Customer Segmentation Analysis: Sun Country Airlines

Allan Almaraz

University of California, Irvine

BANA 200 - Foundations of Business Analytics

Professor Suvesh Balasubramanian

September 10, 2025

Abstract

Sun Country Airlines, a Minneapolis-based low-cost leisure carrier, faced a critical strategic challenge in 2014: leadership lacked validated customer profiles and relied instead on anecdotal assumptions to drive marketing and product decisions. Acting as analytics consultants, this report applies K-Means clustering (k = 5) to 15,144 customer-trip records spanning 2013 to 2014, segmenting travelers across 88 normalized behavioral, demographic, booking, and routing features. Five distinct, business-interpretable segments emerged: MSP Direct Bookers (27.3%), MSP Deal Seekers (25.4%), Inbound Visitors (16.1%), Loyal Frequent Fliers (15.7%), and Returning Home Fliers (15.5%). Analysis reveals that 78.5% of Sun Country customers are unenrolled in the Ufly Rewards program, that a near-complete channel split exists between direct-booking and third-party-booking travelers, and that a round-trip cross-sell opportunity exists between inbound and homebound segments. Three targeted campaigns are proposed, each directly addressing the three stated business goals of Warnken and Vaughan: growing Ufly enrollment, shifting bookings to the higher-margin direct channel, and developing differentiated vacation packages.

The report further demonstrates that exploratory data analysis alone surfaced the most actionable business insight, specifically the 78.5% non-Ufly rate, before any model was run. This reinforces the case that analytical rigor and business literacy, not model complexity, are what drive decision value.

Keywords: customer segmentation, K-Means clustering, loyalty programs, booking channel behavior, and airline marketing.

Introduction

Throughout the 1950s and 60s, a few dozen commercial airlines competed across North American skies. By 2014, a wave of consolidations had reduced that field to three dominant carriers: American Airlines, Delta, and United, each competing on price, scale, and network breadth. Amid this landscape, Sun Country Airlines — a Minneapolis-St. Paul-based leisure carrier founded in 1983 — confronted a distinctly different challenge. The problem was not one of size or scale, but of knowledge. Specifically, Sun Country lacked data-validated customer intelligence.

Sun Country's leadership, represented by Senior Director of e-Commerce Michael Warnken and Director of Customer Digital Experience Roselie Vaughan, recognized in late 2014 that the company's customer assumptions were entirely anecdotal, with no dedicated data analyst on staff to challenge or validate them. Warnken and Vaughan identified three concrete strategic goals they needed customer data to address. First, Vaughan wanted to improve the digital booking experience so that the online channel met modern traveler expectations. Second, both executives recognized that most customers were unenrolled in Sun Country's Ufly Rewards loyalty program, representing a significant missed retention opportunity. Third, Warnken believed Sun Country could not compete on price alone against the major carriers and needed segment-specific vacation package offerings to drive differentiation.

To pursue these goals, Warnken and Vaughan sourced existing data from the Ufly Rewards program database and TSA flight records, representing 1.52 million customers making 1.86 million trips between January 2013 and December 2014. A subset of this data, 15,144 customer-trip records across 90 features, was prepared for analytical modeling and forms the basis of this report.

This report addresses the case questions directly. It evaluates data usability, conducts K-Means clustering to identify five customer segments, visualizes those segments to tell a coherent story about Sun Country's customer base, and translates analytical findings into prioritized business recommendations tied explicitly to Warnken and Vaughan's stated objectives.

Data Overview

One dataset was used in this analysis. Clustering_Data.csv contains 15,144 customer-trip records across 90 features and serves as the sole data source for all segmentation, visualization, and profiling conducted in this report. The dataset arrived in a highly prepared state: all categorical variables had been one-hot encoded, all numeric values were normalized to a 0-to-1 scale, and no missing values were present. For the purposes of clustering, the two identifier columns (uid and PNRLocatorID) were excluded from the feature matrix, leaving 88 usable numeric features.

Those 88 features span six conceptual categories: booking behavior variables including avg_amt, round_trip, group_size, and days_pre_booked; six binary booking channel indicators (SCA Website, Outside/3rd-Party, Reservations, Tour Operator, SY Vacation, and Other); five age group indicators; 28 origin and 37 destination airport dummies; three Ufly membership status indicators (Non-member, Standard, and Elite); and four quarterly seasonality indicators.

While the data is clean and analytically ready, it is limited in ways that matter for longer-term business use. The dataset captures transaction-level behavior but omits ancillary spend such as bag fees or seat upgrades, post-flight satisfaction scores, and return visit frequency. These gaps are addressed in the Long-Term Data Recommendations section.

Exploratory Data Analysis

Before running any clustering model, exploratory data analysis (EDA) was conducted to understand the distribution of key business variables and to identify patterns that would inform segment interpretation.

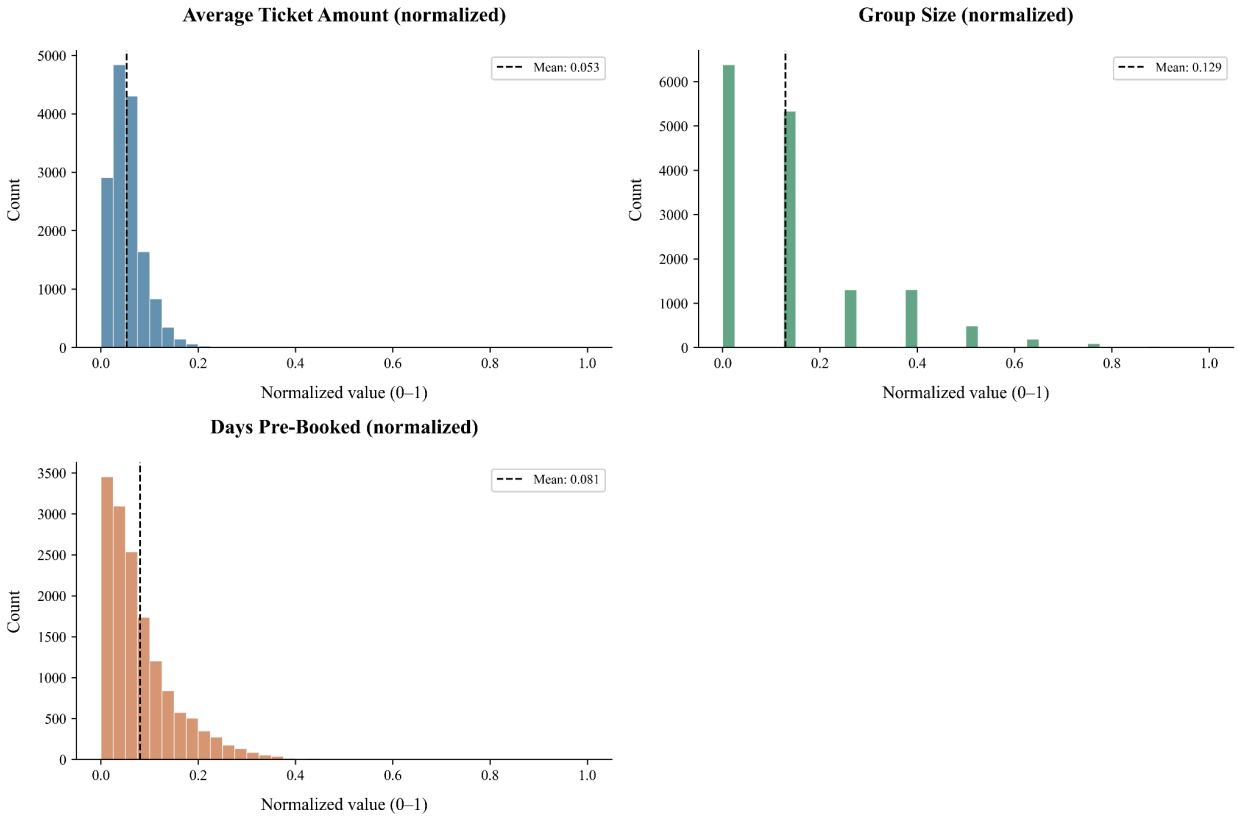
Continuous Variable Distributions

The three primary continuous variables (avg_amt, group_size, and days_pre_booked) were examined via histogram plots, which were normalized between 0 and 1 (see Figure 1). The average ticket amount is heavily concentrated near 0, consistent with Sun Country's positioning as a budget leisure carrier. Group size similarly clusters near 0, with discrete spacing in the histogram reflecting whole-number party sizes and solo or two-person bookings being by far the most common. Days pre-booked shows a right-skewed distribution, with the majority of bookings made relatively close to the departure date (mean = 0.081), though a long tail to the

right indicates that a smaller subset of customers book significantly further in advance. Taken together, these distributions describe a customer base that is largely price-conscious, travels in small groups, and tends to book close to departure — all consistent with Sun Country's core identity as a low-cost leisure carrier.

