

西南石油大學



NAME: KARIMA ZAFAR

ID: 201739060051

MAJOR: SOFTWARE ENGINEERING

GRADE: 2017

AVIATION CUSTOMER VALUE ANALYSIS

1. Target

- Classify customers.
- Analyze the distinctiveness of different customer categories and compare the customer value of different customers

2. Analysis method and process

The goal of this task is: customer value recognition

The most broadly used model for identifying customer value applications is the RFM model.

R: recent consumption time interval recency

F: frequency of consumption

M: consumption amount monetary

But for this model as LRFMC;

L: length of customer relationship;

R: Time interval of recent consumption

F: consumption frequency;

M: Accumulate flight miles within a certain period of time;

C: Average of discount factor

By clustering the five indicators of the airline's LRFMC model, the most valuable customer groups are found.

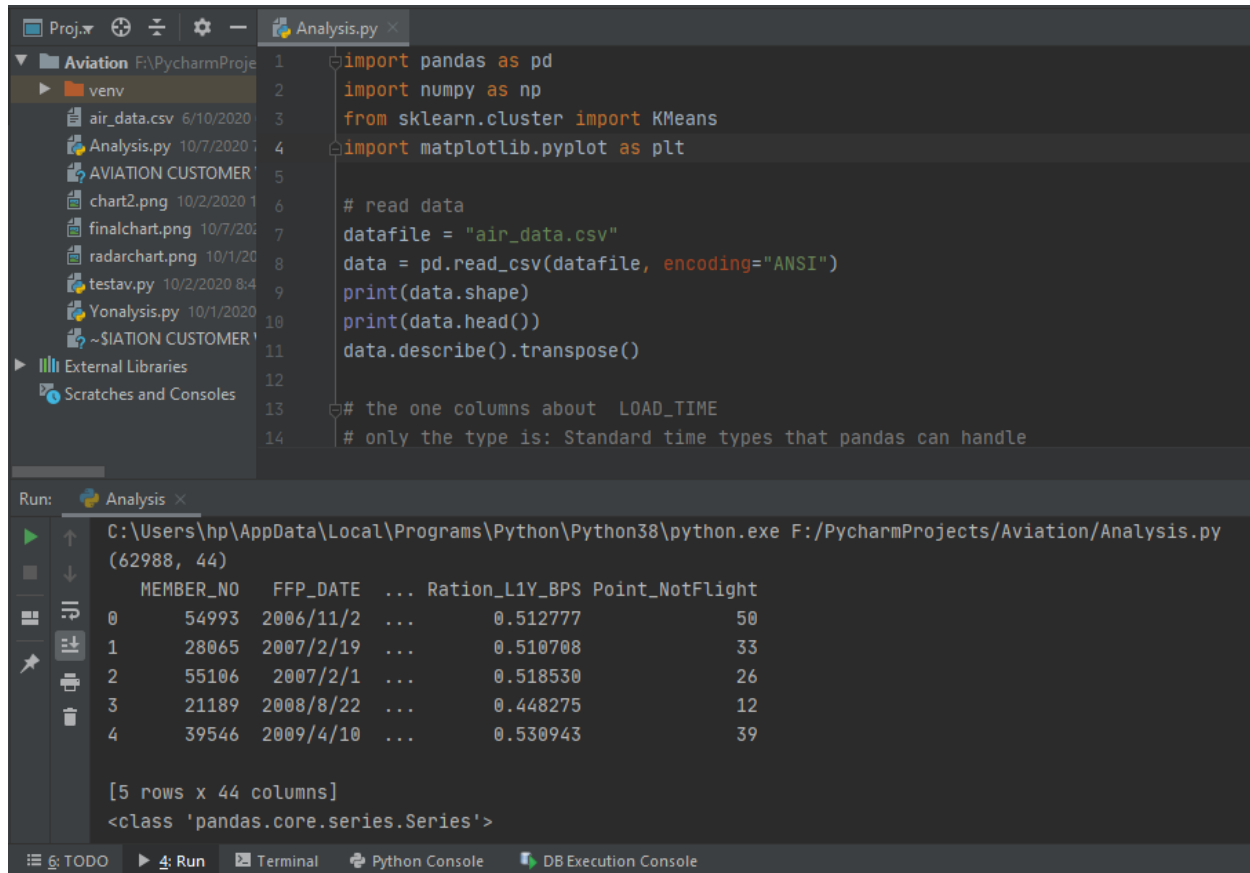
2.1 Analysis process:

1. Extract data from data sources
2. Establish a model, use data to test the model, analyze customer groups, and identify valuable customers.

