



Driving Ufly Rewards Enrollment with Data Analytics

Team Sun Savers

Josh Everly | Narae Kang | Mingming Zhang | Wen-Ling Ku | Shashank Magdi

Business Objective



- How can we help drive the enrollment of customers in ufly reward program by understanding different segments of customer profiles?
- Do online booking channels meet the expectations?

Exploratory Data Analysis

Derive peculiar insights from data

/01

EDA Recommendations

Provide Recommendations w.r.t. the entire data based on EDA

/02

Clustering Results

Observe Results from Clustering on Individual Granularity data

/03

Clustering Recommendations

Provide Recommendations based on Clusters

/04

Exploratory Data Analysis

Derive peculiar insights from data

/01

EDA Recommendations

Provide Recommendations w.r.t. the entire data based on EDA

/02

Clustering Results

Observe Results from Clustering on Individual Granularity data

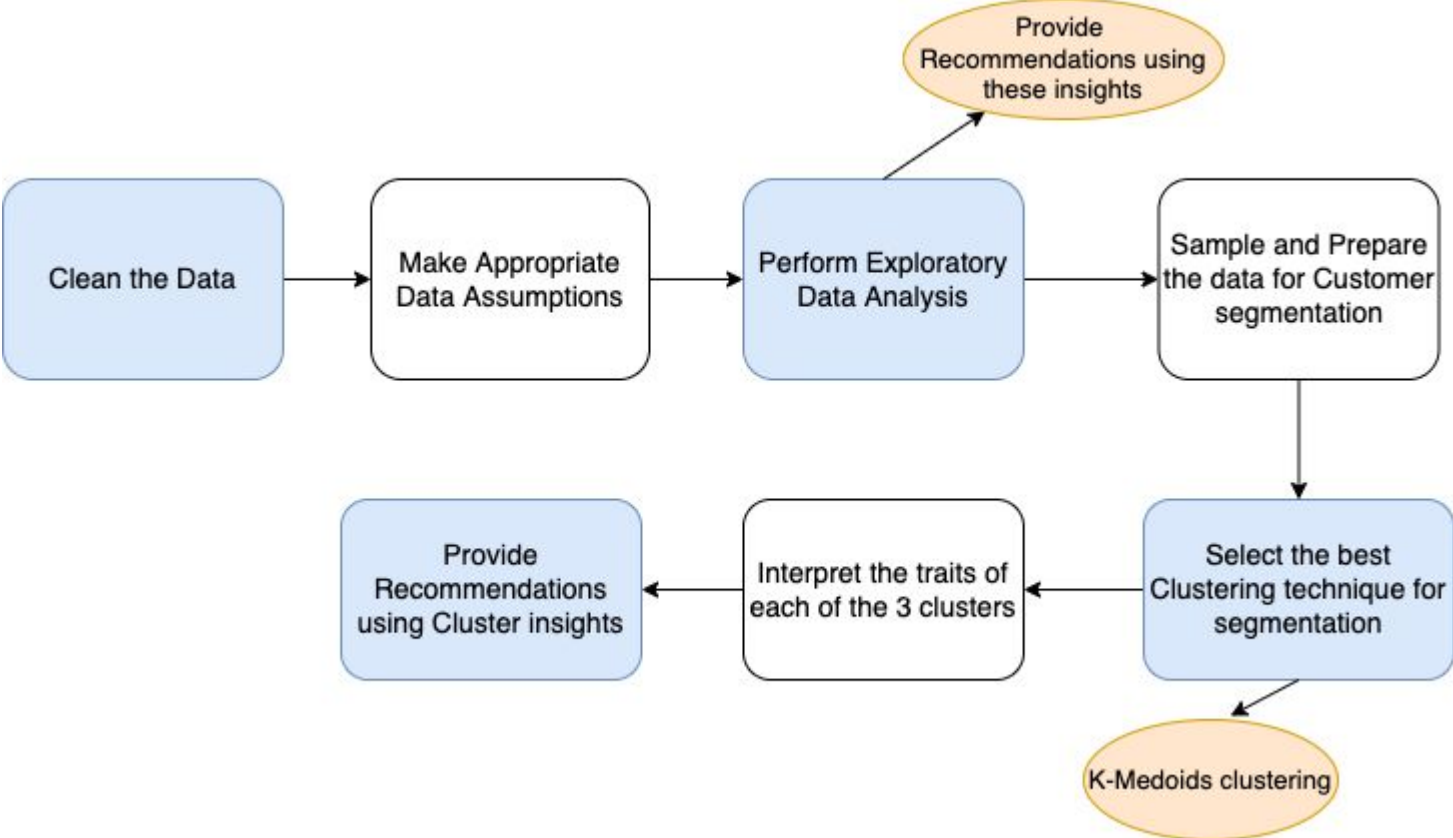
/03

Clustering Recommendations

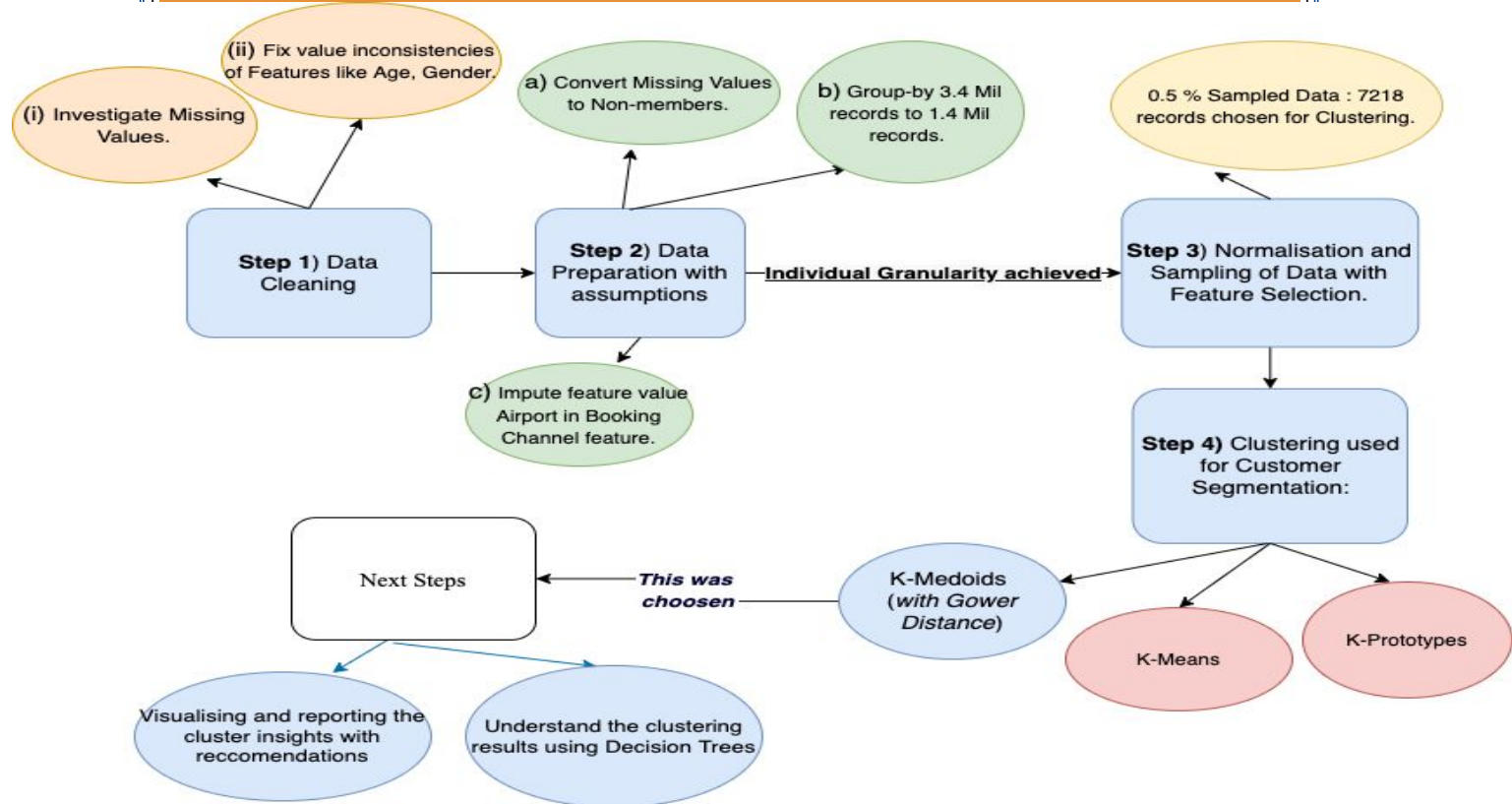
Provide Recommendations based on Clusters

