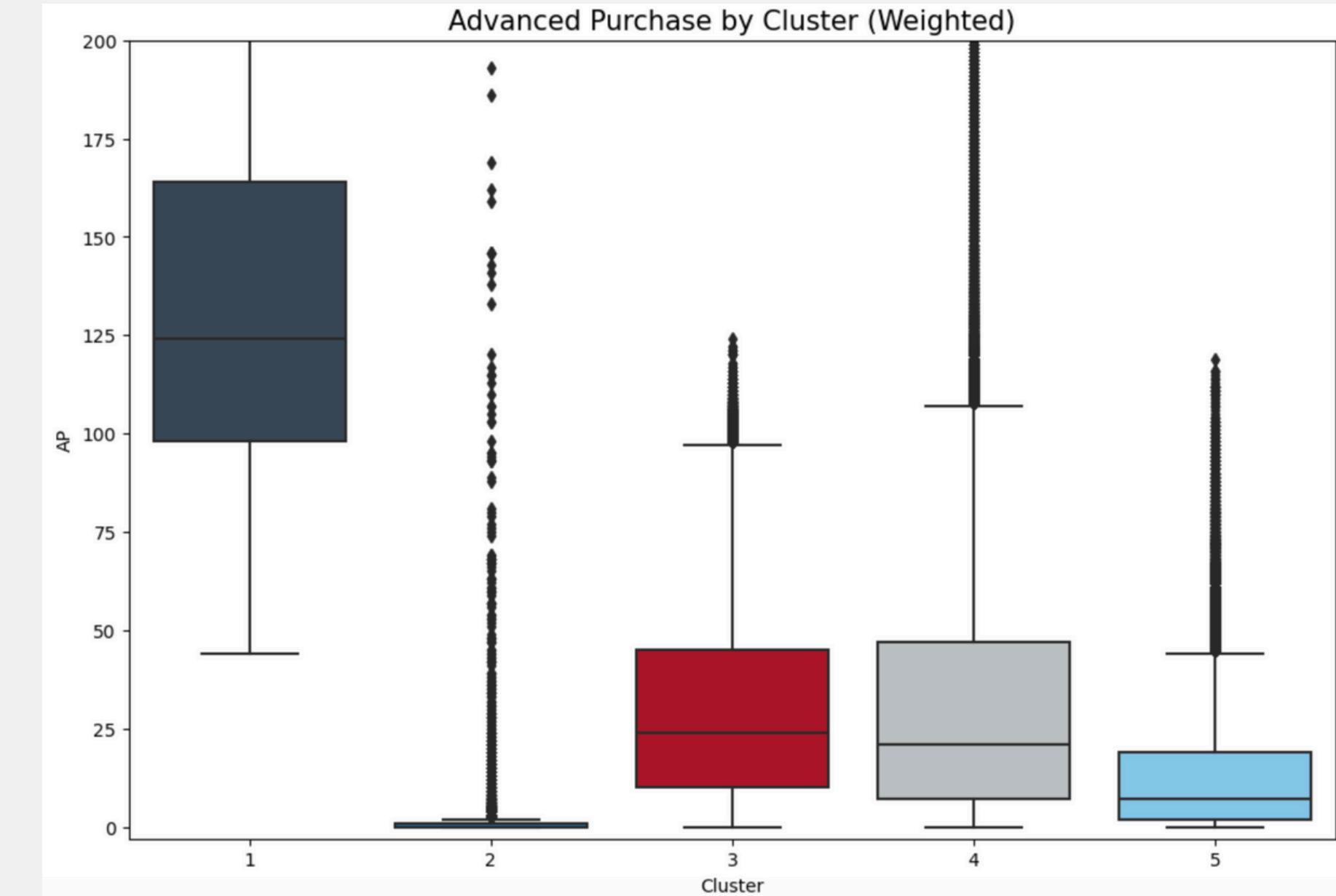
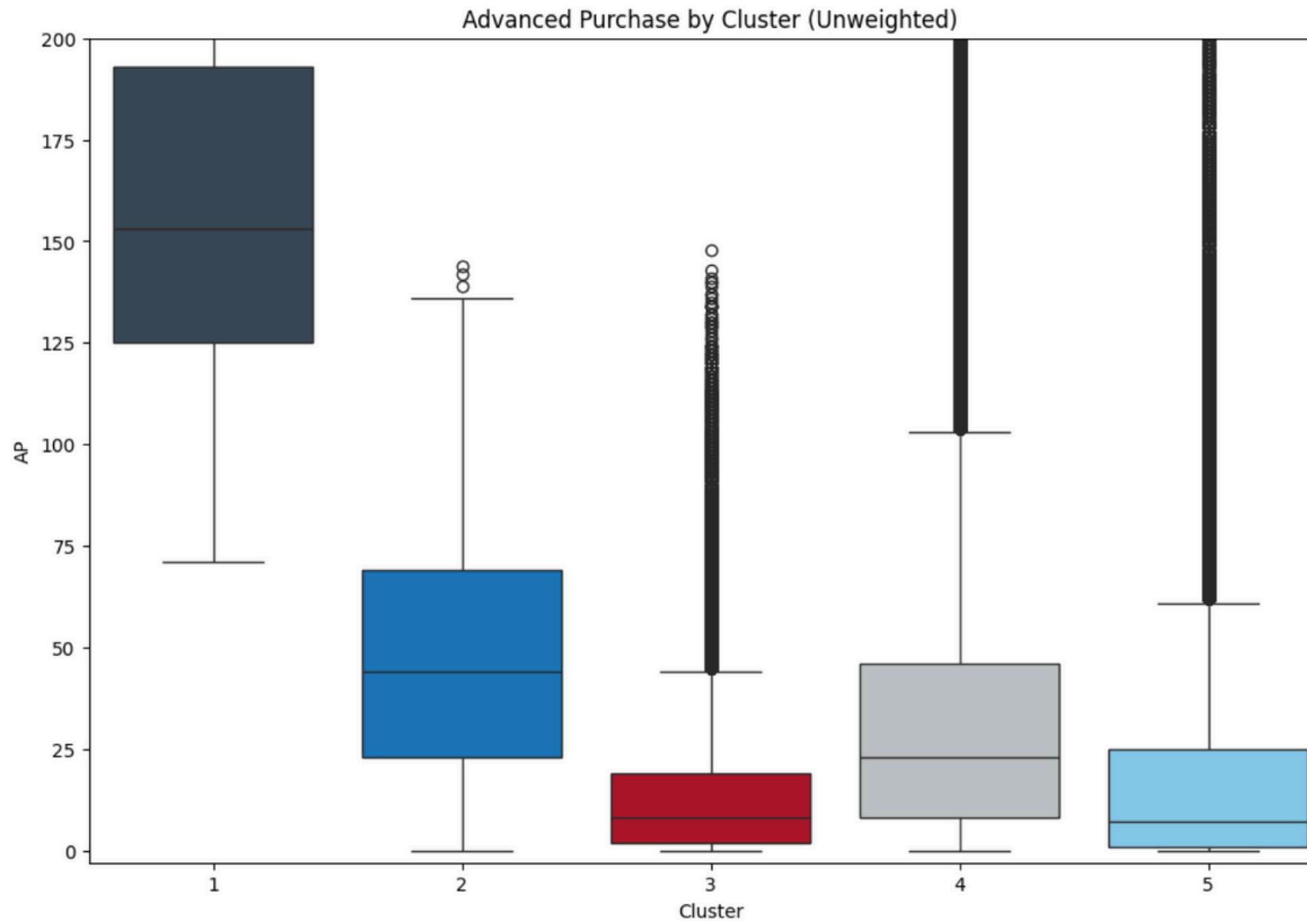
- More spread for Clusters 4 and 2, indicating the possibility of wider revenue variability or the willingness to pay a premium

REVENUE PER PERSON ACROSS THE CLUSTERS



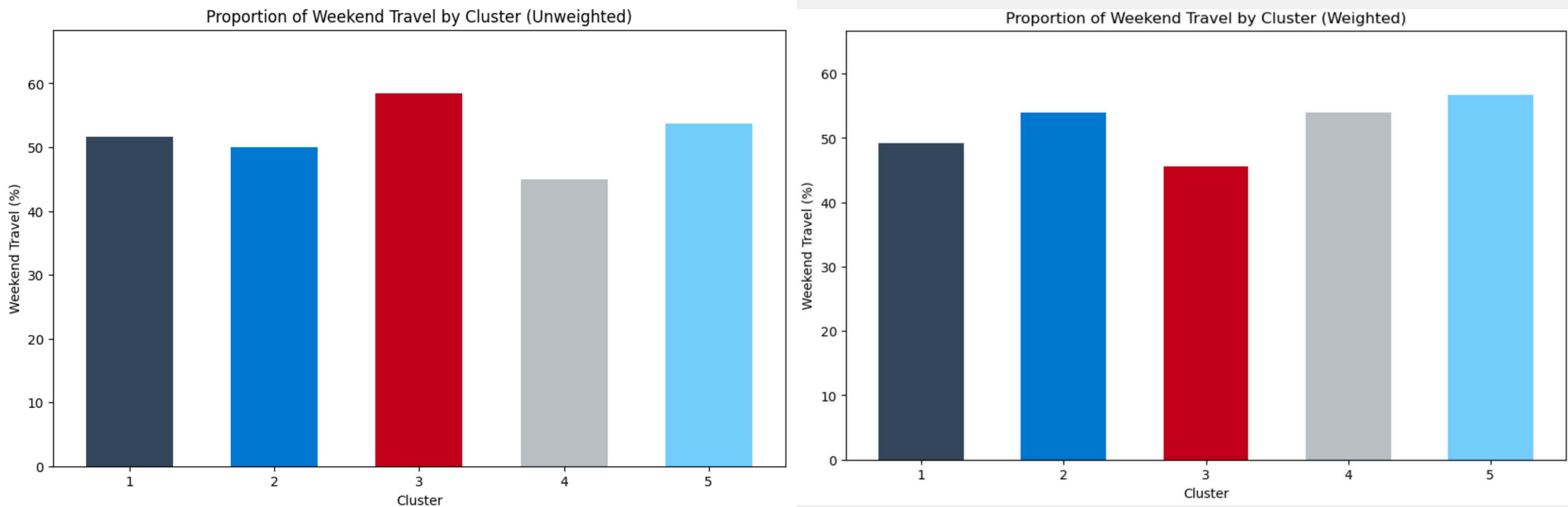
- Weighting highlights high-value customer clusters, such as Clusters 2 and 4
- Cluster 3 median and spread decrease in the weighted plot, indicating that customers may contribute less to high-revenue behaviors once weighting is considered.

