BANA 200A Group Project Report

Sun Country Airlines Case Study

Group 5B - Abby Chung, Ethan Le, Sai Kulkarni,

Yangky Tanuputra Tan

We have been asked by Sun Country Airlines (SCA) to analyze the data they've gathered and to present insights we've found related to customer segments, travel patterns, and other useful points. Through our exploration and analysis of this data, we've grouped the customers into various segments, and have found some actionable insights. In this report, we will explore the process of said exploration and analysis, go over our findings, and report our recommendations.

First, let's take a look at how our customer segments were grouped. SCA's executives have expressed interest in the metrics of Ufly membership status and usage of the official SCA website, and so as we can see in Figure 1, our customers have been grouped accordingly, from Ufly members who book through the website to non-Ufly members who book outside the main website. With our goal of increasing Ufly membership participation and overall making SCA more competitive, a major strategy we can employ is improving our website's search engine optimization (SEO). Having our website come up on the first page of the Google search results will greatly boost customer engagement with our website, and we can use various strategies to target each customer segment. Without further ado, let's dive into our customer segments and tackle them one by one, going over ways

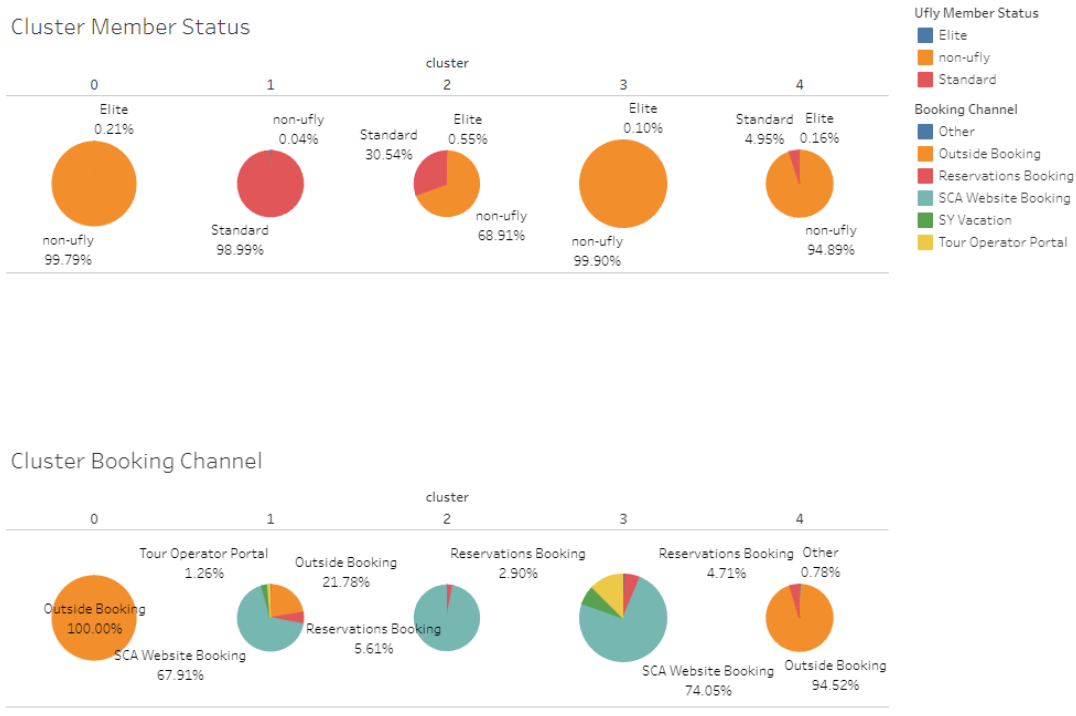


Figure 1: All Clusters UFly Membership Status & Booking Channels

to improve engagement with each one. Below, we have **Figures 2a, 2b, and 2c**, showing an initial comparison between the average ages, days pre-booked, and layover distribution of each cluster. These figures will be referenced in the individual customer segment analysis sections.

