



泰迪智能科技

TipDM Intelligent Technology

# 连锁便利店商品滞销分析



广东泰迪智能科技股份有限公司

专注于大数据&人工智能技术研发及知识传播

[www.tipdm.com](http://www.tipdm.com)

# Part1 背景与需求分析

随着我国一线、超一线城市的零售企业经营成本日渐趋高，业态的逐步缩小已经成为了零售行业求生的法则。但是大部分零售企业或经营者在业态的转化过程中，思考习惯并没有随之变化，产生了诸如库存过量、商品滞销、过期损耗等问题。我市某连锁便利店便存在这样的问题。



**库存过量、商品滞销可能是由哪些原因产生？**

- ◆ 订货不准确
- ◆ 季节变换
- ◆ 竞争对手变多
- ◆ 周围环境
- ◆ 商品陈列位置不当
- ◆ 店员不卖力
- ◆ 缺乏促销手段
- ◆ .....

# 常见的处理滞销商品的方式有哪些？

- ◆ 打折促销
- ◆ 改变陈列位置
- ◆ 店员推销
- ◆ 店铺调配
- ◆ 内部解决
- ◆ .....

# 如果你是业务方，你的需求是什么？



- ◆ 分析商品滞销的原因。
- ◆ 为解决商品滞销提供意见。

业务方提供了1326条商品销售记录，包含以下字段。

字段	描述
商品代号	商品的代号
销售月份	商品销售的月份，如April、August等
销售点类型	商品销售点的类型，如CBD店、社区店
销售额（万元）	商品的销售额，单位：万元

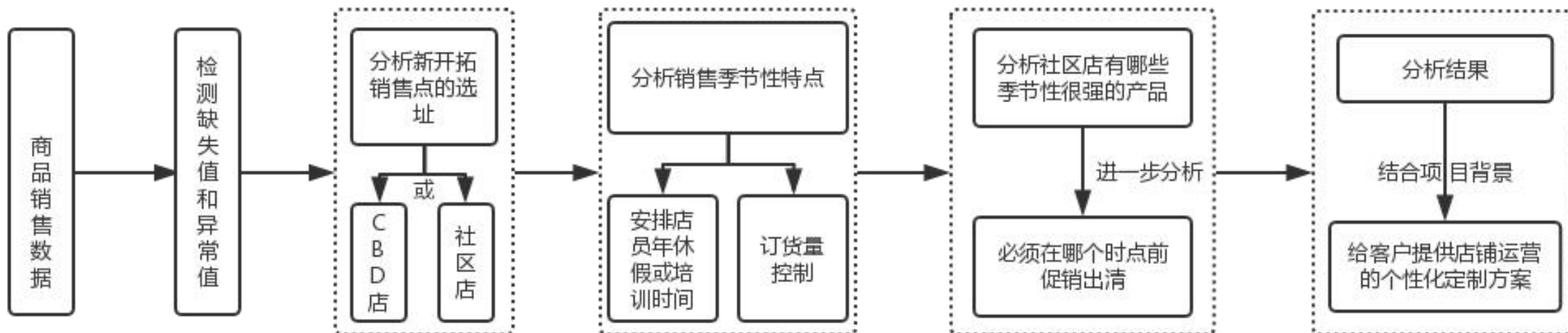
**根据业务需求和数据概况，  
确定项目分析目标和分析步骤。**

本项目根据商品的销售数据，分析不同店铺地址、不同销售季节和不同产品的销售数据，给予该连锁便利店铺货支持和经营策略建议。

连锁便利店商品滞销分析的主要分析步骤如下。

- ◆ 导入相关并读取数据，对数据进行初步理解。
- ◆ 检查数据是否存在缺失或异常，如有则进行处理。
- ◆ 将数据按照CBD店和社区店划分，并对其进行分析，确定选址。
- ◆ 把数据按照季度、月份进行划分，分析后得出安排员工休假及培训时间，同时给出控制订货量方案。
- ◆ 对数据进行可视化分析，选出季节性强的商品并对其进行进一步分析，确定促销时点。
- ◆ 分析以上得出的结论，总结出个性化定制方案提供给客户。