ADVANCED PURCHASE ACROSS THE CLUSTERS



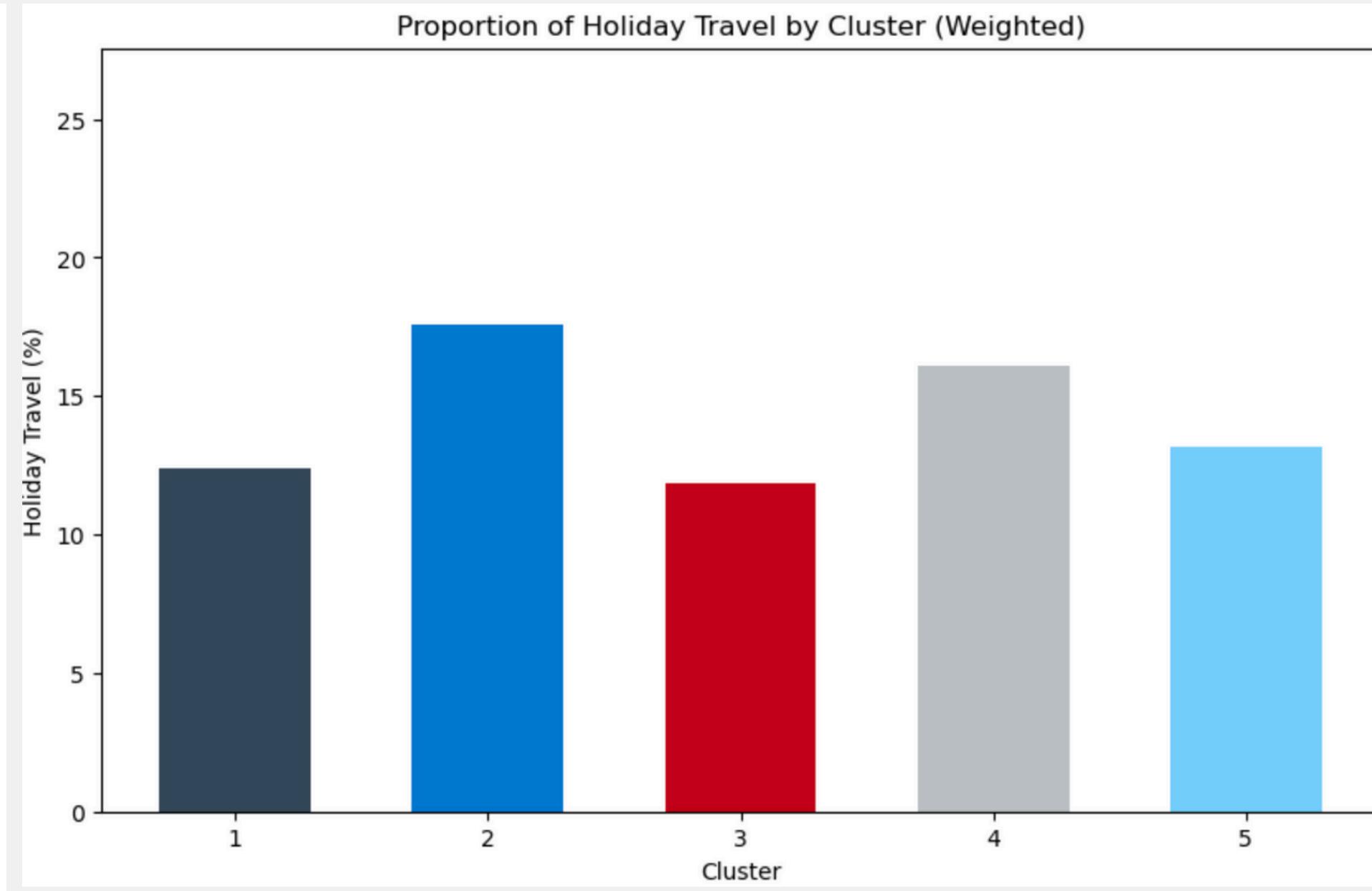
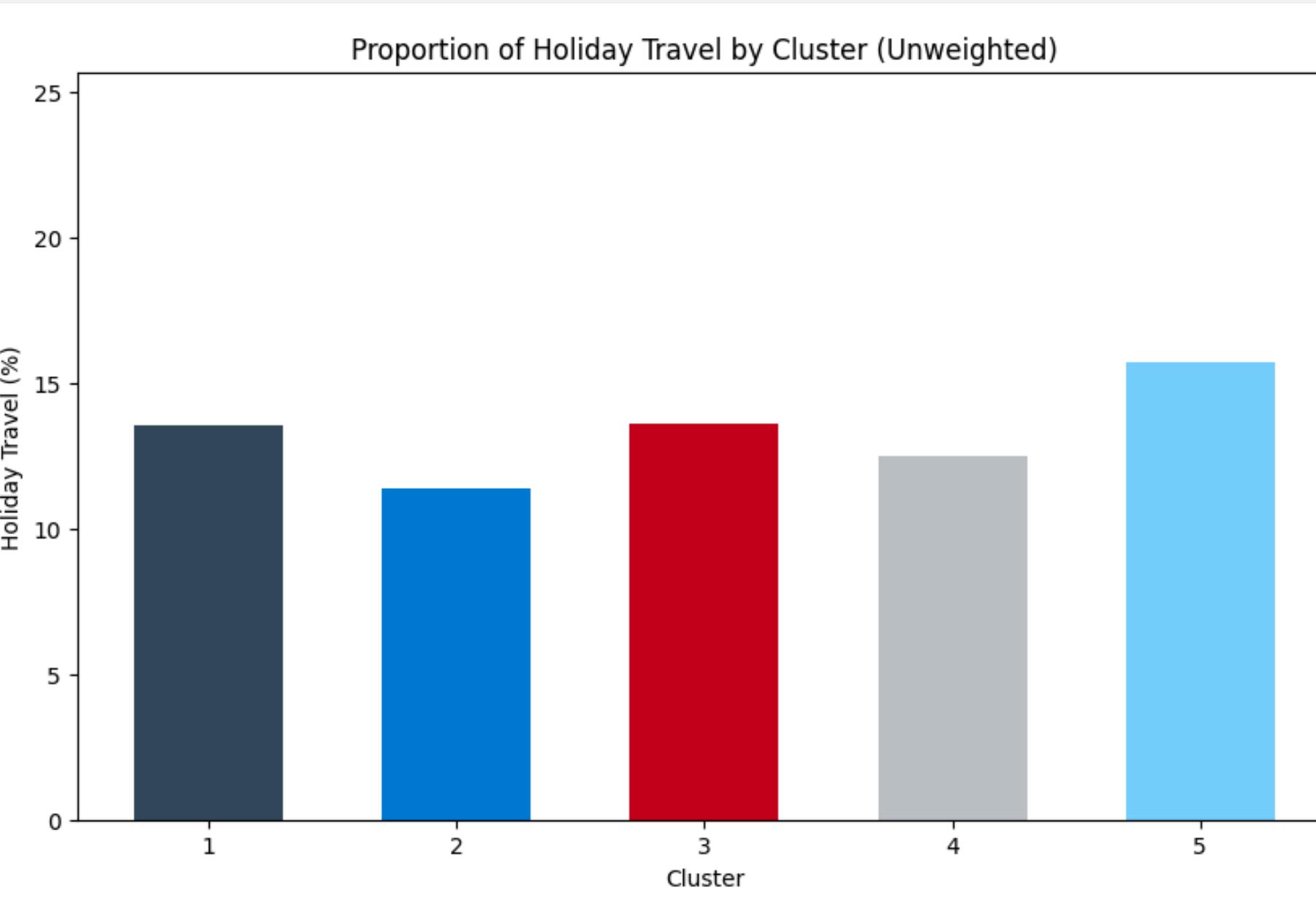
- Lower advanced purchase behavior for Clusters 3, 4, and 5 for both plots, reducing the overall spread and concentrating values toward the lower end
- Cluster 2 exhibits an increase in outliers and a lower median, highlighting more customers making last-minute purchases

