Data Insights, Customer Segments and Recommendations
for Sun Country Airlines

Executive Summary Prepared by: Lakshit Bajaj, Amlendu Kumawat, Shubham Midha, Nikita Saini, Kexin Yang

Background and Business Goal

Sun Country is an American ultra-low-cost airline and has been facing intense competition from large national brands. To make them successful and competitive in the long term, Sun Country would need to amplify its use of data and analytics and find unique ways to differentiate itself apart from the industry.

Sun Country would like to start by knowing more about their passengers and developing robust customer insights, which will be critical for them to be able to improve customer digital experience and advance the business. Data sources readily at hand represented 1.52 million customers making 1.86 million trips on Sun Country between January 2013 and December 2014 can be used to help achieve their goals.

Using the existing data sources, we have analyzed the de-identified data, grouped the 1.5M customers into meaningful segments, teased out insights related to the segments, and developed recommendations to help Sun Country Airlines move closer to the business objectives.

Data Preparation and Explorations

Firstly, we evaluated the quality of the raw data received from Sun Country and realized that it required data cleaning before conducting any analysis. We performed sanity checks to assess the raw data based on accuracy, reliability, and suitability and took necessary data transformations steps to make it usable –

- Accuracy and Reliability: We corrected or filtered out erroneous records such as negative age, booking date after flight date, multiple records or names for a single ticket booking and airlines other than Sun Country. Duplicates and outliers were removed, and missing values were imputed
- Suitability: The data was aggregated to a passenger-level database, to make it clustering-conducive

The data with 1.5M customers was further explored to understand the distributions by customer demographic attributes like gender, age, and Ufly status:

- Gender: 52% (791K) of the customers are Female, and 48% (724K) are Male
- Ufly Status: 17.6% (267K) customers are Members (only 0.1% Elite), and 82.4% are Non-Members
- Age: Customers are equally distributed across the various age groups as expected
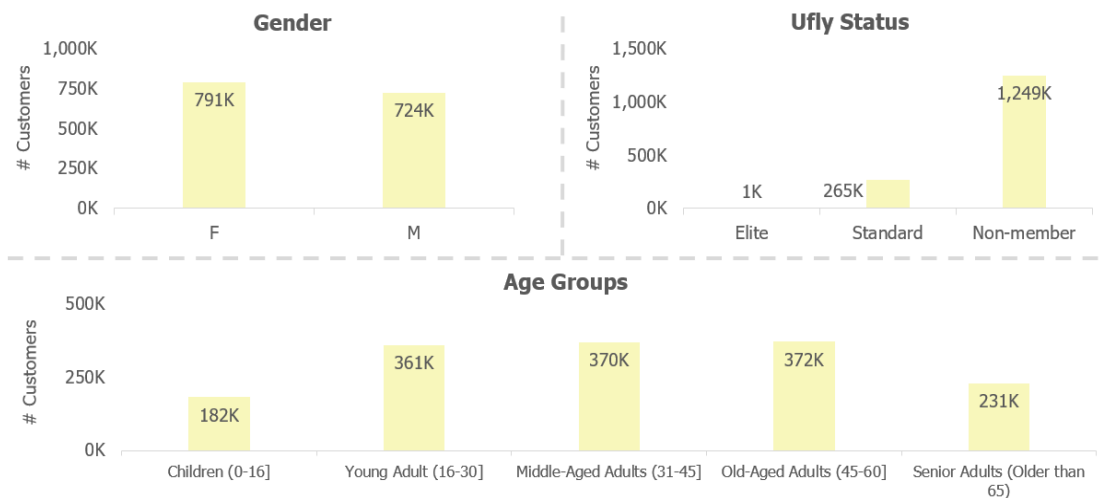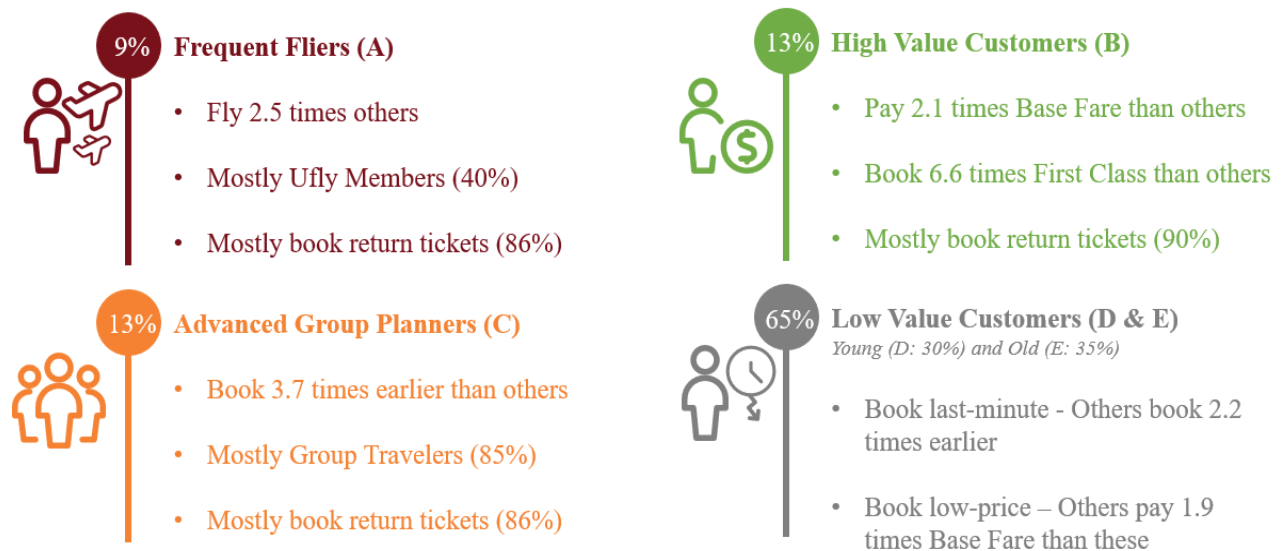
Model Building and Clustering Solution

