Airline Customer Segmentation using LRFMC Model

Data Science Portfolio by Achmad Luckyta Fasyni

Outline

01

OBJECTIVE

02

DATA OVERVIEW

03

EXPLORATORY DATA ANALYSIS

04

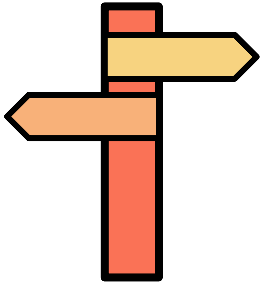
CUSTOMER CLUSTERING

05

RECOMMENDATION

Objective

1. Group customer segment using LRFMC customer lifetime value model
2. Produce actionable insights for proper marketing strategy to each customer segment
3. Learn end-to-end data science project



Customer Segmentation using LRFMC Model

The process where customers of an enterprise are divided into groups based on their purchasing behavior and characteristics.

- **Demographic**
- **Psychographic**
- **Geographic**
- **Behavioral Attributes**

  Past purchasing pattern such as latest purchase and purchase frequency

LRFMC Customer Lifetime Value Model helps the behavioral customer segmentation, utilizing their purchasing records in the airlines transaction. LRFMC Model scores customers according to five attributes:

- **Loyalty**
- **Recency**
- **Frequency**
- **Monetary**
- **Cabin**

This project implement Machine Learning Clustering Model using the K-Means++ Algorithm, where customer records are segmented based on their respective LRFMC values.

Data Overview

Data Overview

Dataset: https://github.com/LuckyFasyni/Airline-Customer-Segmentation-LRFMC-Model/blob/main/1659264665216-flight.csv
Contains demographic and behavior attributes of airline customers.

data cleaning

62988 rows → **61856** rows, 98.2% from original dataset

23 features →

4 Datetime
- ffp_date
- first_flight_date
- load_time
- last_flight_date

5 Categorical
- gender
- ffp_tier
- work_city
- work_province
- work_country

14 Numerical
- member_no
- age
- flight_count
- bp_sum
- sum_yr_1
- sum_yr_2
- seg_km_sum
- last_to_end
- avg_interval
- max_interval
- exchange_count
- avg_discount
- points_sum
- point_notflight

General Workflow

1. Data Understanding
2. Data Cleaning
3. Exploratory Data Analysis
   - Statistical Summary
   - Univariate Analysis
   - Bivariate Analysis
   - Answer Business Questions
4. Data Preprocessing
   - Feature Selection
   - Feature Engineering
   - Outlier Handling
   - Feature Scoring
5. Data Modelling
   - Hyperparameter Tuning
   - K-Means Clustering Modelling
   - Clustering Analysis
6. Business Recommendation

Data Cleaning

lower columns name

convert few columns' data types

standardize datetime columns' format

missing value handling: imputation

drop remaining missing values

check duplicate values

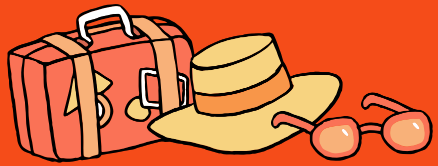
no duplicate values!

impute with 'unknown'

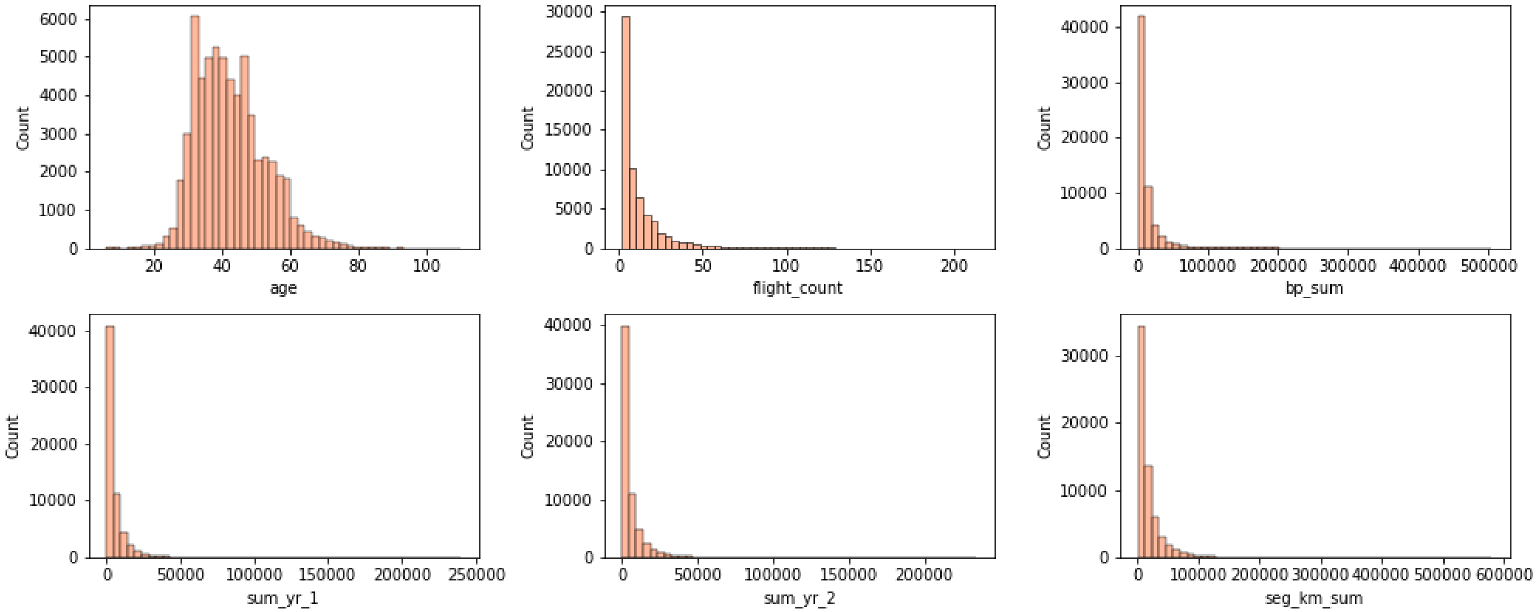
dropna

member_no	0
ffp_date	0
first_flight_date	0
gender	3
ffp_tier	0
work_city	2269
work_province	3248
work_country	26
age	420
load_time	0
flight_count	0
bp_sum	0
sum_yr_1	551
sum_yr_2	138
seg_km_sum	0
last_flight_date	0
last_to_end	0
avg_interval	0
max_interval	0
exchange_count	0
avg_discount	0
points_sum	0
point_notflight	0