/04

Solution Map



Analysis Plan



Data Assumptions

- Before clustering, some columns are deleted if they do not fit individual granularity, we finally keep the data columns of Age, TicketNum, BaseFareAmt, GenderCode, BkdClassOfService, BookingChannel and UflyMemberStatus
- Customers with missing age are assigned mean age of total customers.
- Customers with missing GenderCode are assigned to 'female' (most common gender).
- People with missing UflyMemberStatus are not member.

	Total Null	Percent Null
PostalCode	2744995	0.799035
EnrollDate	2740908	0.797845
CardHolder	2740908	0.797845
UflyMemberStatus	2740908	0.797845
UFLYRewardsNumber	2740908	0.797845
BookedProduct	2209904	0.643276
StopoverCode	1720454	0.500803
GenderCode	43999	0.012808
birthdateid	43999	0.012808
Age	43999	0.012808

Data Cleaning

- People with age over 100 or below 0 are deleted.
- We are only interested in Sun Country flights, so we remove any Marketing Airline Codes that are not 'SY'. This is more than 99% of all flights in the data.
- Replace different airport booking channel codes with 'airport' value
- Create unique customer identifier by combining EncryptedName and birthdateid columns
- Transform dataset into individual granularity

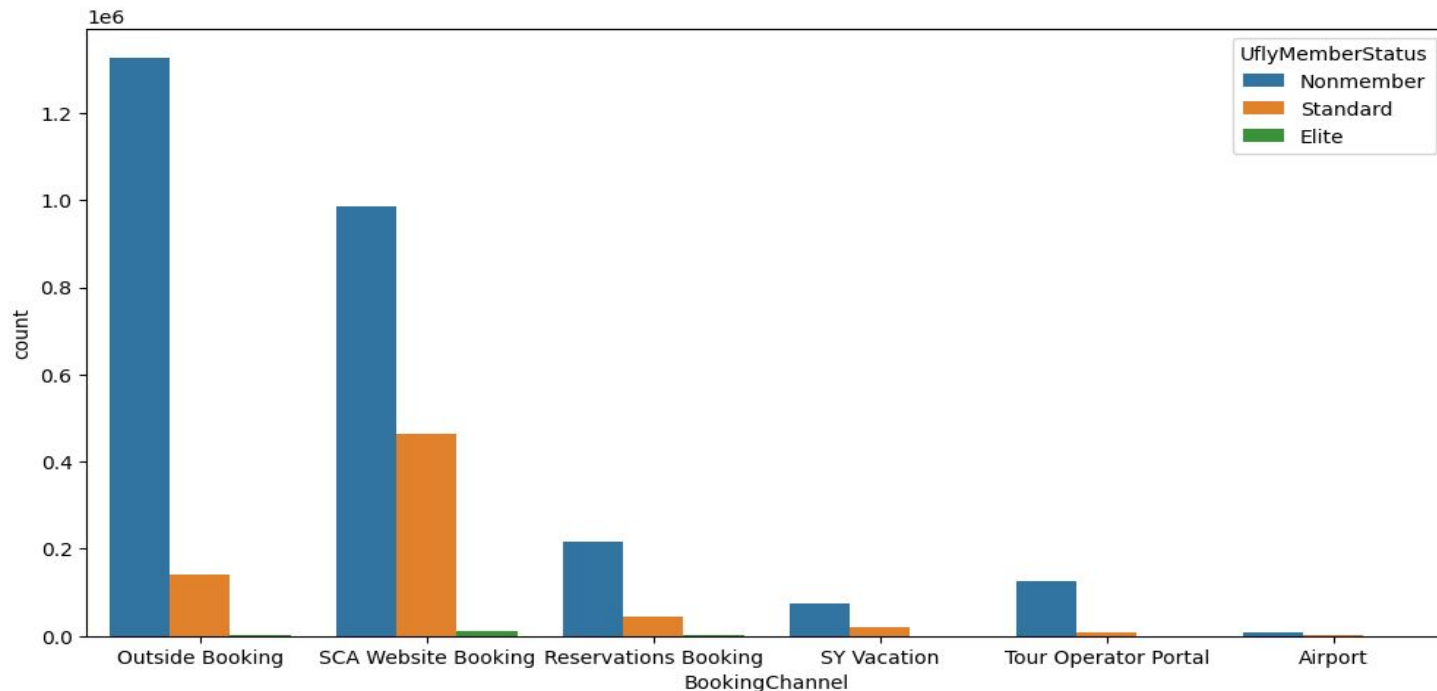
[illegible]

