目录

[一、 背景介绍 2](#_bookmark0)

[1、 行业背景 2](#_bookmark1)

[2、 分析目标 2](#_bookmark2)

[二、 数据说明 2](#_bookmark4)

[三、 数据读取与探索性分析 3](#_bookmark8)

[1、 数据读取 3](#_bookmark9)

[2、 处理菜品名称字符数据 3](#_bookmark10)

1. 构建热销度评分指标 3
2. 绘制条形图热销top10 4

四、 数据预处理 4

1、 统计各状态的订单占比并选取有效订单 4

2、 选取主要特征 5

[五、 构建模型准备 5](#_bookmark8)

[1、 划分测试集训练集 5](#_bookmark9)

[2、 构建菜品二元矩阵的训练集数据 5](#_bookmark10)

六、 模型构建 6

1、 计算菜品间相似度矩阵 6

2、 生成菜品推荐列表 6

六、 模型评价 6

1、 构建测试集客户IP-菜品字典 6

2、 计算推荐准确率 7

# 一、 背景介绍

## 1、 行业背景

在时代飞速发展的背景下，传统餐饮行业的竞争也愈演愈烈，逐渐改变其营业方式。其中，要赢取更多客户的好评，便要弄清楚顾客的喜好，对其提供个性化的推荐服务，智能餐饮推荐服务便应运而生。

智能点餐推荐系统为每位消费者进行大数据分析，精准画像，为消费者个性化的呈现出一个个喜爱甚至惊喜，更能激发起消费者内心的唤醒与消费者体验，可提高店方的宣传效果与转化率，也能增强消费者的愉悦感，是一件互利双赢的事情。

## 2、 分析目标

对某餐饮企业的订单数据进行读取与探索性分析，之后对数据进行预处理，并构建模型，用协同过滤算法为客户进行个性化的智能餐饮推荐服务，最后对模型应用评价指标对分析推荐的效果。

# 二、 数据说明

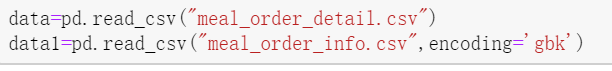
在本项目中，数据集包括meal\_order\_detail.csv和meal\_order\_info.csv两个文件，分别是菜品信息与订单信息，其中具体的数据参数说明如下表所示。