TRAVEL PATTERNS ACROSS THE CLUSTERS FOR WEEKEND TRAVEL



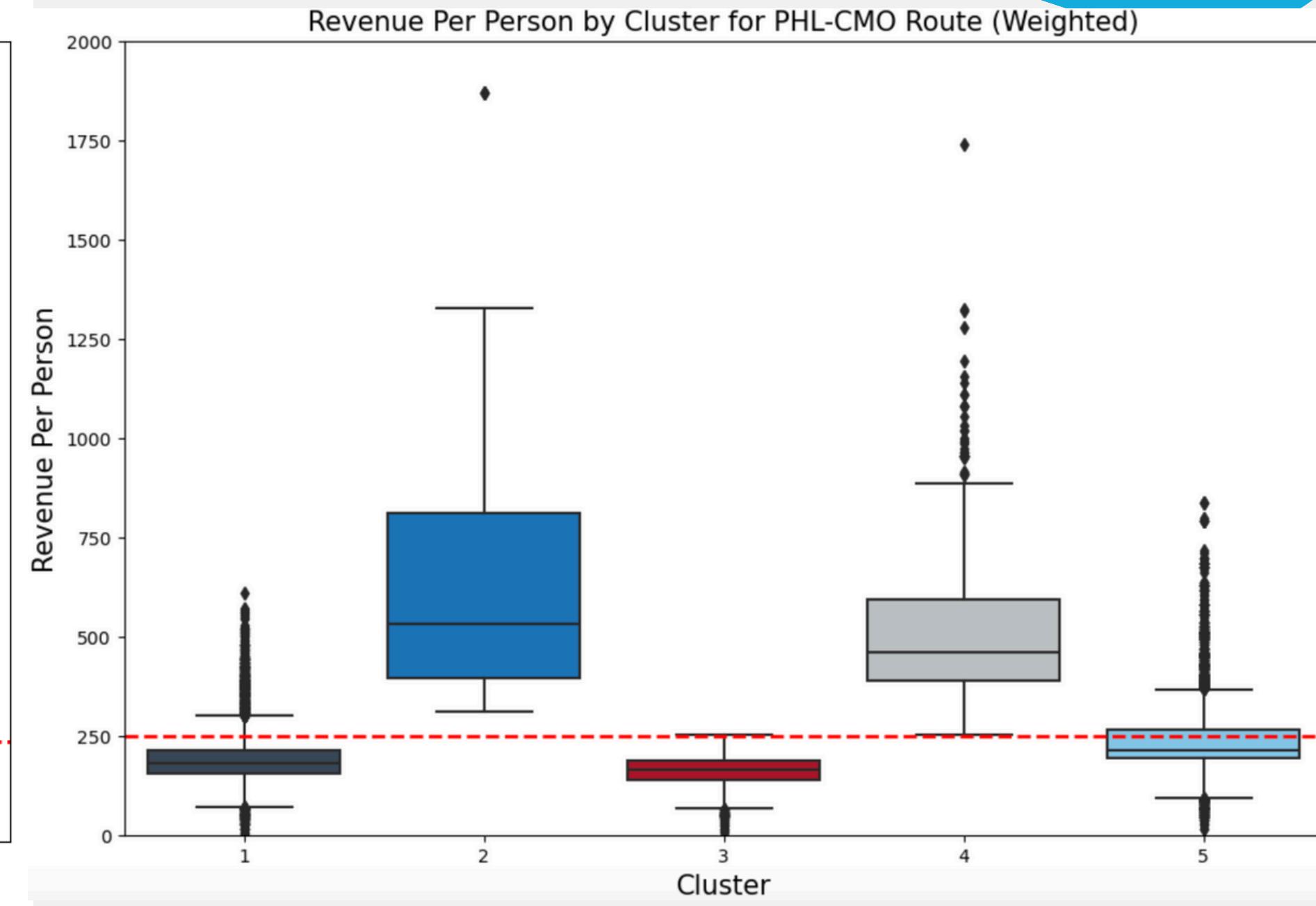
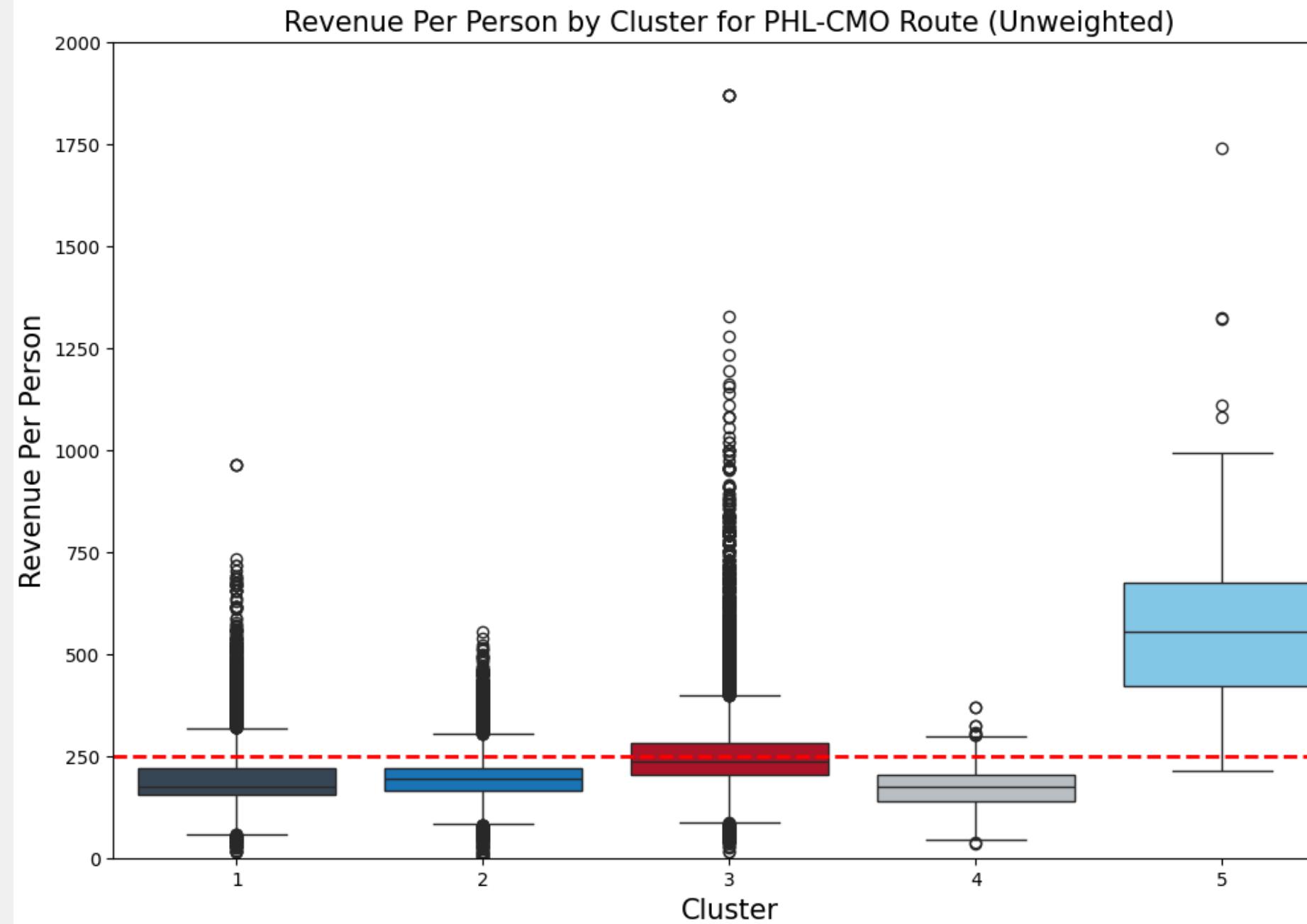
- Cluster 2's and cluster 4's proportion of weekend travel increases compared to its unweighted state, suggesting that these clusters are correlated with a higher willingness to pay
- Cluster 3 shows a reduced proportion of weekend travel in the weighted graph, implying that its weekend travel patterns are less correlated with high willingness to pay.

TRAVEL PATTERNS ACROSS THE CLUSTERS FOR HOLIDAY TRAVEL



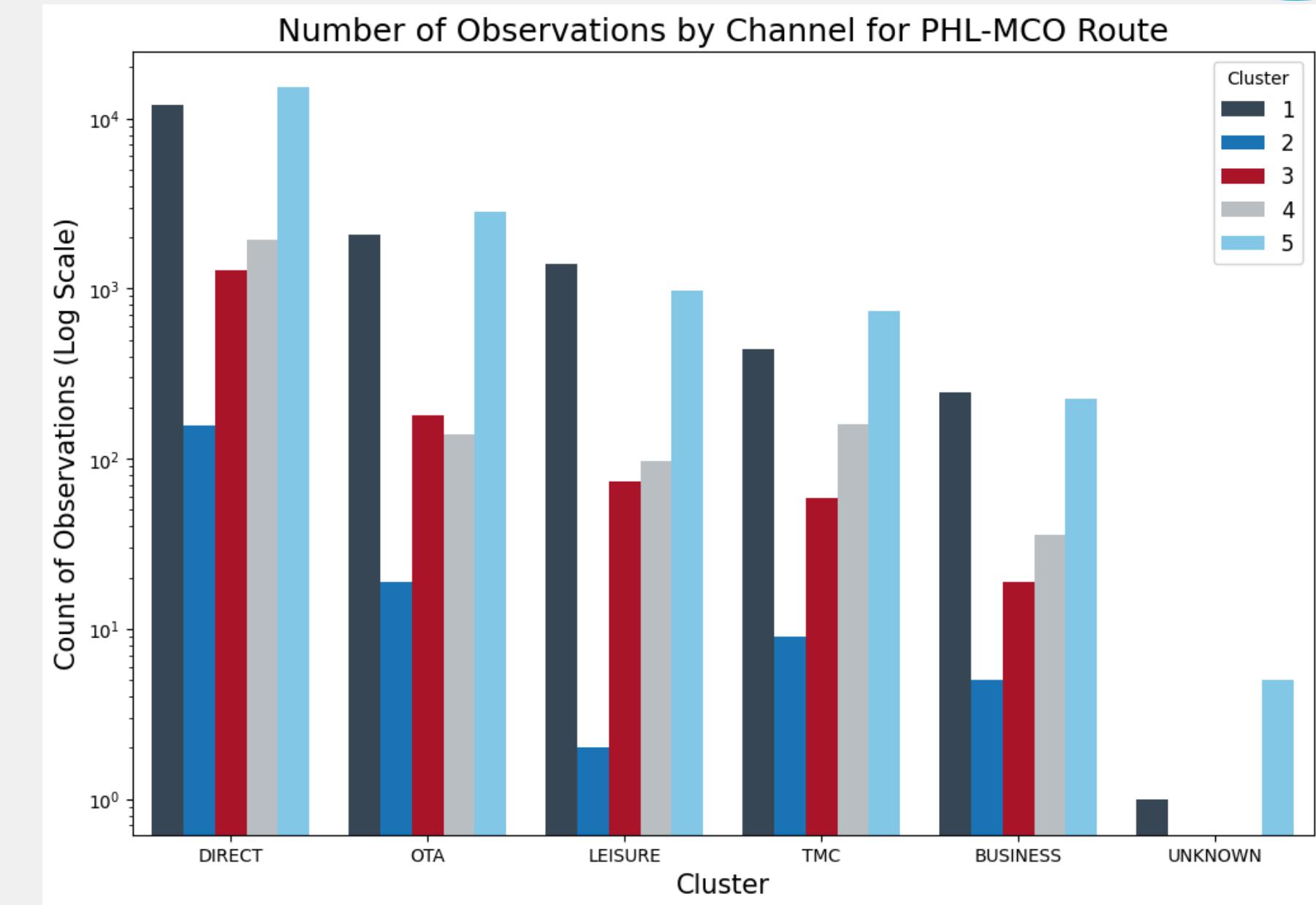
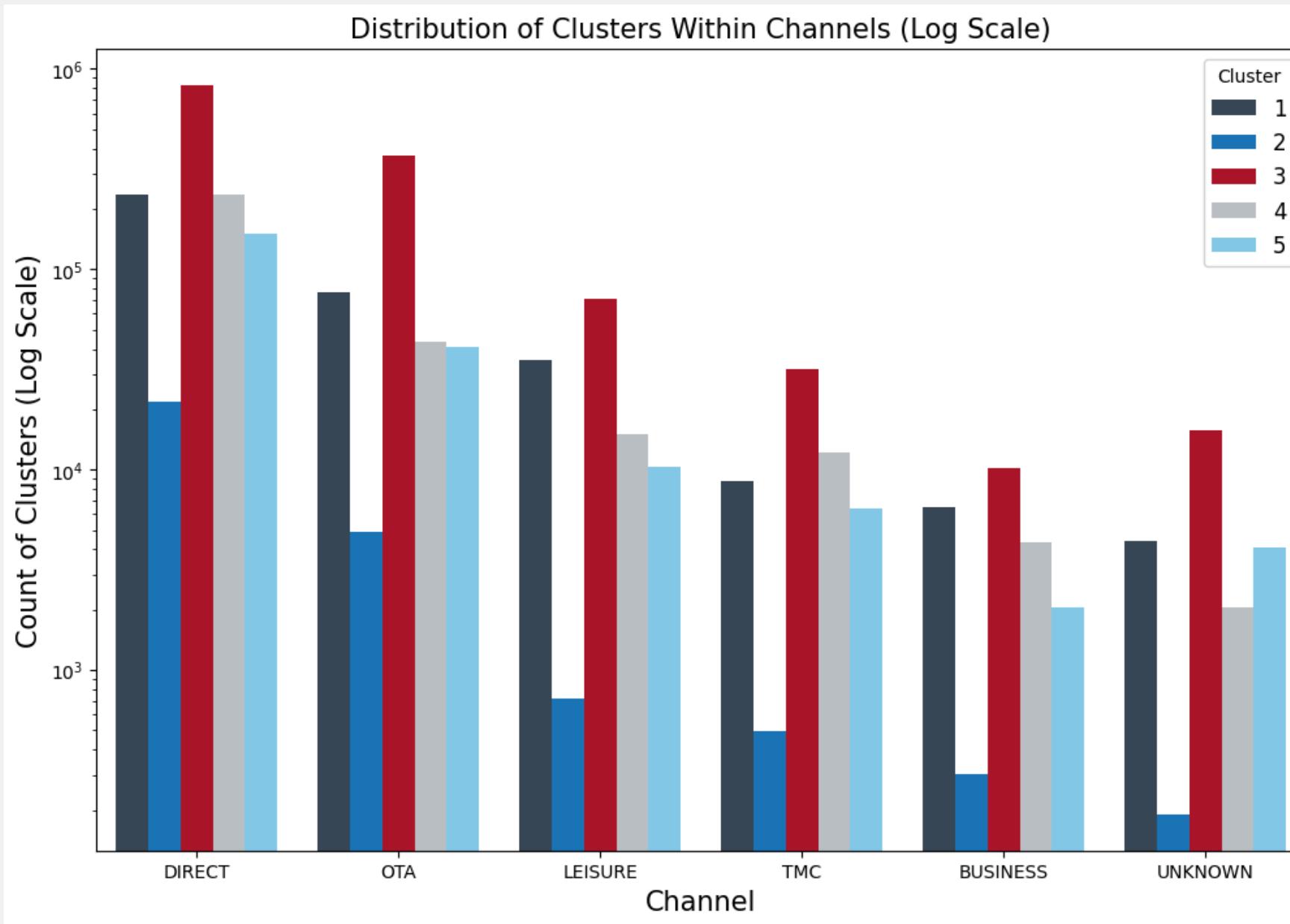
- Cluster 2 and Cluster 4 increase prominence, suggesting potential targeting for holiday promotions
- Cluster 1, 3, and 5 reduced importance may indicate limited value for high-spending holiday campaigns