Data Cleaning

- Normalized Age, TicketNum and BaseFareAmt by combining log transform method and normalization method
- Got Dummy variables from categorical columns: GenderCode, BkdClassOfService, BookingChannel and UflyMemberStatus
- Did random sampling to get 0.5% of data to do analysis (~7200 records)

Age	TicketNum	BaseFareAmt	GenderCode_M	BkdClassOfService_First Class	BookingChannel_Airport	BookingChannel_Outside Booking	BookingChannel_Reservation Bookin
0.162243	-0.047404	2.377485	1.0	0.0	0.0	1.0	0.
-0.942987	1.815508	0.248706	0.0	0.0	0.0	1.0	0.
-0.100907	2.771073	0.030225	0.0	0.0	0.0	1.0	0.
1.583253	-0.047404	-0.531368	1.0	0.0	0.0	0.0	0.
0.214873	-1.456643	-0.142949	0.0	0.0	0.0	1.0	0.

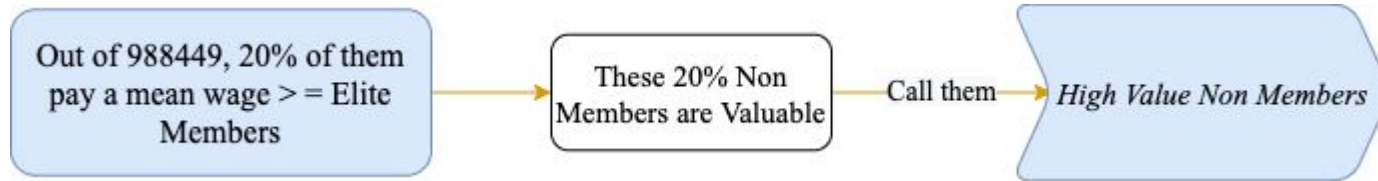
Do Booking channels meet our expectations?



- 67 percent customers, booking via the SCA website are actually non-members. So although these customers are visiting the SCA website, they choose to remain as non-members.
- Let's understand who these customers are, to give appropriate recommendations.

Target Non-members booking via the website

- 35 % (or) 988449 of the total non-members are making their bookings via the website.



Possible Recommendation:

- Sun Country airlines could offer lucrative one - off deals to this group of *High Value Non Members* when they visit the SCA website next time.
- These members have an average age of 41 and frequently fly between MSP, CUN and SFO airports during the months of March, July and December, with 14%(25,174) of them flying in First Class.

Assumption of the Recommendation:

- They will either convert into members, therefore driving enrollment or they will bring more revenue to the airlines.

