Customer Analysis for Sun Country Airlines

BANA 200A Group Project Report

Team 2A

### Introduction

In an increasingly competitive airline industry, Sun Country Airlines stands out as a company with its own advantages and opportunities in this industry. This report aims to provide actionable recommendations for customer behavior by segmenting Sun Country’s customer base into distinct clusters. Using K-mean clustering, we identified five distinct customer segments, each with its own characteristics. Our analysis offers targeted marketing strategies for each segment and recommendations for enhancing long-term strategy for the company. The objective is to enable Sun Country Airlines to become a more data-driven and customer-oriented organization.

### Background

Sun Country Airlines, founded in 1983, has a unique and compelling history that sets it apart from the highly competitive airline industry. The airline initially offered an escape for winter-weary travelers in Minnesota and transported them to sunny destinations in the Southern part of the country. The airline has managed to turn its smaller size into an advantage, offering a more personalized and customer-focused service. As Sun Country looks forward to the future, it aims to leverage its data analytics to gain deeper insights into customer preferences. The goal is to become a more agile and data-driven company, capitalizing on adapting to market trends and customer needs better than its competitors.

### Methodology

We choose K-means clustering for its efficiency in segmenting large datasets into distinct clusters. This unsupervised clustering is ideal for customer segmentation, aiming to identify similarities and differences in customer behavior and traits. K-means divides the dataset into K clusters, in which each data point is assigned to the nearest cluster centroids. (In our case, K = 5). The algorithm iteratively adjusts these centroids and assignments until the optimal clustering is achieved. It helps to provide a transparent and replicable framework for our customer analysis so that we can provide a better conclusion for our client company.

### Data Analysis

* Data Preparation Steps
* Python codes

### Finding and Insights

* Create at least 3 charts/figures for each segment and use 2 different types of charts.
* Characteristics of each segment based on our analysis and visualization.
* “Tell a story” about the unique nature of each segment.

Segmentation:

Cluster 0 - 55+

Cluster 1-35-54, 25-34

Cluster 2- 35-54

Cluster 3- 55+,35-54

Cluster 4- 35-54

Python Group Project Possible Analysis:

Seasonality Analysis by Cluster - Alex

Cluster 0: The All-season Travelers

Cluster 0 shows an even distribution of travel across all four quarters. This segment is not significantly influenced by seasonality and represents a stable customer base that Sun Country can rely on year-round.

Cluster 1: The Winter Escapists

Cluster 1 shows a significant spike in Q1 with over 2,500 travelers, dropping to its lowest in Q3 with only 1500. This segment likely consists of customers looking to escape the winter cold, making them prime targets for winter vacation promotions.

Cluster 2: The Off-Peak Explorers

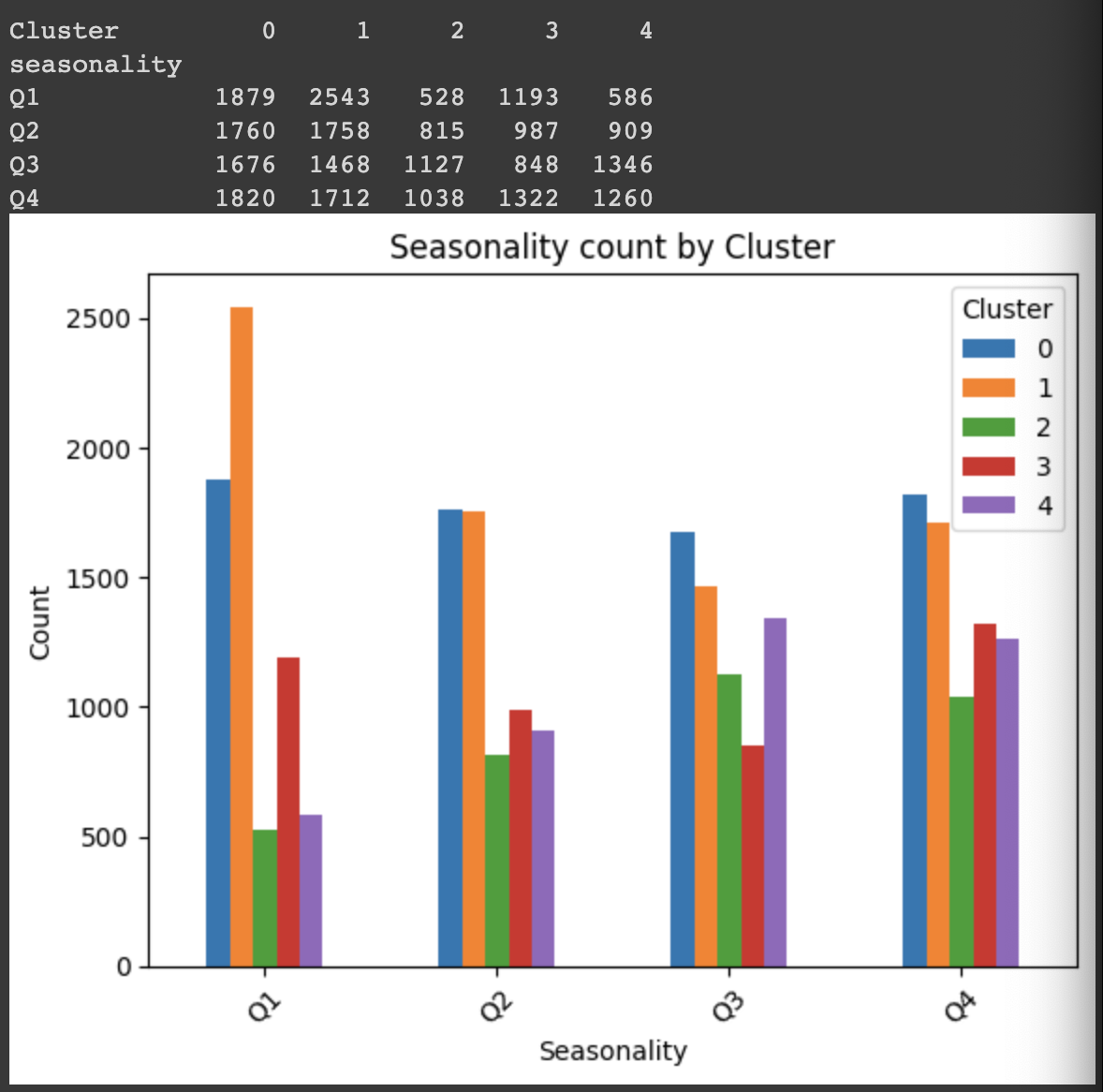
Cluster 2 is the smallest group and is least active in Q1 with about 500 travelers. The segment may consist of customers who prefer to travel during off-peak seasons to avoid crowds.