Exploratory Data Analysis

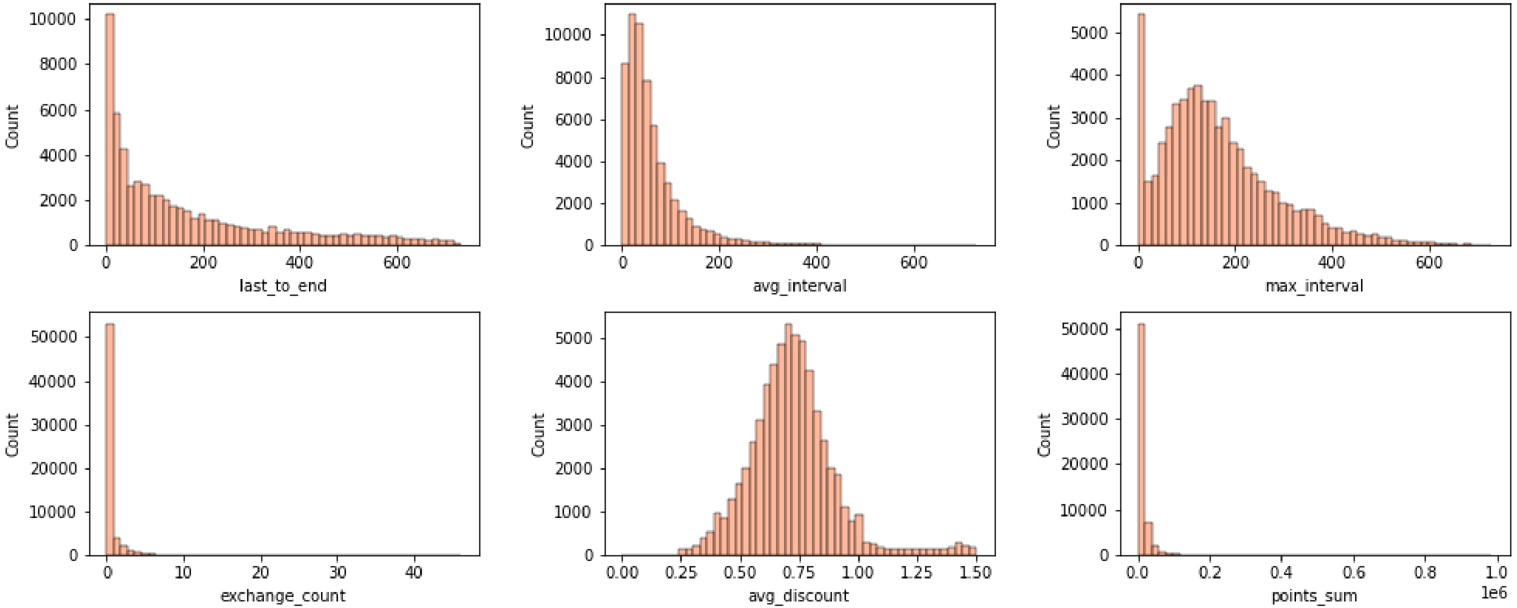


Distribution Plot (1)



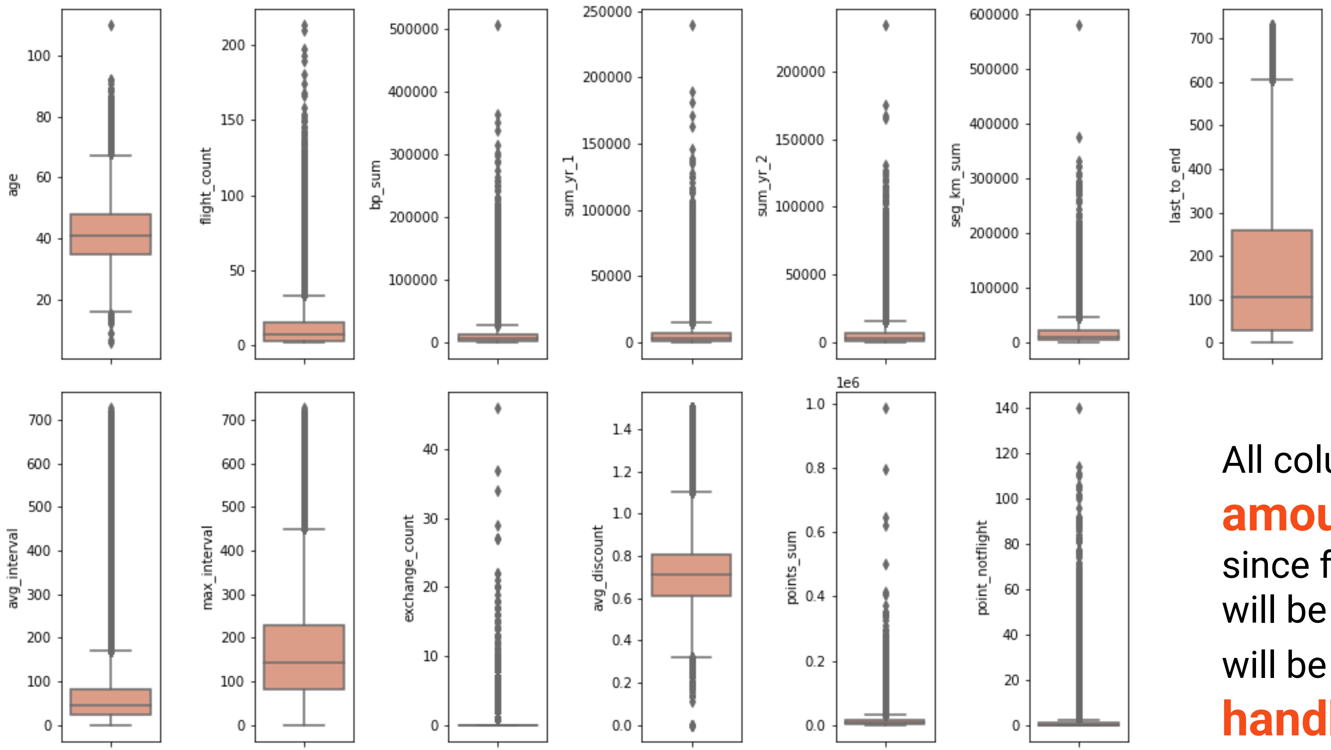
Almost all columns are extremely right-skewed, while age features has slight normal distribution, peak at 30-45 years old.

Distribution Plot (2)



Almost all columns are right-skewed, except avg_discount feature.

Box Plot

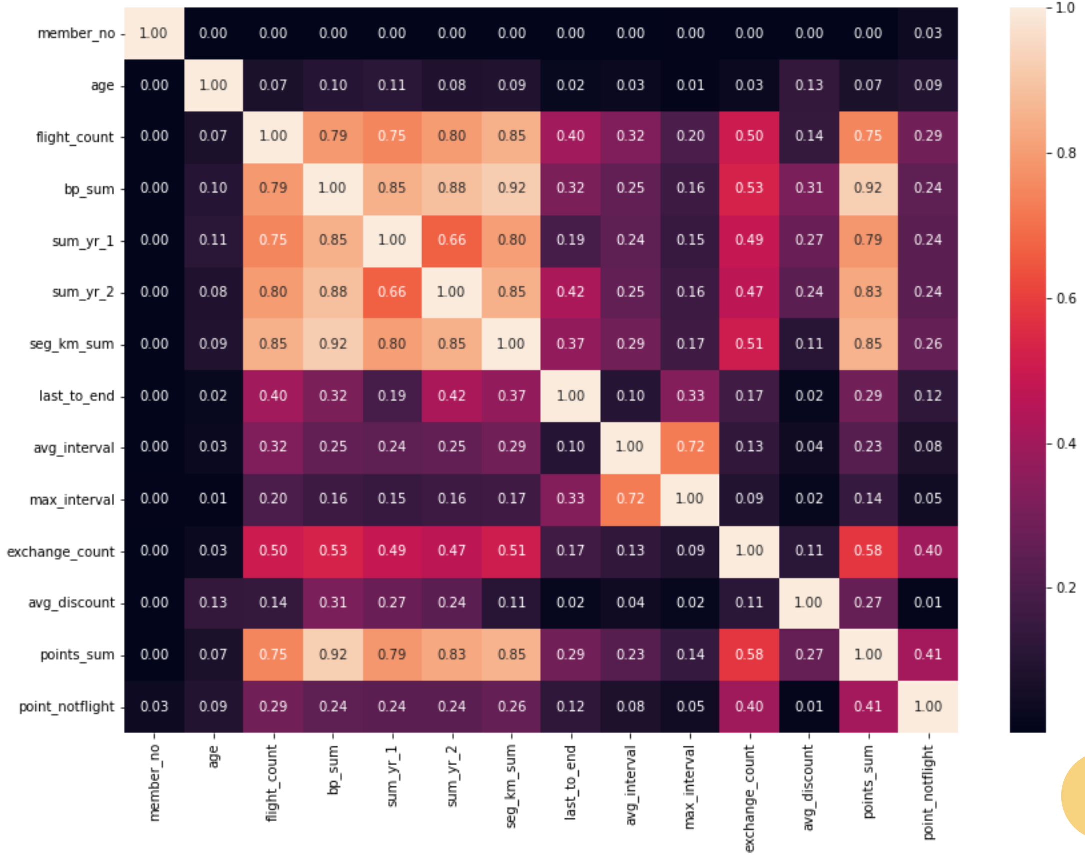


All columns have **fair amount of outliers**, but since feature engineering will be conducted later, there will be **no outliers handling** for now.

