BANA 200: Foundations of Business Analytics

Professor Mingdi Xin

Group Project Report

Team 12B

Becky Wang

Vy Nguyen

Dennis Wu

Hsiang-Han (Cyan) Huang

Kenjee Koh

Introduction + Executive Summary

Objective

In this business analysis report, our group aim to find insights from the reservation data for The Sun Country Airline, discovering the unique characteristics in the customer segments through the relationship between the customer segments and specific variables. With the graphs and insights, our team will provide marketing recommendations for each customer segment.

Key Findings

According to the clustering by Python, we identified 5 customer segments for each cluster and their unique characteristics:

- Cluster 0 - Spontaneous Direct Flyers in Summer
- Cluster 1 - Non-Member Seasonal Group Travelers
- Cluster 2 - Non-Member Middle-Aged Leisure Travelers
- Cluster 3 - Ufly Early-Bird Direct Flyers
- Cluster 4 - Non-Member Business Travelers

Recommendations

1. Offer tailored packages to solo and business travelers, along with personalized itineraries and incentives like tier upgrades to business travelers.
2. Provide more non-digital marketing channels to reach the cluster that creates the most revenue.
3. Design tailored packages for summer and winter/spring travel preference groups with popular seasonal travel destinations.
4. Capitalize Cluster 3's loyalty by expanding Ufly membership through referral and appreciation programs, while offering tier-based perks and premium service discounts.

Data Description

In the dataset, there are variables about detailed information on passenger bookings, including personal details, booking channels, travel specifics, and loyalty program status.

Methodology and Approach

Clustering Approach

K-Means is a clustering algorithm, it will pick out instances in clusters that are like each other but different from other clusters. First, we tested optimum number of clusters in range 1 to 21 and designed an elbow curve plot to find the suitable number (cluster =5). Then we applied the cluster to fit data columns, running 30 times with different centroids. Last, we assigned the cluster column back to the original clustering data by merging on UID number.

Data Preparation

By looking at the original data, there are three kinds of data: continuous variables, categorical variables and Boolean variables. For continuous variables, we scaled the influence of each variable and replaced it with the transformed data; For categorical variables, we chose more influential columns and converted them to dummy variables; For Boolean variables, we just converted them to numerical codes.

Customer Segmentation Analysis

Number + Overview of Segments

As a result of K-Means Clustering method, there are 5 customer segments with different sizes and key distinct characteristics.

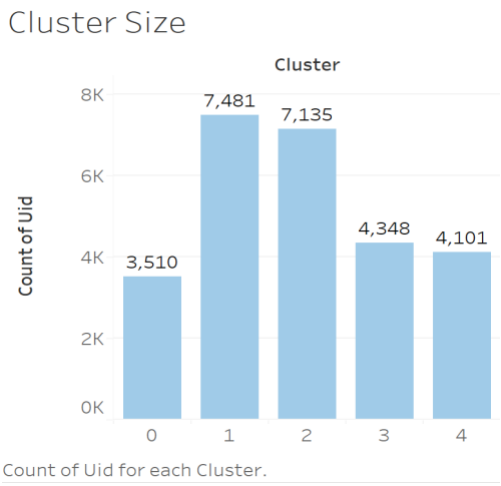

Figure 1: Size of Five Clusters

From the chart, a majority of Sun Country Airlines customers classified into clusters 1 and 2. Meanwhile, clusters 0,3, and 4 have significantly lower numbers of customers compared to the other two. This could be interpreted as clusters 0, 3, and 4 representing smaller, more niche segments of Sun Country customer base.