(Cluster 3: The Consistent Enthusiasts

Cluster 3 shows a consistent range of 800 to 1300 travelers across all seasons. This segment may include business travelers who have to travel irrespective of the season.)

Cluster 4: The Fall Enthusiasts

Cluster 4 has its lowest activity in Q1 with 600 travelers but picks up in Q3 and Q4 with 1200 travelers. This segment may enjoy traveling during the fall and winter seasons, possibly to experience fall seasonal activities.



Group - Hilary (being skipped)

* Overall, most people travel in groups. Only cluster 4 travels more solo
* Cluster 1 has more group size 2, 3,4 . indicating couples or families
* Cluster 0 - 4810 Travel in groups, 2330 travel solo =67% group
* Cluster 1 - 2555 travel in groups, 1799 travel solo =
* Cluster 2 - 5310 in groups, 2171 travel solo
* Cluster 3 - 1776 travel in groups, 1728 solo travel
* ​​Cluster 4 - 1372 travel in groups, 2724 solo travel
  + I suggest advertising a package within late days pre-booked that encourages group travel to incentivize travel

Days pre-booked - Hilary

* (Cluster 0 - Average days between booking and travel day is about 62 days. )
* Cluster 1 - Average days between booking and travel day is about 68
* (Cluster 2 - Average days between booking and travel day is about 60
* Cluster 3 - Average days between booking and travel day is about 51 )
* Cluster 4 - Average days between booking and travel day is about 36
  + I suggest we incentivize more purchases on days closer to the travel date to raise average purchases since they are low.

Class of service - Lalitha percentages

* Cluster 0- From the booked graphs, we can see that cluster 0 was the most number of First class tickets. So cluster 0 has more spending power. Therefore, we should also market on board wifi or the elite membership, paying extra for lounge access etc.
* Cluster 0 showed more discount first class.
* Cluster 1, 4 showed significant increase in first class conversion without the discount.
* Cluster 2 has least economy in both travelled and booked. They want comfort.
* Cluster 4 has highest economy in both booked and traveled.
* Cluster 3 has least amount of first class traveled and booked. Also least amount of discount first class.

Senior discount / Age – Lalitha

* From the age group graph and table, we see that cluster 0 has most customers in age range 55+. Therefore a senior discount should be targeted at cluster 0.Provide additional service to seniors like lounge access.
* Cluster 0,2- . 35-54 This age group is people who are financially more stable in life and have kids. Family group discount is a good idea and we should promote elite membership to these customers.
* Cluster 1 has max customers in ages 25-34. This age group is working professionals who go on work travel or for higher studies and want budget vacations. They are the people that move places frequently maybe. Should promote standard membership and seasonal discounts for vacation.
* 18-24 cluster 1
* 0-17 Cluster 4
* Overall lesser younger people