CUSTOMER SEGMENTS IN THE PHL-MCO SECTOR



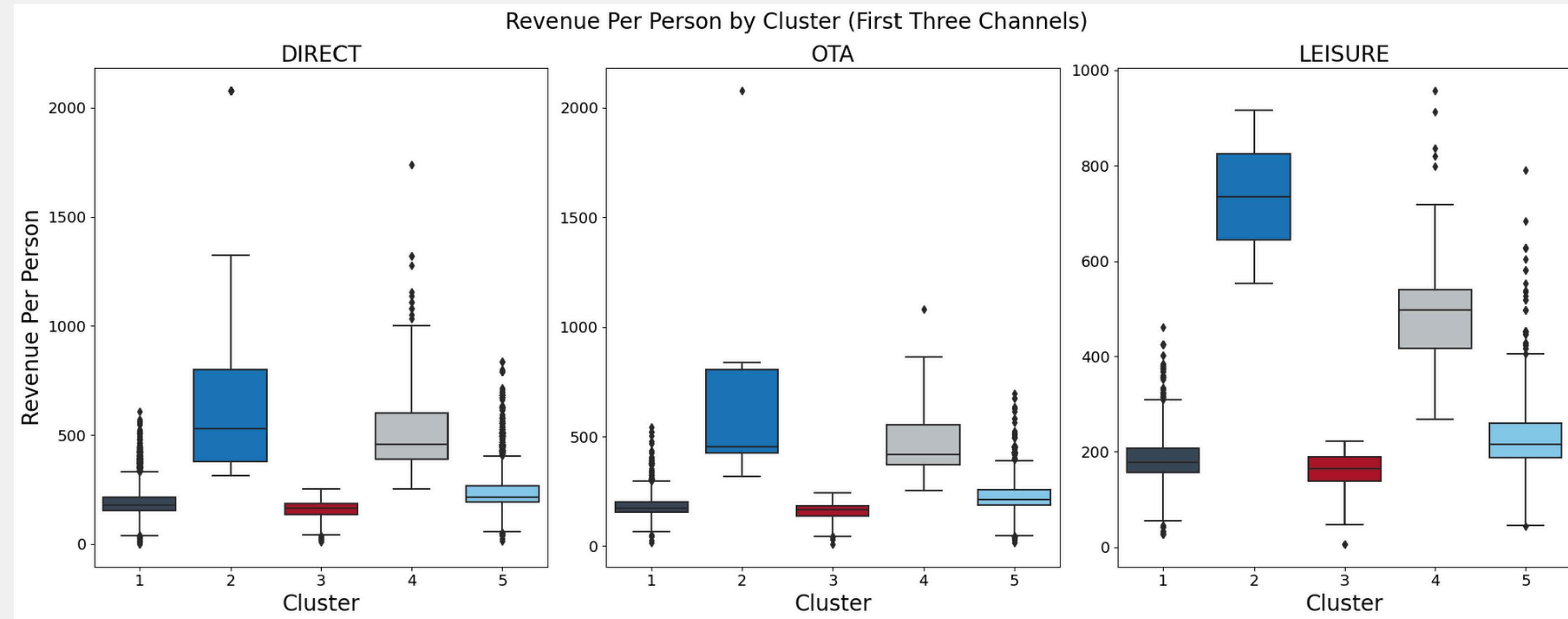
- For Unweighted clustering, Cluster 5 shows highest RPM and Revenue Per Person
- For Weighted clustering, Clusters 1 and 3 have less price points than the average price a customer pays for a PHL-MCO flight

CLUSTER BEHAVIOR BY CHANNEL FOR PHL-MCO ROUTE



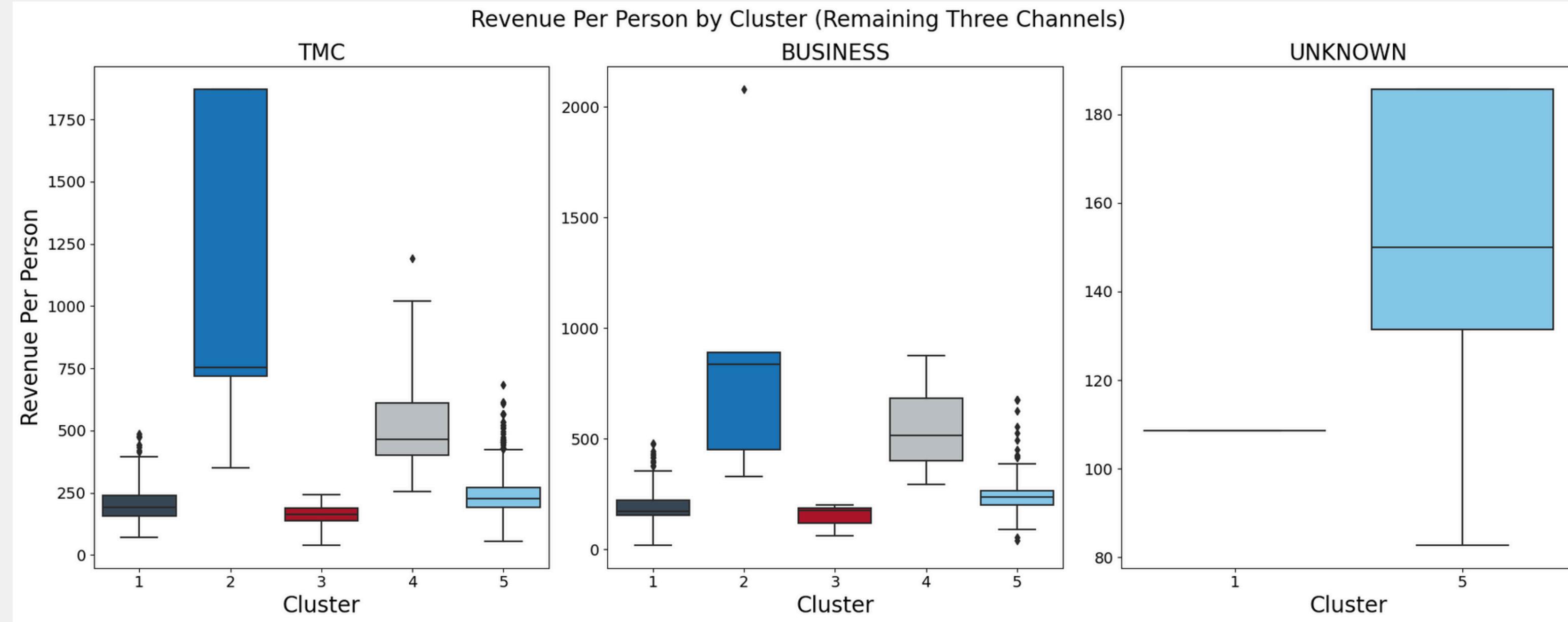
- **High Counts in Direct Channel:** Cluster 1 has the highest concentration in the Direct channel, indicating a strong consumer preference or reliance on this channel.
- **Leisure and OTA Channels:** Clusters 3 and 4 show higher counts in the Leisure and OTA channels, suggesting these clusters are more engaged with indirect booking methods.