Comparison of High Value Non Members vs Elite Members:

S.No	Trait	High Value Non Members(Potential targets)	Elite Members(Actual Members, highest status)
1)	Number of Tickets	181105	14361
2)	Average Base Fare	563	409
3)	Average Age	41	49
4)	Frequently Flown Airports	MSP(47%) , CUN, SFO	MSP(47%), JFK, SFO
5)	Frequently Booking Months (Descending order)	May, June, July	Oct, Sep, August
6)	Frequent Travelling Months (Descending order)	March(18%) , July(13%), December	Oct, Nov, December
7)	How many fly First Class?	14% or 25,174	26 % or 3742

Intuitive Customer Insight Based Questions:

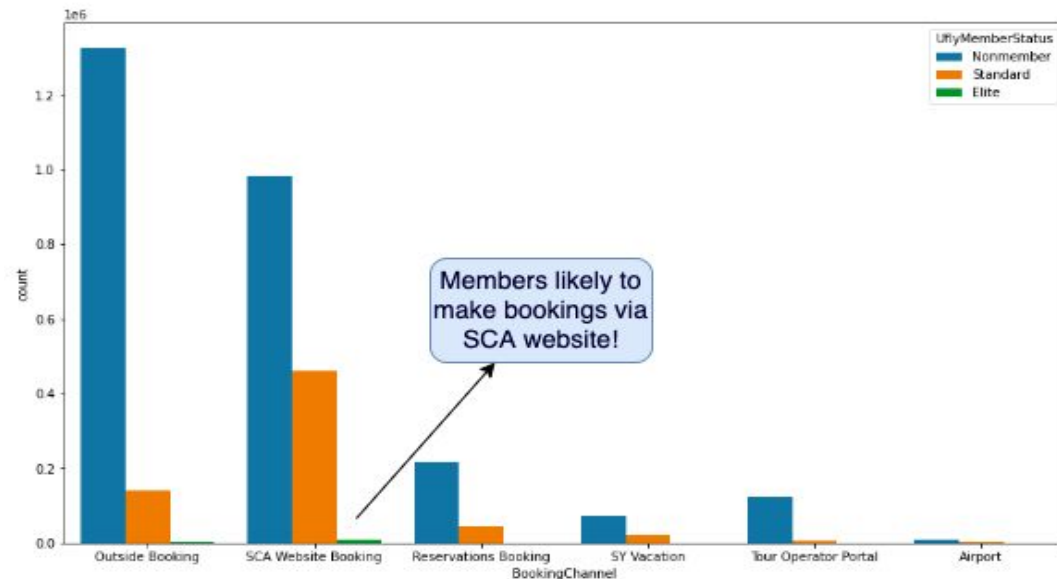
Q1) Is it actually viable to remove the observations having ages greater than 100 or 120?

Ans 1) We found that there were 1604 flight records corresponding to ages between (100,120). However, the number of unique customers aged between 100 and 120 was only 841.

Were they actual passengers? Interestingly there were 80,000 centenarians(100+ people) living in 2013 in the USA so it is plausible that 841 of them traveled using SCA. Removing these customers does not seem like the best step to take!

Q2) Are members most likely to make bookings from the website itself?

Ans2) Yes, thankfully the members are most likely to make bookings from the SCA website rather than the outside booking!



Intuitive Customer Insight Based Questions:

Q3) Which of the Booked Products(Discount code feature) are associated with more discounts?

Ans3) The most successful discount codes were FUNJET, ORB, TVLCTY, DSNY. The N code is just a placeholder for null values.

Discount Codes most often used:		Discount Codes that were actually successful	
N	2204898	FUNJET	23997
SSWMIR	390845	ORB	23257
BSGTIX	137082	N	22809
GRP	129515	TVLCTY	21224
EXP	99621	DSNY	11191
UP	78106	APPLE	7927
SCV	71888	STUNIV	5872
SSWM	32137	CHEOPQ	4250
FUNJET	25281	PCLNOPQ	4181
ORB	23536	EXP	3661

Q4) Any discount codes for Veterans or any intuitive information for this category of fliers?

Ans4) NAVY Booked products code was the only interesting veteran discount code found. 182 records were found with this coupon code. *Interesting fact:* These codes symmetrically belongs to only two flights: [MSP => PNS] and [PNS=>MSP]. The next interesting fact is that PNS is the gulf coast Pensacola airport and there is a Naval Base in Pensacola!