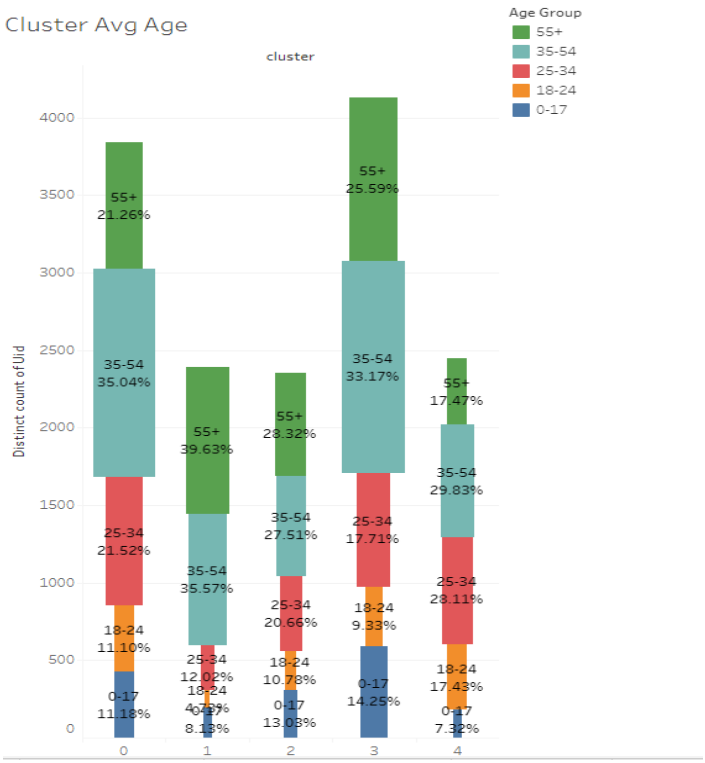


Figure 2a - Age distributions between clusters

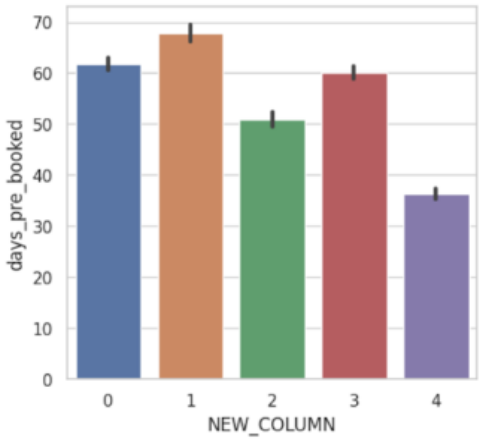


Figure 2b - Days pre-booked between clusters

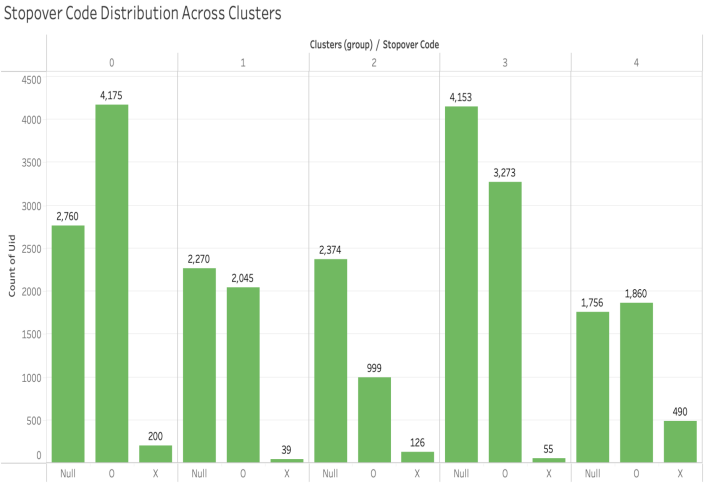
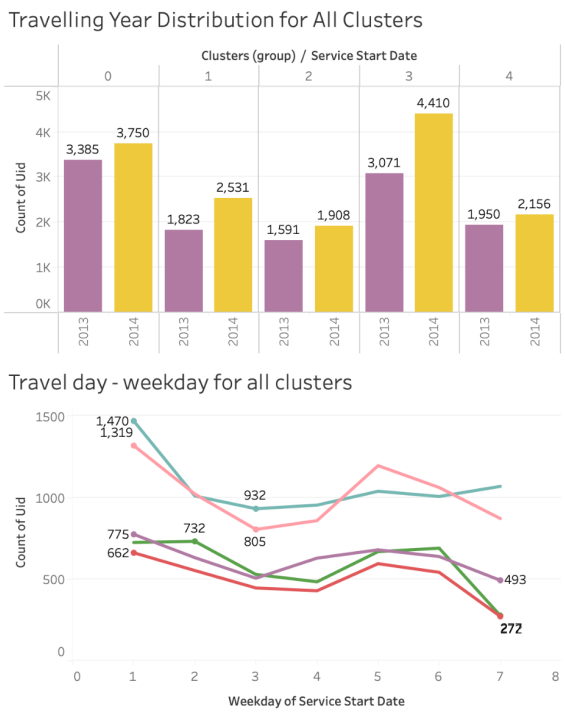


Figure 2c - Stopover code between clusters (Null = No Layover, O = Layover <24hrs, X = Layover > 24hrs)

Below, we also have Figures 3a, 3b, and 3c, displaying the days, months, and years with the most bookings among all clusters. These will also be referenced in individual sections.