Segment Characteristics - Descriptive for All Clusters

Despite some sets of clusters sharing the same size and some common features, all 5 customer segments have key characteristics that set them apart from each other. Below is the summary of each cluster's distinct characteristics and its descriptive title, which will be analyzed later:

1. **Cluster 0 - Spontaneous Direct Flyers in Summer**
    - *Demographics and Membership:* This customer segment is mainly made up of non-cardholders primarily flying.
    - *Travel Behavior:* These customers tend to fly straight to the destination.
    - *Booking Patterns:* They mostly spend and travel in Quarter 3 and tend to book flights closer to the departure date.
2. **Cluster 1 - Non-Member Seasonal Group Travelers**
    - *Demographics and Membership:* The majority of customers are Non-Ufly flyers.
    - *Travel Behavior:* These customers tend to fly in groups and halt in a city for less than 24 hours.
    - *Booking Patterns:* Despite spending equally on flight tickets throughout 4 seasons, they make trips mostly in Quarter 3 and 1.
3. **Cluster 2 - Non-Member Middle-Aged Leisure Travelers**
    - *Demographics and Membership:* Most customers in this segment are over-35-years old and Non-Ufly.
    - *Travel Behavior:* A large proportion of customers have transits in their trips, either layover or stopover.
    - *Booking Patterns:* Similar to Cluster 1, their trips are mostly made in Quarter 3 and 1. However, they only use “Outside Booking” instead of “Website Booking”.
4. **Cluster 3 - Ufly Early-Bird Direct Flyers**
    - *Demographics and Membership:* Almost all customers in this segment are Ufly flyers.
    - *Travel Behavior:* They prefer non-stop flights.
    - *Booking Patterns:* Similarly, their trips are mostly made in Quarter 3 and 1; however, unlike other segments, these customers book tickets days and months in advance for the flights.

5. **Cluster 4 - Non-Member Business Travelers**
    - *Demographics and Membership:* A large proportion of customers in this cluster do not have membership.
    - *Travel Behavior:* They are mostly solo travelers who stop over at multiple cities in a trip for more than 24 hours.
    - *Booking Patterns:* Most of them are last-minute bookers and primarily fly in summer.

Recommendations for Sun Country Airlines

Marketing Recommendation 1

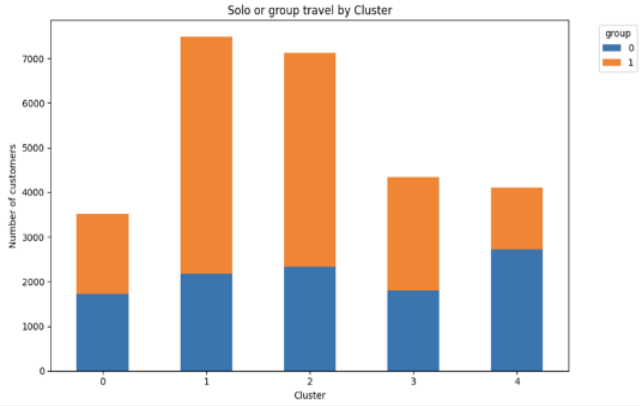

Figure 2: Solo or Group Travel by Cluster

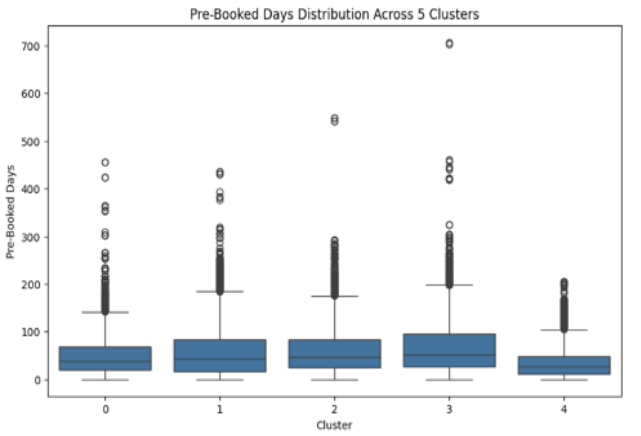

Figure 3: Pre-Booked Days by Cluster

Cluster 4 passengers, primarily solo travelers with last-minute bookings and one-way trips, often have a stopover at MSP before flying to major business cities. This indicates they are likely business travelers. Sun Country Airlines can target this group by offering tailored packages, including flight deals, single-person hotel options, and ride-hailing credits.

Incentives like tier upgrades or first-booking discounts for non-Ufly fliers, combined with emails focused on time-saving benefits, would attract this segment.