连锁便利店商品滞销分析的总体流程如下。



# Part2 数据探索和预处理

## 导入常用的Python第三方库

- numpy : 常用的数值计算库
- matplotlib : 常用的绘图可视化库
- pandas : 常用的结构化数据分析库
- scikit-learn(sklearn) : 常用的机器学习算法库



## 读写文本文件

### 1. 文本文件读取

- 文本文件是一种由若干行字符构成的计算机文件，它是一种典型的顺序文件。
- csv是一种逗号分隔的文件格式，因为其分隔符不一定是逗号，又被称为字符分隔文件，文件以纯文本形式存储表格数据（数字和文本）。

## 读写文本文件

### 1.文本文件读取

`pandas.read_csv(filepath_or_buffer, sep=',', header='infer', names=None, index_col=None, dtype=None, engine=None, nrows=None)`

参数名称	说明
filepath	接收string。代表文件路径。无默认。
sep	接收string。代表分隔符。read_csv默认为“,”，read_table默认为制表符“[Tab]”。
header	接收int或sequence。表示将某行数据作为列名。默认为infer，表示自动识别。
names	接收array。表示列名。默认为None。
index_col	接收int、sequence或False。表示索引列的位置，取值为sequence则代表多重索引。默认为None。
dtype	接收dict。代表写入的数据类型（列名为key，数据格式为values）。默认为None。
engine	接收c或者python。代表数据解析引擎。默认为c。

## 读写文本文件

### 1. 文本文件读取

- sep参数是指定文本的分隔符的，如果分隔符指定错误，在读取数据的时候，每一行数据将连成一片。
- header参数是用来指定列名的，如果是None则会添加一个默认的列名。
- encoding代表文件的编码格式，常用的编码有utf-8、utf-16、gbk、gb2312、gb18030等。如果编码指定错误数据将无法读取，IPython解释器会报解析错误。

## 读写文本文件

### 2. 文本文件储存

文本文件的存储和读取类似，结构化数据可以通过pandas中的to\_csv函数实现以csv文件格式存储文件。

*DataFrame.to\_csv(path\_or\_buf=None, sep=',', na\_rep="", columns=None, header=True, index=True, index\_label=None, mode='w', encoding=None)*

参数名称	说明	参数名称	说明
path_or_buf	接收string。代表文件路径。无默认。	index	接收boolean，代表是否将行名（索引）写出。默认为True。
sep	接收string。代表分隔符。默认为“,”。	index_labels	接收sequence。表示索引名。默认为None。
na_rep	接收string。代表缺失值。默认为“”。	mode	接收特定string。代表数据写入模式。默认为w。
columns	接收list。代表写出的列名。默认为None。	encoding	接收特定string。代表存储文件的编码格式。默认为None。
header	接收boolean，代表是否将列名写出。默认为True。		

## 读写Excel文件

### 1.Excel文件读取

pandas提供了read\_excel函数来读取 “xls”、“xlsx”两种Excel文件。

```
pandas.read_excel(io, sheet_name=0, header=0, index_col=None, names=None, dtype=None)
```

参数名称	说明
io	接收string。表示文件路径。无默认。
sheet_name	接收string、int。代表excel表内数据的分表位置。默认为0。
header	接收int或sequence。表示将某行数据作为列名。默认为infer，表示自动识别。
names	接收int、sequence或者False。表示索引列的位置，取值为sequence则代表多重索引。默认为None。
index_col	接收int、sequence或者False。表示索引列的位置，取值为sequence则代表多重索引。默认为None。
dtype	接收dict。代表写入的数据类型（列名为key，数据格式为values）。默认为None。

## 读写Excel文件

### 2.Excel文件储存

- 将文件存储为Excel文件，可以使用to\_excel方法。其语法格式如下。

*DataFrame.to\_excel(excel\_writer=None, sheet\_name='sheet1', na\_rep='', header=True, index=True, index\_label=None, mode='w', encoding=None)*