# 数据读取与探索性分析

## **3.1数据读取**

将两个csv的信息分别读入dataframe变量data和data1中。



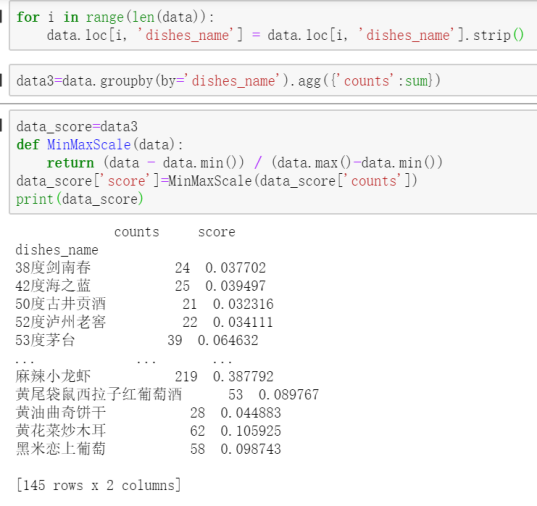
**3.2处理菜品名称字符数据**

在detail文件中，我们想将提取出所有的菜品名称，但是由于菜品名称重复较多，且含有杂乱字符，因此需要对菜品名称字符变量进行处理。



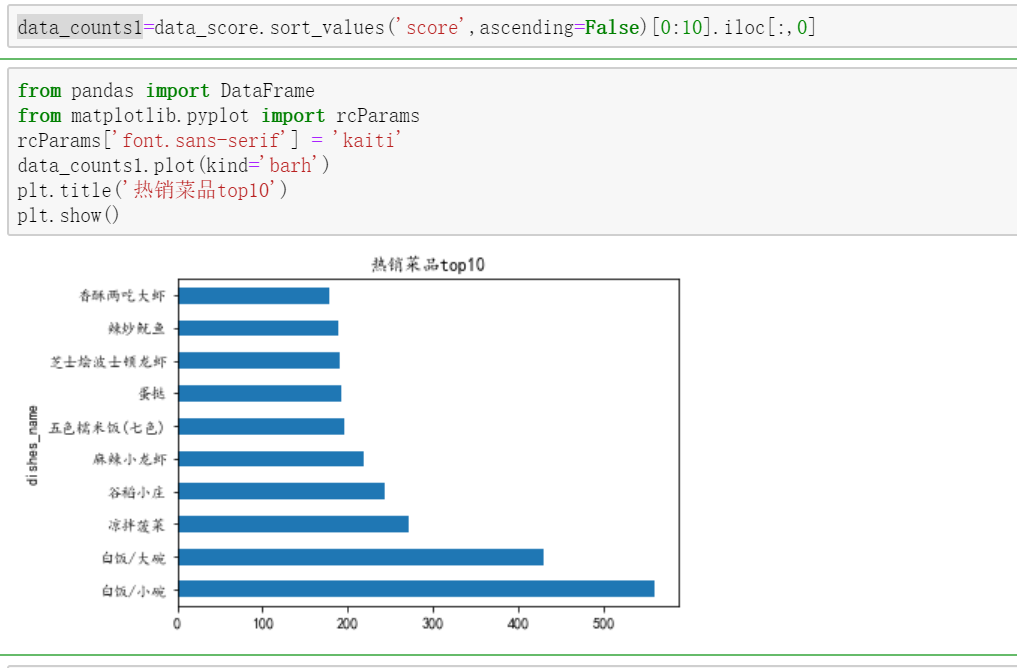
**3.3构建热销度评分指标**

计算每个菜品在统计时间内的热销度，使用MinMax标准化的方式评分，取值范围为0-1。计算公式为。

其中counts为销量，counts为评分。

**3.4绘制条形图展示热销top10**

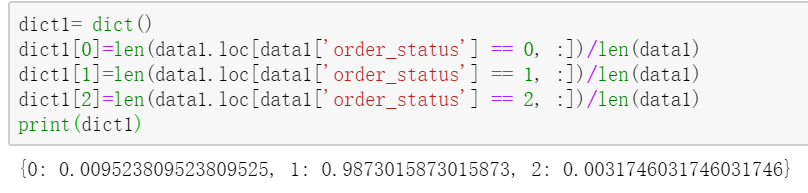
为了方面观察最为热销的菜品，我们选择横向条形条形图的方式展示销量最多的十个。



# 数据预处理

## **4.1统计各状态的订单占比并选取有效订单**

在上一步的数据说明中我们可以从info文件中看到，订单状态有0,1,2三种其中0和2分别是未结算和已锁单的状态，只有1是已结算的效订单，因此需要从订单中观察有效订单的占比，并将所有有效订单选出作为下一步建立模型的数据。

从结果可以看出，有效数据大约98.7%。

下一步便需要剔除无效数据，将有效的订单选取出来。需要剔除的数据有三类，第一类是detail中的白饭数据，因为推荐白饭是没有意义的，第二类是detail中对应在info中订单状态为0和2的数据，第三类是在detail中出现过而在info中没有出现过的数据。

与原数据相比，剔除了近千条。

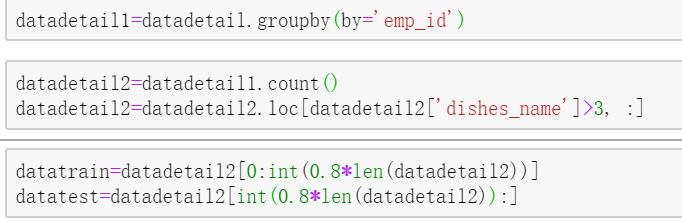
## **4.2选取主要特征**

由于数据的特征较多，在这里要进行主要特征的选取。出于我们对菜品推荐的目的推动，且销量等重要参考标准的数据在对于detail总结的info数据中已经存在，因此我们只需要保留菜品名称和与info表建立联系的客户id两个特征即可。

