



## 大数据课程开发实践期末报告 2

实验题目：\_\_\_\_\_ 期末大作业 \_\_\_\_\_

姓 名：\_\_\_\_\_ 姚光明 \_\_\_\_\_

学 号：\_\_\_\_\_ 20051133 \_\_\_\_\_

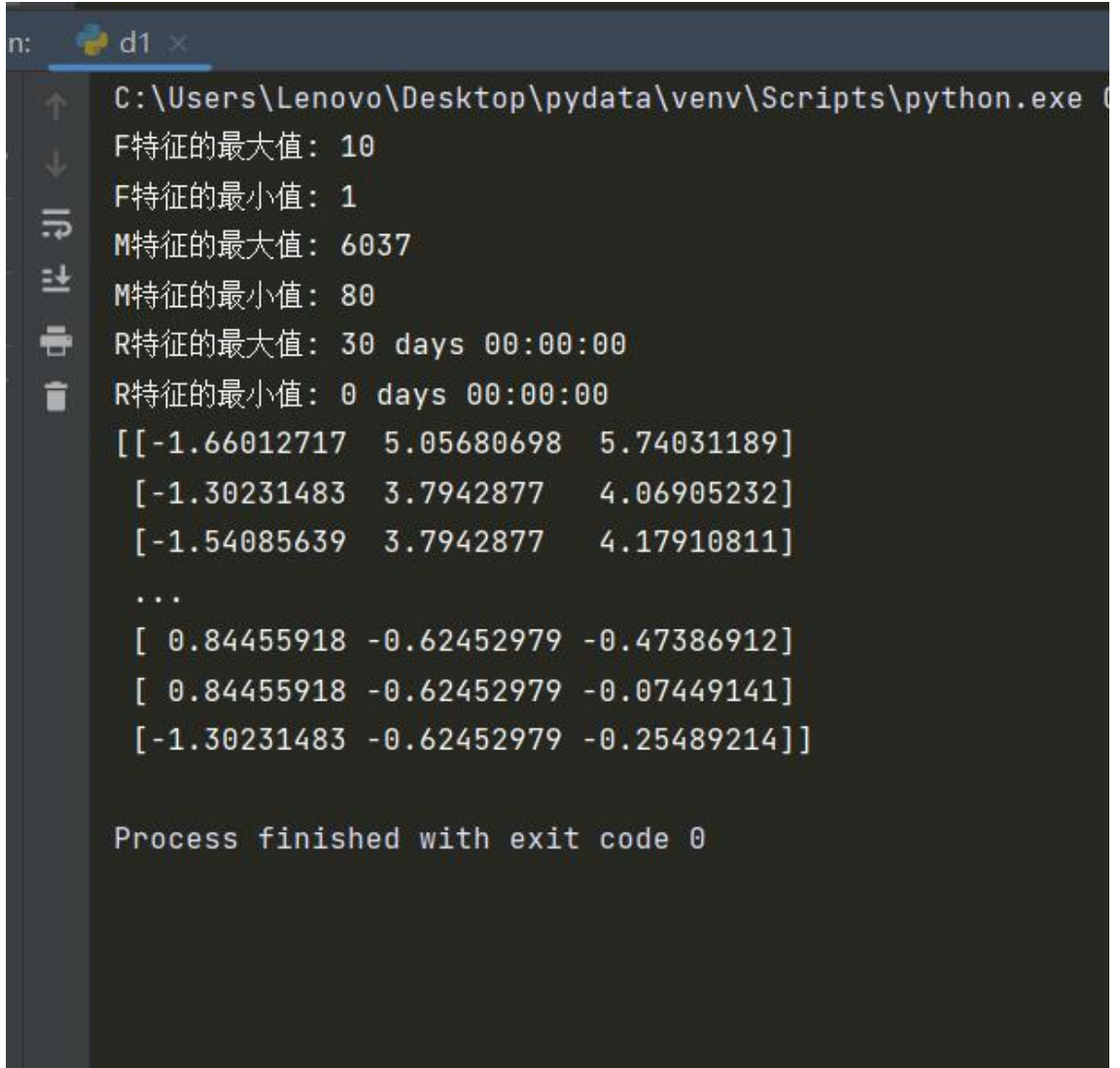
专 业：\_\_\_\_\_ 计算机科学与技术 \_\_\_\_\_

## 目 录

1 餐饮企业客户流失分析 .....	2
2 餐饮企业客户流失预测 .....	3

## 1 餐饮企业客户流失分析

### 1.1 特征值计算



```
n: d1 x
C:\Users\Lenovo\Desktop\pydata\venv\Scripts\python.exe 0
F特征的最大值: 10
F特征的最小值: 1
M特征的最大值: 6037
M特征的最小值: 80
R特征的最大值: 30 days 00:00:00
R特征的最小值: 0 days 00:00:00
[[-1.66012717  5.05680698  5.74031189]
 [-1.30231483  3.7942877  4.06905232]
 [-1.54085639  3.7942877  4.17910811]
 ...
 [ 0.84455918 -0.62452979 -0.47386912]
 [ 0.84455918 -0.62452979 -0.07449141]
 [-1.30231483 -0.62452979 -0.25489214]]

Process finished with exit code 0
```

### 1.2 kmeans 聚类

## 2.1 构建特征

```
C:\Users\Lenovo\Desktop\pydata\venv\Scripts\python.exe C:\Users\Lenovo\Desktop\pyd
历史客户信息表的维数: (2431, 38)
```

```
历史订单表的维数: (6611, 21)
```

	USER_ID	number_consumers	expenditure
0	3556	4	423
1	1874	7	1101
2	3484	5	437
3	3639	2	251
4	3835	2	363

	USER_ID	MYID	ACCOUNT	NAME	ORGANIZE_ID	...	sex	poo	address	age	type
0	3300	NaN	艾朵	a2328	330	...	男	广东汕尾	广州	26	非