- 与to\_csv方法的常用参数基本一致，区别之处在于指定存储文件的文件路径参数名称为excel\_writer，并且没有sep参数，增加了一个sheet\_name参数用来指定存储的Excel sheet的名称，默认为sheet1。

在pandas中，可以使用以下方法查看数据情况。

- 查看数据结构。

*DataFrame.shape*

- 查看数据前n行。

*DataFrame.head()*

- 查看数据后n行。

*DataFrame.tail()*

- 查看数据字段类型。

*DataFrame.dtypes*

- 在pandas中，经常用describe()方法来展示数据的一些描述性统计信息。

*DataFrame.describe()*

- 返回的结果提供了数据集的基本统计指标，如数据个数、均值、标准差、最小值、最大值、中位数和四分位数等。



- 我们读取进来的数据有时需要修改列名。

```
DataFrame.columns = ['x1', 'x2', 'x3', 'x4']
```

- ◆ 1、导入相关Python库
- ◆ 2、读取 “商品销售数据.xlsx” 文件到Python中。
- ◆ 3、对读取进来的数据进行情况查看。
- ◆ 4、对读取进来的数据进行描述性统计分析。
- ◆ 5、对读取进来的数据进行列名修改，修改为 'good' , 'month' , 'sale\_place' , 'sale\_volume' 。

*pandas.read\_excel(io, sheet\_name=0, header=0, index\_col=None, names=None, dtype=None)*

*DataFrame.shape , DataFrame.head() , DataFrame.tail() , DataFrame.dtypes*

*DataFrame.describe()*

*DataFrame.columns = ['x1', 'x2', 'x3', 'x4']*

在商品销售数据中，因为CBD店与社区店所销售的商品种类繁多，为了避免某些缺失值和异常值对分析带来的影响，所以需要就数据进行预处理过程。

- ◆ 首先检查数据中是否存在缺失值的情况。
- ◆ 其次运用箱线图法，以上下四分位作为参考，计算上下边界，判断是否存在异常值的情况。如果样本不在上下边界范围内，则被定义为异常值。

上边界： $UpperLimit = Q_3 + 1.5IQR = 75\%分位数 + (75\%分位数 - 25\%分位数) * 1.5$

下边界： $LowerLimit = Q_1 - 1.5IQR = 25\%分位数 - (75\%分位数 - 25\%分位数) * 1.5$

- ◆ 若存在空值与异常值的情况，则将进一步做替换值处理，若不存在空值与异常值的情况，则进行下一步的数据分析处理。

- 在pandas中，提供了识别缺失值的方法isna()以及识别非缺失值的方法notna()，这两种方法在使用时返回的都是布尔值True和False。

*DataFrame.isna()*

*DataFrame.notna()*

- 对于得到的结果，可以使用sum()方法来统计数据中一共含有多少缺失值。

*DataFrame.isna().sum()*

- isna()和notna()之间结果正好相反，因此使用其中任意一个都可以判断出数据中缺失值的位置。

➤ 在matplotlib.pyplot中，提供了绘制箱线图的方法。

```
matplotlib.pyplot.boxplot(x, notch=None, sym=None, vert=None, whis=None, positions=None, widths=None,
patch_artist=None, bootstrap=None, usermedians=None, conf_intervals=None, meanline=None, showmeans=None,
showcaps=None, showbox=None, showfliers=None, boxprops=None, labels=None, flierprops=None,
medianprops=None, meanprops=None, capprops=None, whiskerprops=None, manage_ticks=True, autorange=False,
zorder=None, *, data=None)
```