Figure 1

Distributions of key continuous variables (average ticket amount, group size, and days pre-booked) across all 15,144 customer-trip records.



Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

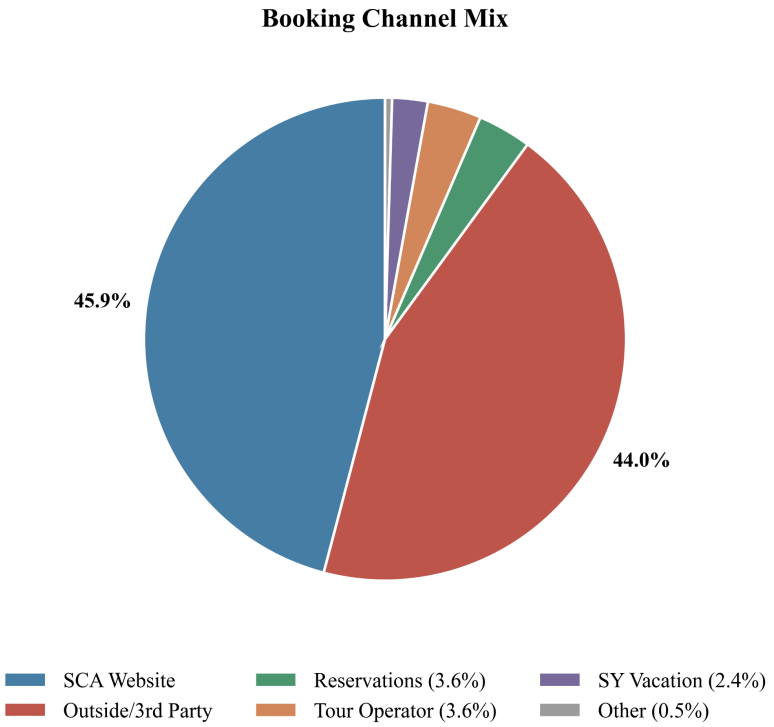
Booking Channel & Loyalty Program Distribution

One of the most consequential findings from EDA was the near-equal split between direct and indirect booking channels. Approximately 45.9% of bookings were made through the SCA website, while 44.0% were made through outside or third-party platforms such as Expedia or Google Flights (see Figure 2). This near-even divide was visible before any model was run and

strongly suggested that booking channel would be a major axis of segmentation — a hypothesis K-Means later confirmed.

Figure 2

Booking channel distribution across all 15,144 customer-trip records.

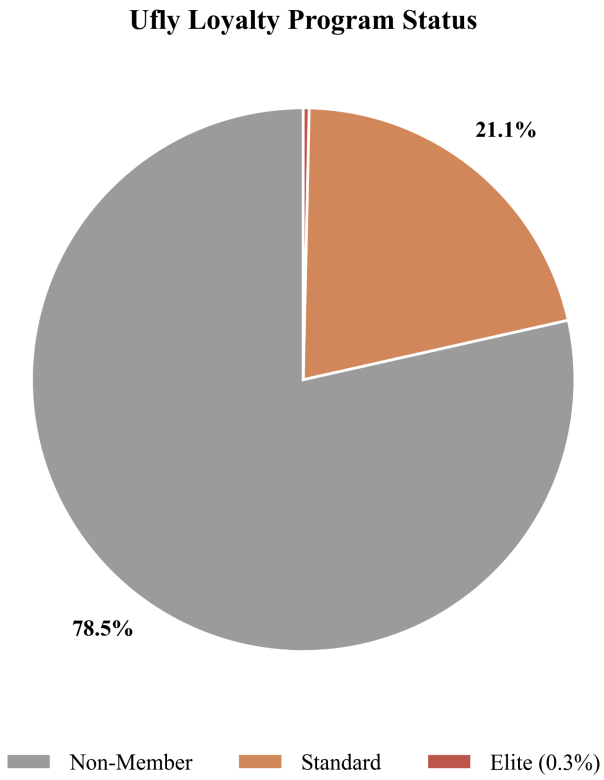


Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

The Ufly loyalty analysis was equally striking (see Figure 3). A full 78.5% of all customers in the dataset are non-Ufly members, with 21.1% holding Standard status and only 0.3% at the Elite tier. This finding fundamentally reframes the primary business question. Rather than asking only who the customers are, the data demands an answer to which segments are most convertible to Ufly membership — and what specific intervention each segment requires.

Figure 3

Ufly loyalty program membership status across all 15,144 customer-trip records.



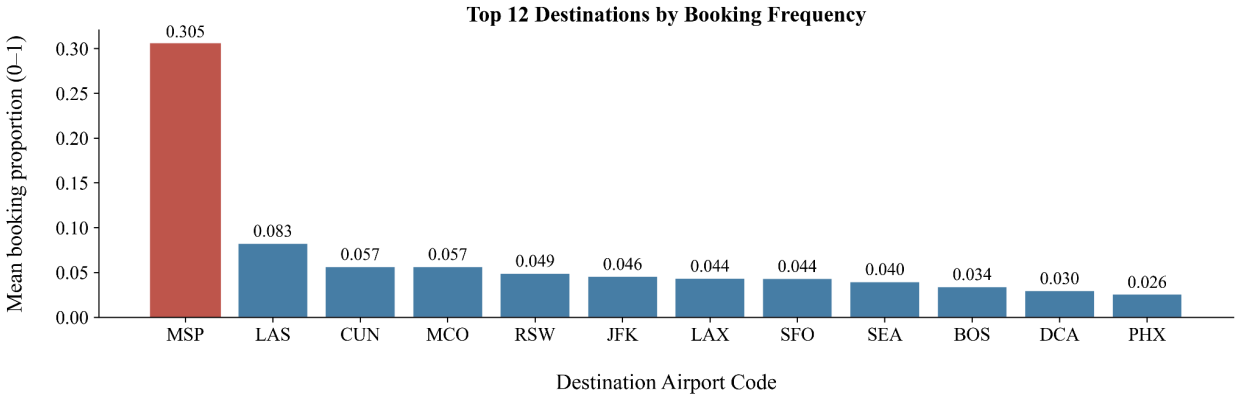
Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Destination Geography

Examination of destination airport frequencies revealed that Minneapolis-St. Paul (MSP) dominates the destination distribution, accounting for approximately 30.5% of all bookings (see Figure 4). This appears counterintuitive for an outbound leisure carrier headquartered in Minneapolis, but is explained by the presence of return-leg bookings and inbound travelers arriving at the MSP hub. Warm-weather leisure destinations including Cancun (CUN), Las Vegas (LAS), Orlando (MCO), and Puerto Vallarta (PVR) follow as the next most frequent, consistent with Sun Country's leisure positioning.

Figure 4

Top 12 destination airports ranked by mean booking proportion across the full dataset.



Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Methodology

K-Means clustering was selected as the primary segmentation method given its computational efficiency, interpretability of centroids, and direct compatibility with the fully numeric, normalized feature matrix. The algorithm works by iteratively assigning each observation to the nearest of k cluster centroids based on Euclidean distance, then updating those centroids until convergence. The result is k clusters, each defined by a centroid vector that translates directly into a customer profile.