Expanding "Stay-and-Fly" partnerships with MSP hotels and lounges, and co-branded deals with Fortune 500 companies, can provide exclusive perks for business travelers with stopovers at MSP.

Marketing Recommendation 2

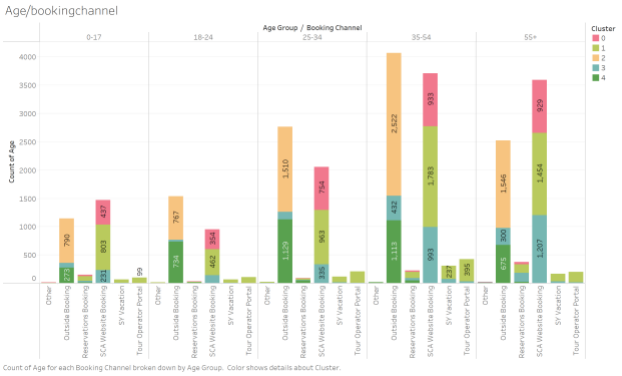

Figure 4: Count of Age by Booking Channel

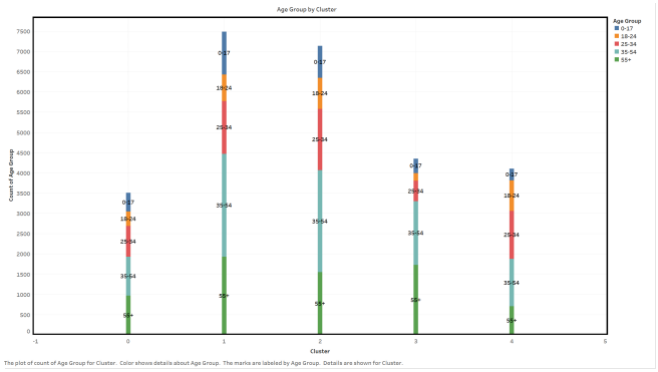

Figure 5: Age Group by Cluster

For Cluster 1 & 2, we can see that they are made up of a significantly larger proportions of fliers in the age groups of 35-54 & 55+. Additionally, we can also observe that Cluster 2 only utilizes " Outside Booking" channels as opposed to digital channels such as the SCA Website and Tour Operator Portal.

Based on our analysis, we've identified that customers in Clusters 1 and 2 are more likely to use non-digital channels when booking their flights. Given this insight, it may be beneficial to invest additional resources into non-digital marketing efforts, such as mailbox fliers, vouchers, and airport booths, to better reach these customers. These channels could provide more effective engagement with these segments, who might rely less on online interactions. By enhancing our presence in offline marketing, we can better align our strategies with the preferences of these customer segments, potentially driving more bookings and increasing brand loyalty.

Marketing Recommendation 3

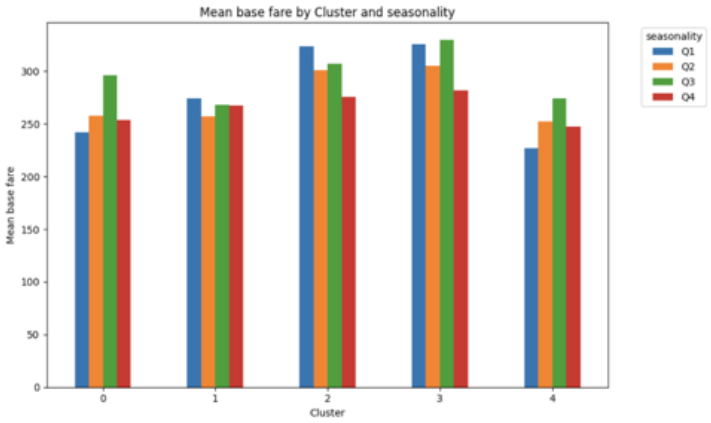

Figure 6: Mean Base Fare by Cluster & Seasonality

According to *Figure 6*, we discovered that each cluster has its own preferred season for traveling. For cluster 0 & 4, they travel more in summer; For cluster 1, 2 & 3, they prefer to travel in the winter/spring. Our team suggest marketing campaigns that are based on the popular destination for the two different groups.