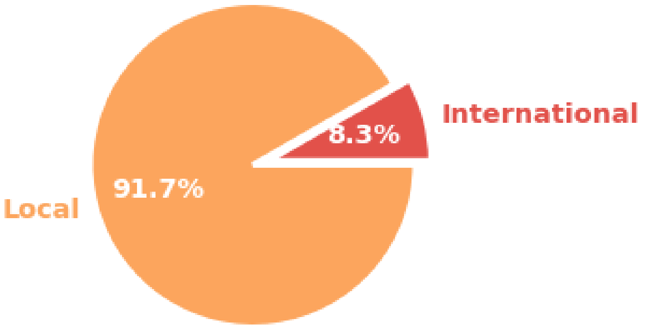
Bivariate Analysis using Correlation Plot

Column **“flight_count”, “bp_sum”, “sum_yr_1”, “sum_yr_2”, “seg_km_sum”, “points_sum” are highly correlated** each other, which is make sense.

→ The more frequent the customer purchases or flies, the further the flight distance covers, the more the fare revenues, and the more points the customers get.



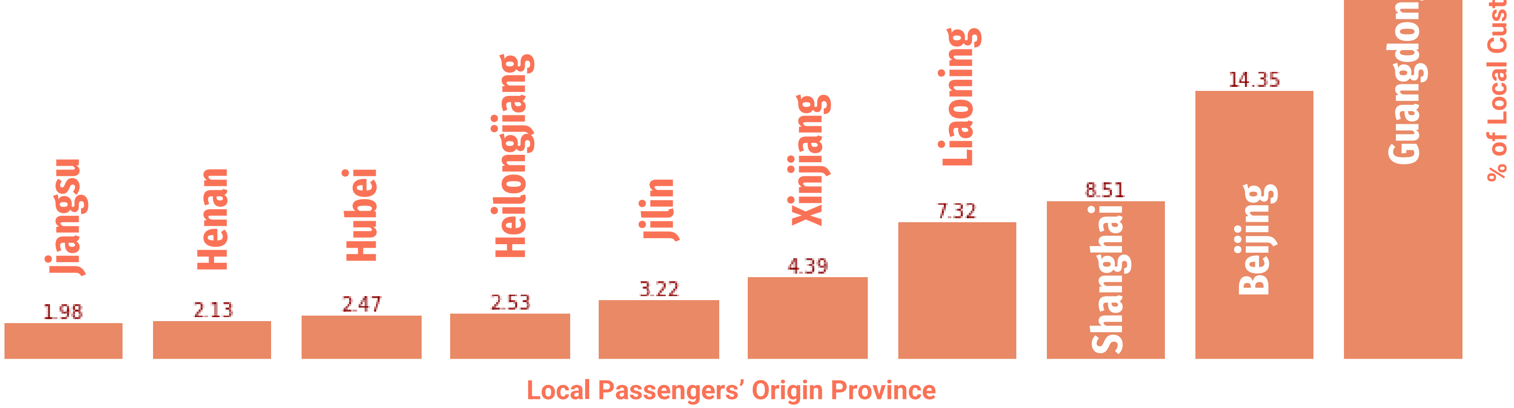
How is the Proportion of Local Passengers?



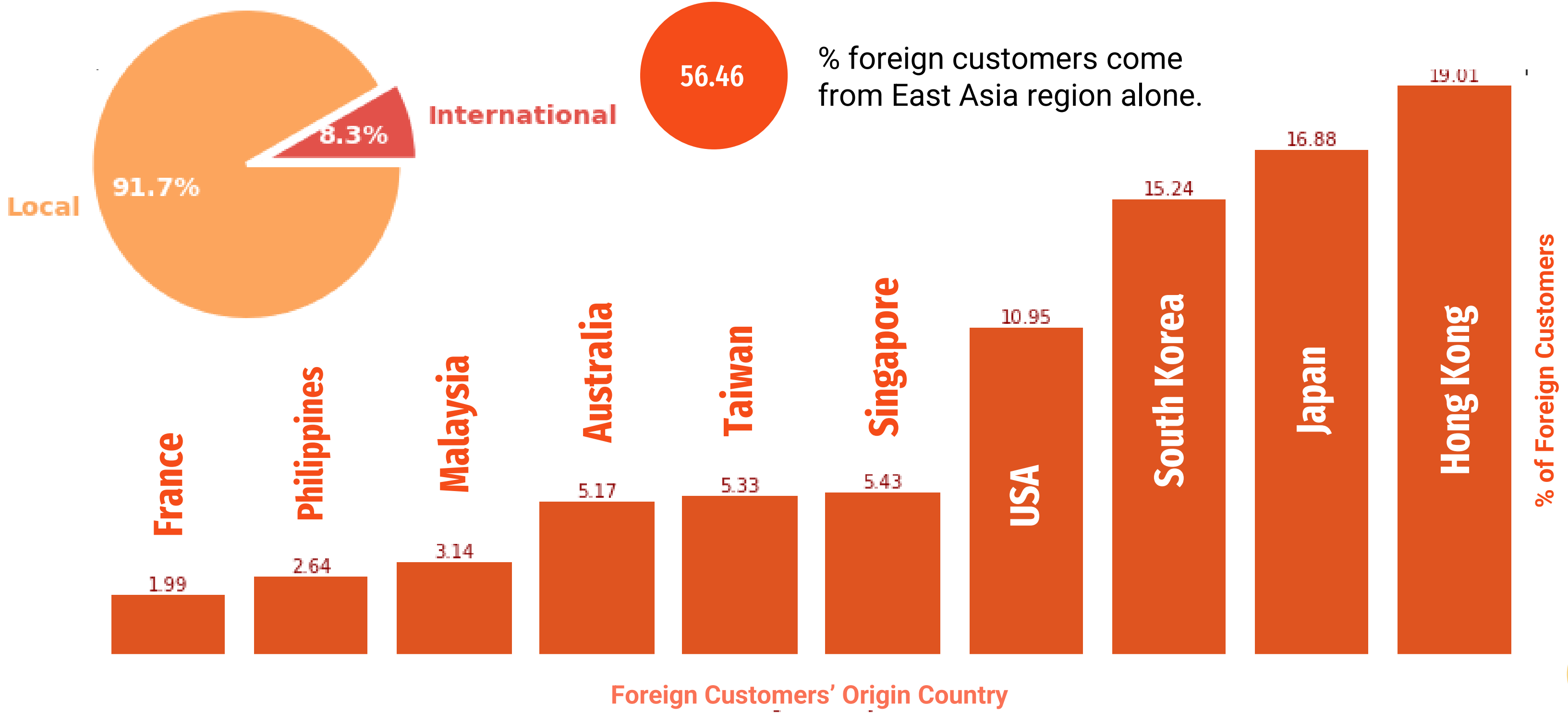
Local customers take up to 91.7% from total passengers.

49.7

% total overall customers come from Guangdong, Beijing, and Shanghai



International Passengers?