2.2 Data extraction

With 2014-3-31 as the end time, a two-year period is selected as the analysis observation data to form historical data, and the data from 2014-3-31 until now is used as incremental data.

Read data, print shape and describe:



The screenshot shows a PyCharm IDE with a project named 'Aviation'. The file explorer on the left shows a 'venv' directory and several files including 'air_data.csv', 'Analysis.py', 'AVIATION CUSTOMER', 'chart2.png', 'finalchart.png', 'radarchart.png', 'testav.py', 'Yonalysis.py', and '~SIATION CUSTOMER'. The main editor displays the 'Analysis.py' script with the following code:

```

1 import pandas as pd
2 import numpy as np
3 from sklearn.cluster import KMeans
4 import matplotlib.pyplot as plt
5
6 # read data
7 datafile = "air_data.csv"
8 data = pd.read_csv(datafile, encoding="ANSI")
9 print(data.shape)
10 print(data.head())
11 data.describe().transpose()
12
13 # the one columns about LOAD_TIME
14 # only the type is: Standard time types that pandas can handle

```

The Run console at the bottom shows the execution of the script. It displays the file path, the shape of the data (62988, 44), and a preview of the first five rows of the data:

	MEMBER_NO	FFP_DATE	...	Ration_L1Y_BPS	Point_NotFlight
0	54993	2006/11/2	...	0.512777	50
1	28065	2007/2/19	...	0.510708	33
2	55106	2007/2/1	...	0.518530	26
3	21189	2008/8/22	...	0.448275	12
4	39546	2009/4/10	...	0.530943	39

Below the table, the console shows the output of the 'print(data.shape)' statement: [5 rows x 44 columns] and the output of 'print(data.head())': <class 'pandas.core.series.Series'>.

2.2.1 Attribute Specification

According to the model LRFMC, the relevant indicators are selected and the attributes are constructed.

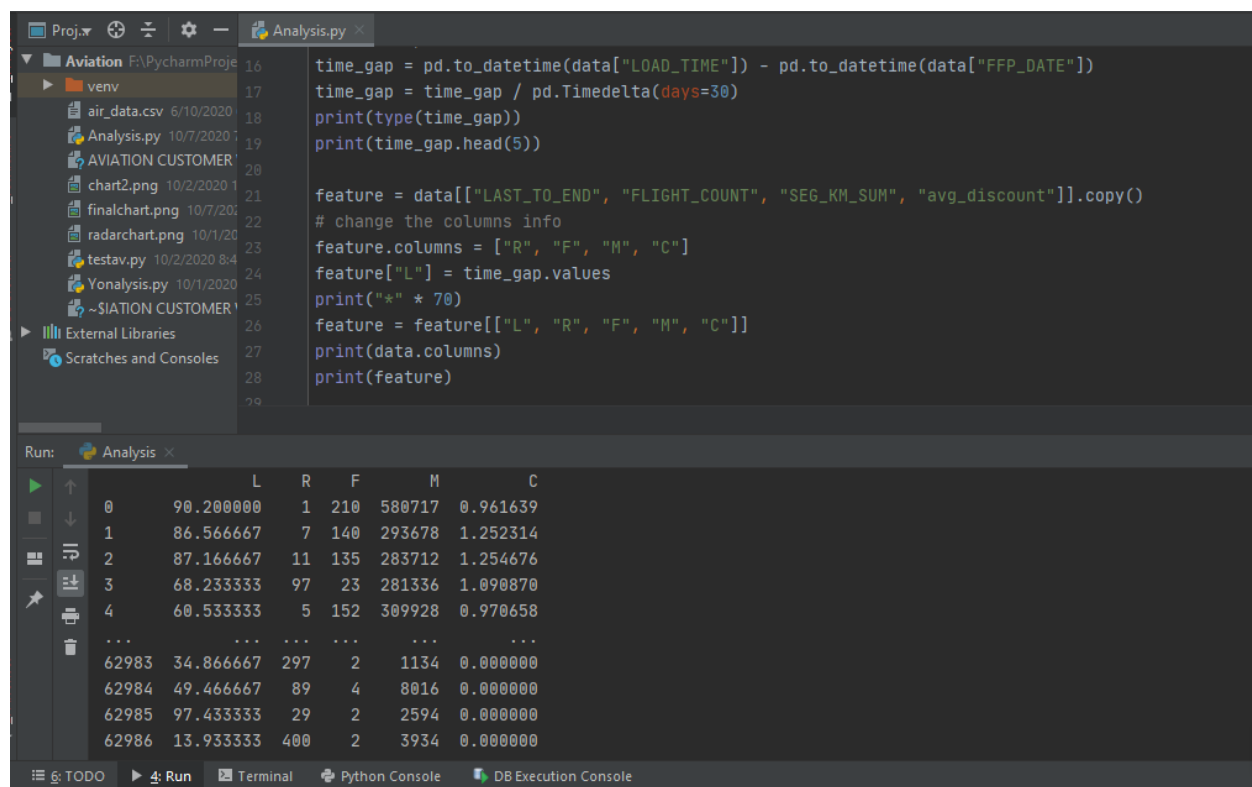
$L = \text{LOAD_TIME} - \text{FFP_DATE}$; the number of months from the member's membership time to the observation window = observation end time - membership time