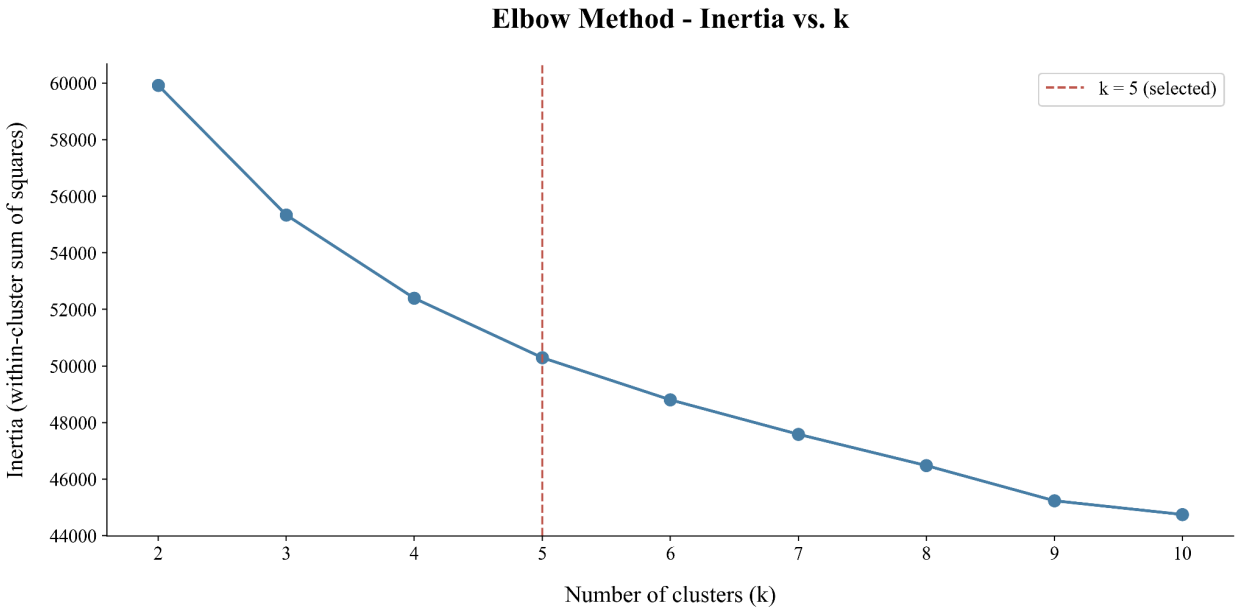
The feature matrix consisted of 88 numeric variables extracted from Clustering_Data.csv after removing the uid and PNRLocatorID identifiers. All features were already normalized to a 0-to-1 scale, making distance-based clustering applicable without additional standardization. The final model was fit using k = 5, random_state = 42, and n_init = 10 to ensure reproducibility and stable initialization. After fitting, cluster labels were reordered by cluster size so that the largest cluster received the label of Cluster 0.

To determine the optimal number of clusters, the elbow method was applied across k = 2 through k = 10. The resulting inertia plot shows a steep decline from k = 2 to k = 5, after which the rate of improvement slows considerably (see Figure 5). This inflection point indicates that adding clusters beyond five yields diminishing returns in within-cluster cohesion, making k = 5 the natural stopping point. Five clusters is specific enough to reveal distinct customer behaviors across booking channel, loyalty status, and travel direction, yet simple enough that each segment translates directly into a targeted marketing campaign. A smaller k would obscure meaningful

differences between customer types, while a larger k risks producing segments that are statistically distinct but operationally indistinguishable.

Figure 5

Elbow method analysis for k = 2 through k = 10.



Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Analysis: The Five Customer Segments

K-Means clustering identified five distinct, business-interpretable customer segments. Table 1 provides a high-level summary of each segment's size and primary characteristics.

Table 1

Summary of five customer segments identified by K-Means clustering (k = 5).

Cluster	Name	N	%	Key Distinguishing Features
C0	MSP Direct Bookers	4,132	27.3%	74% SCA website; 91% MSP origin; 99.9% non-Ufly; Q1 peak
C1	MSP Deal Seekers	3,845	25.4%	100% 3rd-party channel; 89% MSP origin; 99.9% non-Ufly
C2	Inbound Visitors	2,439	16.1%	95% MSP destination; low spend; one-way dominant; non-Ufly

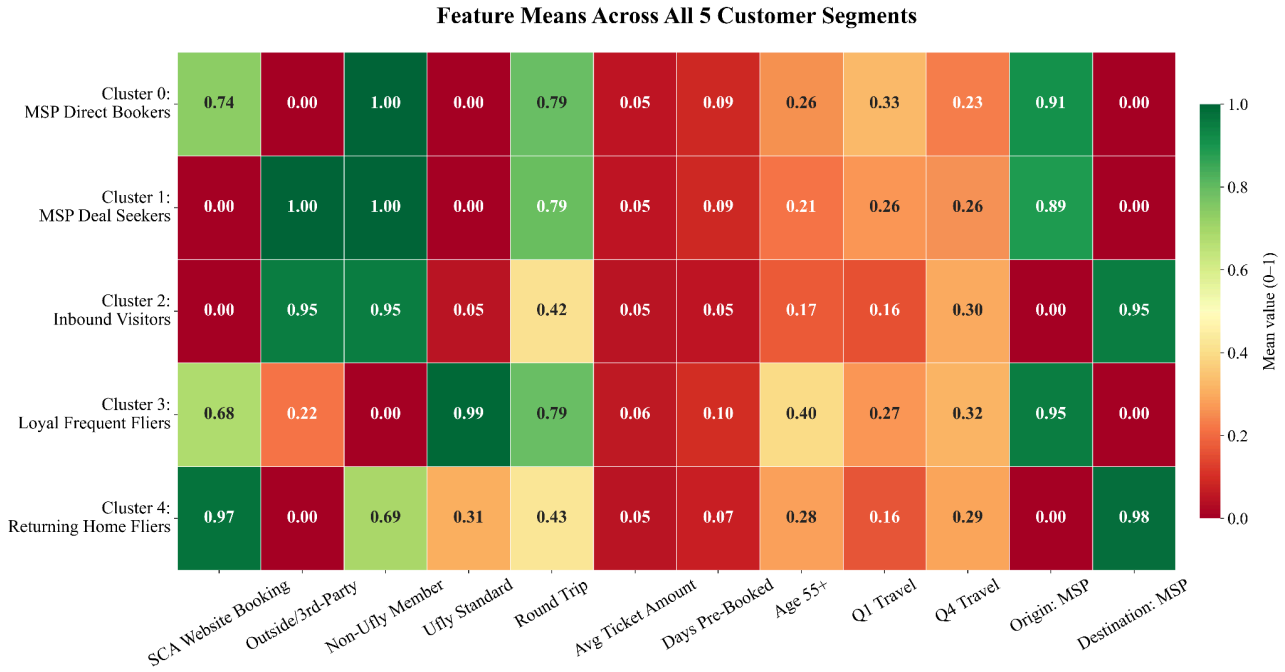
Cluster	Name	N	%	Key Distinguishing Features
C3	Loyal Frequent Fliers	2,377	15.7%	99% Ufly Standard; highest avg ticket; longest lead time; Q4 peak
C4	Returning Home Fliers	2,351	15.5%	98% MSP destination; 30% Ufly Standard; one-way; cross-sell opportunity

Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

To compare all five segments simultaneously, Figure 6 displays the mean value of each key feature across clusters. The heatmap makes three structural patterns immediately visible. First, booking channel splits Clusters 0 and 1 in opposite directions: Cluster 0 shows 0.74 on SCA Website Booking while Cluster 1 shows a perfect 1.00 on Outside/3rd-Party with 0.00 on SCA Website, confirming that these two segments are behaviorally identical except for where they choose to book. Second, loyalty status is almost entirely concentrated in Cluster 3, the only row with a meaningfully green Ufly Standard cell at 0.99 and a fully red Non-Ufly cell at 0.00; Clusters 0, 1, and 2 all show 1.00 on Non-Ufly Member, while Cluster 4 shows a partial split at 0.69 Non-Ufly and 0.31 Ufly Standard, suggesting some existing brand affinity. Third, travel direction separates Clusters 2 and 4 from the rest: both show 0.00 on Origin: MSP but 0.95 and 0.98 respectively on Destination: MSP, visually confirming their inbound identity. Each of these patterns is explored in detail in the segment profiles below.