1	3497	NaN	艾国真	a2525	330	...	男	广东汕尾	广州	43	非
2	2863	NaN	艾锦程	a1891	330	...	男	广东汕尾	佛山	58	已
3	3006	NaN	艾穹	a2034	330	...	女	四川广安	广州	22	已
4	3160	NaN	艾融乐	a2188	330	...	女	广东中山	广州	20	非

```
[5 rows x 38 columns]
```

	USER_ID	ACCOUNT	LAST_VISITS	type
0	3300	艾朵	2016-05-23 21:14:00	非流失
1	3497	艾国真	2016-07-18 13:40:00	非流失
2	2863	艾锦程	2016-04-16 20:51:00	已流失
3	3006	艾穹	2016-03-26 20:47:00	已流失
4	3160	艾融乐	2016-07-17 11:40:00	非流失

	USER_ID	ACCOUNT	LAST_VISITS	type	number_consumers	expenditure
0	3300	艾朵	2016-05-23 21:14:00	非流失	10.0	1782.0
1	3300	艾朵	2016-05-23 21:14:00	非流失	2.0	345.0
2	3300	艾朵	2016-05-23 21:14:00	非流失	10.0	1295.0
3	3300	艾朵	2016-05-23 21:14:00	非流失	6.0	869.0
4	3497	艾国真	2016-07-18 13:40:00	非流失	3.0	589.0

	USER_ID	ACCOUNT	LAST_VISITS	type	number_consumers	expenditure
--	---------	---------	-------------	------	------------------	-------------

```

0    2361    41    237.0 ... 非流失    7.0    1029.0
1    3478    37    231.0 ... 非流失    5.0    833.0
2    3430    34    224.0 ... 非流失    5.0    866.0
3    3762    33    208.0 ... 非流失   10.0   1303.0
4    3307    33    199.0 ... 非流失    2.0    371.0