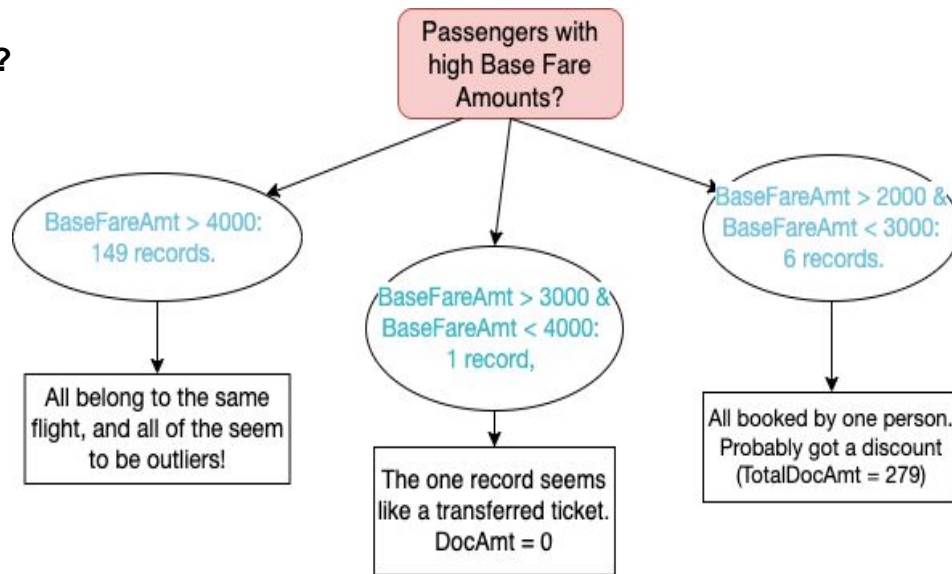
Probable recommendation:

Can give NAVY Code to that particular journey Itinerary, as promoting this to such a location like PNS() which has a coast can help drive the sales for this particular flight / or these kinds of flights in the future and also provide discounts to the much deserved veterans of the country!

Intuitive Customer Insight Based Questions:

Q5) Any passengers having unusually high base fare amount?

- The flow chart succinctly summarizes the information. The intuitive fact is there is only one unique value of Base fare above 4000, i.e. 4342\$. But the total amount they paid was only 440\$.
- However, there are 149 such records with the same fare. This could only mean that the flight no. 530 was an alternative flight that these 149 passengers were provided as they used some prior account credit to reduce the fare. **The passengers from flight, paying this fare, could be considered as outliers, skewing the data!**
- Are there any other passengers on the flight paying lesser? Yes, there are 15 people having a lesser base fare amount than 4342. These are normal records, not outliers.



Q6) Any trends in flight booking patterns of different customer age groups?

Ans 6) Yes, when we dug through the senior citizens age group, we found that there were 19231 unique travelers above the Age 80. There were more even valued bookings than odd, and after analysis, we concluded that older travelers (above the age 80) prefer booking round trips more often than not.

This information can be used to understand that segment of the customers, when adding additional Sun Country hospitality services across locations.

Exploratory Data Analysis

Derive peculiar insights from data

/01

EDA Recommendations

Provide Recommendations w.r.t. the entire data based on EDA

/02

Clustering Results

Observe Results from Clustering on Individual Granularity data

/03

Clustering Recommendations

Provide Recommendations based on Clusters

/04

Data Preparation for Clustering

Aggregate

- Individual Granularity

Filtering

- Age : 0 – 100
- Airline : Sun Country

Replace Data Value

- Various airports -> 'Airport'

Impute Missing Values

- Age : Mean
- Gender: Female
- Membership : Non-member

Clustering Feature Selection

Age

Ticket
Number

Base Fare

Gender

Class of
Booking

Channel of
Booking

Membership
Status

Clustering Results

Cluster 1: Retention Customers

- Ufly Rewards Members (Standard, Elite)
- Book on SCA website

Cluster 2: Prospective Members

- Non-members
- Book on SCA website

Cluster 3: “Tough Nut” Customers

- Non-members
- Booking on outside channels
- Converting these users maybe a more difficult task

Clustering Results cont.