Figure 3a - Flights by year between clusters



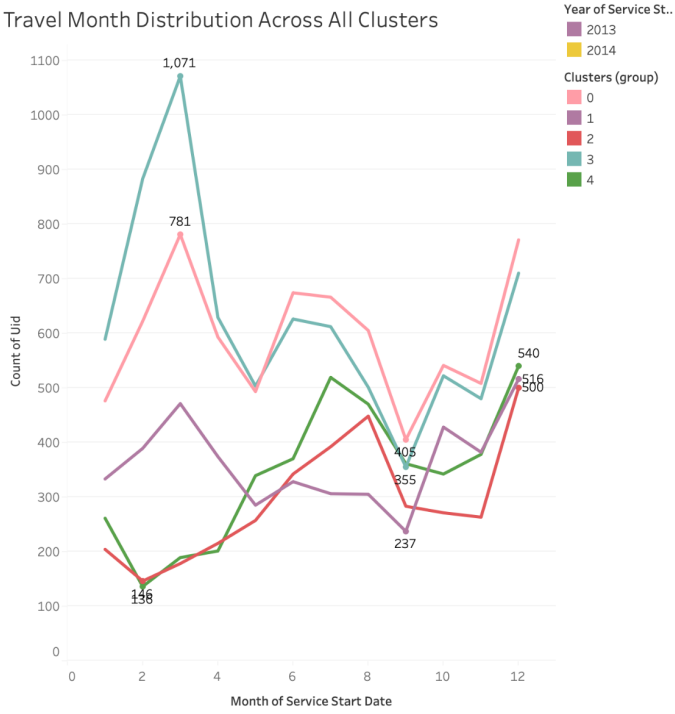


Figure 3b - Number of flights by day between clusters

Figure 3c - Number of flights by month between clusters

Cluster 0 - Middle-aged independents, and convenient flyers

First of all, the customer segment of cluster 0 can be defined as middle-aged independent and convenient travelers. The most common group size is 2 people **(Figure 4a)**, which implies that people are flying with their spouses, friends, or colleagues. This group's dominant age group is middle-aged, with 35% of people being 35-54 **(Figure 2a)**. They are

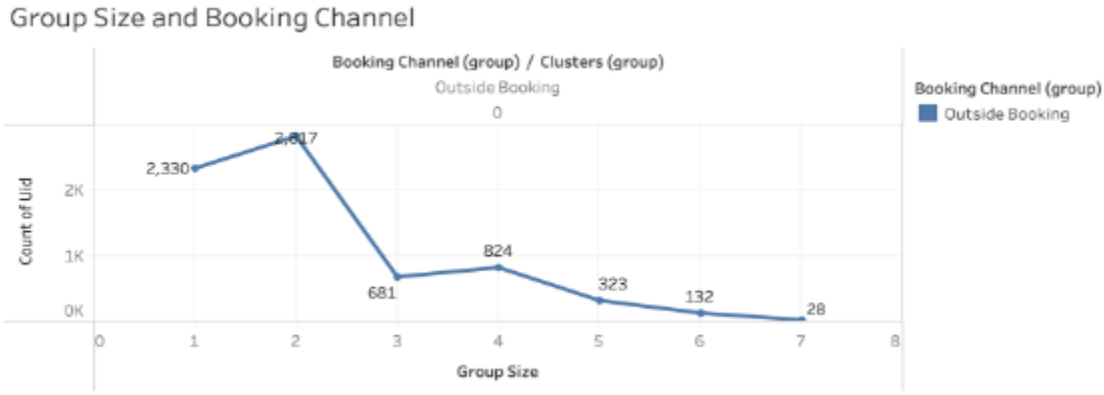


Figure 4a - Cluster 0 number of groups of each size

almost entirely non-Ufly members **(Figure 1a)** and all of them book outside**(Figure 1b)**. These two pieces of evidence indicate that cluster 0 travelers are convenient travelers because they

Figure 5a - Cluster 0 Origin Airport

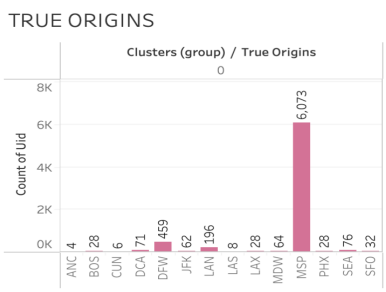
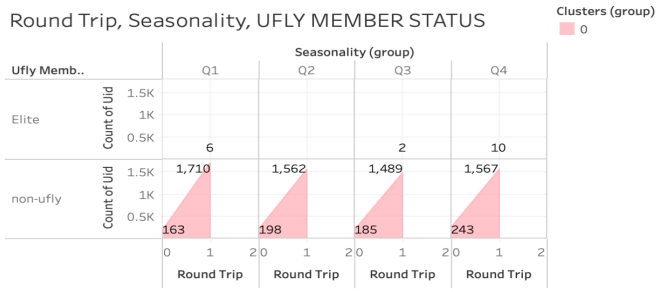


Figure 5b - Cluster 0 round trips (0 for one-way, 1 for round trip) by quarter



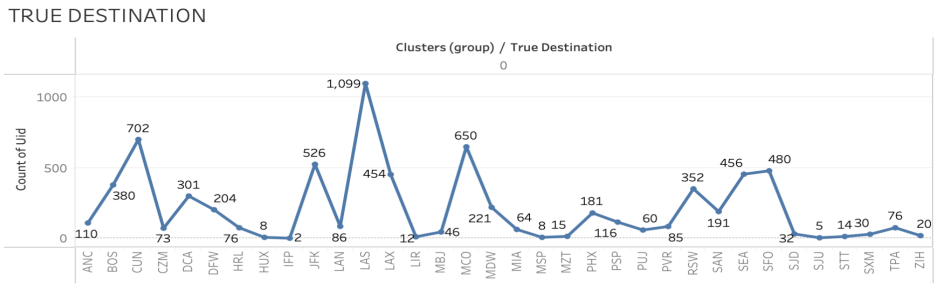


Figure 5c - Cluster 0 Destination Airport

may search for other airlines to find a suitable flight time on the travel applications for their own convenience. Additionally, most travelers from cluster 0 depart from Minneapolis **(Figure 5a)** and travel all over the U.S., mostly in Las Vegas **(Figure 5c)**. October and January are the times when cluster 0 groups book tickets the most; August is the least. The two hottest travel months for cluster 0 are March and December **(Figure 3c)**. These travelers also mostly booked round trip tickets, and they usually book their tickets 60 days before the flight **(Figure 5b & 2c)**. The majority of cluster 0 members are able to stay in layover city less than 24 hours, which means they can accept a long trip with a transfer flight in the middle stated in **Figure 2b**.