参数名称	说明	参数名称	说明
x	指定要绘制箱线图的数据。	widths	接收float。指定箱线图的宽度。默认为0.5。
notch	接收boolean。表示是否是凹口的形式展现箱线图。默认为False。	patch_artist	接收boolean。表示是否填充箱体的颜色。默认为False。
sym	接收string。指定异常点的形状。默认为 “+” 。	labels	表示为箱线图添加图例。默认为None。
whis	接收float。指定上下边界与上下四分位的距离。默认为1.5倍四分位差。	vert	接收boolean。表示是否需要将箱线图垂直摆放。默认为True。

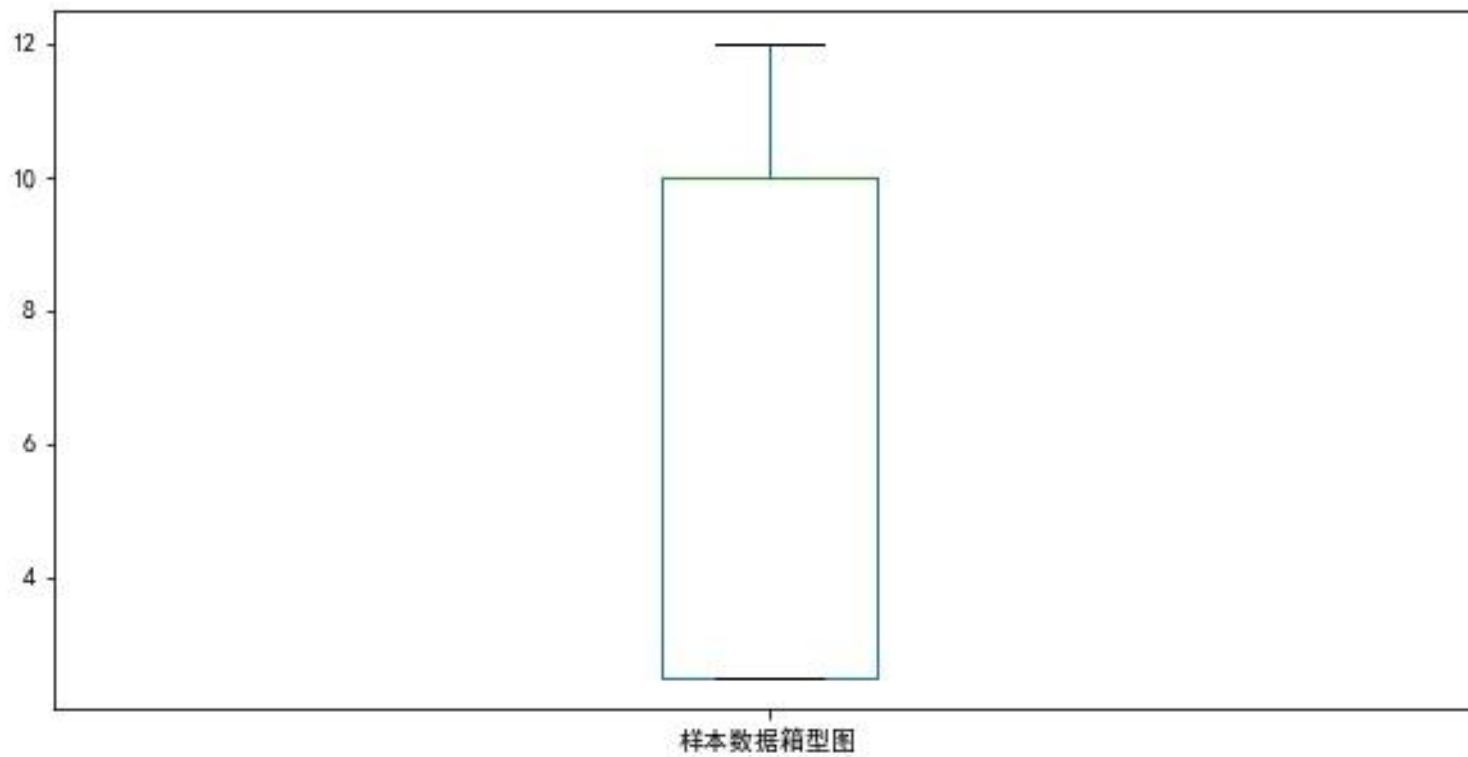
- ◆ 1、对读取进来的数据进缺失值检查。
- ◆ 2、对读取进来的数据进异常值判断