★ Retention Customer Cluster

- **Important Features:** SCA booking, ufly rewards members
- Group of customers who are members (both standard and elite) and book on SCA. Our favorite customers!
- ~18% of our sample

★ Prospective Members Cluster

- **Important Features:** SCA booking, non-members
- Group of customers booking on SCA who are not yet members. Can we convert these customers to the rewards program while they are booking on the website?
- ~40% of our sample

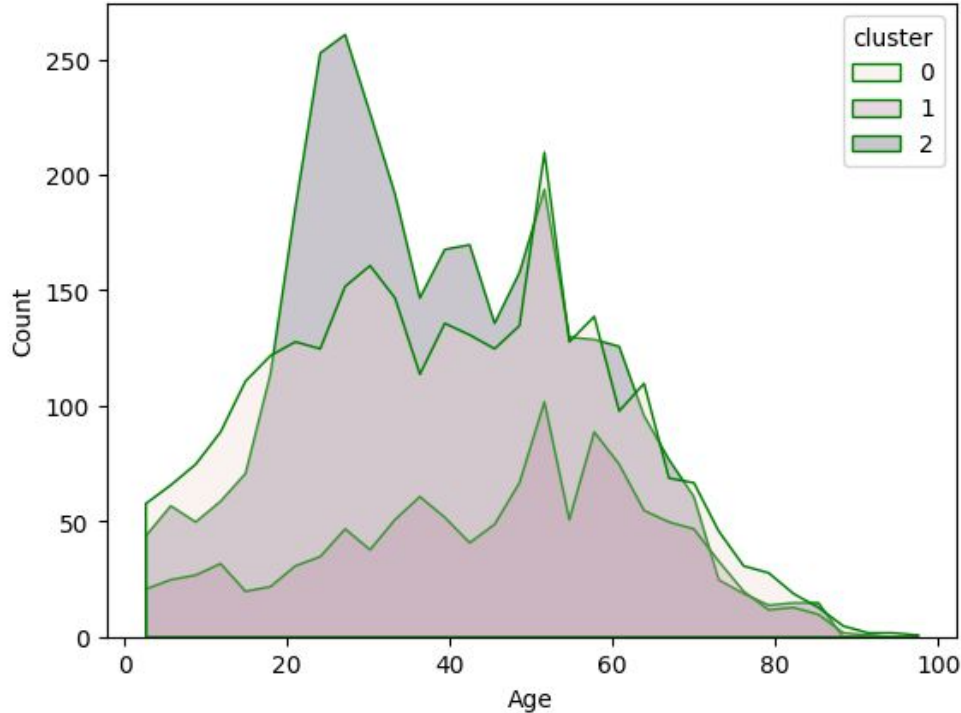
★ “Tough Nut” Cluster

- **Important Features:** outside booking, non-members
- Group of customers who utilize outside booking and are not members. We have little opportunity to convert these customers.
- ~42% of our sample