Figure 6

Feature means heatmap across all five customer segments.



Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Cluster 0: MSP Direct Bookers (27.3%)

Cluster 0 is the largest segment, comprising 4,132 customers (27.3% of the dataset). This group is characterized by a high SCA website direct-booking rate (74%), near-total MSP origin (91%), a high round-trip rate (79%), and moderate group sizes. Travel peaks in Q1, consistent with Sun Country's winter-escape leisure identity. These are Minneapolis residents booking warm-weather escapes directly on Sun Country's website.

Most critically, 99.9% of this segment are non-Ufly members — and this is the single most important finding in the entire analysis. Sun Country's largest, most digitally engaged, and most channel-loyal customer group has near-zero loyalty program participation. The enrollment opportunity is immediate, high-volume, and low-friction given these customers' existing comfort with the SCA digital platform.

Cluster 1: MSP Deal Seekers (25.4%)

Cluster 1 contains 3,845 customers (25.4%) who are, in behavioral terms, nearly identical to Cluster 0. Both segments share the same Minneapolis origin (89% for Cluster 1), similar round-trip rates, and similar age and seasonal patterns. The one critical difference is that 100% of Cluster 1 bookings are made through outside or third-party channels. These are comparison shoppers who find Sun Country through Expedia, Google Flights, or similar aggregators rather than navigating to SCA.com directly.

This 100% third-party booking rate is a margin problem with a structural solution. Every booking through a third-party platform costs Sun Country distribution fees and surrenders direct customer relationship data. Migrating even a fraction of this segment to the direct channel represents a meaningful revenue and data opportunity — and the structural similarity to Cluster 0 means that the same leisure traveler profile is within reach.

Cluster 2: Inbound Visitors (16.1%)

Cluster 2 is the most geographically distinctive segment. A full 95% of bookings in this cluster have MSP as the true destination, meaning these travelers are flying into Minneapolis rather than out of it. They are inbound visitors traveling to the Twin Cities from other markets. This group shows lower average ticket prices, predominantly one-way travel (42% round-trip versus 79–89% in Clusters 0 and 1), and overwhelmingly non-Ufly status (95% non-member). While this segment is less commercially valuable than the outbound leisure segments in isolation, it forms half of a strategically important pairing with Cluster 4.

Cluster 3: Loyal Frequent Fliers (15.7%)

Cluster 3 is Sun Country's most valuable existing customer segment. This group of 2,377 customers (15.7%) is 99% Ufly Standard members, books with the longest advance lead times of any cluster, carries the highest average ticket prices, and shows strong direct-channel booking behavior (68% SCA website). Travel peaks in Q4 during the holiday and year-end season, and the segment skews older, with the 55+ age group representing 40% of the cluster. This profile is consistent with a deliberate, brand-loyal, experienced traveler.

The business priority here is retention, not acquisition. These customers are already deeply engaged. However, Sun Country currently offers no Elite tier pathway to lock them in

further. A competitor offering a status-match or Elite upgrade campaign to this segment could rapidly erode the loyalty Sun Country has built.

Cluster 4: Returning Home Fliers (15.5%)

Cluster 4 contains 2,351 customers (15.5%) who are overwhelmingly flying home to Minneapolis. A full 98% of their bookings have MSP as the true destination, and the segment is dominated by one-way travel. Approximately 30% hold Ufly Standard status — the second highest Ufly engagement rate after Cluster 3 — suggesting a group with partial loyalty affinity that could be deepened.

The most strategic insight about this segment emerges when it is paired with Cluster 2. Cluster 2 flies into Minneapolis; Cluster 4 flies home to Minneapolis. Together they represent a round-trip cross-sell opportunity. A customer completing a one-way trip home via Sun Country has just experienced the product firsthand and is a natural target for an outbound vacation package offer at the point of checkout.

Visualizations

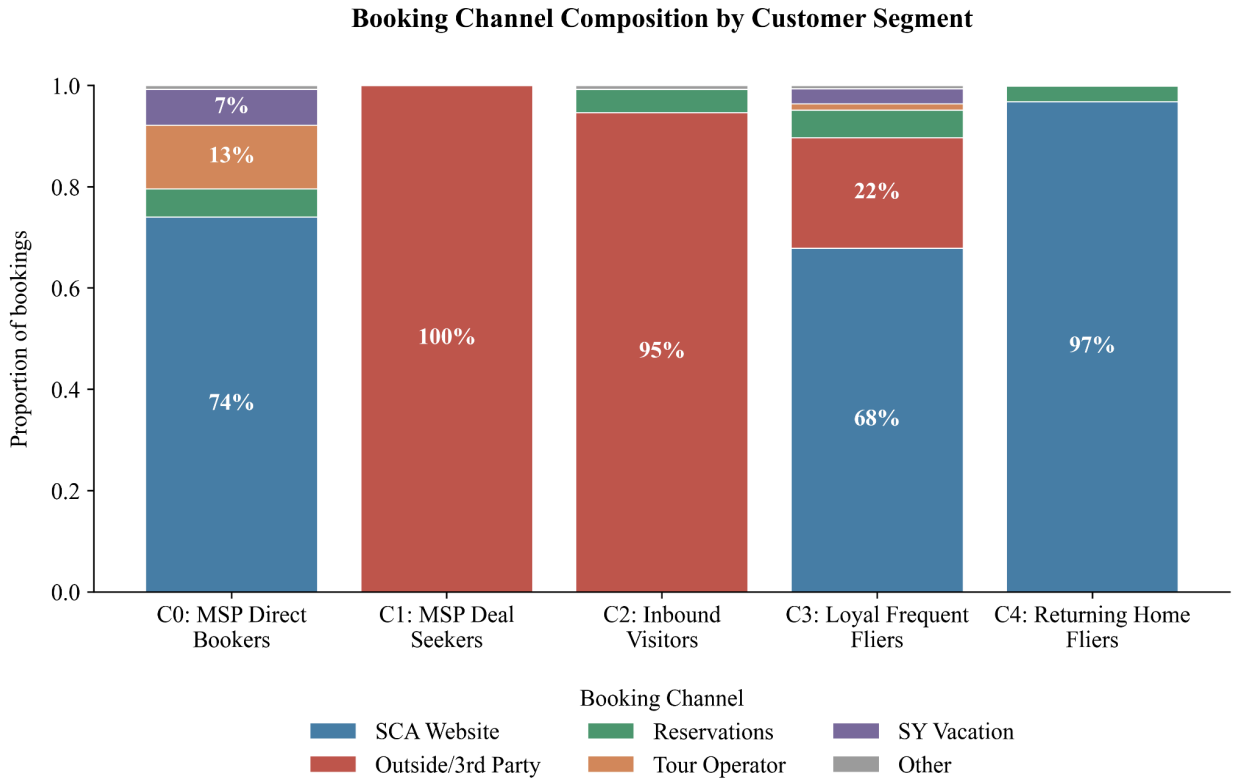
The following visualizations build on the EDA findings by showing how key behavioral patterns differ across the five identified segments. Where the EDA charts described the full dataset, the segment-level charts below reveal which customer groups are driving each pattern. Figures 7 through 13 are presented in sequence; together they constitute a coherent data portrait of Sun Country's customer base.