Stratified Sampling – Owing to the enormous size of data (1.5M customers), clustering on the whole population was computationally expensive. We explored various approaches and technologies, and finally used stratified sampling on the customers based on demographic features to get a representative sample of the population and perform clustering

Clustering Process –
1. Feature Selection and Engineering: Selected features based on domain knowledge and correlation analysis –
   a. Features from Raw Data – Gender, Age, Number of Flights, Base Fare Amount, Ufly Member Status
   b. Engineered Features – % Flights booked using Website, % Round Trips, Date Difference (Days), Passenger Type (Group v/s Individual), % Flights booked during Holidays, % Flights booked using Discounts, % First-Class Tickets booked
2. Clustering Model:
   a. Since the customer-level data consisted of both categorical and numeric variables, we tried multiple approaches that incorporated both data types effectively
   b. We compared the suitability of the following – K-prototype, K-medoids and Gaussian Mixture Model (GMM) and found the most interpretable clusters from the results of K-prototype
3. Choosing the right number of clusters: 5 cluster solution provided the right balance between explainability and inter-cluster differentiability

Clustering Output – Customer Profiles and Insights

Based on clustering, we have four customer groups, with two low-value clusters grouped into one. We have been able to group customers of Sun Country based on flying frequency, average base fare, passenger type (group v/s individual) and return trip booking:

9% Frequent Fliers (A)
- Fly 2.5 times others
- Mostly Ufly Members (40%)
- Mostly book return tickets (86%)

13% High Value Customers (B)
- Pay 2.1 times Base Fare than others
- Book 6.6 times First Class than others
- Mostly book return tickets (90%)

13% Advanced Group Planners (C)
- Book 3.7 times earlier than others
- Mostly Group Travelers (85%)
- Mostly book return tickets (86%)

65% Low Value Customers (D & E)
Young (D: 30%) and Old (E: 35%)
- Book last-minute - Others book 2.2 times earlier
- Book low-price – Others pay 1.9 times Base Fare than these

Recommendations

Target segments A, B, and C –
- They showcase high potential such as flying frequently, paying more, and booking early and round trips, while only 30% of customers in these segments are members
- These segments are 35% of overall customers and drive 55% of the base fare revenue. Their average Base Fare is $400 vs $270 for the overall customer population
- Targeting them would be rewarding for them as well as targeting would yield the highest return on investment for Sun Country Airlines

Ufly Membership Advertising – Customized membership advertising based on customer profiles showcasing –
- Frequent flying benefits to Cluster A
- First-class upgrade benefits to Cluster B
- Group bookings and advance booking benefits to Cluster C

Improve Digital Customer Experience –
- Customers belonging to Cluster C book in groups and ~6 months in advance, mostly for vacation destinations (Las Vegas, Cancún, and Florida)
- They should be pitched for advanced booking with attractive discounts