R=LAST_TO_END; the number of months from the customer's last boarding time to the end of the observation window = the time from the last boarding time to the end of the observation window

F=FLIGHT_COUNT; The number of planes taken in the observation window = the number of flights in the observation window

M=SEG_KM_SUM; customer's accumulated flight mileage during the observation time = total flight kilometers in the observation window

C=avg_discount; average discount coefficient = average discount rate



```
16 time_gap = pd.to_datetime(data["LOAD_TIME"]) - pd.to_datetime(data["FFP_DATE"])
17 time_gap = time_gap / pd.Timedelta(days=30)
18 print(type(time_gap))
19 print(time_gap.head(5))
20
21 feature = data[["LAST_TO_END", "FLIGHT_COUNT", "SEG_KM_SUM", "avg_discount"]].copy()
22 # change the columns info
23 feature.columns = ["R", "F", "M", "C"]
24 feature["L"] = time_gap.values
25 print("*" * 70)
26 feature = feature[["L", "R", "F", "M", "C"]]
27 print(data.columns)
28 print(feature)
29
```

Run: Analysis

	L	R	F	M	C
0	90.200000	1	210	580717	0.961639
1	86.566667	7	140	293678	1.252314
2	87.166667	11	135	283712	1.254676
3	68.233333	97	23	281336	1.090870
4	60.533333	5	152	309928	0.970658
...
62983	34.866667	297	2	1134	0.000000
62984	49.466667	89	4	8016	0.000000
62985	97.433333	29	2	2594	0.000000
62986	13.933333	400	2	3934	0.000000

2.3 Cluster analysis K-Mean method

The construction of customer value analysis model is mainly composed of two parts:

1. Based on the data of 5 indicators of airline customers, cluster analysis of customer groups
2. Analyze the characteristics of each customer group in combination with the business, analyze customer value, and rank the customer group

The K-mean clustering algorithm is used to group customer data, and the number of cluster categories is 5.

Print mean/count, etc. and standardize data:

The screenshot shows a PyCharm IDE window with a project named 'Aviation'. The file explorer on the left shows a directory structure with files like 'air_data.csv', 'Analysis.py', 'AVIATION CUSTOMER', 'chart2.png', 'finalchart.png', 'radarchart.png', 'testav.py', 'Yonalysis.py', and '~SIATION CUSTOMER'. The main editor displays the code in 'Analysis.py'.

```

26 feature = feature[["L", "R", "F", "M", "C"]]
27 print(data.columns)
28 print(feature)
29
30 # find mean count etc
31 feature_summary = feature.describe(percentiles=[], include='all')
32 print(feature_summary)
33
34 # standardize
35 feature = (feature - feature.mean(axis=0)) / (feature.std(axis=0))
36 print(feature.head())
37
38
39 k = 5

```

The 'Run' output window shows the following data:

	L	R	F	M	C
count	62988.000000	62988.000000	62988.000000	62988.000000	62988.000000
mean	49.500028	176.120102	11.839414	17123.878691	0.721558
std	28.240765	183.822223	14.049471	20960.844623	0.185427
min	12.166667	1.000000	2.000000	368.000000	0.000000
50%	42.333333	108.000000	7.000000	9994.000000	0.711856
max	114.566667	731.000000	213.000000	580717.000000	1.500000