Booking Channel and Loyalty Status (Figures 7 and 8)

The segment-level channel breakdown in Figure 7 confirms the sharp split observed in EDA. Cluster 0 books 74% directly through SCA.com, while Cluster 1 shows a perfect 100% third-party rate with zero direct bookings. Cluster 3 (Loyal Frequent Fliers) also skews strongly toward the direct channel at 68%, suggesting that loyalty and direct booking behavior are correlated. Cluster 4 (Returning Home Fliers) shows the highest direct booking rate of any segment at 97%, though this likely reflects the limited availability of one-way homebound routes on third-party platforms rather than deliberate brand preference.

Figure 7

Booking channel composition by customer segment.

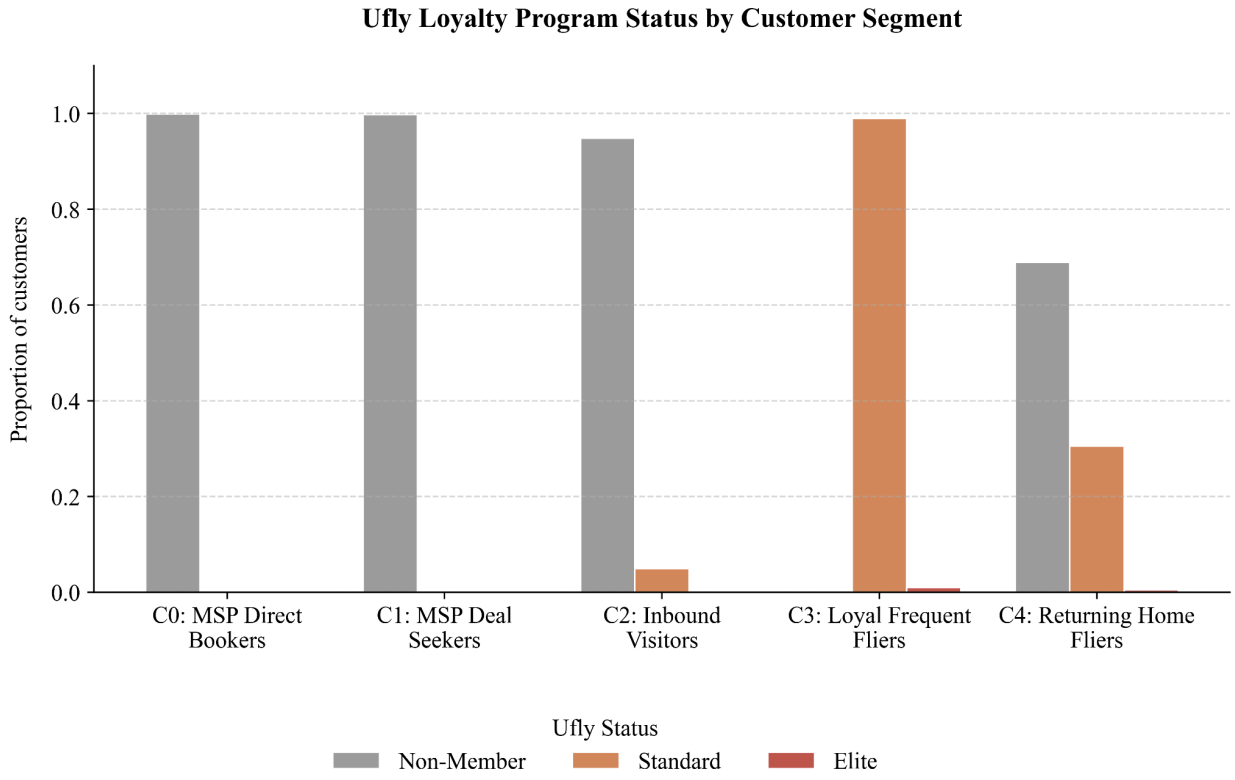


Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

The loyalty breakdown in Figure 8 makes the enrollment opportunity concrete. Clusters 0, 1, and 2 are at or near 100% non-member; Cluster 3 is nearly 100% Standard. Cluster 4 is the only segment with a meaningful mixed profile at 69% non-member and 31% Standard, suggesting partial loyalty conversion has already occurred naturally in this group. The absence of any meaningful Elite membership across all five segments highlights a structural gap in Sun Country's current loyalty program.

Figure 8

Ufly loyalty program membership status by customer segment.



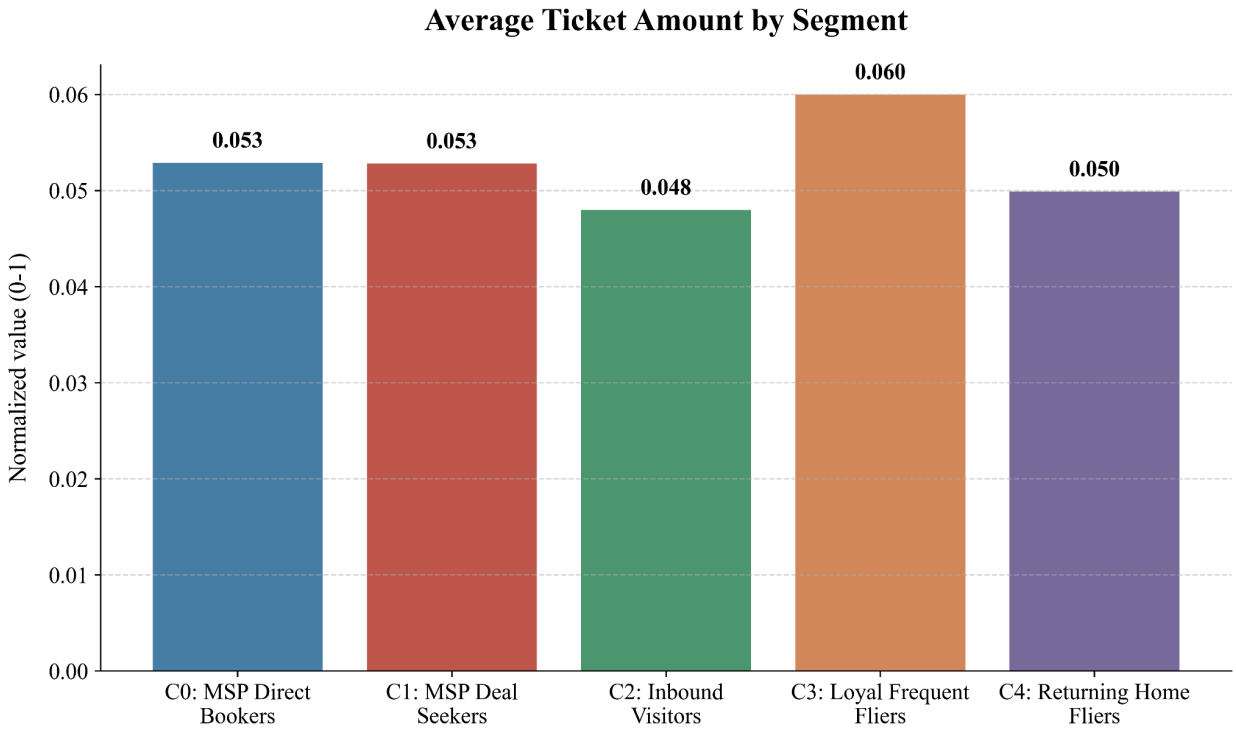
Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Ticket Amount, Booking Lead Time, and Destinations (Figures 9, 10, and 11)

Cluster 3 records the highest average ticket amount at 0.060, while Cluster 2 records the lowest at 0.048 (Figure 9). The differences across segments are modest in absolute terms given the normalization, but the consistent ordering supports the segment narrative: Cluster 3 is the premium spend group, Clusters 0 and 1 are mid-range, and Cluster 2 is the most price-sensitive.

Figure 9

Average ticket amount by customer segment.