REVENUE ANALYSIS ACROSS CLUSTERS AND CHANNELS



- Clusters 2 and 4: Higher revenue per person in Direct and OTA channels and even higher in Leisure channel, highlighting high-value travelers.
- Clusters 1 and 3: Consistently lower revenue across Direct, OTA, and Leisure channels, indicating lower spenders.

REVENUE ANALYSIS ACROSS CLUSTERS AND CHANNELS

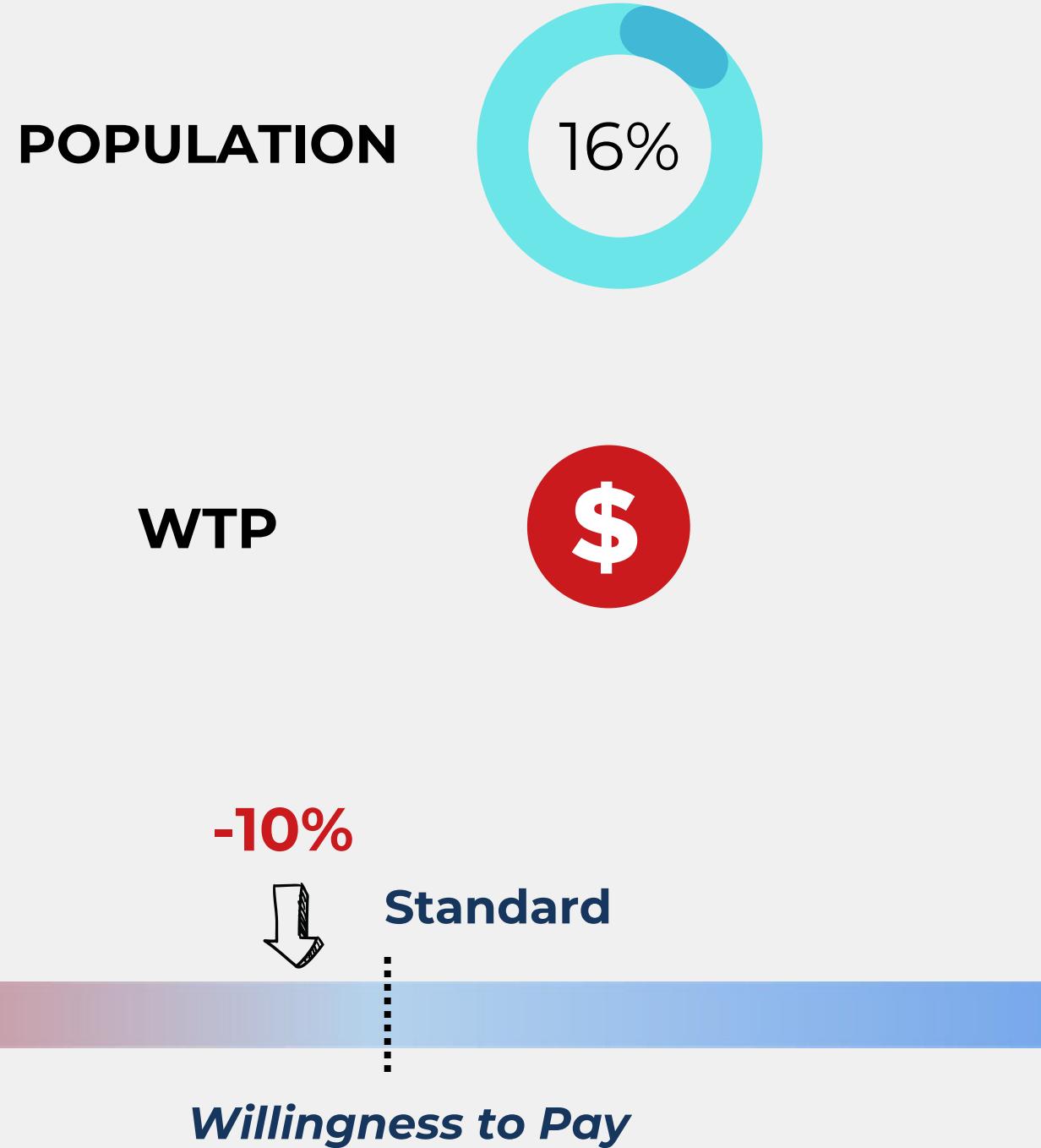


- **TMC Channel: High revenue variability, with Cluster 4 showing potential high-value travelers.**
- **Stable Revenue in Business Channel: Cluster 2 leads in the Business channel with consistent revenue, reflecting reliable engagement from this segment.**

CLUSTER 1

Budget Planners

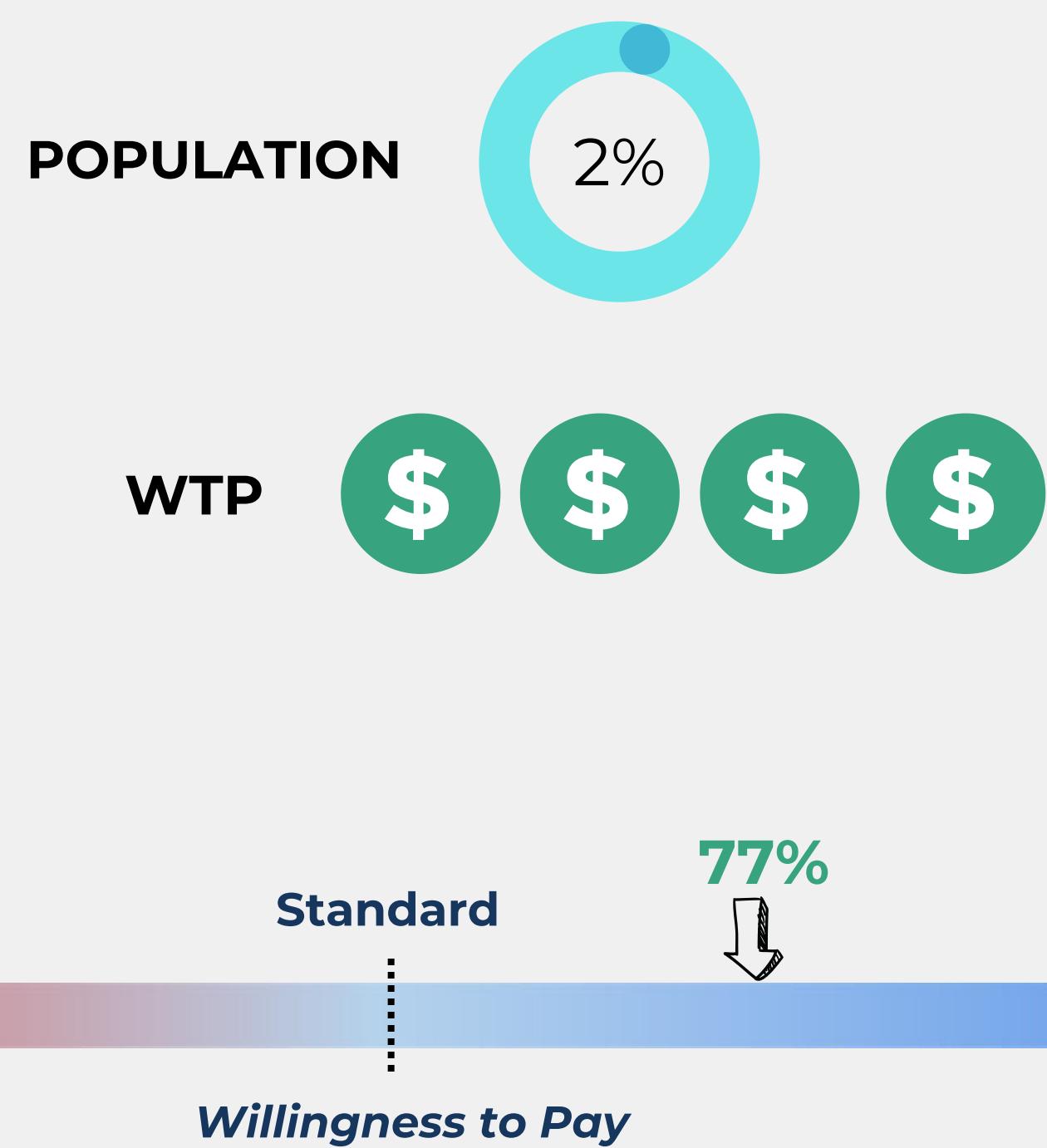
- **Characteristics:** Budget-conscious, prioritize affordability, and book in advance. Biggest segment of advance booking
- **Revenue Contribution:** Low revenue per mile and per person.
- **Travel Patterns:** Occasionally travel on weekends, but avoid peak holiday periods to minimize costs.
- **Willingness to Pay (WTP):** -10% below standard, indicating a price sensitivity.
- **Channel Pattern:** Prefers Direct channel for low-cost options.