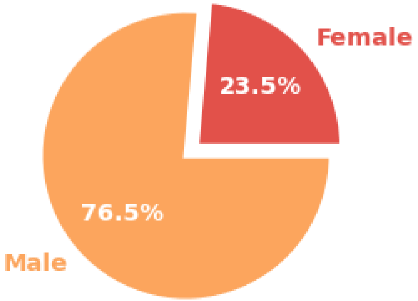


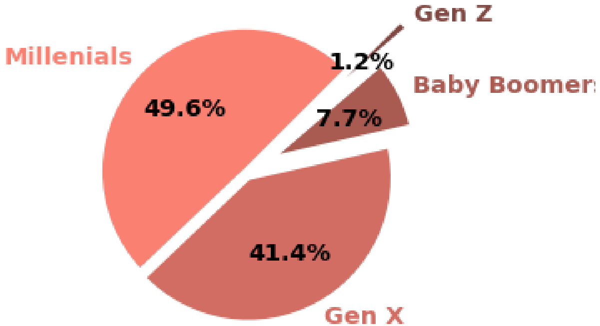
56.46

% foreign customers come from East Asia region alone.

Demographic (Gender and Age)

Most customers are **Man, three times from Woman**

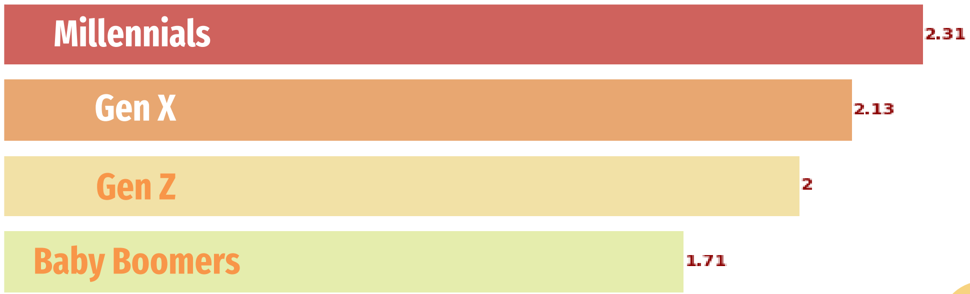




Age Groups:

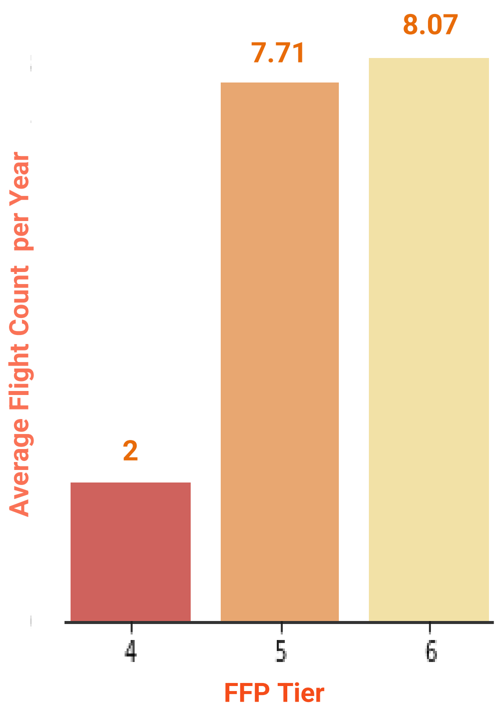
- **Baby Boomers** >=58 years old
- **Generation X** 42-57 years old
- **Millennials** 26-41 years old
- **Generation Z** <25 years old

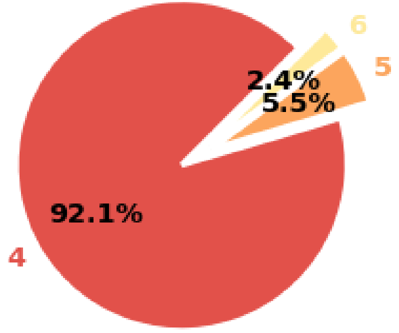
Half of customers come from Millennials, by 49.6%. While the least is Gen Z, which is understandable since they have much less purchase power by age now.
Millennials also travel 2-3 times per year in average. Most than others, although not quite significant.



Average Flight Count per Year

Frequent Flyer Program





Almost all customers still hold Tier 4 of FFP, by 92.1% which should be encouraged more to upgrade their level.

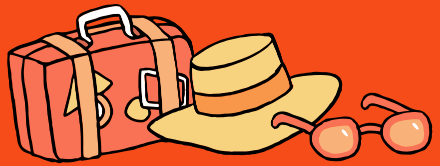
8 Tier 5 and 6 holder travels 7-8 times per year.

5969
Tier 4 passengers has flown 7 times per yr

1625
Tier 5 passengers has flown 8 times per yr

There are **7594 customers** in total that **should be encouraged to level up their FFP!**

Customer Clustering



Feature Selection & Engineering

According to the LRFMC model of airline customer value[1], 6 features related to the LRFMC model indexes:

- **ffp_date:** frequent flyer program join date
- **load_time:** observation date
- **flight_count:** flight count
- **avg_discount:** average discount value
- **seg_km_sum:** total flight distance
- **last_to_end:** last flight time from observation time

Loyalty = load_time - ffp_date
The number of months between the time of membership and the end of observation window (unit: month).

Recency = last_to_end
The number of months from the last time the customer took the company's aircraft to the end of the observation window (unit: month).

Frequency = flight_count
Number of times the customer takes the company's aircraft in the observation window (unit: Times).

Monetary = seg_km_sum
Accumulated flight history of the customer in observation time (unit: km).

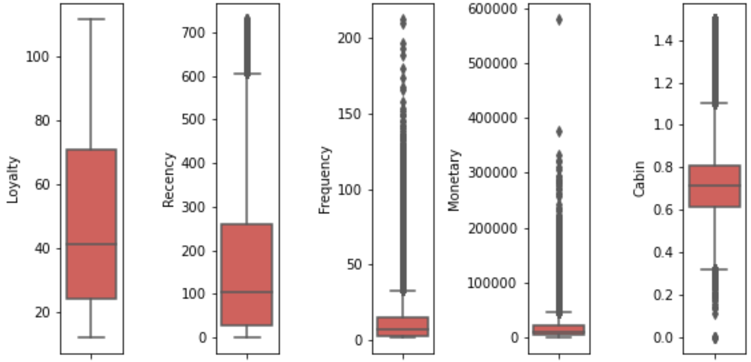
Cabin = avg_discount
Average value of the discount coefficient corresponding to the passenger space during the observation time (unit: none).

[1] Tao, Y.: Analysis Method for Customer Value of Aviation Big Data Based on LFRMC Model. ICPCSEE 2020, CCIS 1257. 89-100(2020).

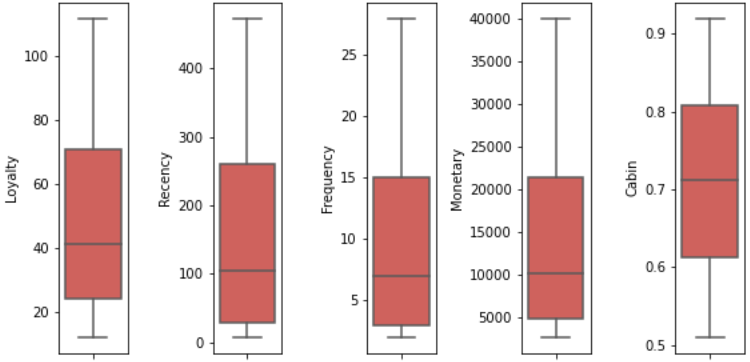
Outliers Handling

K-Means Clustering is sensitive with range, we should handle the outlier wisely to optimize our modelling process and improve the result.

Only Loyalty feature that has no outlier.