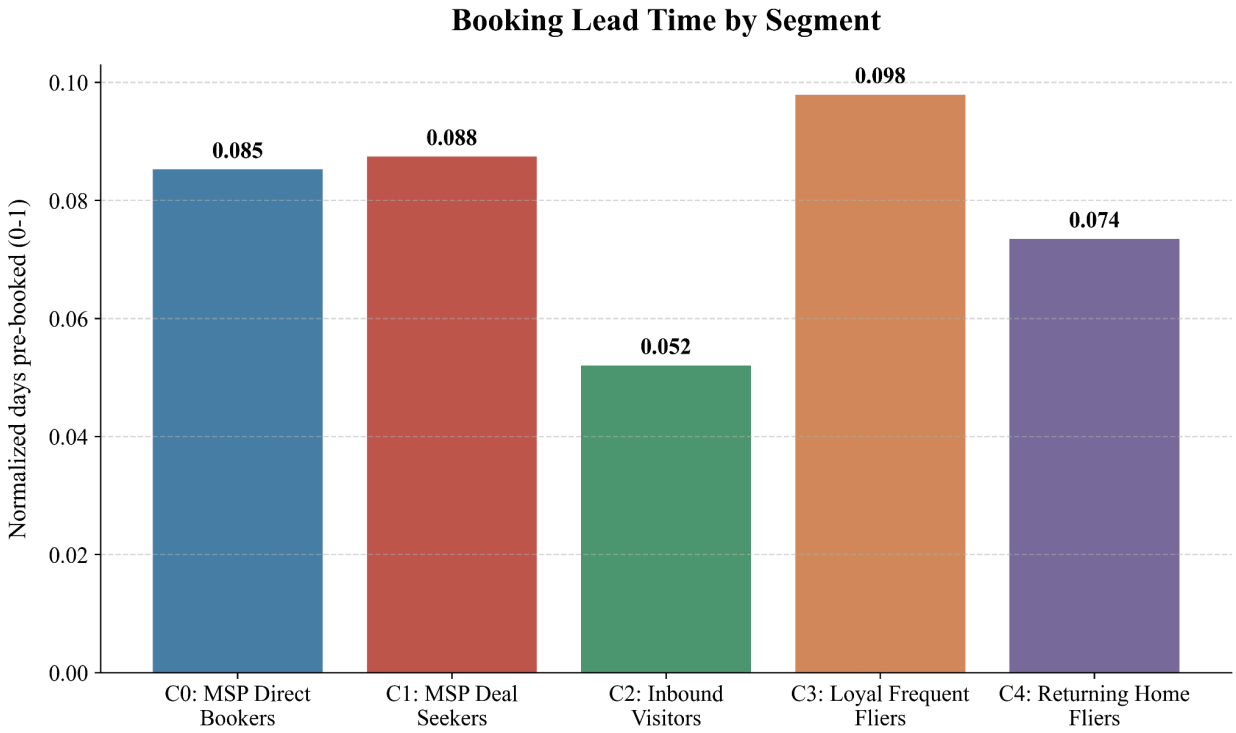


Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Cluster 3 also shows the longest average booking lead time at 0.098, while Cluster 2 books latest at 0.052 (Figure 10). This reinforces the behavioral contrast between a deliberate, planning-oriented Cluster 3 and the more transactional, reactive profile of the inbound segments.

Figure 10

Booking lead time by customer segment.

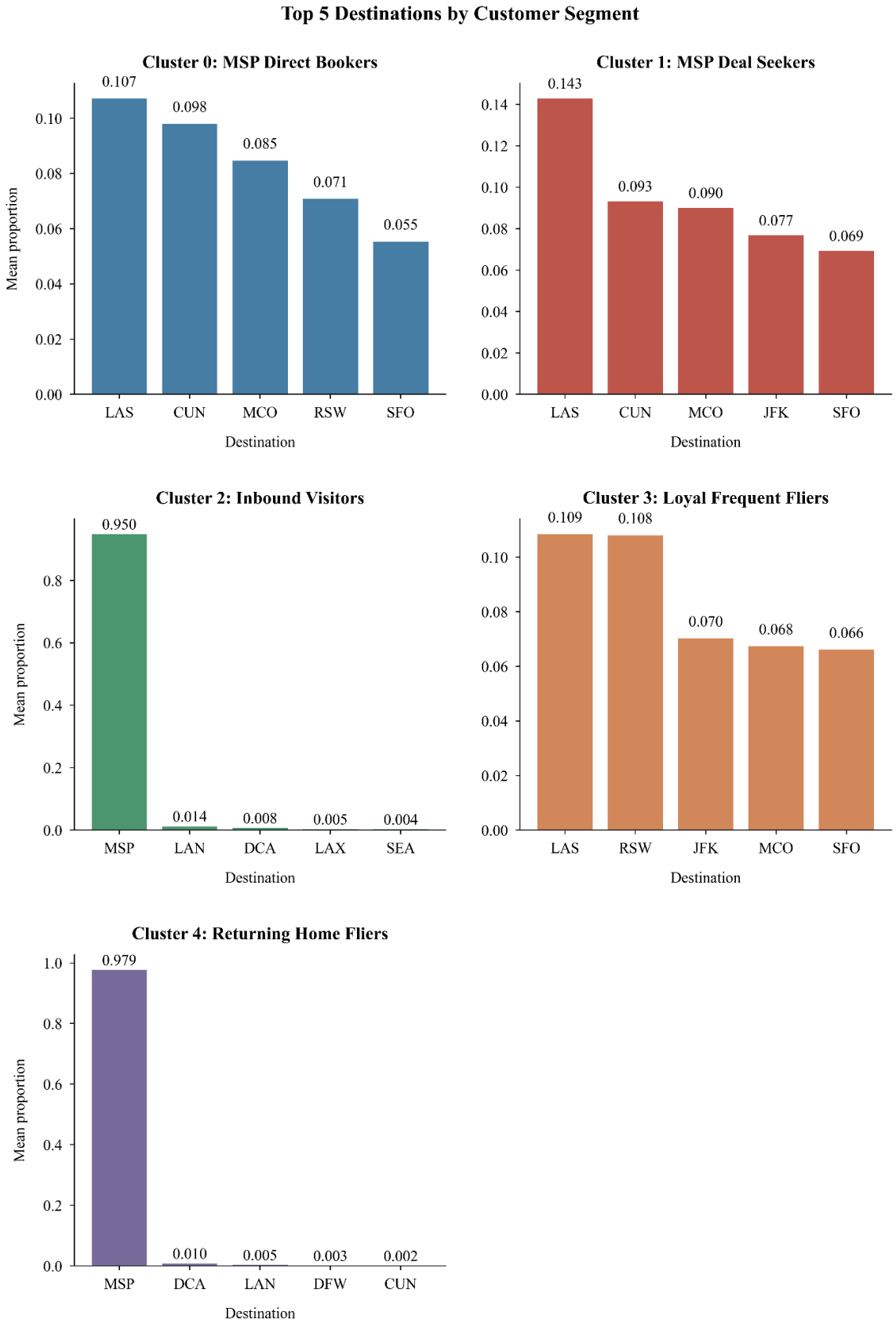


Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

The destination charts in Figure 11 tell the geographic story most directly. Clusters 0 and 1 are leisure travelers whose top destinations are Las Vegas, Cancun, and Orlando — consistent with Sun Country's core product. Clusters 2 and 4 both show MSP as the overwhelming top destination at 0.950 and 0.979 respectively, confirming their inbound identity. Cluster 3 shows a more distributed destination profile with no single airport dominating, consistent with an experienced traveler base that is not locked into one vacation market.

Figure 11

Top 5 destinations by customer segment.