In order to attract more customers that also have cluster 0's characteristics, Sun Country airline should promote more during the most booking month, October and in Minneapolis, Minnesota region(Marketing in large events and fairs). Sun Country could give discounts to people who buy from the travel agencies or travel booking applications. Since the travelers usually book in October and pre-booked their flight 2 months before the flight, it means that

they are preparing for their winter break or Christmas. Thus, Sun Country should advertise winter travel in October. For example, an advertisement like a married couple fly from Minneapolis to Las Vegas to celebrate their Christmas party with their friends. Similar to December, March is also when the cluster 0 customers fly frequently. According to Wood, March is one of the most visited months due to the warm weather. Many visitors come to enjoy the desert or casino, which is adequate for a globally known casino city. To increase the sales in March, Sun Country can provide their customers a Las Vegas travel guideline, so they can be introduced to the activities that are available to them during their stays. Sun Country can also collaborate with the travel agencies and booking applications to spread the messages like “Are you ready for March Vegas” or “Christmas is coming.”Not only travel agencies, Sun Country can have an alliance with the hotel near the airport as well. Since cluster 0 travelers can accept to stay in layover city for less than 24 hours, Sun Country could offer a combo booking plan of a flight ticket and a one day accommodation in layover city for the customers. With the accommodation, the travelers may have incentive to plan a one day local trip in the layover city; for instance, shopping, visiting the tourist stops, and so on.

Cluster 1 - Minnesotan Business Network

As for cluster 1, as shown from **Figure 6a**, Ufly members have booked their flights via SCA website. Especially the non-Ufly members, out of 4248 number of people, a grand portion of 2869 chose the website as their main booking

Figure 6a - Cluster 1 member status & booking channel

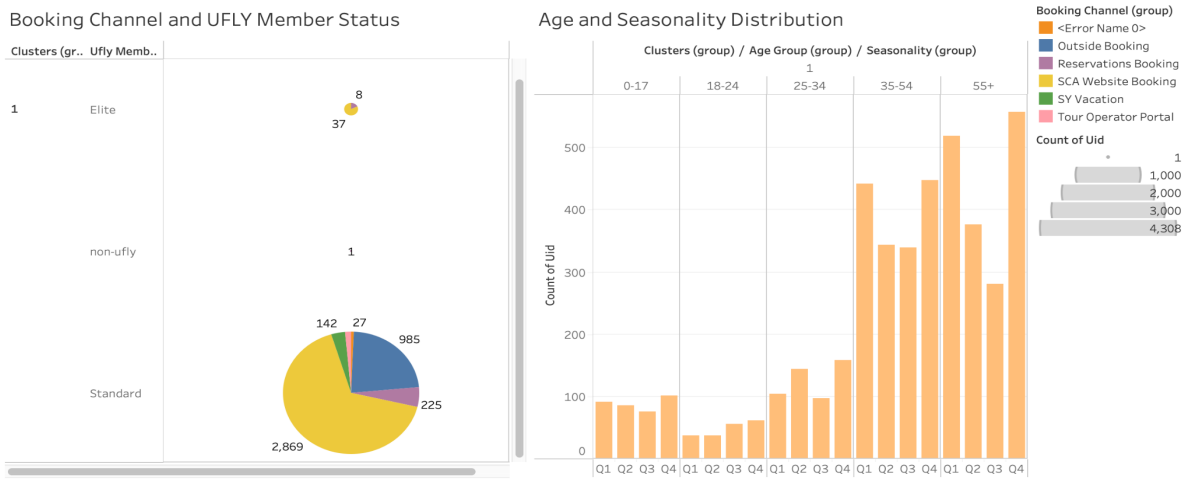


Figure 6b - Cluster 1 booking season by age group

channel. It seems that the majority of the members prefer to stick with the website that they are most comfortable and familiar with. Given that most customers within this range are from ages 35-54 and 55+, it is safe to assume that this customer group would not adapt to changes very quickly. That is to say, major changes in the website could potentially lead to decrease in sales. While a lot of younger generations can migrate to different channels quickly and easily when other channels provide a cheaper option, this age group is most likely to be devoted to the SCA website because of the familiarity.

This age group also shows a high reservation rate in Q1 and Q4. **(Figure 6b)** As most clients consist of businessmen and businesswomen, the start and end of each year is the time for planning the upcoming year's business trip. In cases where companies require their employees to book the flights in advance, they would still pay a significant amount if Sun Country Airlines were to increase the cost for Business Class or First Class during this period. Sun Country Airlines could select a specific week in Q1 and Q4 to sell their tickets at a higher price to target the pre-planning business companies.

