

目 录

目录

目 录.....	1
1. 用 python 读取餐饮数据，熟悉订单数据。	2
2. 数据中有重复的订单，所以对订单进行去重操作。	2
3. 统计每一单的费用.....	2
4. 查看费用不超过 1000 的订单号	3
5. 计算菜品热销度.....	3
6. 热销 TOP10.....	3
7. 统计每个订单的状态占比.....	4
8. 选取特征.....	4
9. 划分训练集、测试集.....	5
10. 构建客户——菜品二元矩阵.....	6
11. 相似度.....	6
12. 进行推荐.....	6
13. 准确率.....	6

1. 用 python 读取餐饮数据，熟悉订单数据。

可以看出有 10037 条信息，每条信息有 19 列。

detail_id	order_id	dishes_id	logicprn_name	parent_class_name	dishes_name	itemis_add	counts	amounts	cost	place_order_time	discount_amt	di
0	754	137	610047	NaN	NaN	西瓜胡萝卜沙拉	0	1	26	NaN	2016/8/3 14:01:13	NaN
1	753	137	609970	NaN	NaN	麻辣小龙虾	0	1	99	NaN	2016/8/3 14:01:47	NaN
2	758	137	609945	NaN	NaN	农夫山泉NFC果汁100%橙汁	0	1	6	NaN	2016/8/3 14:02:11	NaN
3	755	137	609992	NaN	NaN	番茄炖牛腩	0	1	35	NaN	2016/8/3 14:02:37	NaN
4	759	137	610010	NaN	NaN	白饭小碗	0	4	1	NaN	2016/8/3 14:04:55	NaN
...
10032	5077	1324	609991	NaN	NaN	香烤牛排	0	1	55	NaN	2016/8/20 18:57:27	NaN
10033	5260	1324	609938	NaN	NaN	小米南瓜粥	0	1	13	NaN	2016/8/20 18:58:08	NaN
10034	5259	1324	609940	NaN	NaN	清蒸鲷鱼	0	1	56	NaN	2016/8/20 18:58:36	NaN
10035	5258	1324	609963	NaN	NaN	盘蟹蒸蛋	0	1	55	NaN	2016/8/20 18:58:37	NaN
10036	5261	1324	609935	NaN	NaN	山药养生粥	0	1	19	NaN	2016/8/20 18:59:30	NaN

10037 rows x 19 columns

2. 数据中有重复的订单，所以对订单进行去重操作。

```
[933 rows x 3 columns]
```

去重后有 933 笔订单

3. 统计每一单的费用

```
order_id
137      3288
162     23370
163      2930
165     67320
166      2805
...
1319     5490
1321    13566
1322     7620
1323    14475
1324     5694
```

4. 查看费用不超过 1000 的订单号

```
order_id
137      False
162      False
163      False
165      False
166      False
...
1319     False
1321     False
1322     False
1323     False
1324     False
Name: accounts_payable, Length: 933, dtype: bool
```

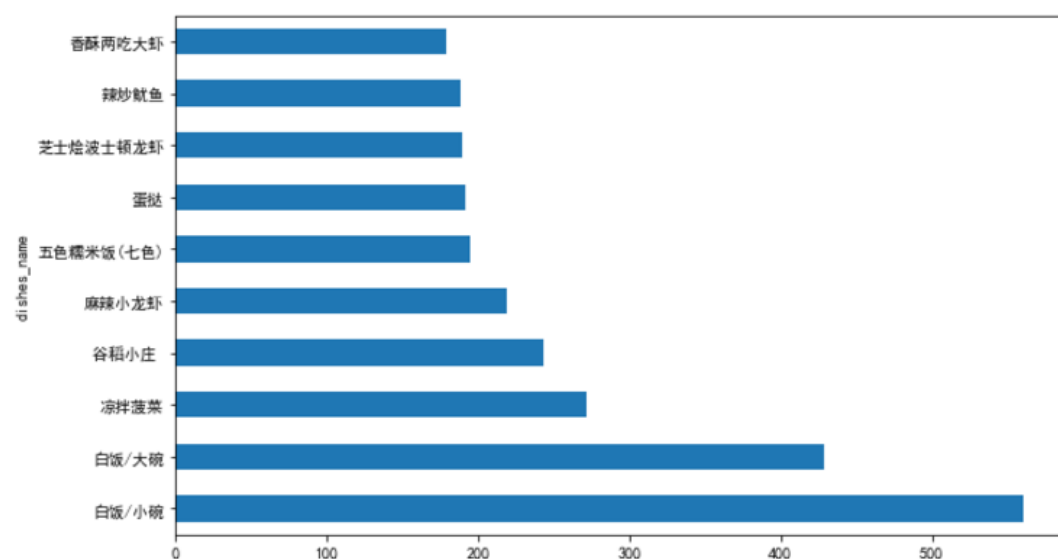
每笔订单都超过了 1000 元

5. 计算菜品热销度

菜品的热销度为：

	dishes_name	counts
0	西瓜胡萝卜沙拉	0.000000
1	麻辣小龙虾	0.000000
2	农夫山泉NFC果汁100%橙汁	0.000000
3	番茄炖牛腩	0.000000
4	白饭/小碗	0.333333