```

[5 rows x 9 columns]

(2431, 9)

合并后表中的空值数目: USER\_ID

0

frequence 0

numbers 0

amount 0

ACCOUNT 0

LAST\_VISITS 46

type 0

number\_consumers 4

expenditure 4

dtype: int64

(2383, 9)

(2383, 9)

```

    USER_ID  frequence  numbers ... number_consumers  expenditure  average
0    2361         41    237.0 ...              7.0        1029.0   146.77
1    3478         37    231.0 ...              5.0         833.0   145.32
2    3430         34    224.0 ...              5.0         866.0   142.42
3    3762         33    208.0 ...             10.0        1303.0   146.12
4    3307         33    199.0 ...              2.0         371.0   152.76

```

[5 rows x 10 columns]

```

    USER_ID  frequence  numbers ...  expenditure  average  recently
0    2361         41    237.0 ...        1029.0   146.77         0

```

## 2.2 预测



```

'准流失' '非流失' '准流失' '准流失' '准流失' '准流失' '准流失' '非流失' '准流失' '非流失' '准流失' '准流失'
'非流失' '非流失' '准流失' '非流失' '准流失' '准流失' '准流失' '非流失' '准流失' '准流失' '非流失' '非流失'
'准流失' '准流失' '准流失' '准流失' '非流失' '非流失' '准流失' '准流失' '非流失' '准流失' '准流失' '准流失'
'准流失' '准流失' '准流失' '非流失' '准流失' '非流失' '非流失' '准流失' '准流失' '准流失' '准流失' '非流失'
'准流失' '准流失' '非流失' '非流失' '准流失' '准流失' '准流失' '准流失' '准流失' '准流失' '非流失' '准流失'
'非流失' '非流失' '准流失' '非流失' '非流失' '准流失' '非流失' '准流失' '非流失' '非流失' '准流失' '准流失'
'准流失' '准流失' '非流失' '准流失' '准流失' '准流失' '准流失' '准流失' '非流失' '准流失' '非流失' '准流失'
'准流失' '准流失' '准流失' '准流失' '准流失' '非流失' '准流失' '准流失' '非流失' '准流失' '非流失' '非流失'
'准流失' '非流失' '非流失' '非流失' '准流失' '非流失' '非流失' '准流失' '准流失' '非流失' '准流失' '非流失'
'非流失' '非流失' '准流失' '准流失' '非流失' '非流失' '准流失' '准流失' '非流失' '准流失' '准流失' '准流失'
'准流失' '非流失' '准流失' '准流失' '准流失' '非流失' '非流失' '准流失' '准流失' '非流失' '准流失' '准流失'
'准流失' '非流失' '准流失' '准流失' '非流失' '非流失' '准流失' '准流失' '非流失' '准流失' '准流失' '非流失'
'非流失' '非流失' '非流失' '准流失' '非流失' '非流失' '准流失' '非流失' '非流失' '准流失' '准流失' '准流失'
'准流失' '非流失' '准流失' '准流失' '非流失' '准流失' '非流失' '准流失' '非流失' '准流失' '准流失' '准流失'
'非流失' '非流失' '非流失' '非流失' '非流失' '准流失' '准流失' '准流失' '非流失' '准流失' '准流失' '准流失'
'准流失' '准流失' '非流失' '非流失' '非流失' '准流失' '准流失' '准流失' '非流失' '准流失' '准流失' '非流失'
'非流失' '非流失' '非流失' '非流失' '准流失' '非流失' '准流失' '准流失' '非流失' '准流失' '非流失' '非流失'
'准流失' '准流失' '非流失' '非流失' '准流失' '非流失' '准流失' '准流失' '非流失' '准流失' '准流失' '非流失'
]

混淆矩阵:
[[139  38]
 [ 21 184]]
精确率:
0.829
召回率:
0.898
F1值: 0.862

Process finished with exit code 0

```