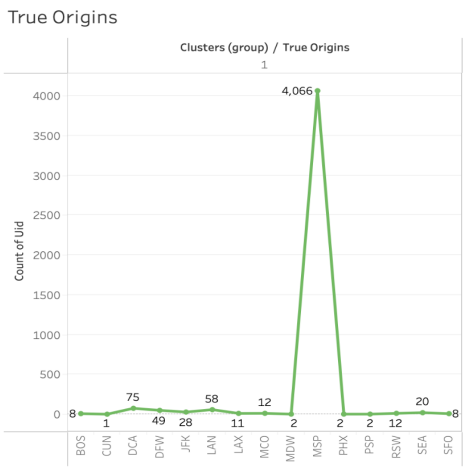


Figure 7a - Cluster 1 origin airport

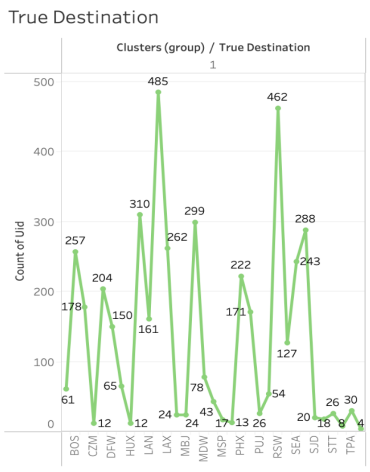


Figure 7b - Cluster 1 destination airport

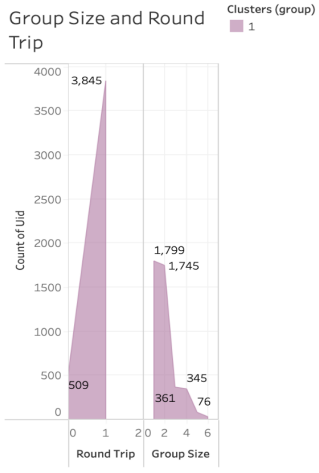


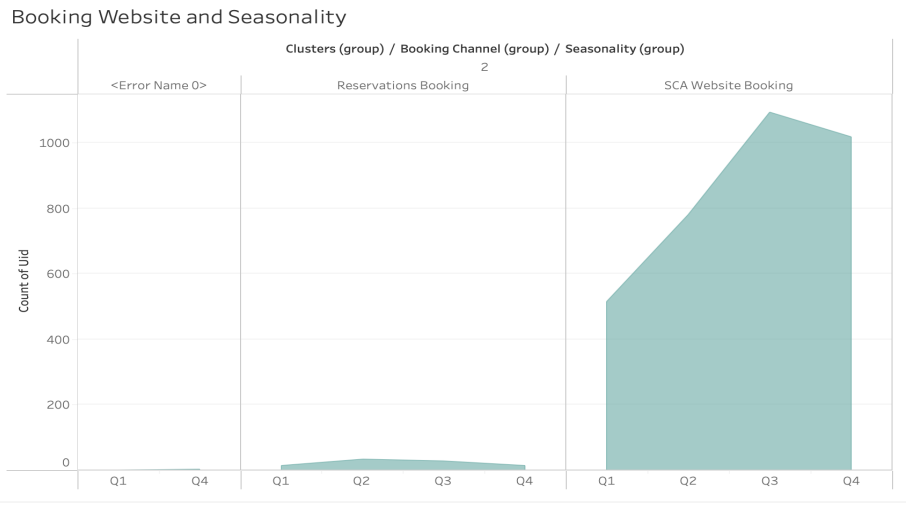
Figure 7c - Cluster 1 group size & round trip distribution

According to **Figure 7a**, this group's major departure point was Minneapolis. Most traveled to different cities, with LAX being the highest destination. People tend to travel either alone or with a single companion. Since both of these airports are international airports and

provide exclusive lounge service, Sun Country Airlines could provide 1 extra lounge coupon for the accompanying traveler if the flight was booked as a SY flight with high booking code.

Cluster 2 - Urgent Travelers

Figure 8 - Cluster 2 booking season & booking channel



Cluster 2 can be described as Middle-aged Urgent travelers. Urgent travelers refer to customers who do not fly frequently and only fly when they have an emergency. In addition, Urgent travelers usually only have a small portion of the market segment for airline business. By looking at **Figure 3a**, cluster 2 has the least number of PNR created, which means cluster 2 has the lowest population among clusters. Furthermore, **Figure 2b** shows that the travelers have an urgent task since most of them do not stay in layover city. Similar to cluster 0, middle-aged group also has a dominant