	L	R	F	M	C
0	1.441178	-0.952660	14.104488	26.887901	1.294751
1	1.312523	-0.920020	9.122093	13.193844	2.862354
2	1.333768	-0.898260	8.766208	12.718386	2.875087

Clustering:

The screenshot shows a PyCharm IDE window with a project named 'Aviation'. The file explorer on the left shows a directory structure with files like 'air_data.csv', 'Analysis.py', 'AVIATION CUSTOMER', 'chart2.png', 'finalchart.png', 'radarchart.png', 'testav.py', 'Yonalysis.py', and '~SIATION CUSTOMER'. The main editor displays the code in 'Analysis.py'.

```

39 k = 5
40 x = feature[['L', 'R', 'F', 'M', 'C']]
41 kms = KMeans(n_clusters=k)
42 kms.fit(x) # training model
43 print('Cluster centers:', kms.cluster_centers_)
44 print('Category:', kms.labels_)
45 r1 = pd.Series(kms.labels_)
46 r1 = r1.value_counts()
47 r2 = pd.DataFrame(kms.cluster_centers_)
48 r = pd.concat([r2, r1], axis=1)
49 r.columns = [u'number of clusters', u'R', u'F', u'M', u'C', u'L']
50 r.index.name = 'Clustering category'
51 r.index = ([u'Customer Group 1', u'Customer Group 2', u'Customer Group 3', u'Customer Group 4', u'Customer Group 5'])

```

Drawing radar chart:

The screenshot shows a PyCharm IDE with a project named 'Aviation'. The file explorer on the left shows a directory structure with files like 'air_data.csv', 'Analysis.py', 'AVIATION CUSTOMER', 'chart2.png', 'finalchart.png', 'radarchart.png', 'testav.py', 'Yonalysis.py', and '~SATION CUSTOMER'. The main editor displays the code for 'Analysis.py'.

```

53 # Plot chart
54 labels = r.columns
55 plot_data = kms.cluster_centers_
56 color = ['b', 'g', 'r', 'c', 'y']
57 angles = np.linspace(0, 2*np.pi, k, endpoint=False)
58 plot_data = np.concatenate((plot_data, plot_data[:, [0]]), axis=1)
59 angles = np.concatenate((angles, [angles[0]]))
60 fig = plt.figure(figsize=(8, 6))
61 ax = fig.add_subplot(111, polar=True) # polar parameter
62 for i in range(len(plot_data)):
63     ax.plot(angles, plot_data[i], 'bo-', color=color[i], label=r.index[i], linewidth=2)
64     ax.set_thetagrids(angles * 180/np.pi, labels)
65 plt.legend(loc=1)
66 plt.show()
67
for i in range(len(plot_data))

```

The Run console at the bottom shows the output of the script:

```

3 0.663343 -0.430416 0.794378 12.605052 1.991687
4 0.390687 -0.930900 9.976218 13.969099 1.343389
Cluster centers: [[ 1.17216961 -0.38403311 -0.07949906 -0.08688036 -0.15782906]
[-0.31806098  1.67413359 -0.5736022  -0.53589517 -0.21141023]
[-0.69666215 -0.42889889 -0.15035832 -0.14886325 -0.28203463]
[-0.03573013  0.03393228 -0.27349702 -0.28702318  1.96290207]
[ 0.48975399 -0.80844185  2.49762918  2.43896643  0.32458468]]
Category: [4 4 4 ... 0 1 1]