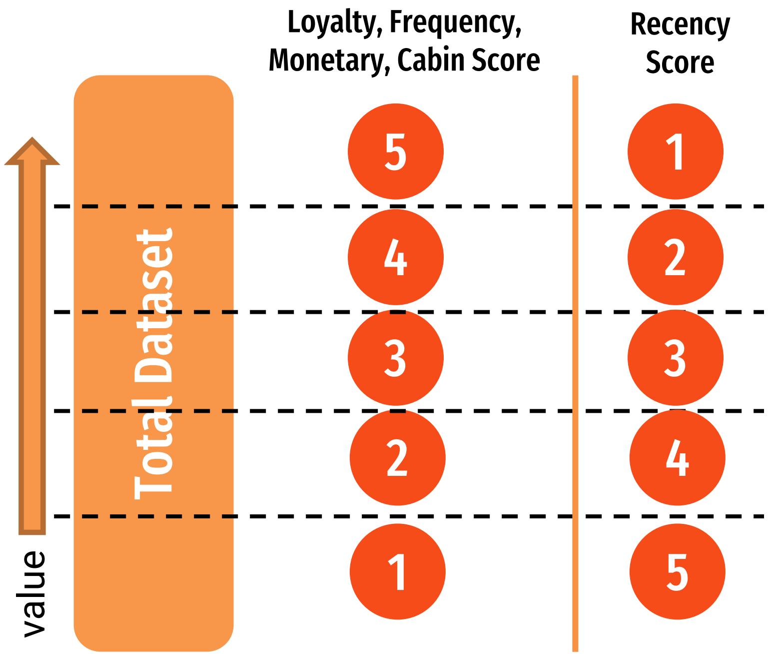
replace outliers with 10th and 90th percentile



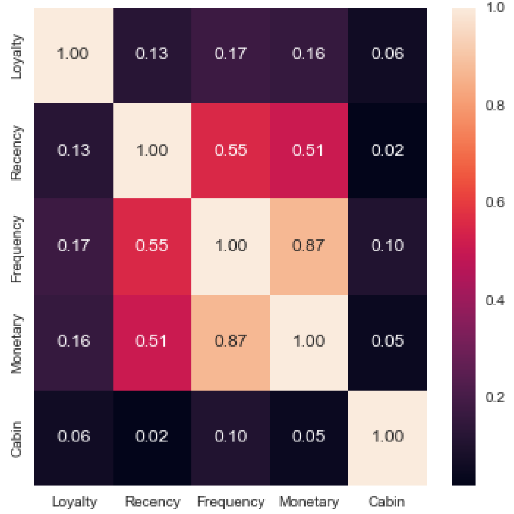
There is no outlier!

Feature Scoring

Scale 1–5 based on dataset quintile.

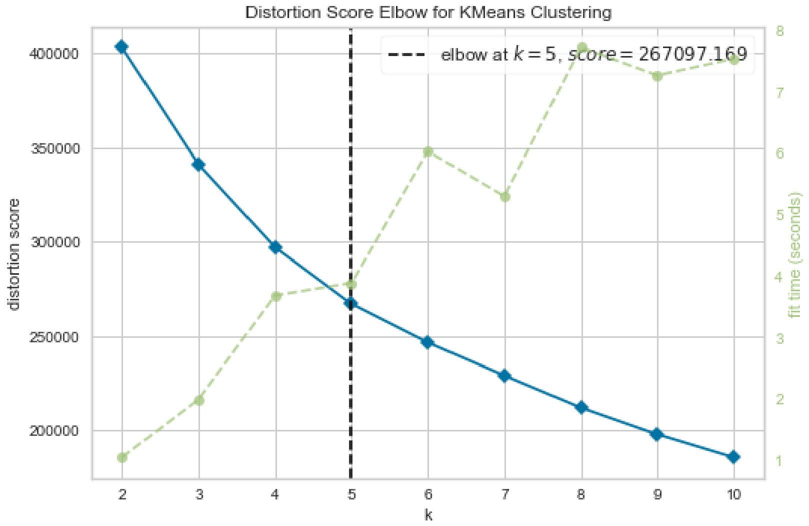


Correlation Plot

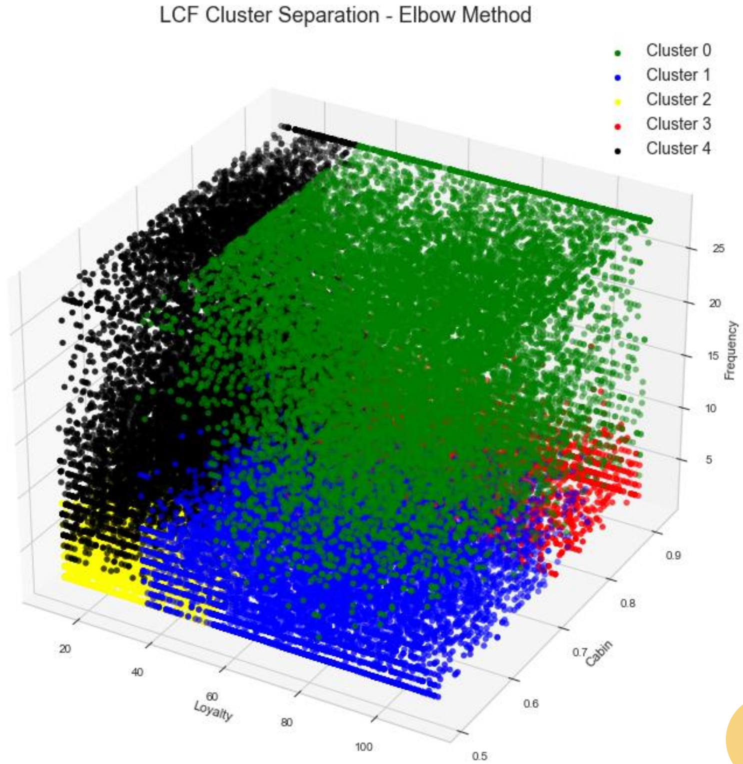


Feature F (Frequency) and M (Monetary) is highly correlated, which is make sense. The more frequent a customer flies, the further flight distance is.

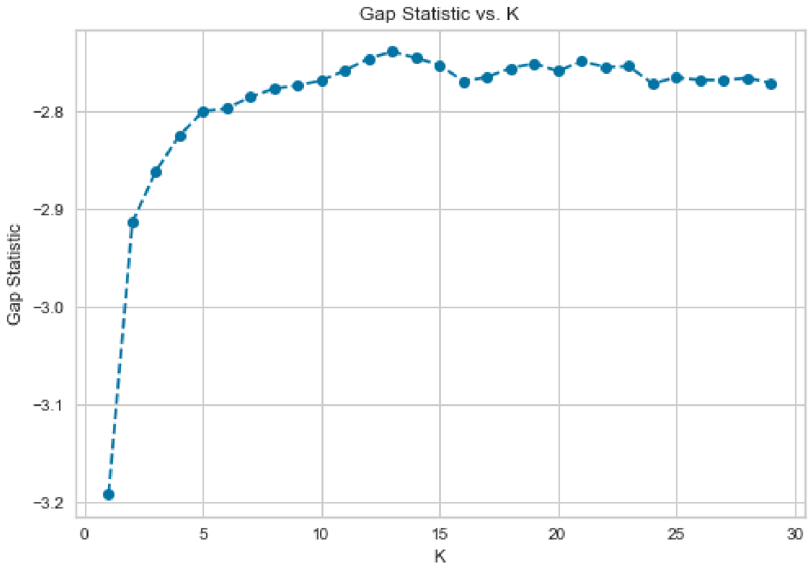
Hyperparameter Tuning (Elbow Method)