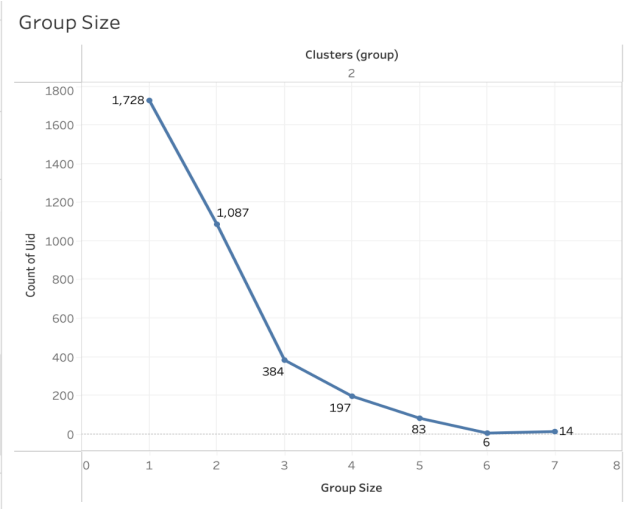


Figure 9a - Cluster 2 group size distribution

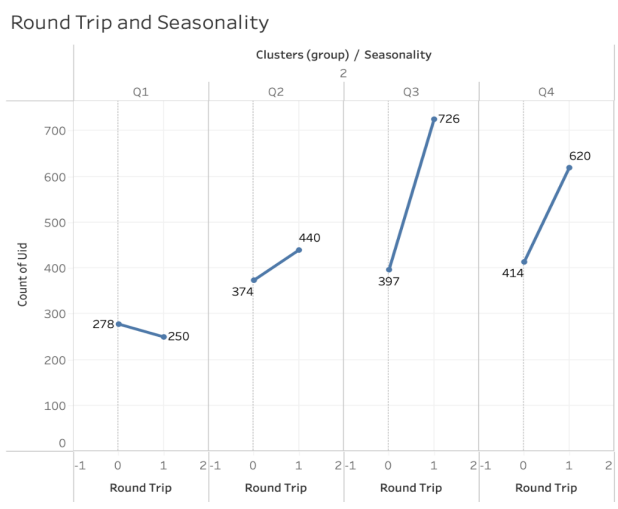


Figure 9b - Cluster 2 round trips (0 = one-way, 1 = round-trip) by season

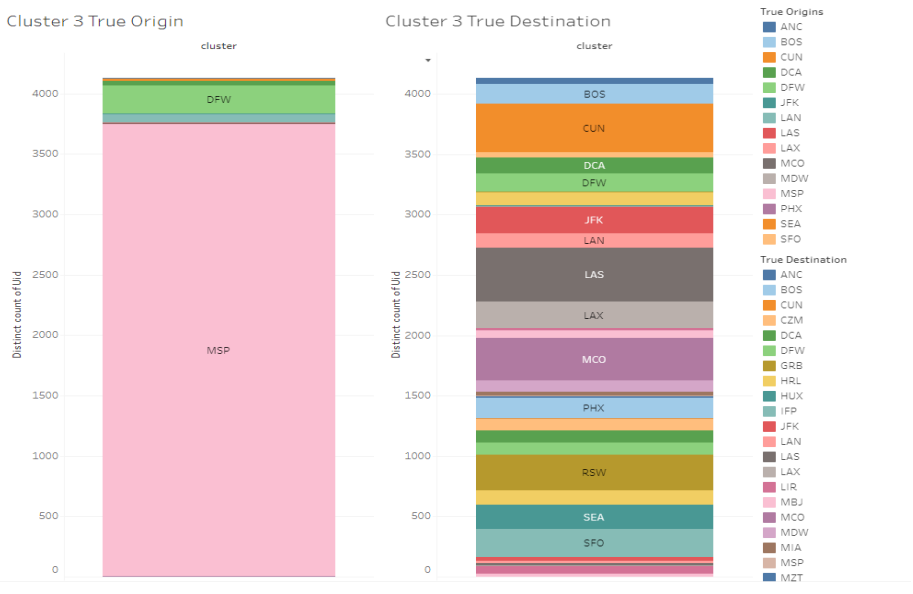
percentage in cluster 2 aged groups **(Figure 2a)**. From 2c, cluster 2 book their ticket 50 days before their flight **(Figure2c)**, which can be assumed that they may have a sudden leisure plan. Besides pre-books days, 69% of cluster 2 population are non-Ufly members, and the rest is the standard member **(Figure 1a)**. Additionally, they mostly order round trip tickets in Fall and Winter based on what is stated on **Figure 8 and 9b**. Surprisingly, SCA website is the majority tool for cluster 2 to purchase the ticket according to **Figure 1b**. The common group size in cluster 1 is one person per group.

Since the visualization shows that cluster 2 travelers are most likely to have an urgent trip in August and December, Sun Country could schedule more flights during these two months. In addition, 2/3 of the cluster 2 population is non-Ufly members; hence, Sun Country should emphasize the significance of being a membership and the difference between standard and non-Ufly members on their official website. For example, customers may get a discount on the ticket, have access to the comfort lounge, and so on. Moreover, due to numerous round trips booking during fall and winter, Sun Country can offer a discount on the fall or winter round trip ticket for Ufly members; therefore, cluster 2 might have incentive to be a part of the members. Lastly, the most important is that the official website has become the purchasing ticket tool for cluster 2 members, so the website is a great resource to increase brand awareness and sales to cluster 2 customers. For instance, publishing posts that mention "Surprising trip" or "Emergency flight" that catch urgent travelers' attention.