# 构建模型准备

## **5.1划分测试集训练集**

以8:2的比例划分测试集与训练集，由于train\_test\_split 函数只能随机划分，同一个客户id对应的菜品数据可能同时被划分到训练集和测试集当中，因此在训练时某个客户id的订单信息可能是不够完整的。因此我们需要根据客户id对训练集测试集进行划分，且为保障训练的准确性，只保留点了三个菜以上的客户id的订单。



## **5.2****构建菜品二元矩阵的训练集数据**

每个客户id为一行，如果该订单点了某菜品，则该行对应的列值为1，否则为0。



# 模型构建

## **6.1计算菜品间相似度矩阵**

我们使用Jaccard系数计算各个菜品之间的相似度，公式如下。

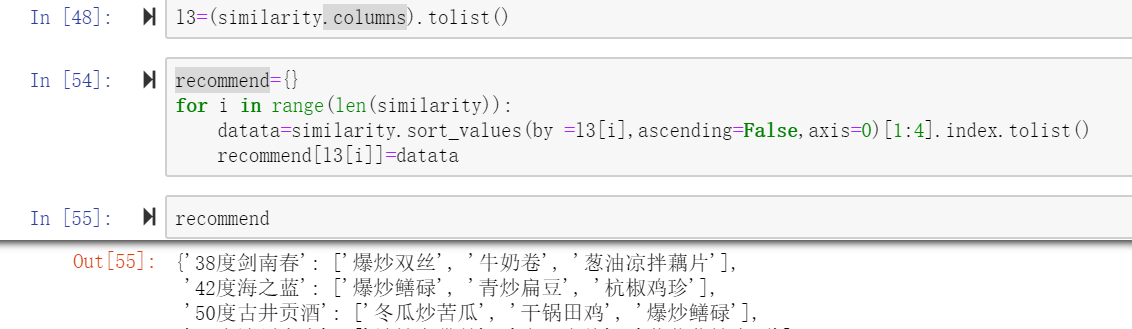
即需要我们计算每两个菜品，分别点了的客户人数，以及都点了的客户人数。



矩阵的对角线元素全部为1，即菜品自身相关度为1。

## **6.2生成菜品推荐列表**

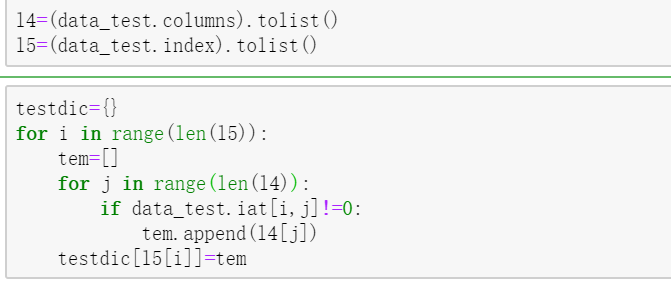
我们的目的是根据顾客点了某个菜，根据相似度矩阵推荐给他与这个菜品相似度最高的三个菜品（除了这个菜品本身）。



# 模型评价

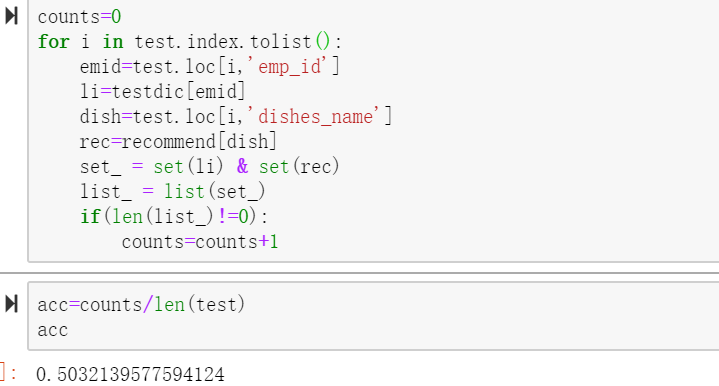
## **7.1构建测试集客户IP-菜品字典**

要测试推荐的准确率，我们需要知道对于每一位客户，我们推荐的菜品是否是该客户点过的，即是否是合这位顾客的胃口的，若是则推荐成功，因此需要先构建测试集字典。



## **7.2计算推荐准确率**

对于每位顾客点的每份菜品，我们推荐三个相似度最高的，若该位顾客点过这饭菜，则此次菜品推荐成功。



推荐的准确率约为百分之五十。