Customer Segmentation for XYZ Airlines & Recommendations

Executive Summary

Aditya Tomar, Jhan-Syuan Lin, Niharika Agnihotri, Saloni Jadhav, Yash Bhirud

---

Business Goal: Understand various types of customers based on their behavior patterns
Our data analysis team analyzed XYZ Airlines' customer data to form meaningful customer segmentations and provide insights based on their characteristics. These insights allow XYZ to better understand its customers based on their needs, wants & behavior.

Initial Findings on exploring the dataset:
The data consists of a mix of XYZ's flight records as well as information from their UFly Rewards program database. From this, the following are our initial findings:
- In a total of 1.4 million users, approximately 82% of them are non-members
- On average, the total bookings done by members is 1.2 times more than the non-members
- The top 10% of customers based on dollars spent contribute to 50% of XYZ's revenue
- Top destinations are Las Vegas, New York & Florida.
- There is also a significant amount of customers who had their flight tickets upgraded or downgraded: For example, around 5% of customers who booked 'coach' did not end up traveling in it, and around 35% of customers who booked Discount First Class ended up getting upgraded to 'Coach', and approximately 3% of First class passengers got upgraded.

**Approach: To use clustering methods to group and understand similar customers**
- Since the main goal for XYZ is to understand its various types of customers, we have taken the approach of grouping similar customers together using clustering analysis. Improved awareness of the customer base can help refine marketing and expansion strategies which yield a better return on investments. Efficient resource management can reduce costs, raise efficiency and boost productivity.
.
- With the help of this analysis, we understand the impact of online booking channels on customer experience, as well as generate ideas as to how to onboard customers on their UFly rewards program through targeted marketing.

Clustering   Analysis
Ultimately, we utilized the Gaussian Mixture Model to cluster the customers into 5 groups. This number is decided by the BIC metric, which the lower score is better. Before we came to our final model, we actually tried various clustering algorithms. The first technique that comes into our mind is K-Prototypes, given its capability to handle mixed data directly. However, we found that it is not very efficient to train such a vast amount of data. This pattern appears again when we tried DBSCAN and K-Medoids, which are both not practical choices for this project. As for k-means, even though it generates the clustering result in a reasonable amount of time, the problem is that there is no clear elbow point for choosing the natural cluster number.

SunCountry Member by Cluster
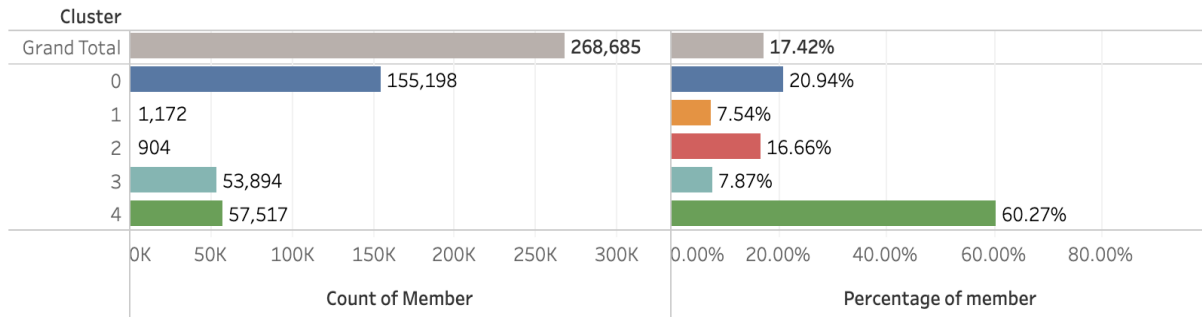
Member Distribution by Cluster