```

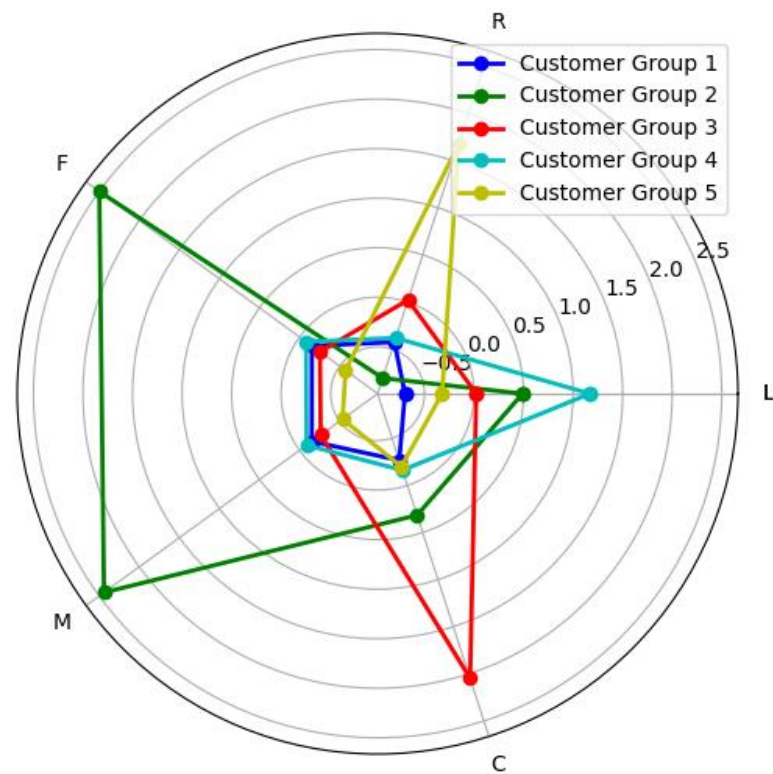


Figure 1 Radar chart of customer value analysis (top)

Customer group 1: Smallest on L and C

Customer group 2: Largest on F and M, smallest on R

Customer group 3: The largest in C

Customer group 4: Largest on L

Customer group 5: R is the largest; F and M are the smallest

Analyze the actual situation: evaluate and analyze a certain customer group through the distribution of the size of a certain indicator in the customer group.

For example, customer group 2 is the largest on F and M, and is the smallest on R, indicating that customer group 2 has the largest number of flights, the largest mileage, and the time from the last flight to the observation window is the shortest. It shows that FMR is a dominant feature in customer group 2.

The index characteristics of customer group 5 are exactly the opposite of those of customer group 2, which shows that FMR is a disadvantaged characteristic of customer group 5.

After analysis, the five customer groups have significantly different performance characteristics. According to the characteristics, five different levels of customer groups are defined: important-keep customers, important-develop customers, important-retain customers, general-value customers, and low -Value customers.

Analyze the customer value ranking based on customer group characteristics:

Customer Value Ranking:

CUSTOMER GROUP	RANK	SIGNIFICANCE
Customer Group 2	1	Important – keep
Customer Group 3	2	Important – developing
Customer Group 4	3	Important – retain
Customer Group 1	4	General – value
Customer Group 5	5	Low – value

3. Customer value analysis:

Important-Keep customers: This type of customer has a long membership time (L), frequent flights (F), the longest total mileage (M), and the average discount rate (C) is relatively large, and the most recent time (R) have taken the company's flight. Such customers have the highest value to the company and are the most ideal customers for airlines.

Recommendation: For this type of customer group, airlines should put resources on them as much as possible, develop one-to-one precision marketing, and improve customer satisfaction, so that this group of people has a longer consumption cycle.

Important-Developing customers: This type of customer has a relatively short membership time (L), a relatively low number of flights (F), and a relatively short total flight mileage (M). The time spent on company flights (R) recently is relatively long, and the average discount rate (C) Higher level. The current value generated by such customers is not high, but they are potential value customers of airlines.

Suggestion: airlines must develop marketing strategies. For example, if customers choose their airlines for a certain number of times, they will upgrade their membership and give them more preferential methods to attract such customers to spend on their airlines and provide high-quality services. Increase the satisfaction of such customers and make them become loyal customers of airlines.

Important-Retaining customers: This type of customer has a long membership time (L) and is a loyal customer group of airlines. This type of customer has high flight frequency (F) and total flight mileage (M) in the past, but took flight time recently (R) Longer. This type of customer value transformation is highly uncertain.

Suggestion: The airlines, based on the recent consumption time and frequency of these customers, infer customer consumption changes, and formulate corresponding strategies to recover such customers as much as possible and increase the value of such customers.

General-value customers and low-value customers: These types of customers have relatively small values for the joining time (L) and the average discount rate (C), the flight mileage (M) is short, the flight frequency (F) is low, and the last flight is taken Time (R) is relatively long. Such customers only choose the airline when it sells low-cost tickets.

Recommendation: Frequently release special discount air tickets and make full use of aviation resources to make these customers pay more attention to the airline.