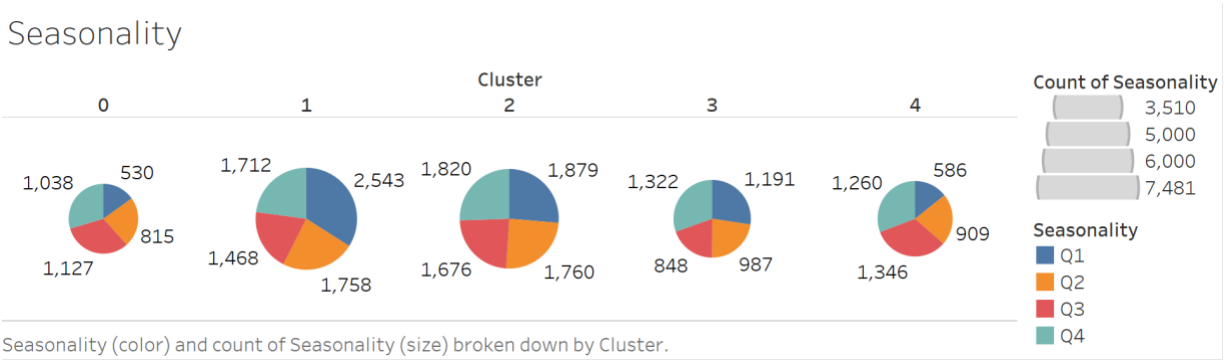

Figure 7: Cluster and Seasonality

For Summer Travelers, Cluster 0 & 4, we can promote Fort Myers' pristine beaches and outdoor activities; create exclusive VIP packages that include flight, hotel stays at luxury resorts, and tickets to popular shows or events in Las Vegas; also, feature unique cultural experiences like guided tours of coffee plantations, local cooking classes, or traditional festivals in Costa Rica.

For winter/spring travelers, cluster 1,2 & 3, we can offer flight and hotel packages that include tickets to iconic Seattle attractions such as the Space Needle, Pike Place Market, and the Museum of Pop Culture; we can offer exclusive deals on flights, luxury beachside hotels, and resort amenities in Miami.

According to *Figure 2,* we find out that cluster 1 & 2 are group-travelers. Our team suggests that Sun Country Airlines can promote "Family Packages" and "Group-oriented" packages for them, creating family-friendly deals that include kid-friendly resorts, activities, and excursions.

