**项目分析报告**

**一、餐饮业行业背景与行业特色**

中国餐饮行业从2012年开始转型，2016年中国餐饮市场规模突破3.5万亿大关。从全国范围看，按照餐饮消费额，广东是全国餐饮市场规模最大省份。前十大餐饮大省分别是：广东省、山东省、江苏省、河南省、浙江省、四川省、湖南省、湖北省、福建省、安徽省。而按照城市对比，餐饮消费力指数排名前15名的城市分别是：上海、北京、苏州、深圳、厦门、广州、杭州、南京、天津、长沙、青岛、成都、武汉、郑州、重庆。

现如今，消费者迭代带来需求变化，使消费环境日益复杂化，一线城市市场发展难度增大，增速放缓;中小城市市场成熟度相对较低，但消费挖掘潜力较大。这些需求多样而多变的年轻消费群不仅是餐饮行业的挑战，亦是改革的原动力，能够顺应他们的升级需求进行转型的餐饮企业便是推动行业变革的加速剂。变得越来越“挑剔”的消费者，也催生了新的餐饮消费趋势。

首先是餐饮行业呈现出快时尚化的趋势。在生活节奏快速的当下，快时尚正成为一种趋势，餐饮行业也不例外的呈现快时尚化。人气餐饮品牌TOP50中，大众休闲餐饮最受好评。在餐饮行业回归大众市场并且主要受年轻人推动的大背景之下，价格适中、体验舒适、菜品符合大众口味的快时尚化新型休闲餐饮日益受到广泛欢迎。开放较早的厦门、广州、天津、上海、武汉是饮食最“洋气”的城市，北京是对新颖菜系接受度最高的城市，而武汉、青岛则是餐饮口味相对“传统”的城市。整体上，南方城市对休闲饮品偏好度比较高，尤其是厦门对咖啡的喜爱非常明显，而上海、长沙对奶茶也是爱得“深沉”。

其次是特色化。中国作为一个餐饮文化大国，八大菜系在全国各地各领风骚，呈现显著的特色化。与此同时，随着新锐菜品和外来菜系的加入，也丰富了消费者的就餐选择，城市间的“偏爱”与“众爱”在文化与流行间共存。作为城市迁入主力，18-25岁人群贡献了大量的餐饮消费，从菜系看，在外地也可以受到欢迎的为川菜和粤菜，而湘菜和江浙菜则主要依靠区域性偏好。同时，随着新一代消费者逐渐成长起来，他们对各地口味的接受度越来越高，数据显示，以东北菜为代表的北方菜在南方已有了大范围渗透；而港式料理除了一线城市以外，也已经打入了大部分华东地区城市。相对而言，分量足、实惠、口味亲民的西北菜近年来增长迅猛；反观江浙菜则还是主要在华东地区比较流行。

近年来，餐饮健康化的也趋势十分明显，这都源于国民健康意识的不断提升。北京、厦门、杭州是对健康化饮食偏好度相对较高的城市，其中北京、厦门是对素食接受度较高的两大城市；长沙、杭州、武汉则是对水果生鲜偏好度较高的三座城市。随着年纪的增长，人们对健康的关注度也会随之提高，不仅80后对其偏好度较高，70后也是重要的客群组成部分。另一方面，身份的改变也会影响消费者的饮食习惯，对比学生和刚毕业的年轻白领的中餐饮食结构，可以发现相对健康清淡的江浙菜、粤菜是学生转化为年轻白领后首先增加的中餐消费品类。

随着菜品的不断细分以及小品类的崛起，餐饮选择日益丰富，呈现极致化的特点。比如火锅类目就日渐细分出云南火锅、豆捞、韩式火锅、潮汕牛肉锅等品类，虽然目前市场体量尚小，但仍然为很多喜欢尝新的80后、90后们提供了多元化的餐饮选择。以小龙虾为例，这样一个小入口的相关菜系日渐增多，例如小龙虾饭、熟食品牌、各种口味的小龙虾以及以小龙虾为材料的其他菜品，把一个单品类做到极致化成为新的趋势。

最后是餐饮行业的潮流化。当“网红”渗入了人们的生活之后，以“网红店”为代表的潮流化品牌也日益增多，潮流化也成为餐饮行业一个显著的趋势。要打造一个现象级的潮流品牌，除了品牌自身的创新外，品类的选择也十分重要。例如，奶茶本身就是年轻人高偏好的一个品类，尤其是90后年轻女生，而这个群体又是最容易受潮流影响以及形成自发传播的，因而品类就相对容易打造流行品牌。与此同时，年轻人的消费习惯与观念也在引领着餐饮行业的潮流。