Cluster 3 - Minnesotan Families

Onto Cluster 3, this cluster is composed of non-Ufly members from the Minnesota area traveling to vacation destinations, such as Cancun, Las Vegas, Florida, and New York as shown by **Figures 10a and 10b**, displaying the origin airport and

Figure 10a - Cluster 3 origin airport **Figure 10b - Cluster 3 destination airport**



destination airports of this cluster. What we've found about this group of customers is that there are proportionally more young travelers (0-17) than other groups **(Figure 2a, pg 2)**, the average group size was the largest, and this cluster was most actively booking around 2 months in advance for flights in March, December, April, June and July, in that order. Thus, we believe this segment is composed largely of families going on vacation. Henceforth Cluster 3 will be referred to as "Minnesotan Families". We've noticed through our exploration of the data that while Minnesotan Families are booking their trips through the SCA website, they are largely not members of the Ufly rewards program. As Minnesotan Families are booking most of their trips in Q1 and Q4 **(Figure 11)**, we recommend focusing advertising efforts on the SCA website targeting this group during those months, specifically in January if we were to choose a single month to focus on. Banner ads or email ads advertising cheaper group fares would likely be an effective way to secure membership from Minnesotan Families. Additionally, expanding Ufly benefits to include bundles in these vacation destinations, such as with resorts in Cancun or show tickets in Vegas, would likely lead to a rise in membership among this group. Offering discounts on grouped seats to our Ufly member benefits would likely lead to a rise in membership among this customer segment as well. As these families are from the Minnesota area, increasing advertising efforts geographically targeted in this area, such as on local television and radio, will likely lead to growth in sales. Summarily, expanded Ufly member benefits and timely targeted ads are our major suggestions for this group.

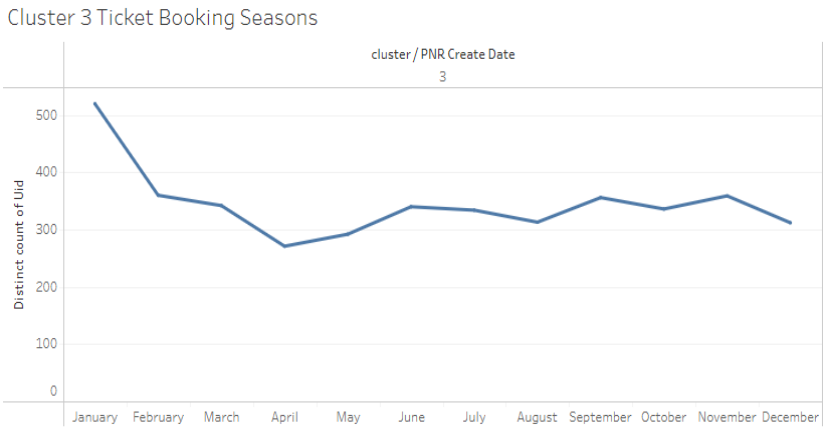


Figure 11 - Cluster 3 ticket bookings by month

Cluster 4 - Spontaneous Tourists

Lastly is Cluster 4. Looking at **Figures 12a and 12b**, we can see that the origin and destination charts are the inverse of the Minnesotan Families - this cluster is composed of

people from various origins traveling to Minnesota. It consists of the largest proportion of 18-24 year olds, and overall 75% of the travelers in this group are between the ages of 18-54. Additionally, this group tends to have the least time between booking their flights and traveling, with an average of around 36 days between booking and flying. Thus, cluster 4 will be known as our “Spontaneous Tourists”. These Spontaneous Tourists are traveling largely in Q3 and Q4 **(Figure 13)** , and have the smallest average group size of 1.55. An average group size of less than 2 suggests that the majority of these travelers are traveling alone, and as they seem to be mostly tourists spending the summer in the area. By adding package deals with local hotels, and specifically deals for solo travelers to the benefits of Ufly membership and advertising such to our Spontaneous Travelers, we should see Ufly membership rise and overall sales with this customer segment rise.