Top 5 Destinations by Customer Segment
Cluster 0: MSP Direct Bookers
Mean proportion
0.107
0.098
0.085
0.071
0.055
LAS
CUN
MCO
RSW
SFO
Destination
Cluster 1: MSP Deal Seekers
0.143
0.093
0.090
0.077
0.069
LAS
CUN
MCO
JFK
SFO
Destination
Cluster 2: Inbound Visitors
0.950
0.014
0.008
0.005
0.004
MSP
LAN
DCA
LAX
SEA
Destination
Cluster 3: Loyal Frequent Fliers
0.109
0.108
0.070
0.068
0.066
LAS
RSW
JFK
MCO
SFO
Destination
Cluster 4: Returning Home Fliers
0.979
0.010
0.005
0.003
0.002
MSP
DCA
LAN
DFW
CUN
Destination

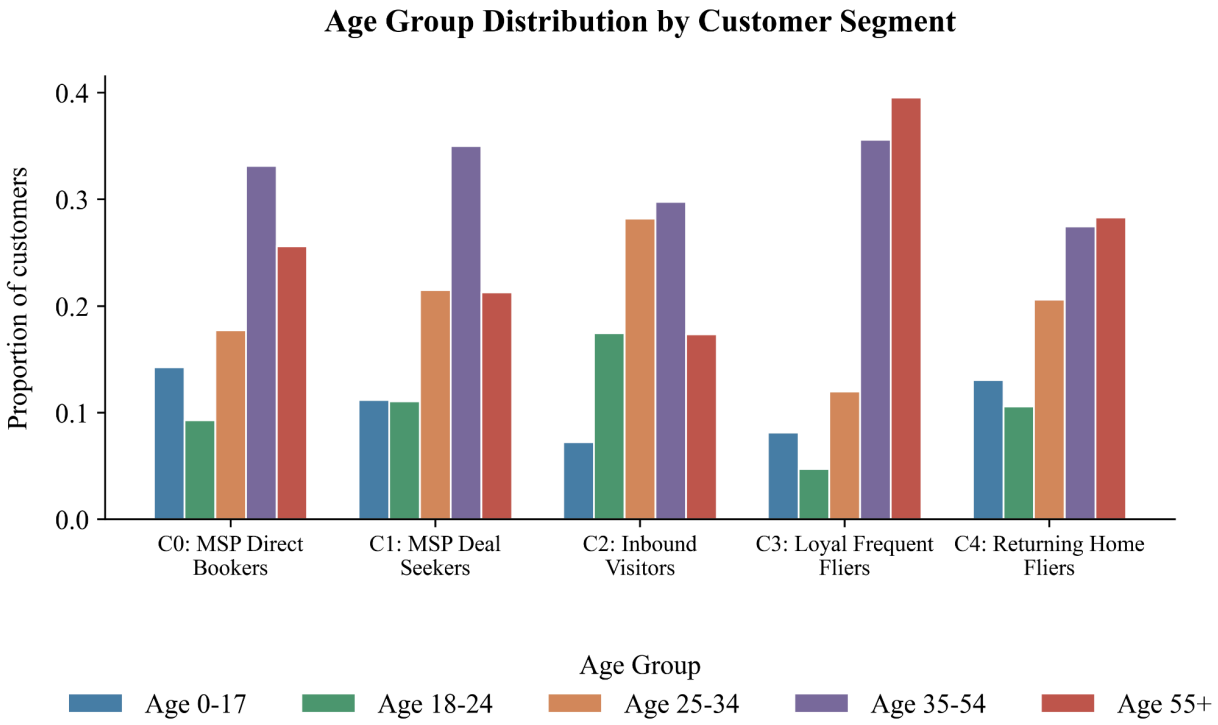
Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Age Distribution and Seasonality (Figures 12 and 13)

Cluster 3 shows the most skewed age profile, with the 55+ group representing approximately 40% of the segment — the highest of any cluster (Figure 12). All other segments are dominated by the 35–54 bracket. Cluster 2 (Inbound Visitors) has the youngest overall profile, with a relatively higher proportion of 18–24 travelers. These age patterns are relevant to campaign design: older Cluster 3 travelers may respond differently to digital outreach than younger segments.

Figure 12

Age group distribution by customer segment.



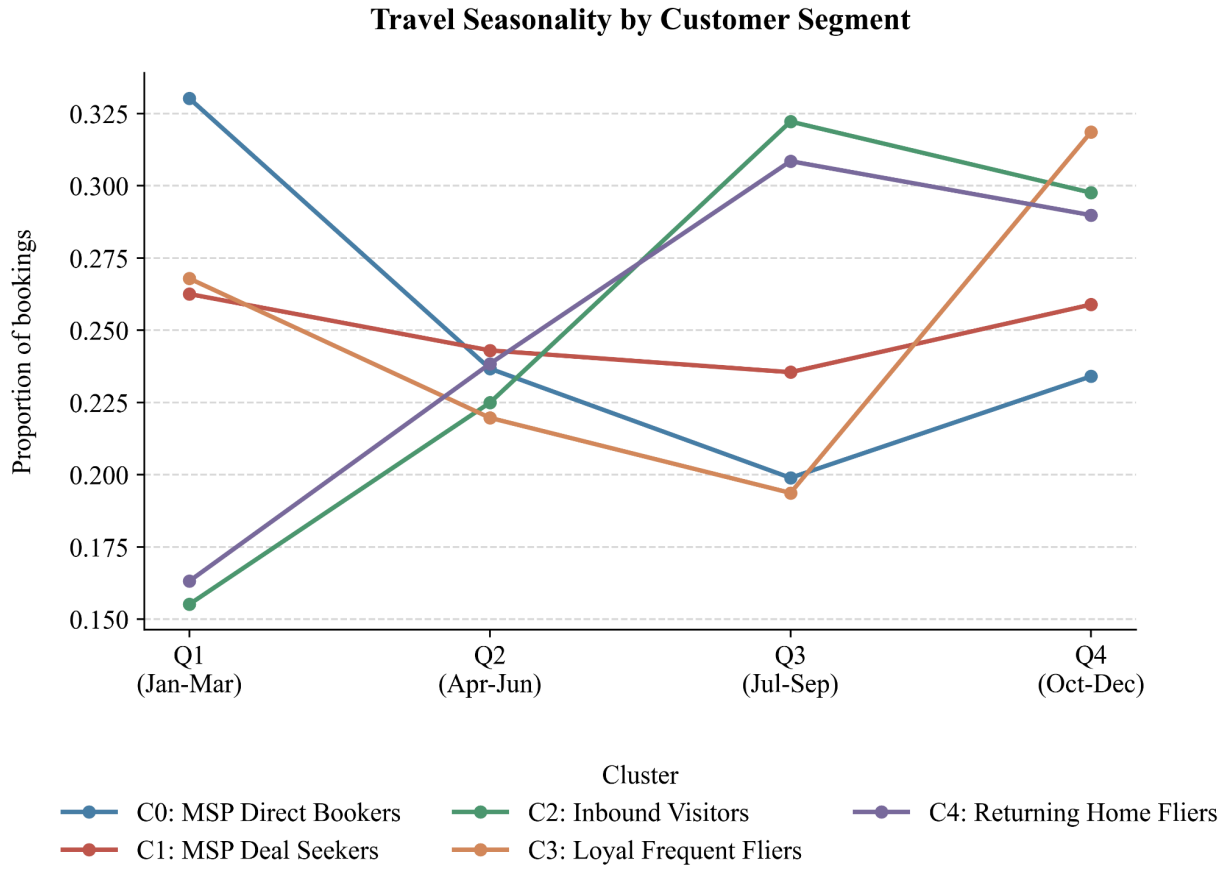
Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Seasonality patterns in Figure 13 reinforce the segment narratives. Clusters 0 and 1 peak strongly in Q1 (January through March), consistent with the winter-escape leisure traveler profile. Cluster 3 follows a different pattern, peaking in Q4 (October through December),

pointing to holiday and year-end travel as a defining characteristic of this group. Cluster 2 peaks in Q3 (July through September), suggesting summer visits to Minneapolis as the primary travel occasion for inbound travelers.

Figure 13

Travel seasonality by customer segment.



Note. Data from Balasubramanian, S. (2025). BANA 200 course materials [Dataset]. University of California, Irvine.

Business Insights & Recommendations

The following recommendations are each structured using a data result, motivation inference, recommendation, and business goal framework. Each maps directly to one of Warnken and Vaughan's three stated business objectives.

Recommendation 1: Launch Direct Booking & Ufly Enrollment Campaign

Clusters 0 and 1 together represent 52.7% of total customers. Cluster 0 books 74% directly on SCA.com while Cluster 1 books 100% through third-party platforms; both segments are 99%+ non-Ufly members, both originate primarily from MSP, and both show identical leisure travel patterns across seasonality, round-trip rate, and destination preferences.