Marketing Recommendation 4

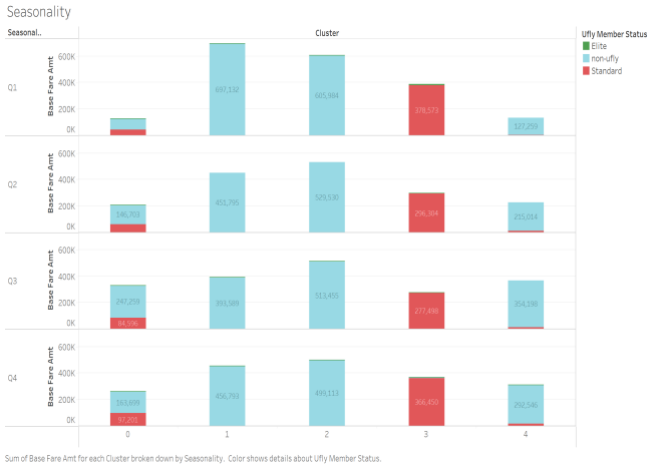

Figure 8: Membership Status by Cluster

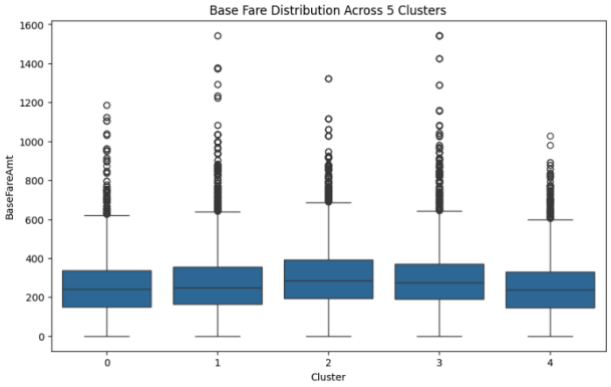

Figure 9: Base Fare Distribution by Cluster

Cluster 3 stands out from the *Figure 8* as the only group with nearly all customers enrolled in the Ufly Rewards progra. Sun Country can leverage this segment's loyalty to expand their Ufly membership through Referral and Appreciation programs. These could include exclusive referral bonuses (lounge access, free baggage, extra miles, etc.), partnering with businesses or organizations where customers in Segment 3 frequently travel for work to promote membership as a perk for corporate travelers.

Besides, the largest median and variation in Base Fare, as shown in *Figure 9,* indicate that customers in Cluster 3 are willing to pay more for additional services or premium options. Combining with early-booking tendency from *Figure 3,* Sun Country can further capitalize on this Cluster by introducing tier-based loyalty perks to upgrade tier for customers who consistently book in advance or offering discounts for Ufly travelers who purchase premium adds-on services. This could encourage long-term commitment for existing Ufly members in addition to attracting more members.

Conclusion

Summary of Findings

- The K-Means clustering method identified five distinct customer segments, each with unique characteristics. Clusters 1 and 2 represent most of Sun Country Airlines' customers, while Clusters 0, 3, and 4 account for smaller, niche segments.
- Cluster 0 consists of spontaneous direct flyers who primarily travel in summer and book at the last minute. Cluster 1, composed of non-member seasonal group travelers, and Cluster 2, non-member middle-aged leisure travelers, both prefer group travel, with Cluster 2 relying solely on non-digital booking channels. Cluster 3 features Ufly members who book flights well in advance and prefer non-stop trips. Cluster 4 comprises non-member business travelers,

Overall Recommendations

Based on the segmentation analysis, we recommend Sun Country Airlines implement the below strategies to better serve distinct customer groups and maximize revenue potential.

- **Business Travelers (Cluster 4)**: To attract solo business travelers with last-minute bookings, offer specialized packages that include flight deals, single-person hotel options, and ride-hailing credits. Incentives like tier upgrades and first-booking discounts, combined with personalized emails focusing on time-saving benefits, would further enhance appeal. Expanding "Stay-and-Fly" partnerships with MSP hotels and lounges, and establishing co-branded deals with Fortune 500 companies, would provide exclusive perks and strengthen loyalty.
- **Older and Group Travelers (Clusters 1 & 2)**: Customers in these clusters prefer non-digital booking channels. Sun Country Airlines should invest in non-digital marketing efforts, such as mailbox fliers, vouchers, and airport booths, to engage with these segments. Promoting family and group-oriented packages, with kid-friendly resorts and activities, would appeal to group travelers, especially in Clusters 1 and 2.

- **Seasonal Travelers (Clusters 0, 1, 2, 3 & 4)**: Tailor marketing campaigns to target seasonal preferences. Promote summer destinations like Fort Myers, Las Vegas, and Costa Rica to travelers in Clusters 0 and 4. For winter/spring travelers in Clusters 1, 2, and 3, focus on destinations like Seattle and Miami, offering exclusive flight and hotel packages to align with travel trends.
- **Ufly Members (Cluster 3)**: Leverage the strong loyalty of Ufly members by expanding Referral and Appreciation programs. Offer exclusive referral bonuses like lounge access, free baggage, and seat upgrades. Additionally, capitalize on the higher spending patterns and early booking tendencies in this cluster by introducing tier-based loyalty perks and discounts on premium add-ons to incentivize long-term commitment.

Future Considerations

The current segmentation analysis is based on 2013-2014 data, and customer behavior may have shifted significantly since then. Regular re-clustering and updated segmentation models will be essential to track evolving preferences. In addition, diving deeper into customer satisfaction and feedback could provide valuable insights. Gathering detailed post-flight surveys or satisfaction data for each segment could reveal customer preferences, pain points, and areas for improvement.