6. 热销 TOP10



热销度最高的为白饭，主食销量比较高。六月可能因为天气原因，凉拌菠菜更受欢迎。水产品成熟，龙虾类的菜品销量更高。

7.统计每个订单的状态占比

```
1    0.987302
0    0.009524
2    0.003175
```

可以看出有 0, 1, 2 三种状态，1 为正常订单，占比最多，0 和 2 为未完成订单或者无效订单，占比较少。

8.选取特征

	dishes_name	emp_id
0	西瓜胡萝卜沙拉	1248
2	麻辣小龙虾	1248
4	农夫山泉NFC果汁100%橙汁	1248
6	番茄炖牛腩	1248
10	凉拌菠菜	1248
...
33165	香烤牛排	1380
33166	小米南瓜粥	1380
33167	清蒸鲈鱼	1380
33168	盘蟹蒸蛋	1380
33169	山药养生粥	1380

8675 rows × 2 columns

9.划分训练集、测试集

训练集为:

	dishes_name	emp_id	dishes_id	rating
718	孜然羊排	1538	106	1
1450	香菇鹌鹑蛋	1538	115	1
1066	芝士焗波士顿龙虾	1538	111	1
2049	香菇鸡肉粥	1538	123	1
1540	拌土豆丝	1538	116	1
1766	53度茅台	1538	118	1
1841	意文柠檬汁	1538	120	1
105	麻辣小龙虾	1538	101	1
1798	路易拉菲红酒干红	1538	119	1
2026	黄油曲奇饼干	1538	122	1
5200	倒立蒸梭子蟹	990	172	1
5689	蒜香辣花甲	990	179	1
1541	拌土豆丝	990	116	1
5103	葱姜炒蟹	990	169	1
2844	啤酒鸭	990	136	1
5360	独家薄荷鲜虾牛肉卷	990	175	1
5318	露露无糖杏仁露	990	174	1
3228	葡萄奶酥(超酥香的饼干)	990	145	1
2746	爆炒猪肝	990	135	1
1317	水煮鱼	990	113	1
6951	花蛤蒸蛋	1141	201	1
6989	五彩豆	1141	202	1

测试集为:

	dishes_name	emp_id	dishes_id	rating
0	西瓜胡萝卜沙拉	1248	100	1
1	西瓜胡萝卜沙拉	1097	100	1
2	西瓜胡萝卜沙拉	1141	100	1
3	西瓜胡萝卜沙拉	996	100	1
4	西瓜胡萝卜沙拉	1002	100	1
...
8344	土豆西红柿汤面	1536	241	1
8345	土豆西红柿汤面	1489	241	1
8346	土豆西红柿汤面	1477	241	1
8347	铁板牛肉	1119	242	1
8348	铁板牛肉	1125	242	1

[8319 rows x 4 columns]

10. 构建客户——菜品二元矩阵

```
dishes_id 100 101 102 103 104 105 106 ... 236 237 238 239 240 241 242
emp_id
982      1   0   0   1   0   0   0   ...   0   0   0   0   0   0   0
984      0   0   0   0   0   0   0   ...   0   0   0   0   0   1   0
986      0   0   0   0   1   0   0   ...   0   0   0   0   0   0   0
987      0   1   0   1   1   0   0   ...   0   0   0   0   0   0   0
988      0   1   0   0   1   0   0   ...   0   0   0   0   0   0   0
...
1605     0   1   0   0   1   0   1   ...   0   0   0   0   0   0   0
1606     0   0   0   0   0   0   0   ...   0   0   0   0   0   0   0
1607     0   0   0   0   1   0   1   ...   0   0   0   0   0   0   0
1609     0   1   0   1   1   0   1   ...   0   0   0   0   0   0   0
1610     0   0   0   0   1   0   0   ...   0   0   0   0   0   0   0

[455 rows x 143 columns]
```

11. 相似度

```
[[0.02857143 0.09090909 ... 0.08333334 0.13043478
 0.13333334]
 [0.02857143 0.13333334 ... 0.09090909 0.125
 0.03846154]
 [0.09090909 0.13333334 0.04347826 0.2
 0.15384616]
 ...
 [0.08333334 0.09090909 0.04347826 ... 0.13043478
 0.0625]
 [0.13043478 0.125 0.2 ... 0.13043478 0.0625]
 [0.13333334 0.03846154 0.15384616 ... 0.0625 0.0625 0.]
]]
```

12. 进行推荐

1538号客户推荐中正确的菜品有：
百里香奶油烤红酒牛肉，

13. 准确率

正确率为0.125000

总结

根据该餐饮企业该月的菜品销售对菜品进行推荐的准确率较低，准确率为 0.25，对客户进行菜品推荐时，收获的满意度可能较低。