These two clusters represent the same traveler type, Minneapolis-based leisure travelers, differentiated solely by where they choose to book. Cluster 0 has already adopted Sun Country's direct channel, meaning the primary barrier to Ufly enrollment is simply a lack of prompting. Cluster 1 has not adopted the direct channel, which indicates that price comparison is their primary decision driver.

For Cluster 0, the recommended intervention is a post-booking Ufly enrollment prompt on the booking confirmation page, paired with a first-flight miles bonus. Because these customers already trust the SCA.com experience, conversion friction is minimal. For Cluster 1, the recommended intervention is a guaranteed 5–10% discount available exclusively through direct bookings on SCA.com. This removes the primary reason to use third-party aggregators while simultaneously introducing this segment to the direct booking experience, where a Ufly enrollment prompt can then be served. This single recommendation addresses both Ufly Rewards enrollment and direct channel booking growth.

Recommendation 2: Introduce a Ufly Elite Tier for Loyal Frequent Fliers

Cluster 3 contains 2,377 customers (15.7% of total). Ninety-nine percent are Ufly Standard members. This segment carries the highest average ticket spend of any cluster, the longest advance booking lead time, a 68% SCA website booking rate, and a strong Q4 travel peak.

These are Sun Country's most valuable existing customers: older, experienced, deliberate travelers who plan well in advance and spend more per trip. They are already invested in the Ufly program but currently have no Elite tier to aspire toward. Without an upgrade pathway, a competitor offering a status-match or Elite upgrade campaign could rapidly erode the loyalty Sun Country has built.

The recommended action is to introduce an Elite tier gated behind an annual spend or trip-frequency threshold, with perks including priority boarding, one free checked bag, and

bonus miles on Q4 travel. A progress-based messaging approach, framed as ‘You're almost Elite,’ creates a concrete retention mechanism that rewards existing loyalty and gives customers an incentive to continue booking with Sun Country over competitors. This recommendation directly addresses the goal of developing differentiated, premium product offerings and strengthens the depth of the Ufly program.

Recommendation 3: Cross-Sell Outbound Vacation Packages at Checkout

Clusters 2 and 4 together represent 31.6% of total customers and form a geographic mirror: Cluster 2 (Inbound Visitors) flies into MSP from various origins, while Cluster 4 (Returning Home Fliers) flies home to MSP one-way. Cluster 4 shows 30% Ufly Standard membership — the second highest rate of any segment — alongside one-way travel dominance.

A customer completing a one-way trip home to Minneapolis via Sun Country has just experienced the Sun Country product firsthand. At the point of checkout or post-flight, they are well-positioned for an outbound leisure offer — precisely the type of winter-escape product that defines Clusters 0 and 1. Cluster 4's partial Ufly enrollment suggests some existing brand affinity that can be leveraged. For Cluster 2 (Inbound Visitors), a round-trip bundle offer at the time of booking tests whether inbound visitors can be converted to full round-trip customers.

The recommended action is to deploy a 'Plan your next escape' checkout widget targeting Cluster 4 customers immediately following their homebound booking, converting one-way home buyers into round-trip vacation package customers. This recommendation directly addresses Warnken's goal of developing differentiated vacation packages and drives Ufly enrollment through the checkout interaction.

Reflection: EDA and Clustering as Complementary Tools

The most important business insight in this analysis was visible before any clustering model was run. A single frequency count revealed that 78.5% of Sun Country customers are unenrolled in the Ufly Rewards program, and the near-equal channel split between SCA.com and third-party platforms immediately suggested that booking behavior would be a primary axis of differentiation. These findings did not require an algorithm.

However, EDA alone cannot answer the question that matters most for action: which customers should be targeted, and how? Knowing that 78.5% of customers are non-Ufly is an

important finding, but it does not indicate whether to pursue the direct bookers, the third-party shoppers, or the inbound visitors. K-Means is what separated the 78.5% into actionable groups — specifically revealing that Clusters 0 and 1, representing 52.7% of all customers, share the same origin, travel pattern, and leisure motivation but differ only in where they book. That distinction is invisible in any EDA chart and only surfaces through segmentation. The result is two fundamentally different interventions: a post-booking prompt for Cluster 0 and a direct-booking discount for Cluster 1.

This is also why model complexity should be kept in check. K-Means centroids are directly readable as segment averages and translate cleanly to a non-technical audience. A more complex algorithm might improve some internal fit metric while producing segments that cannot be mapped to a marketing campaign — statistically distinct but operationally indistinguishable. The standard for a useful model is not sophistication; it is whether a business decision can follow from it.

Long-Term Data Recommendations

The current dataset, while analytically usable, captures transaction behavior without the richness that would enable deeper and more actionable segmentation over time. Four enhancements are recommended for Sun Country's longer-term data strategy.

The highest-priority addition is ancillary spend tracking. Linking bag fees, seat upgrades, priority boarding purchases, and travel insurance selections to individual customer records would add a revenue-sensitivity dimension currently invisible in the data, allowing Sun Country to distinguish high-value from budget-conscious travelers within each segment — a distinction that matters directly for vacation package pricing and upsell strategy.

Post-flight satisfaction capture (NPS or CSAT) would reveal which segments are at churn risk versus which are brand advocates. Cluster 3, for example, may contain a subset of highly satisfied customers alongside a subset of habitual but dissatisfied ones — a distinction that is currently invisible but that would enable proactive retention before a competitor makes an offer.

Tracking loyalty conversion events — specifically, whether Ufly sign-ups occur following post-booking emails, in-flight offers, or post-trip surveys — would identify which touchpoints are most effective for each segment. This closes the feedback loop on

Recommendation 1 and enables Sun Country to refine its enrollment campaign based on actual conversion data rather than assumptions.

Finally, enriching the demographic and behavioral data with income proxies (zip code linked to census-tract median income), device type at booking, and return visit frequency would sharpen segment distinctiveness and enable more precise personalization. Device type is particularly relevant to Vaughan's digital experience goal: a customer completing a complex vacation package booking on a mobile device faces meaningfully different user experience challenges than one booking on a desktop.

Conclusion

This analysis answers the central question posed to Sun Country's analytics consultants: who are Sun Country's customers, and what does the data say about how to serve them better? K-Means clustering of 15,144 customer-trip records identified five actionable segments whose defining characteristics map directly onto the three strategic goals Warnken and Vaughan articulated.

The most immediate opportunity is a Ufly enrollment campaign targeting Clusters 0 and 1, who represent 52.7% of customers with near-zero loyalty penetration. This campaign requires no further data collection — only a post-booking prompt and a first-flight miles incentive. The most structurally important opportunity is migrating Cluster 1 from third-party platforms to SCA.com, which addresses both a margin problem and a data problem simultaneously. The most long-term valuable opportunity is retaining Cluster 3 through an Elite tier upgrade pathway, protecting the customer segment most at risk of competitive poaching.

Collectively, these findings demonstrate that Sun Country's data, even in its current state, is highly usable for generating decision-ready business intelligence. The dataset required minimal cleaning, and the most actionable insights were visible before any model was run. This reinforces the project's core analytical lesson: data literacy and business context are the highest-leverage analytical skills. Sophisticated models amplify insight, but they do not create it.

Warnken and Vaughan now have validated customer profiles to replace the anecdotal assumptions that previously drove Sun Country's marketing. The path to the company's fourth decade of operation, and to sustainable competitive differentiation against Delta, United, and American, runs directly through the segmentation framework presented here.

References

Thomas, L. C., & Bapna, R. (2016). Data-driven customer insights at Sun Country Airlines [Teaching case]. Carlson School of Management, University of Minnesota.

Sun Country Airlines. (2025). clustering_data.csv [Dataset]. University of California, Irvine - BANA 200.

Sun Country Airlines. (2025). customer_data_dictionary.xlsx [Data dictionary]. University of California, Irvine - BANA 200.