CLUSTER 2

Premium last minute

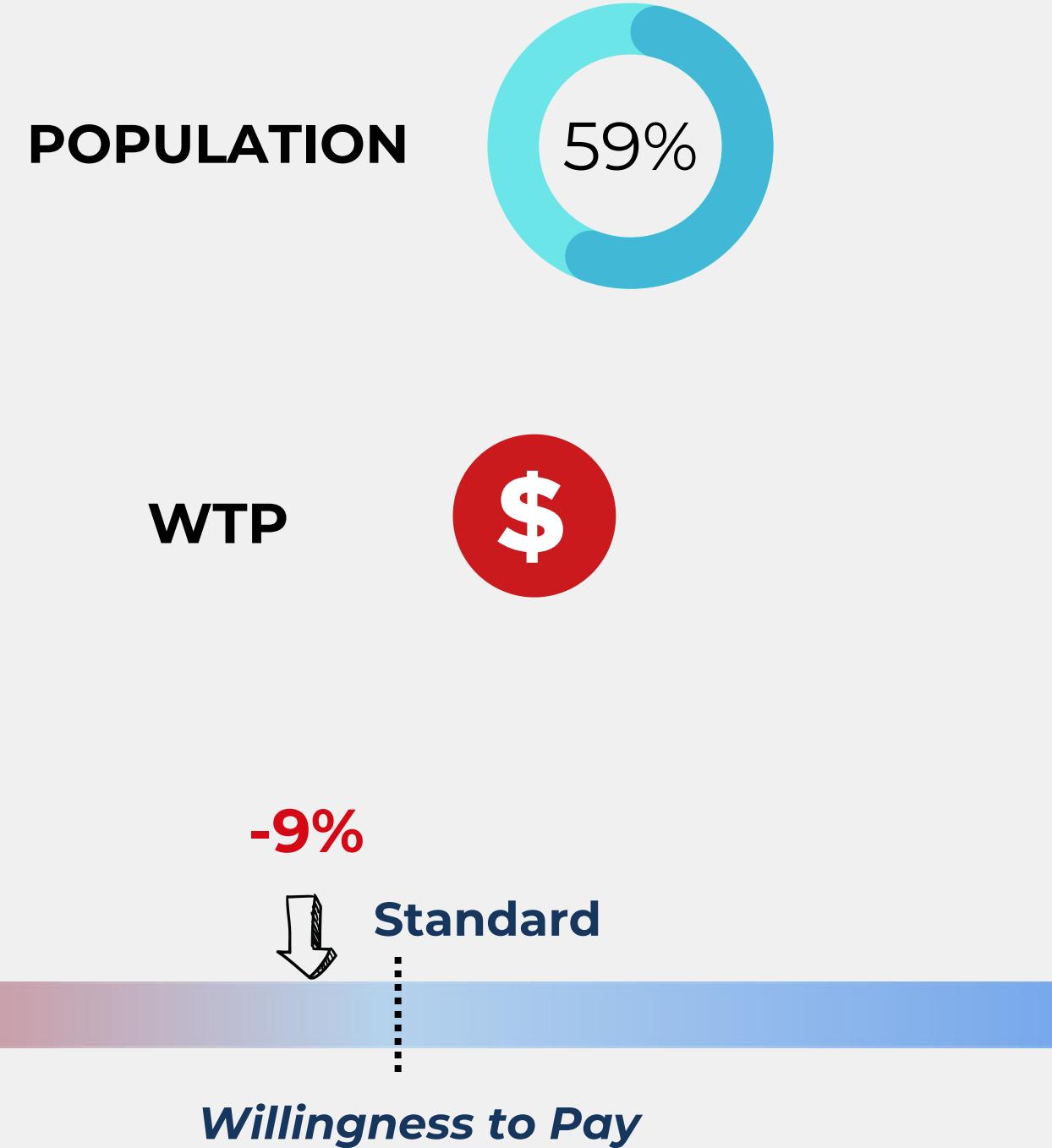
- **Characteristics:** High-spending, book close to departure and travel during peak times.
- **Revenue Contribution:** High revenue per mile and per person, willingly paying above the average, with significant spending on add-ons and upgrades.
- **Travel Patterns:** Prefer weekend and holiday travel and book extremely late
- **Willingness to Pay (WTP):** 77% above standard, the most lucrative segment for high-yield bookings.
- **Channel Pattern:** High engagement with OTA and Leisure channels for flexible bookings.



CLUSTER 3

Careful Planners

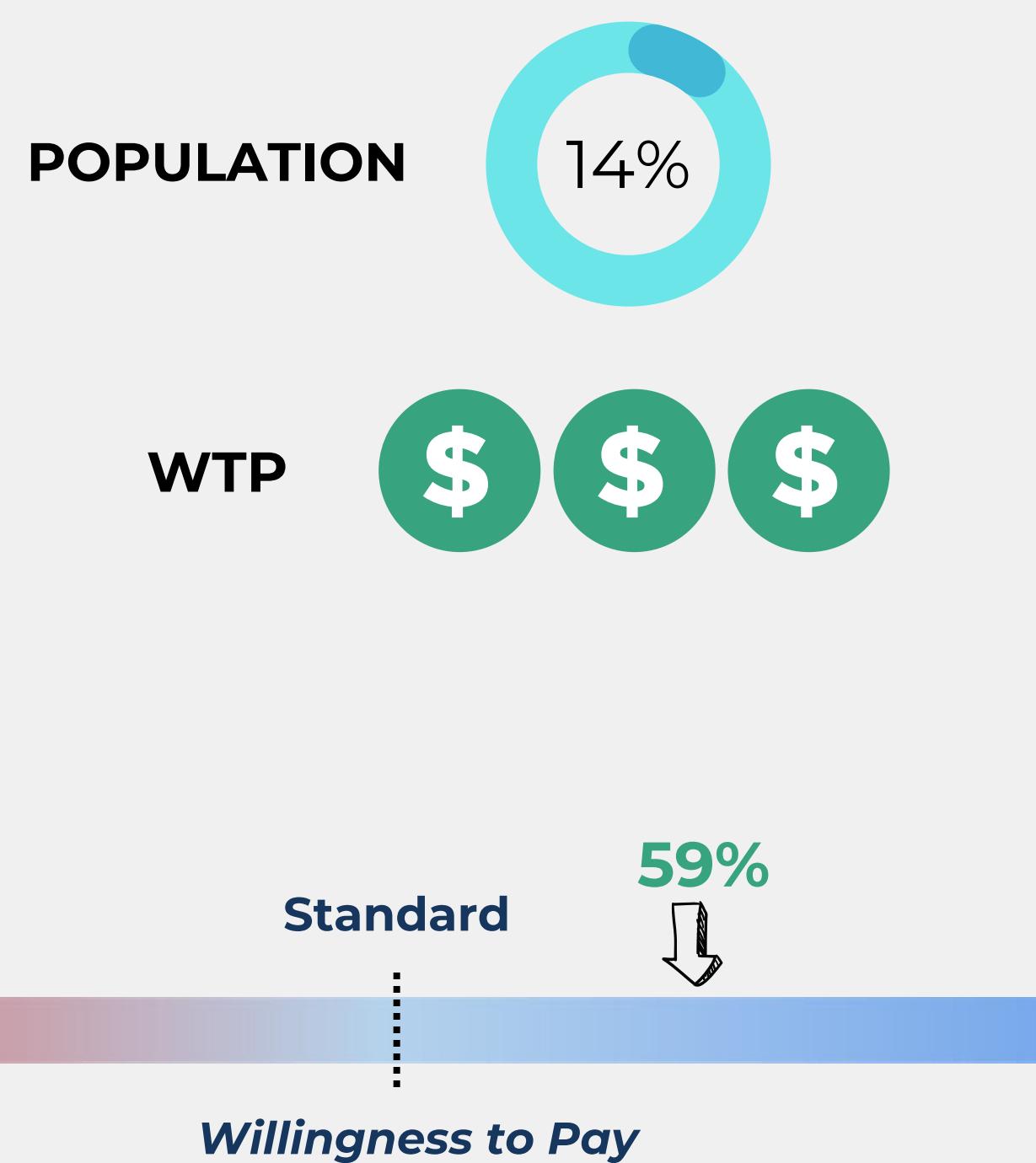
- **Characteristics:** Price-sensitive, flexible but avoid premium costs, and plan closer to the travel date.
- **Revenue Contribution:** Low to moderate revenue, avoiding high-cost travel options.
- **Travel Patterns:** Avoid weekend travel, indicating a preference for less busy travel periods.
- **Willingness to Pay (WTP):** -9% below standard, reflecting a reluctance to pay premium prices.
- **Channel Pattern:** Limited use of OTA, favors planned bookings.



CLUSTER 4

Premium Payers

- **Characteristics:** Varied behaviors, mix of advance and last-minute bookings, willing to pay high nevertheless
- **Revenue Contribution:** High revenue per mile and per person
- **Travel Patterns:** Prefer weekend and holiday travel, showing some willingness to pay extra.
- **Willingness to Pay (WTP):** 59% above standard, suggesting strong premium purchases.
- **Channel Pattern:** Balanced use of Direct, OTA, Leisure and Business channels.

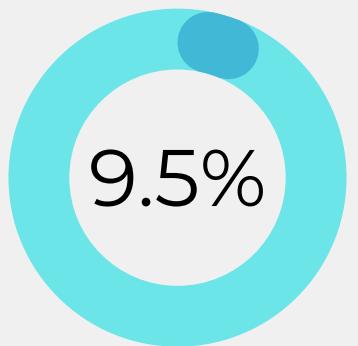


CLUSTER 5

Weekend Travelers, Moderate Spenders

- **Characteristics:** Prioritize convenience, make last-minute bookings, and frequently choose premium options.
- **Revenue Contribution:** Mixed revenue per mile and per person, occasionally opting for premium services.
- **Travel Patterns:** Largest segment of weekend travelers, frequently traveling during peak times.
- **Willingness to Pay (WTP):** 13% above standard, reflecting occasional preference for convenience and flexibility.
- **Channel Pattern:** Strong preference for premium and flexible bookings through TMC and Leisure.