Ufly member status – Chun Hui

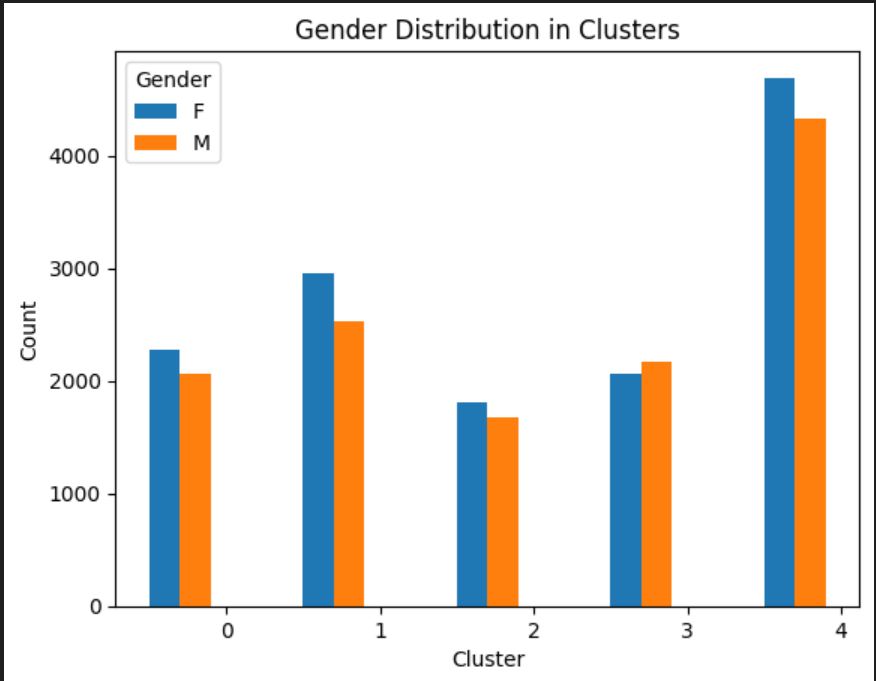
* Cluster 0 represents most of the customers who joined the ufly program, most of which owns a standard status.
* Cluster 1 and 2 has a combination of members and non-members, but the proportion of members in cluster 1 is nearly ⅓, much more than that for cluster 2.
* We can see that from cluster 0 to 4, there is a gradual increase in the number of non-members.



Round Trip - Lalitha

* Cluster 0 has least one way. This group are probably natives who live there.
* Cluster 3 has most one way. Could be they are moving out of the city for work/study. Or someone who wants flexibility in their travel plans .
* Cluster 1 and 4 have most round trip Families, Vacationers, work trips etc
* Cluster 2 has least round trips. Visiting family for extended period. Or Just wealthy and don’t care about budget

Gender - Chun Hui



Credit Card - Abhinav

Stopover Code - Abhinav

* Cluster 0- From the booked graphs, we can see that cluster 0 was the most number of First class tickets. So cluster 0 has more spending power. Therefore, we should also market on board wifi or the elite membership, paying extra for lounge access etc.
* Cluster 0 showed more discount first class.
* Cluster 1, 4 showed significant increase in first class conversion without the discount.
* Cluster 2 has least economy in both travelled and booked. They want comfort.
* Cluster 4 has highest economy in both booked and travelled.
* Cluster 3 has least amount of first class travelled and booked. Also least amount of discount first class.

Coupon Sequence Number - Hilary

* No important data

\*\*\*

### Case Questions and Conclusions

* **Based on your analysis, how many different customer segments (i.e., clusters of similar customers) are there? What are the key characteristics of each segment? Can you come up with descriptive titles for the segments?**

We identified five different customer segment segments through K-means clustering. (Need more)

* **What advice would you offer the executives of Sun Country Airlines (Warnken and Vaughan), based on your segmentation analysis, so they can better achieve their business objectives?**
* How usable is the data you received from the client (Sun Country)? Given the state of the data, what steps would you take to make it usable and useful—for the initial assigned project and for Sun Country longer term?

The data received from Sun Country is highly usable for segmentation and customer analysis. For longer-term utility, we recommend regular data cleaning and updating the dataset with new variables like customer feedback or specific spending patterns.

* Based on what you see in the data, how many different ways can you view the information to paint a picture of different customer segments? Which ways offer Sun Country the most insights? How could the insights be used?

The customer segments’ data can be viewed in multiple ways, including age group, gender, seasonality, booking channel, etc. (Need more)

* How would you visualize and present the insights you found for the client?

We would use bar charts to show the distribution of each cluster across different objectives, and line charts to show ages over time. (Need more) These visualizations will make it easier for us to understand the data at a glance.

* What will you do in order to show Warnken and Vaughan how the insights you derived connect to their business objectives?

To show how the insights connect to business objectives, we would map each cluster’s characteristics to Sun Country’s strategic goals like increasing (Need more) and targeting new customer demographics.

* How much power lies in simply understanding and exploring your data? Consider the downside of focusing too much on fancy models.

Understanding and exploring the data provides invaluable insights and suggestions that can directly impact our business decisions. A thorough data analysis can reveal any patterns and trends that are crucial for any subsequent analysis. While advanced models can offer deeper insights, the downside is that they can be time-consuming and complex to run. By using K-means clustering, we attempt to strike a balance between providing actionable insights without the need for overly complex models.