- **Cluster 0** are your **average customers** who encounter nearly **48% of the total cohort.** Nearly **78%** of these customers have booked their tickets **through the SCA website** and their booking/travelling distribution is uniform across all quarters. Also, 20% of these customers are **non-members having no multi-city trips.**
- **Cluster 1** are your **advance bookers** who book nearly **4-5 months in advance** of their travel. These customers have mostly done their bookings during Q3 & Q4 - between July and December, and fly in Q1 - between January and March.
- **Cluster 2** are the **spontaneous/last minute bookers** who have purchased their **ticket at the airport** itself. Even though their distribution is uniform throughout the year, they have the **least average expenditure per person.**
- **Cluster 3** are the **price sensitive bookers** who account for **44% of the cohort** and have made their bookings through **outside channels only**. Almost **93%** of these customers are **non-members** as well.
- **Cluster 4** are your **frequent flyers** who **consist of SC members** and spend the most among all groups. They have the highest average total amount per person with most of their bookings via the SCA website. They also have the highest number of trips with max upgrades. This cluster group also accounts for all credit card holders.
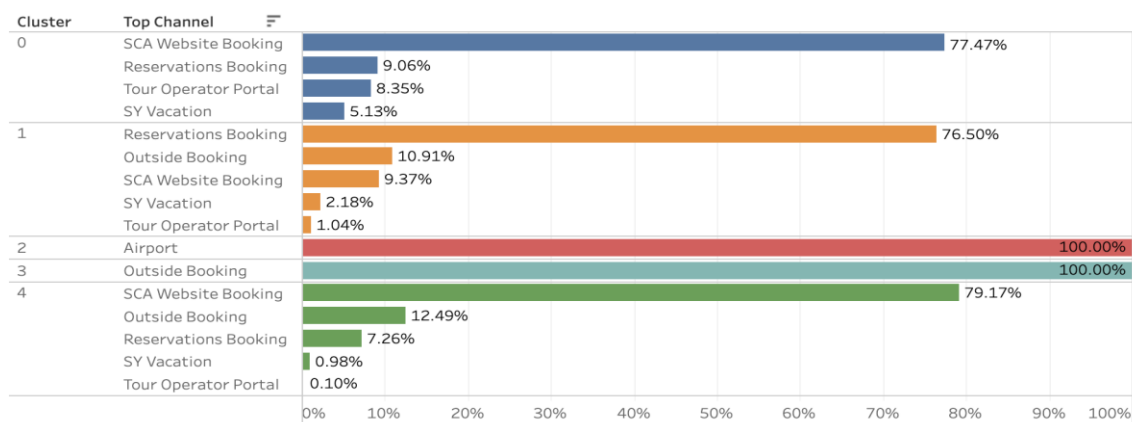
Channel Distribution by Clusters

Recommendations:

- **Average Customers (C0):** Since almost 80% of people book through SCA website but just 20% of them are members, we can **incentivize converting to UFly Membership** during checkout.
- **Advance Bookers (C1):** We can create a **sense of urgency for booking**, as these customers like to research and a time constraint might drive their decision to book quicker.
- **Last Minute Bookers (C2):** For these customers, we can **lower the discounts** since they have time constraints and hence lower bargaining power.
- **Price Sensitive Customers (C3)**: For price-sensitive customers who are booking via external websites, we can **offer them benefits by enrolling in the UFly program** by sending them emails such that if they enroll, they will receive discounted rates in the long term.
- **Frequent Flyers (C4):** We can **offer seat upgrades and rewards** for frequently flying customers to increase overall satisfaction and cement their loyalty to XYZ Airlines.
- **In general,**
  - We can allow customers to gain **Referral points** for friends and family in case they enroll.
  - XYZ can build and **collaborate with companies** by affiliating special business memberships with discounts for **large bookings** and certain routes.
  - Points received from booking via **SCA website should be higher** than if booked through external websites.
  - **Increase awareness** and the benefits of UFly Rewards Program. S**implify the signup** process for the UFly rewards program.
  - UFly Rewards program should have different **membership tiers,** customers in a higher tier should receive more points. This would encourage even price sensitive customers to stay loyal to SCA.
  - Points can be **redeemed to receive benefits** such as free upgrades, seat selection and to move bookings.