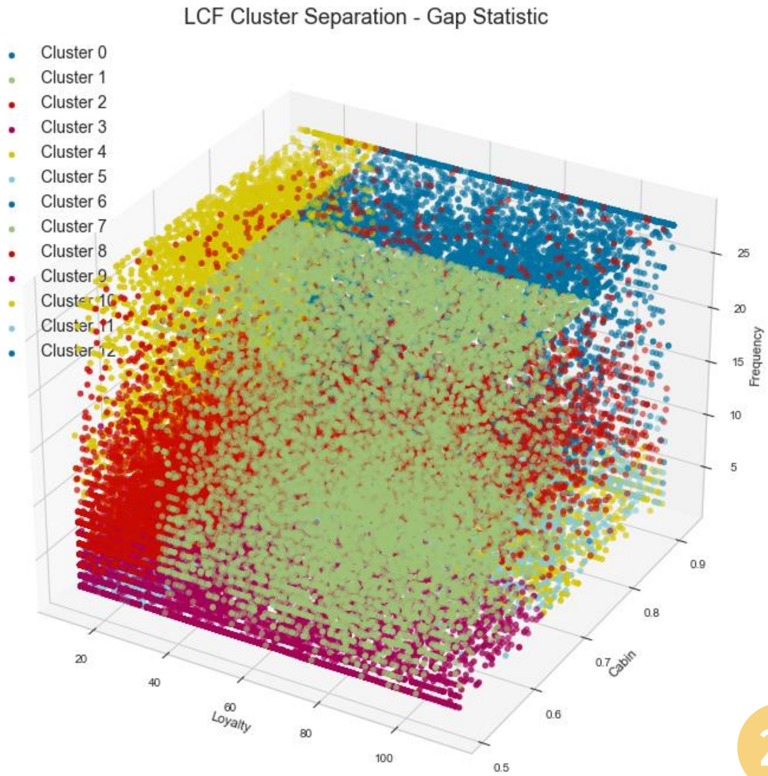


Using K Elbow Visualizer package, the graph shows that
the optimum K number is 5



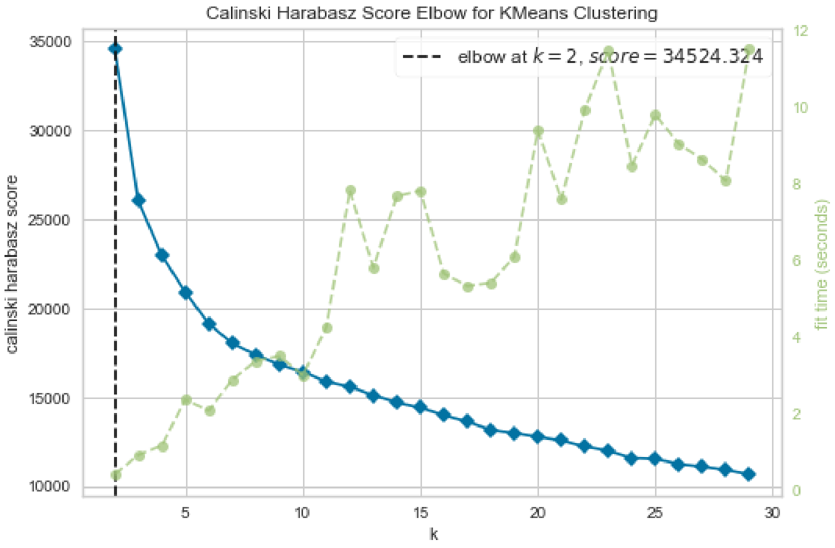
Hyperparameter Tuning (Gap Statistic)



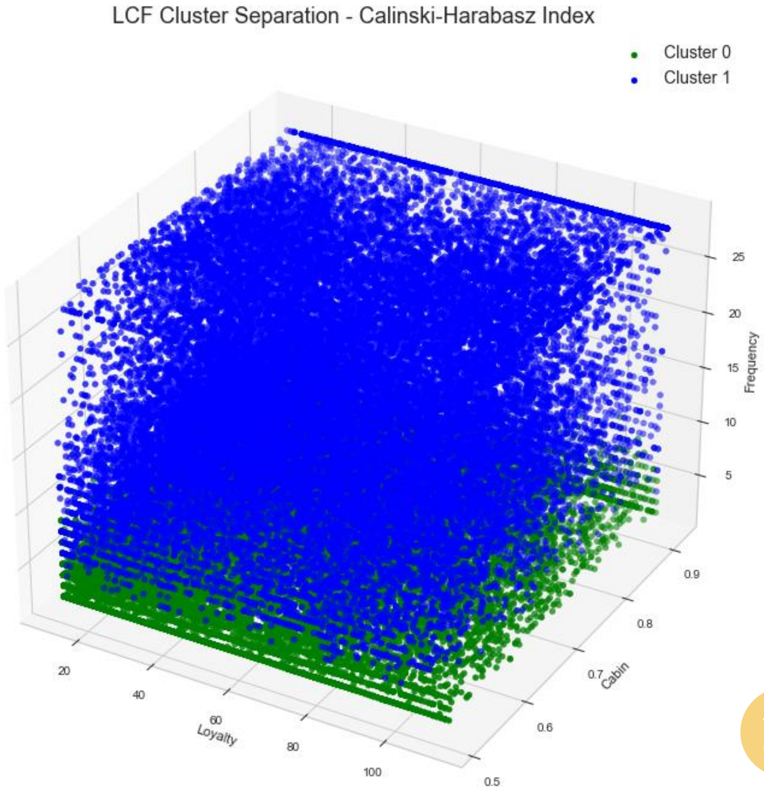


Using Gap Statistic, the graph shows that
the optimum K number is 13

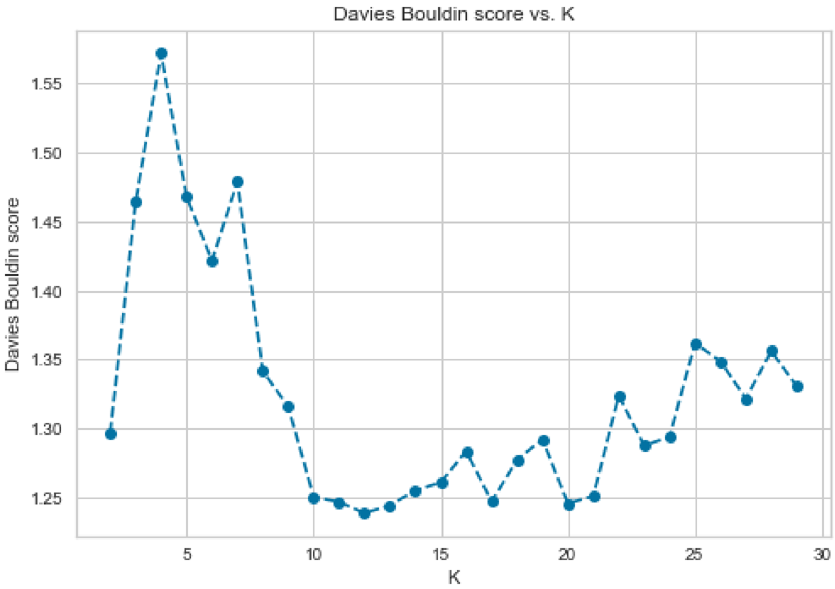
Hyperparameter Tuning (Calinski-Harabasz Index)