POPULATION



WTP



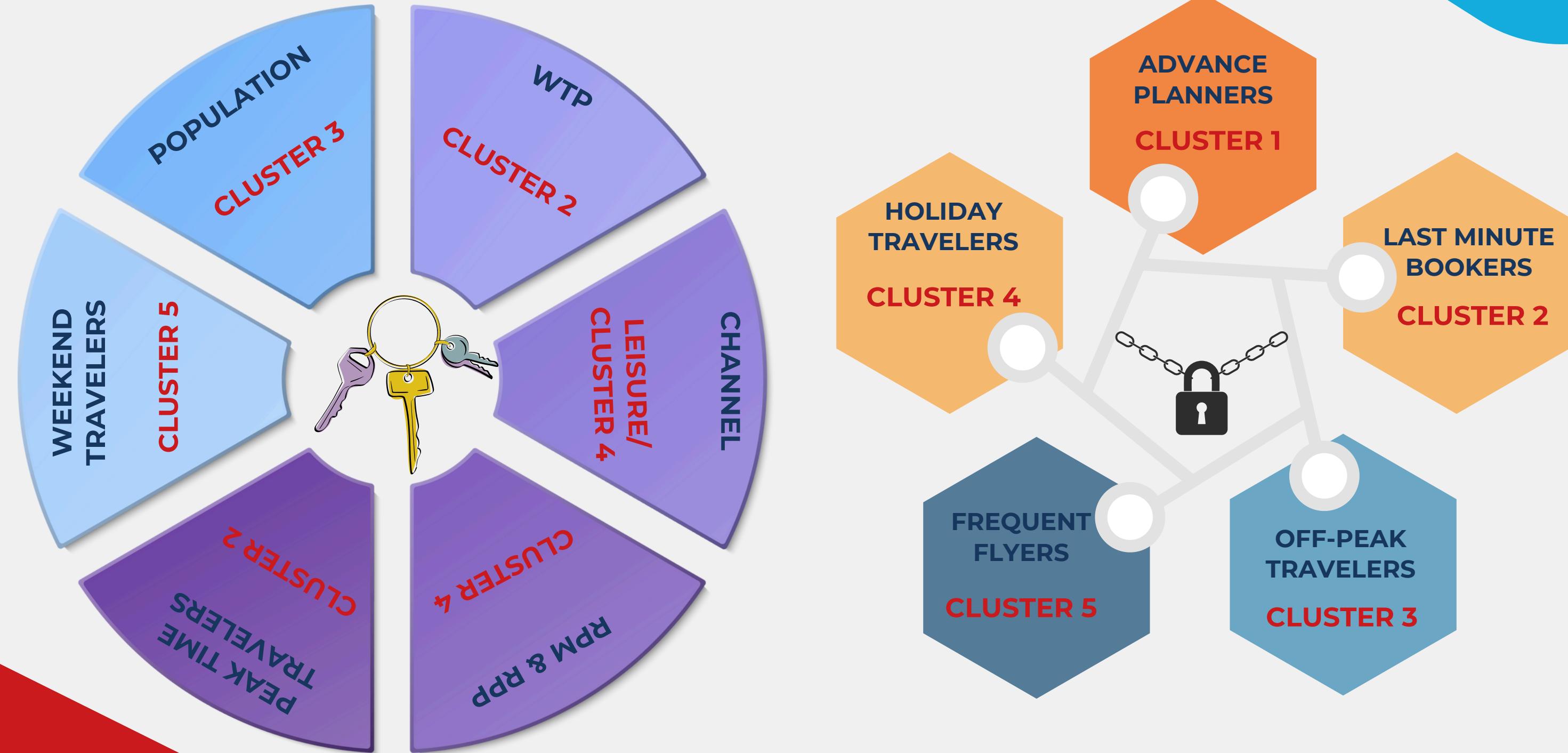
13%

Standard



Willingness to Pay

WHAT ARE ALL CLUSTERS TRYING TO TELL?



WHAT'S IN IT FOR DIFFERENT STAKEHOLDERS?

STRATEGIC RECOMMENDATIONS

1. Marketing Stakeholder:

- Target Clusters 2, 4 and 5 with premium service promotions and holiday campaigns.
- Develop tailored engagement strategies for high-value travelers, leveraging flexibility and convenience.

2. Business Ops Stakeholder:

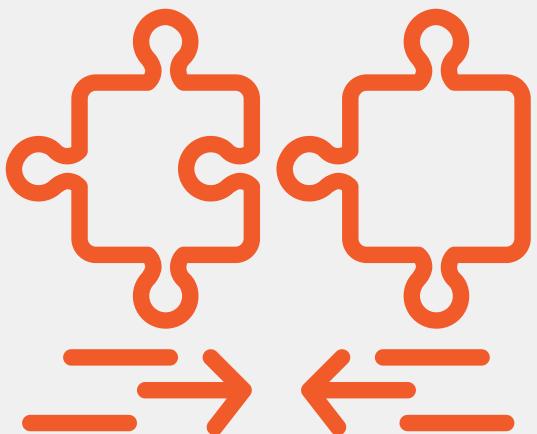
- Optimize seat inventory for peak times, focusing on high-revenue segments.
- Implement process improvements to support efficient booking and travel management.

3. Revenue Ops Stakeholder:

- Adjust pricing models to maximize yield from high WTP segments.
- Explore dynamic pricing for last-minute and premium bookings.

4. Long Term Planning:

- Develop a roadmap for ongoing customer segmentation updates, plan for future shifts in consumer behavior, ensuring relevance in changing market conditions.





FlyAhead Consultants

Thank You For All Your
Time & Attention!

 Search Destination

LET'S FLY!

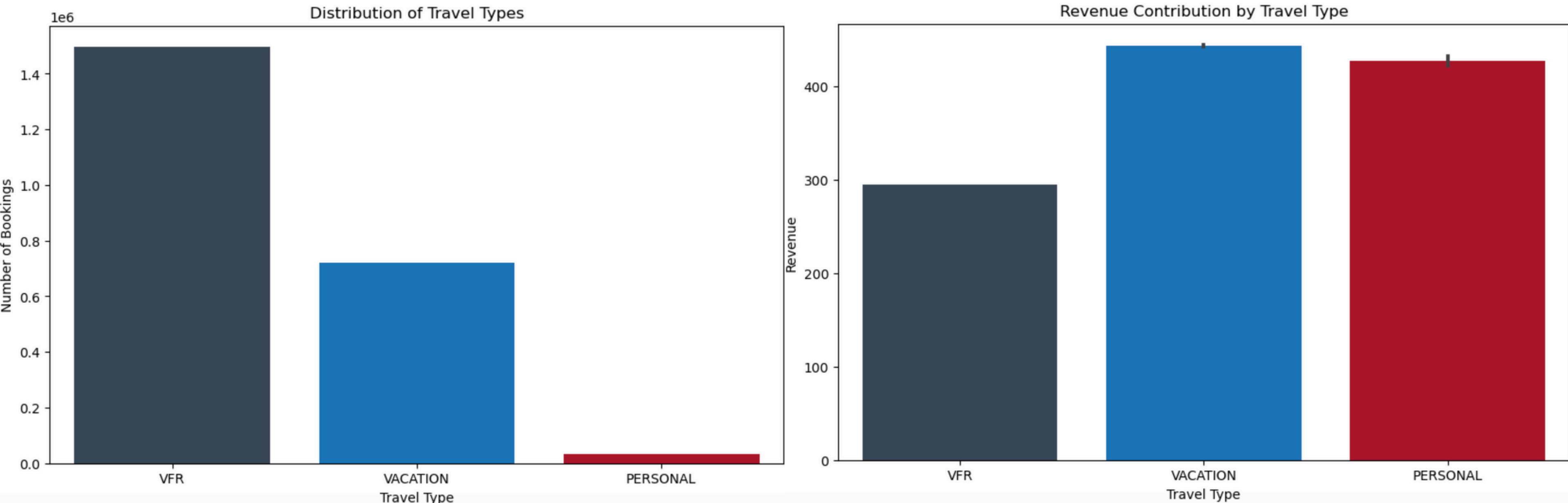
APPENDICES



Appendix 1:



Distributions: Travel Types & Revenue by Type (EDA)

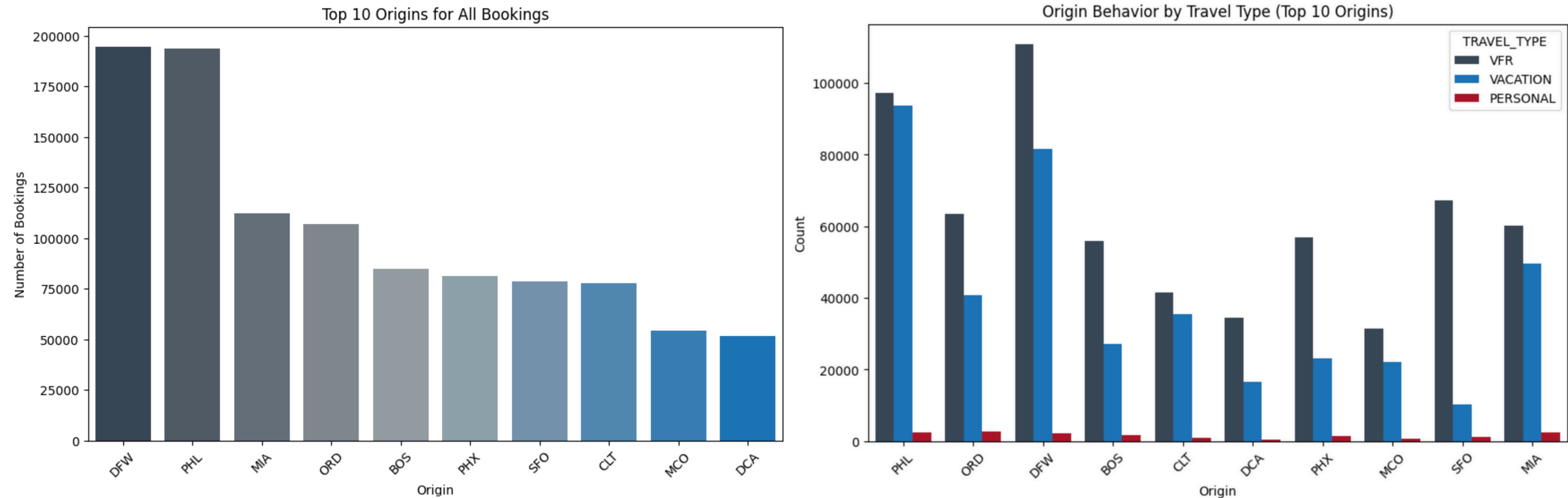


- **High Volume, Lower Revenue Per Booking**
- **Low Volume, High Revenue Per Booking**

Appendix 2:



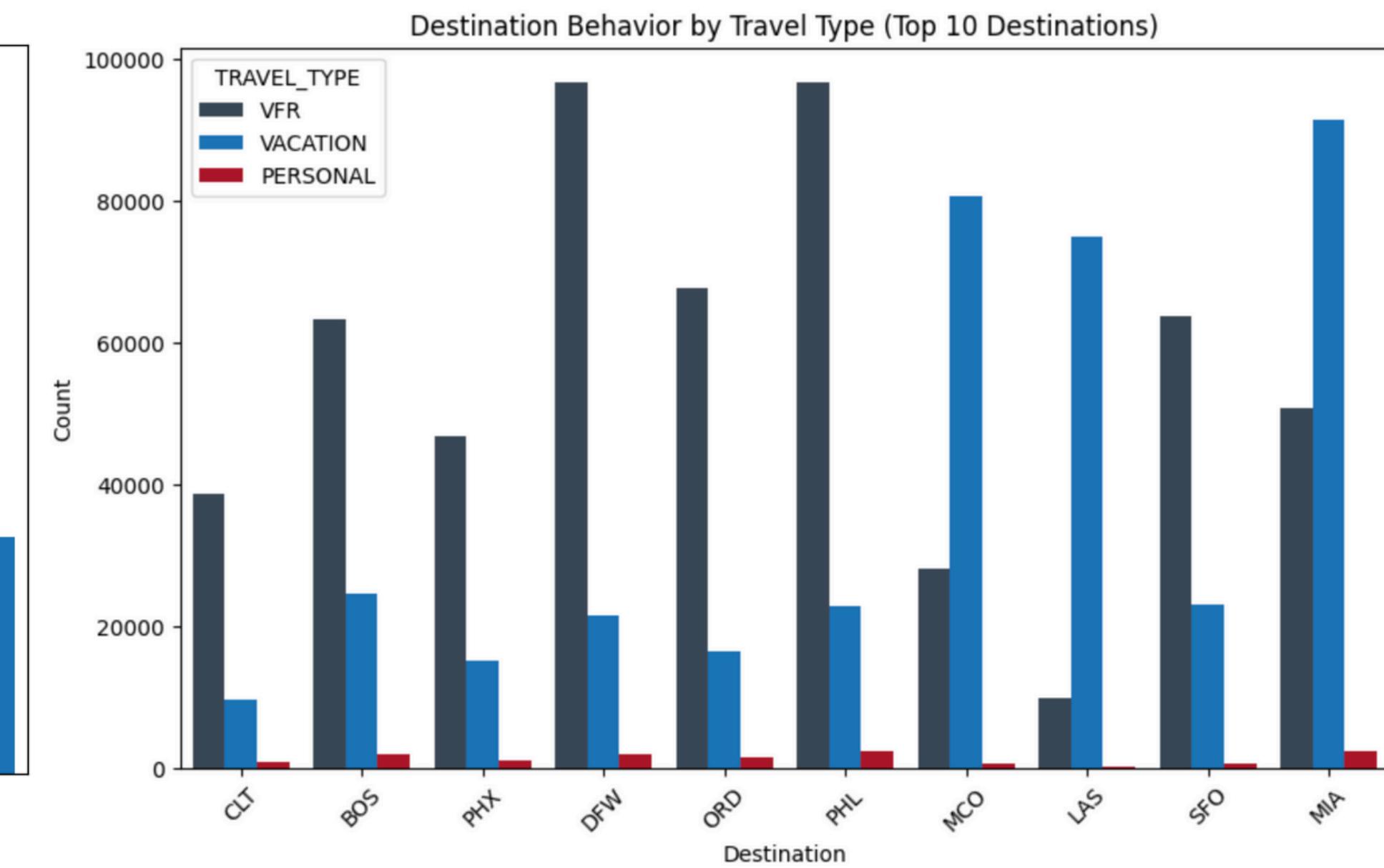
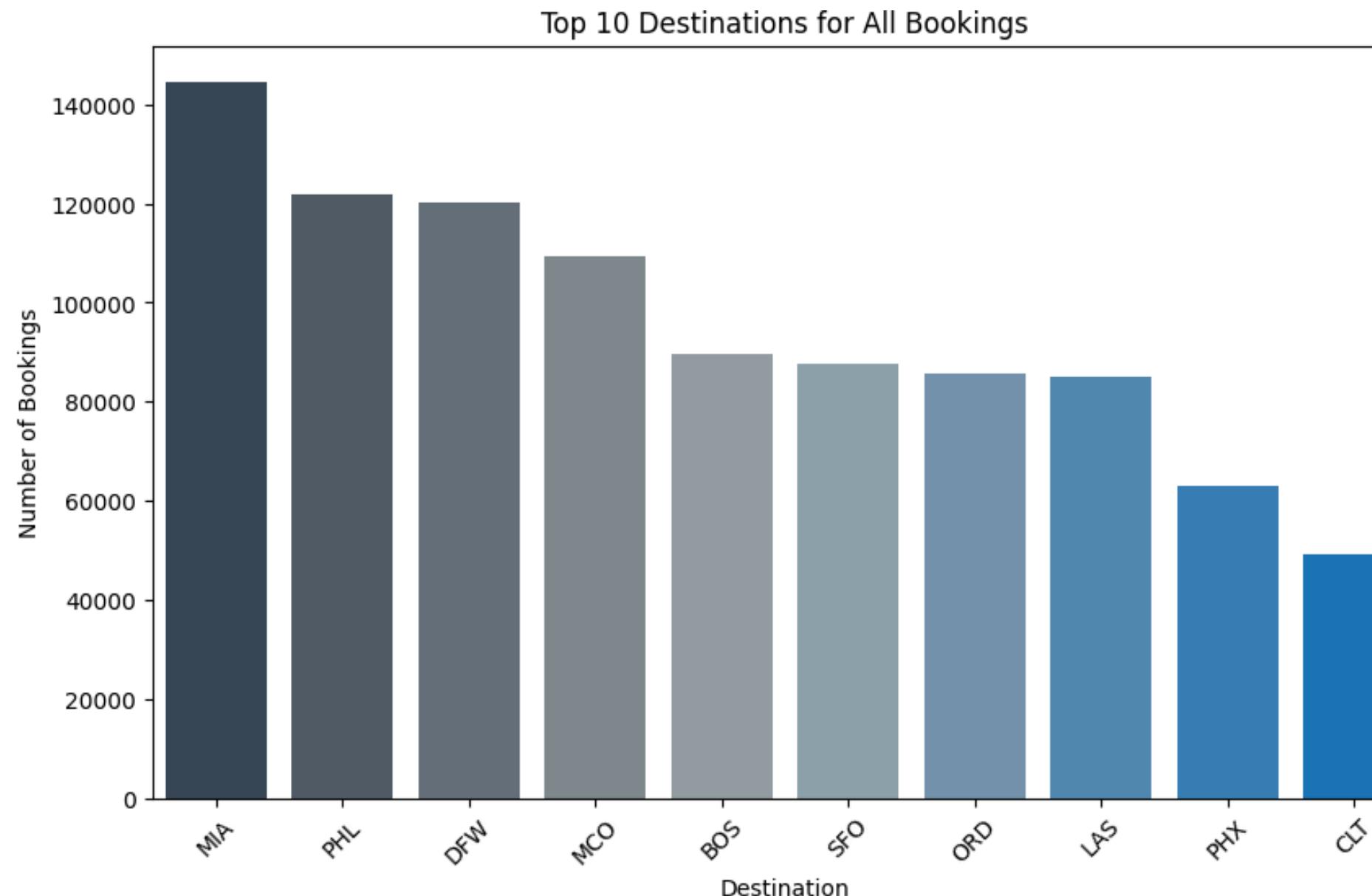
Bookings by Flight Origins (EDA)



- **High booking volume: Dallas and Philadelphia**
- **Dominance of VFR travel → these hubs are central to routes**

Appendix 3:

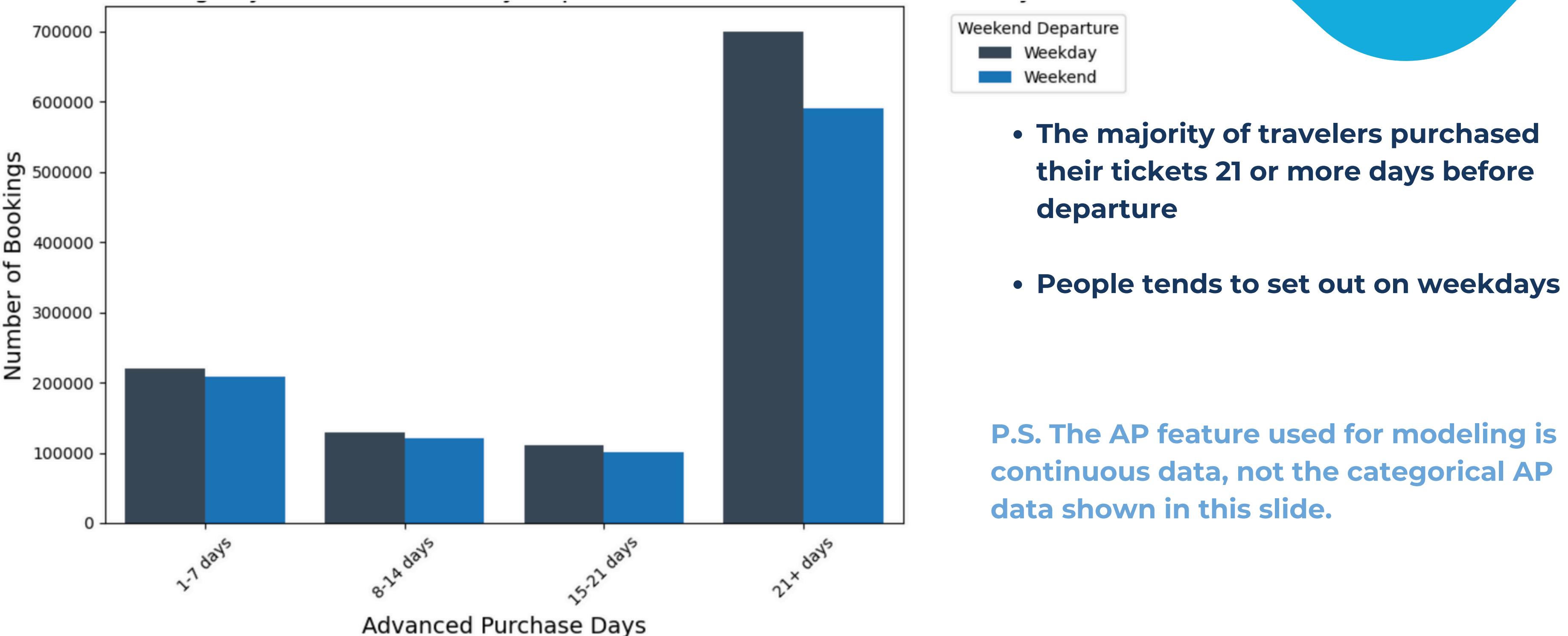
Bookings by Destinations (EDA)



- **Miami: prime hub for vacation**
- **VFR travel dominates most top destinations → family and friend connections**
- **Orlando and Las Vegas: strong vacation destinations**

Appendix 4:

Weekend/Weekday Travel by Advanced Purchase



Appendix 5: CLUSTER 2 - HIGHEST PAYERS

CLUSTER 2 - Highest Payers

Occurrence

0

500