Python Code for K-Means clustering

[1]:
```
#Loading Dataset
import pandas as pd
data = pd.read_csv('Clustering Data.csv')
pd.set_option('display.max_columns', 100)
print(f'The dataset has {data.shape[0]} rows and {data.shape[1]} columns')
data.head()
```

```
The dataset has 15144 rows and 90 columns
```

[3]:
```
#Missing values check
missing_values = data.isnull().sum()
missing_values
```

[5]:
```
#Prepare columns for K-Means
clustering_data = data.copy()
clustering_data.head(10)
```

[7]:
```
#Droping two string column(*Not sure if I can drop columns)
columns_to_cluster = clustering_data.iloc[:, 2:]
columns_to_cluster.head(10)
```

[9]:
```
#Find an optimum cluster we should use in K-Mean
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

inertia = []                  # sum of squared distances between each data point and its closest cluster centroid (wcss)
cluster_range = range(1, 21)


for cluster_num in cluster_range:

    print(f'Iteration Number: {cluster_num}')

    kmeans = KMeans(n_clusters=cluster_num, n_init=10)
    kmeans.fit(columns_to_cluster)
    inertia.append(kmeans.inertia_)


#Design plot figures
plt.figure(figsize=(10,6))
plt.plot(cluster_range, inertia, marker='o', linestyle='--')
plt.title('Elbow Curve')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia')
plt.grid(True)
plt.show()
```

[13]:
```
#Applying K-Means
kmeans = KMeans(n_clusters=5, n_init=30)
kmeans.fit(columns_to_cluster)

####### For analysis, do NOT assign the labels to the transformed data!
data['Cluster'] = kmeans.labels_

#print cluster(result)
data.head(10000)
data.to_csv('clustered_data.csv', index=False)
```

[15]:
```
#Summary of cluster size
number_of_clusters = data['Cluster'].nunique()

for i in range(number_of_clusters):
    print(f"The size of the Cluster {i}  |  {data[data['Cluster'] == i].shape[0]}")
```

[17]:
```
#Dataframe joined based on the 'uid' column
transformed_data = pd.read_csv('sample_data_transformed.csv',low_memory=False)
final_dataframe = transformed_data.merge(data[['uid', 'Cluster']], on='uid', how='left')
final_dataframe.head(10000)
final_dataframe.to_csv('final_clustered_data.csv', index=False)
```

Python Code for Visualization

```
# Number of customers by Cluster and groupVisualize
customers_by_group.plot(kind='bar', stacked=True, figsize=(10, 7))
plt.title('Number of customers by Cluster and group')
plt.xlabel('Cluster')
plt.ylabel('Number of Customers')
plt.xticks(rotation=0)
plt.legend(title='group', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```

```
# Solo or group travel by Cluster
group_by_cluster = final_clustered_data.groupby('Cluster')['group'].value_counts().unstack()
group_by_cluster.plot(kind='bar', stacked=True, figsize=(10, 6))
plt.title('Solo or group travel by Cluster')
plt.xlabel('Cluster')
plt.ylabel('Number of customers')
plt.xticks(rotation=0)
plt.legend(title='group', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```

```
# Base Fare Distribution Across 5 Clusters
plt.figure(figsize=(10, 6))
sns.boxplot(x='Cluster', y='BaseFareAmt', data= final_clustered_data)
plt.title('Base Fare Distribution Across 5 Clusters')
plt.xlabel('Cluster')
plt.ylabel('BaseFareAmt')
plt.show()
```

```
# Pre-Booked Days Distribution Across 5 Clusters
plt.figure(figsize=(10, 6))
sns.boxplot(x='Cluster', y='days_pre_booked', data= final_clustered_data)
plt.title('Pre-Booked Days Distribution Across 5 Clusters')
plt.xlabel('Cluster')
plt.ylabel('Pre-Booked Days')
plt.show()
```