**二、数据获取与探索分析**

**1.1读取订单数据，并进行探索**

使用pandas读取订单数据：

import pandas as pd

csv\_data = pd.read\_csv("./stock\_day.csv") #读取整个csv文件

csv\_data = pd.read\_csv("./stock\_day.csv", usecols=['open', 'close']) #读取指定列索引字段的数据

csv\_data.to\_csv('demo.csv') #将修改完的csv的文件保存到新的路径下

**1.2 处理菜品名称字符数据**

import re

f=open(‘meal\_order\_detail.csv’,’r’)

csv=f.read()

f.close()

csv=re.sub(‘[,.!?:”\’-]’, ’’ ,csv)

dishes\_name = csv.split()

dishes\_name\_sq={}

For i in dishes\_name:

If i not in dishes\_name\_sq.keys():

dishes\_name\_sq[i]=1

else:

dishes\_name\_sq[i] += 1

res=sorted(dishes\_name\_sq.items(),key=lambda x:x[1],reverse=true)

print(res)

**1.3 构建热销度评分指标**



：某项菜品的销售份数

：该企业最近30天内有销售记录的菜品中的最大销售份数

：该企业最近30天内有销售记录的菜品中的最小销售份数

被下单数最多菜品：

data.groupby(by=’dishes\_name’).agg({‘counts’:sum}).idxmax()

被下单数最少菜品：

data.groupby(by=’dishes\_name’).agg({‘counts’:sum}).idxmin()

计算营业额：

sum(data[‘counts’]\*data[‘counts’])

**1.4绘制条形图展示热销TOP10**

import matplotlib.pyplot as plt

from numpy import arange

from numpy.random import rand

def gbar(ax, x, y, width=0.5, bottom=0):

X = [[.6, .6], [.7, .7]]

for left, top in zip(x, y):

right = left + width

ax.imshow(X, interpolation='bicubic', cmap=plt.cm.Blues,

extent=(left, right, bottom, top), alpha=1)

fig = plt.figure()

xmin, xmax = xlim = 0, 10

ymin, ymax = ylim = 0, 1

ax = fig.add\_subplot(111, xlim=xlim, ylim=ylim,

autoscale\_on=False)

X = [[.6, .6], [.7, .7]]

ax.imshow(X, interpolation='bicubic', cmap=plt.cm.copper,

extent=(xmin, xmax, ymin, ymax), alpha=1)

ax.set\_aspect('auto')

plt.show()

**任务二：数据预处理**

**2.1统计每个订单状态的占比**

先读取订单信息表这个文件

**2.2选取有效的订单数据**

注意不需要推荐白饭，这没有意义

所以要把记录中的白饭去除

联系两张表，删除订单详情表中有但是订单表中没有的

**2.3选取主要特征**

只需要保留客户的ID和菜品的名称即可，只需保留这两个字段，其他删除即可

**任务三：模型构建准备**

**3.1将数据划分为训练集和测试集**

训练集：测试集=8：2

可以从客户ID入手，先对客户ID进行划分，而后获取每个客户的订单情况

这里只保留了点了三个菜以上的客户

**3.2构建训练集数据客户—菜品二元矩阵**

纵坐标：菜品名称

横坐标：客户ID

1代表客户点过这道菜

0代表没点过

2代表锁单

**任务四：模型构建**

算法：协同过滤推荐算法

项目中选择基于物品的协同过滤算法，个性化推荐主要分为两步：

1. 计算菜品与菜品之间的相似度
2. 根据菜品相似度和客户的历史行为给客户生成推荐列表

**4.1计算相似度矩阵**

JACCARD相似度：A∩B：同时点了A和B的人数统计出来（在做出的客户—菜品二元矩阵中进行这一操作）；A∪B：点了A或B的（还是在二元矩阵中进行）

**4.2 生成推荐列表**

P=SIM\*R

P为目标客户对所有菜品的感兴趣程度

SIM为所有菜品之间的相似度

R代表了客户对物品的兴趣

R只存在0或1，代表客户是否用餐

先生成推荐列表，再推荐前多少名菜品即可

**任务五：模型评价**

**5.1 构建测试集客户IP—菜品字典**

**5.2构建评价指标，并分析推荐效果**

根据正确推荐的菜品数和所有菜品数的比值，来计算所有菜品推荐准确率

准确率