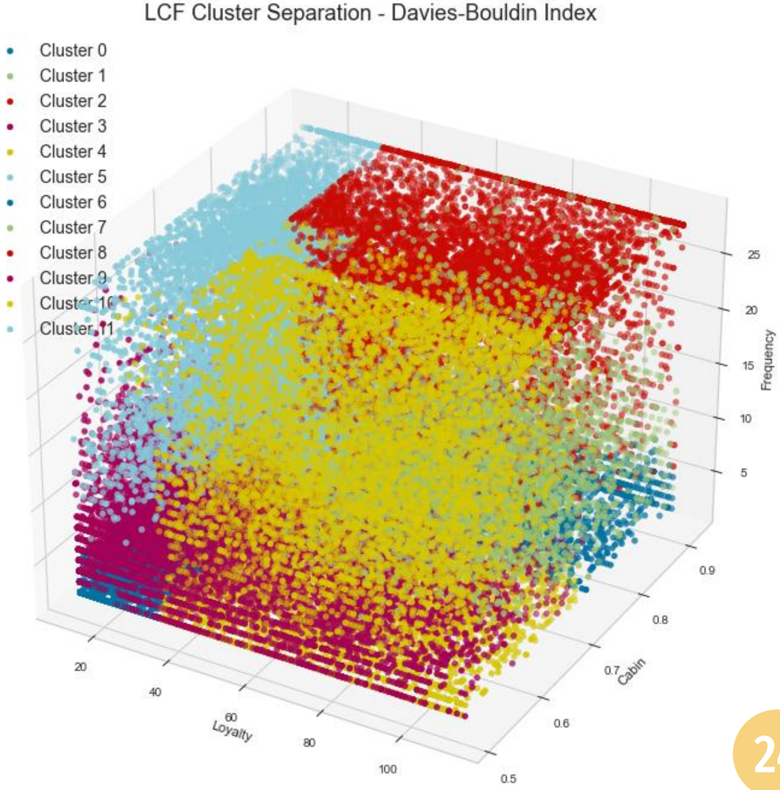


Using K Elbow Visualizer package, the graph shows that
the optimum K number is 2



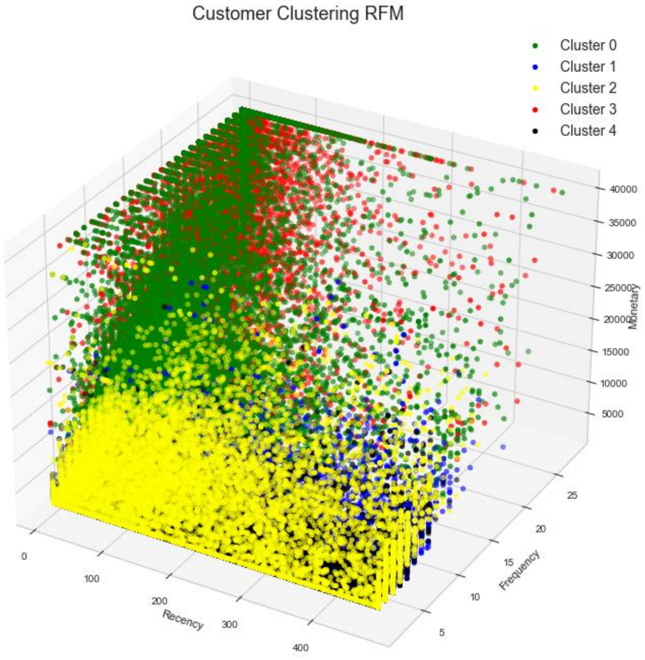
Hyperparameter Tuning (Davies-Bouldin Index)

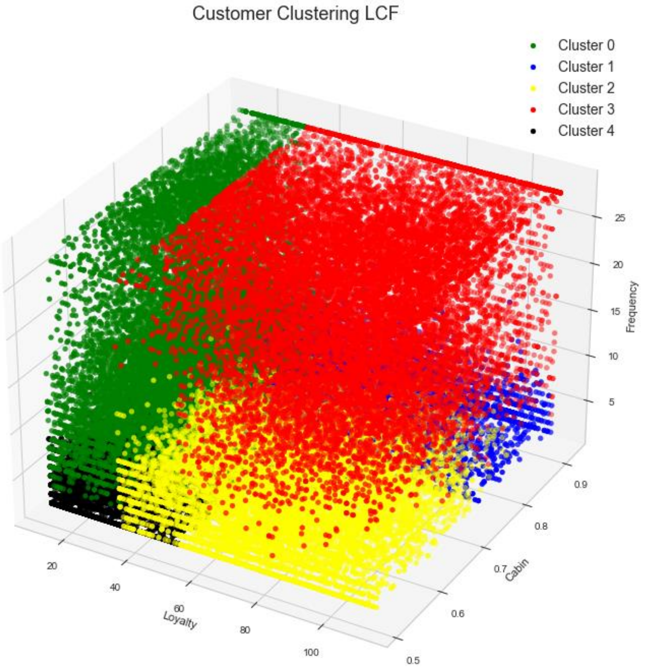




Using Davies-Bouldin Score, the graph shows that
the optimum K number is 12

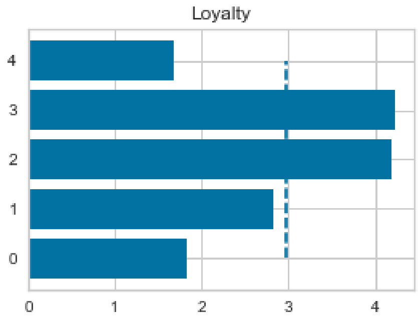
Clustering Result



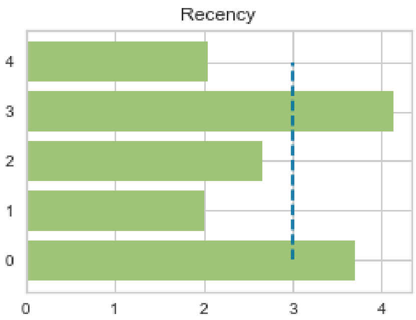


Based on its compactness and separation shown by the clustering visualizations,
the optimum K number is 5

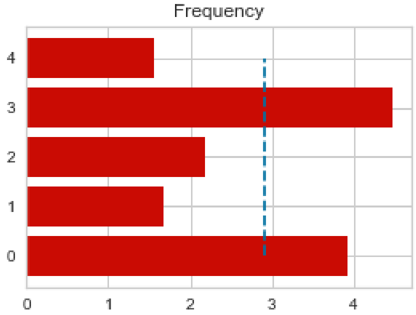
Feature Overview per Cluster



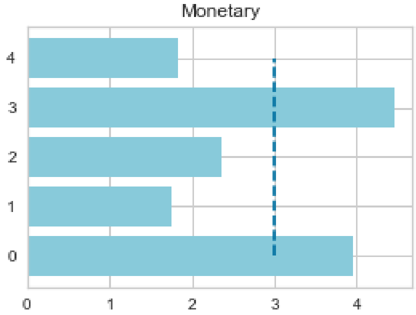
High Cluster 2 and 3 **Average** Cluster 1 **Low** Cluster 0 and 4



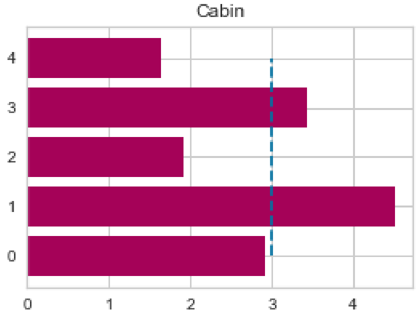
High Cluster 0 and 3 **Average** Cluster 2 **Low** Cluster 1 and 4



High Cluster 0 and 3 **Low** Cluster 1, 2, and 4

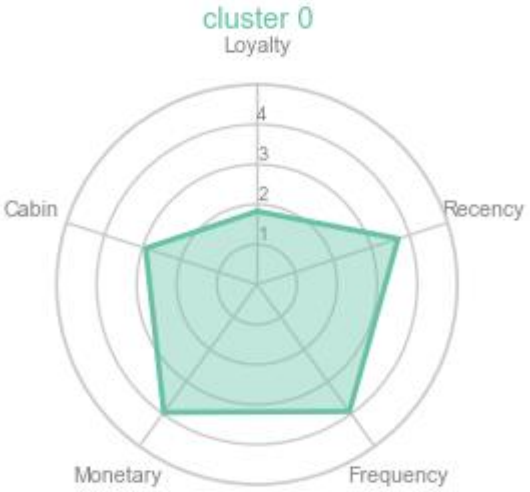


High Cluster 0 and 3 **Low** Cluster 1, 2, and 4



High Cluster 1 and 3 **Average** Cluster 0 **Low** Cluster 2 and 4

Cluster 0 (Potential Loyalist/High Prospect)



22.3 % from total customers

Average Lifetime Value

25 months relation length

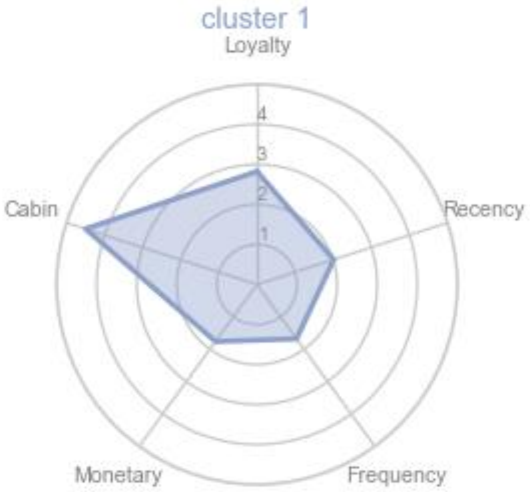
47 months from latest flight

6 flights per year

Behavior Description:
New member with high consumption due to frequent purchase and longer distance. Fresh, near from the latest purchase and often sit in standard class of cabin.