Clustering Report: Results of K-medoids with 3 clusters & 7218 records

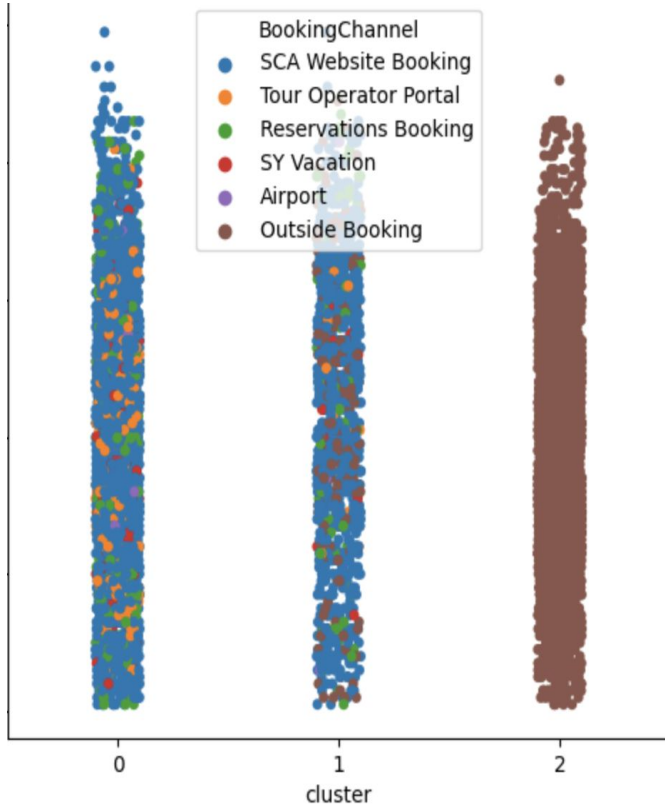
	Prospective members Cluster	Retention Members Cluster	Tough Nut customer Cluster
Who are they?	Non members, booking tickets via SCA website.	Higher spending(mean fare of 290), older customers with SCA membership.	Younger(mean age 36), harder to convert Non Members, booking tickets via outside bookings.
Booking Channel trends	Primarily book through SCA website, tour operator portal	Mixed booking channels, no real trends	Everyone(3195) used outside booking as the chanel
Age trends	Avg age = 39; Median age = 39	Avg age = 45; Median age = 48	Avg age = 38; Median age = 36
Booked Class trends	First class: 59	First class: 74	First class: 17
Ufly Member status trends	Median fare: 240, but 17 yr old F outlier is present.	Median Fare: 265(highest)	Median Fare: 258
No. of Tickets trends	Similar No. of Tickets distribution as the last cluster. (Mostly 1-5 ticket numbers)	Mostly 1-5 but although, few instances of 10-30, 40-50. (other clusters do not have this)	Similar No. of Tickets distribution as the first cluster. (Mostly 1-5 Ticket Numbers)

Comparing cluster wise Age distribution :



- This plot confirms that the Cluster(0), i.e. the **Tough Nut Cluster** consists of younger(mean age 36) customers than their older counterparts.
- The mean age also looks higher for Cluster(1), i.e. **Retention Members** cluster(with an average age 45) , confirming our previous detailed report on cluster characteristics.

Comparing cluster wise booking channel distribution



- This strip plot confirms that the tough nut cluster(2) contains all the bookings as outside bookings
- Also, the prospective cluster(0) primarily has bookings from SCA website.

Possible Recommendation:

- If SCA plans to offer lucrative deals to convert non-members to members, SCA possibly should not prioritize customers belonging to the tough nut cluster, i.e. customers having a mean age of 36, mostly flying coach, buying between 1-5 tickets.
- This is because everyone in that cluster, 3195 of them have booked via outside bookings, and it will be harder to convince these customers to not only visit the website but also become members.

Business Steps/Actions



Increasing visibility of Ufly rewards on the website.

- Redesign the website to feature the program, ensuring users on the site are cognizant of its existence.

Offering discounts during checkout.

- Before users finish their purchase and leave the website, incentivize signing up for ufly rewards by offering a % discount on tickets if users sign up for a membership.

Both these steps could be launched as an **experiment using A/B tests to monitor the efficacy of the recommendation.**

Based on our clustering results, ~40% of customers are booking on SCA website but are not ufly rewards members.

How can we drive enrollment of ufly rewards while users are on the site?

Further Steps



Utilize Entire Data

Use heavy computational resources to extrapolate result to full data set.



Machine Learning Algorithm

Develop an algorithm to predict which customers are most similar to our “prospective members” cluster and target those customers using target marketing.

Conclusion

EDA ANALYSIS

- SCA Website is the favorite booking channel for members
- Target Non-members booking via SCA website

Clustering

- Clustering result
 - Prospective members Cluster
 - Retention Members Cluster
 - Tough Nut customer Cluster

Clustering recommendation

- Drive enrollment of Ufly rewards while users are on the site