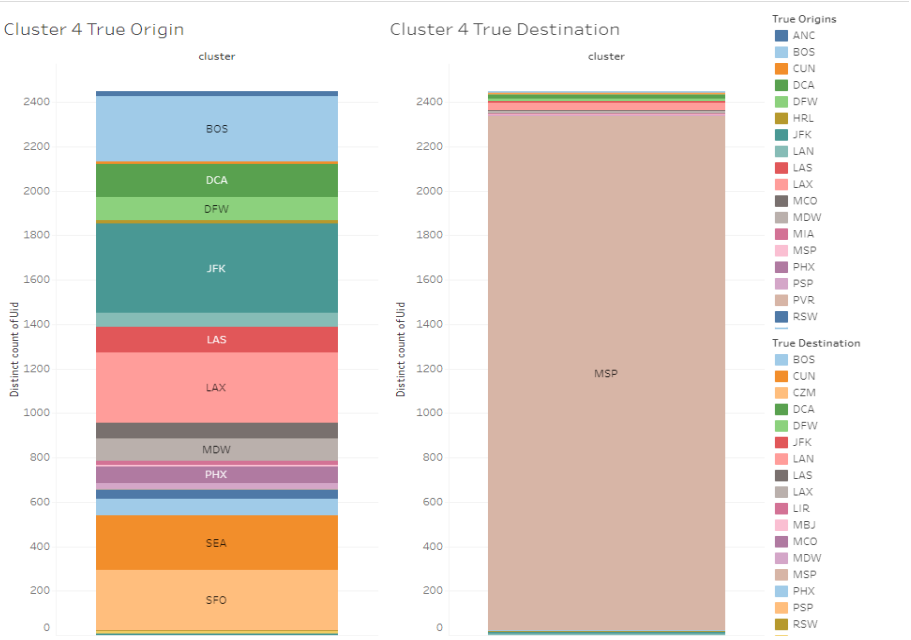


Figure 12a - Cluster 4 origin airport

Figure 12b - Cluster 4 true destination

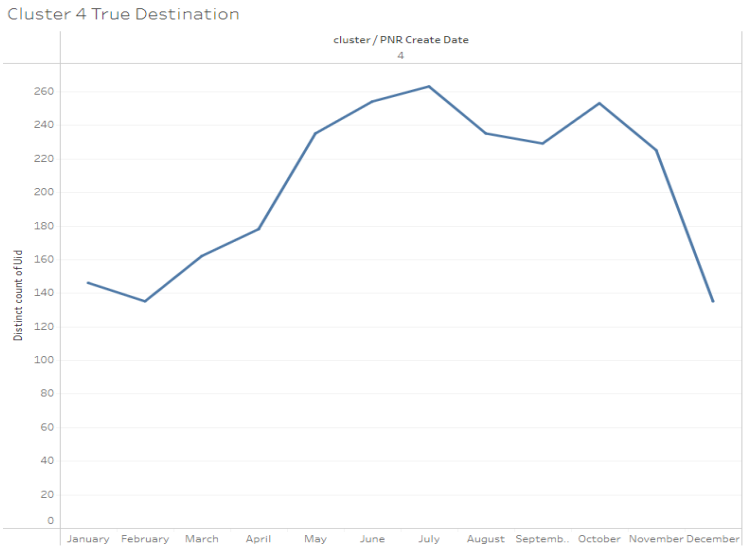


Figure 13 - Cluster 4 ticket bookings by month

Appendix

PART 1:

You will first identify customer segments in the data, using the K-Means Clustering method (in Python) assuming there are 5 centroids. You will use Clustering Data.csv for this clustering exercise.

Code written in Google colab :

```
[1]  import numpy as np
     import pandas as pd
     from sklearn.cluster import KMeans
```

```
[2]  from google.colab import files
     uploaded = files.upload()
```

```
Choose files  Clustering Data.csv
• Clustering Data.csv(text/csv) - 4330134 bytes, last modified: 30/08/2023 - 100% done
Saving Clustering Data.csv to Clustering Data.csv
```

```
[3]  df = pd.read_csv('Clustering Data.csv')
```

```
[4]  df_1 = df.iloc[:,2:90]
```

Used kmeans to generate clusters

```
[5]  kmeans = KMeans(n_clusters = 5, init='k-means++',random_state=0)
     kmeans.fit(df_1)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
  warnings.warn(
KMeans
KMeans(n_clusters=5, random_state=0)
```

```
[7]  predict=kmeans.predict(df_1)

[8]  predict

     array([1, 1, 2, ..., 2, 2, 4], dtype=int32)

[9]  df['NEW_COLUMN'] = pd.Series(predict, index=df.index)
```

In the dataframe ‘df’, added a new column with the name ‘NEW_COLUMN’ to label each row(or UID) with specific cluster number(0,1,2,3,4).

PART 2:

Second, you need to extract the cluster "Assignments" from the cluster results data set and add it to the reservations data in the sample_data_transformed.csv file. You can use the following Python code to merge the data, assuming the reservation data is in the dataframe named customer_data and the clustering results are in the dataframe named clustered_data:

final_dataframe = customer_data.merge(clustered_data[['uid', 'cluster']], on='uid', how='left')

The above code performs a left join between two dataframes, in which all rows from the customer_data dataframe are retained, and matching rows from the clustered_data dataframe are joined based on the 'uid' column.

Code written in Google Colab:

```
[11] from google.colab import files
     uploaded = files.upload()

     Choose files  sample_dat...sformed.csv
     • sample_data_transformed.csv(text/csv) - 8702953 bytes, last modified: 30/08/2023 - 100% done
     Saving sample_data_transformed.csv to sample_data_transformed.csv

[12] customer_data = pd.read_csv('sample_data_transformed.csv')

     <ipython-input-12-15d3700b9d23>:1: DtypeWarning: Columns (13) have mixed types. Specify dtype option on import or set low_memory=False.
       customer_data = pd.read_csv('sample_data_transformed.csv')

     final_dataframe = customer_data.merge(df[['uid', 'NEW_COLUMN']], on='uid', how='left')
```

In this code we have used ‘NEW_COLUMN’ instead of ‘cluster’ as we renamed the column in code in a different way.

Finally, we converted this dataframe to CSV file. And downloaded this CSV file.

```
[14] final_dataframe.to_csv('final_dataframe.csv')

from google.colab import files
files.download("final_dataframe.csv")
```

Part 3:

Third, visualize the segment data, using Python or other tools of your choice. You should use the combined dataset that you have obtained in the previous step. You should create at least 3 charts/figures for each segment and use at least 2 different types of charts (for example, line, bar, or box chart, etc.) in the report, in order to "tell a story" about the unique nature of each segment.

For visualization, we used Python (Seaborn, Matplotlib) and Tableau. We created visualizations for each customer segment(each cluster) and some visualizations across all the clusters to get the idea of trend or distribution across all clusters.