Recommendation:
We should **retain these customers** as long as possible. We can **offer extra discount or reward points** after some period of membership or mileage, even one free ticket. Offer them **affiliate program** that could be redeemed to upgrade for higher class of cabin to give them experience

Cluster 1 (Hibernating VIP)



21.1 % from total customers

Average Lifetime Value

39 months relation length

261 months from latest flight

1 flight per year

Behavior Description:
Have been with us in a medium term but low consumption due to infrequent purchase and shorter distance. Haven't purchased our service lately, but often book for higher class of cabin.

Recommendation:
We should **attract them to re-purchase**. Assuming this type of customer comes from higher class, bring them back with **flight promotions that is bundled with destination event vouchers or signatured souvenirs**. **Run surveys** to find out what went wrong and avoid losing them to a competitor.

Cluster 2 (Low Consumer)

16.8 % from total customers

Average Lifetime Value

71 months relation length

145 months from latest flight

1 flight per year

Behavior Description:
Long term customer but low consumption, occasional traveler and shorter distance (less purchase). Not so far from the latest purchase and often book for low class of cabin.

Recommendation:
We should **attract them to re-purchase** by providing special offers for regular customers such as **discount for first two flights in a year** and **free voucher for affiliated product/event thereafter**.

Cluster 3 (Loyalist/High Value)

23.6 % from total customers

Average Lifetime Value

73 months relation length

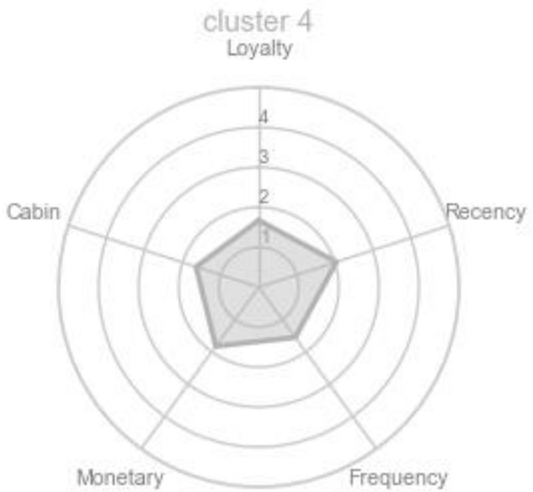
25 months from latest flight

3.5 flight per year

Behavior Description:
Most valuable, core customers. Long-term customer with high consumption due to frequent purchase and longer distance, loyal, often fly with higher class of cabin.

Recommendation:
We should **maintain the relationship and reward** these customers by giving **extra discount or free ticket, chance to win prize of flight to popular destination, and special souvenirs.** We can also give them **early access for our newest product/service** and offer them **affiliate program** to help our promotion.

Cluster 4 (Uncertain Lost/Low Value)



16.2 % from total customers

Average Lifetime Value

22 months relation length

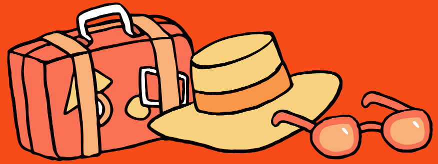
259 months from latest flight

1.75 flight per year

Behavior Description:
New member with very low consumption, rare purchase and shorter distance. Haven't used our service lately and used to book for low class of cabin.

Recommendation:
We should promote them **first time member buyer**. Provide **free vouchers and starting reward points to spend on tickets and helpful product in airline platform**. **Promote cheap tickets for short flights** in public holidays. Send them campaigns/promo in mainstream platforms to reconnect.

Recommendation



Business Recommendation

- **Increase the airport facilities**, like more airline counters, gates, and hangars **in several local provinces: Guangdong, Beijing, and Shanghai**. **Provide more flights** in those areas to accommodate more customers.
- To expand the brand globally, we need to increase the proportion of international customers. The earliest regions to increase the flight service is in East Asian neighbor countries: Hong Kong, Japan, South Korean, and Taiwan, plus US. **Build 3 branch offices** (if currently not available): 1). **Japan/South Korea**; 2). **Taiwan**, also covers Hong Kong, or ask Guangdong to cover these areas; and 3). **US**. This office should provide more exposure and suitable country-based marketing strategy and flight service.
- **The main marketing content should be tailored for millennials audience.**
- Create a distinctive flight service and requirement for each FFP Tier based on discount, point rewards, and on-flight service. Requirements and reward points should be fulfilled by purchase amount and flight frequency. Promote them back to the customers. There are **7594 customers in total that should be encouraged to level up their FFP Tier.**

Segmented Marketing Strategy

Cluster 3 - Loyalist

focus **on maintaining relationship and reward**. Give extra discount or free ticket, chance to win prize of flight to popular destination, and special souvenirs. Assign as first adopter for newest product/service and offer affiliate program.

Cluster 0 – Potential Loyalist/High Prospect

focus on **retention strategy**, by offer extra discount or reward points based on membership time and hook with affiliate program.

Cluster 1 – Hibernating VIP

focus **on attracting them to re-purchase**. Offer flight promotions that is bundled with destination event vouchers or signatured souvenirs. Run satisfaction surveys.

Cluster 2 – Low Consumer

focus **on attracting them to re-purchase**. Provide special offers such as discount for first two flights in a year and free voucher for affiliated product/event thereafter.

Cluster 4 – Uncertain Lost

promote **first time member buyer**. Provide free vouchers and starting reward points to spend on tickets and product in airline platform. Promote cheap tickets for short flights in public holidays.

Reference

[1] Tao, Y.: Analysis Method for Customer Value of Aviation Big Data Based on LFRMC Model. ICPCSEE 2020, CCIS 1257. 89-100(2020)

[2] Kandeil, D., Saad, A. and Youssef, S. M.: A Two-phase Clustering Analysis for B2B Customer Segmentation. INCoS 2014, IEEE. 221-228(2014)

[3] Wer, J. T., et. al: Applying Data Mining and RFM Model to Analyze Customers' Values of A Veterinary Hospital. IS3C 2016, IEEE, 481-484(2016)

[4] https://www.kaggle.com/code/gilangpanduparase/air-line-customer-segmentation/notebook?scriptVersionId=64809717

[5] https://clevertap.com/blog/rfm-analysis/

[6] https://towardsdatascience.com/cheat-sheet-to-implementing-7-methods-for-selecting-optimal-number-of-clusters-in-python-898241e1d6ad

[7] https://towardsdatascience.com/cheat-sheet-to-implementing-7-methods-for-selecting-optimal-number-of-clusters-in-python-898241e1d6ad

[8] https://github.com/mrafifrbbn/airline_customer_segmentation/blob/main/Airline%20Customer%20Segmentation.ipynb

Thank You

Visit my page:
https://github.com/LuckyFasyni/Airline-Customer-Segmentation-LRFMC-Model

luckyfasyni

LuckyFasyni