*DataFrame.isna() , DataFrame.notna()*

*DataFrame.isna().sum()*

*matplotlib.pyplot.boxplot(x, notch=None, sym=None, vert=None, whis=None, positions=None, widths=None, patch\_artist=None, bootstrap=None, usermedians=None, conf\_intervals=None, meanline=None, showmeans=None, showcaps=None, showbox=None, showfliers=None, boxprops=None, labels=None, flierprops=None, medianprops=None, meanprops=None, capprops=None, whiskerprops=None, manage\_ticks=True, autorange=False, zorder=None, \*, data=None)*

- 对数据进行预处理并绘制箱线图可知，数据中不存在缺失值和异常值，因此不需要对数据进行额外处理，可直接进行下一步的数据分析阶段。



# Part3 统计分析与可视化

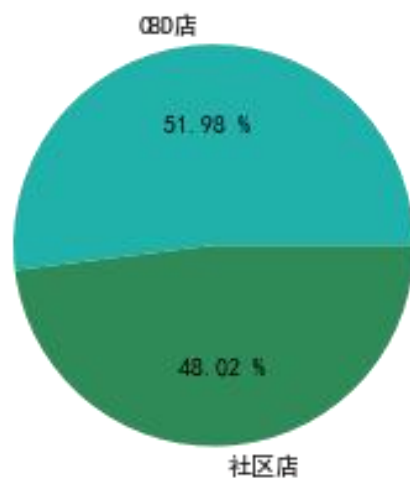


- 新开拓销售点的选址，直接关系到一个店的整体盈利，正确选址店铺地址，在一定程度上可以达到事半功倍的效果。
- 根据已有的数据，可以得到备选址方案有CBD店与社区店，那么将利用已有的数据，从CBD店和社区店的总体利益出发，通过比较两个店之间的总销售额、各商品的销售情况以及CBD店和社区店的收益稳定性，最终从优选址，得到心仪店铺地址。

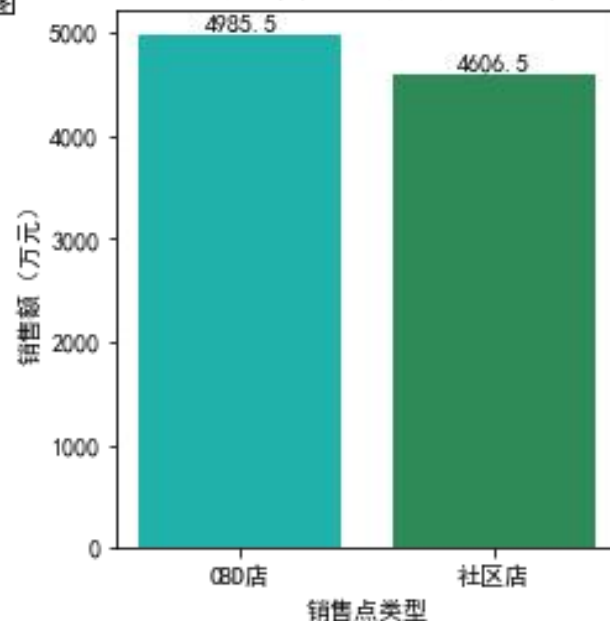
## CBD店与社区店总销售额比较

- 店铺总销售额的高低是衡量该店铺收益情况的重要标准，同时也是从优选择店铺地址的重要组成部分。于是先从CBD店与社区店的总销售额出发，通过比较两个店之间的总销售额，从而判断两店间的总体收益情况。
- CBD店总体销售额为4985.50（万元），社区店的总体销售额为4606.50（万元）。同时可以得出，在相同时时间段里，两店之间的总体销售额之差为379（万元）。

不同店铺类型的销售额占比情况展示图

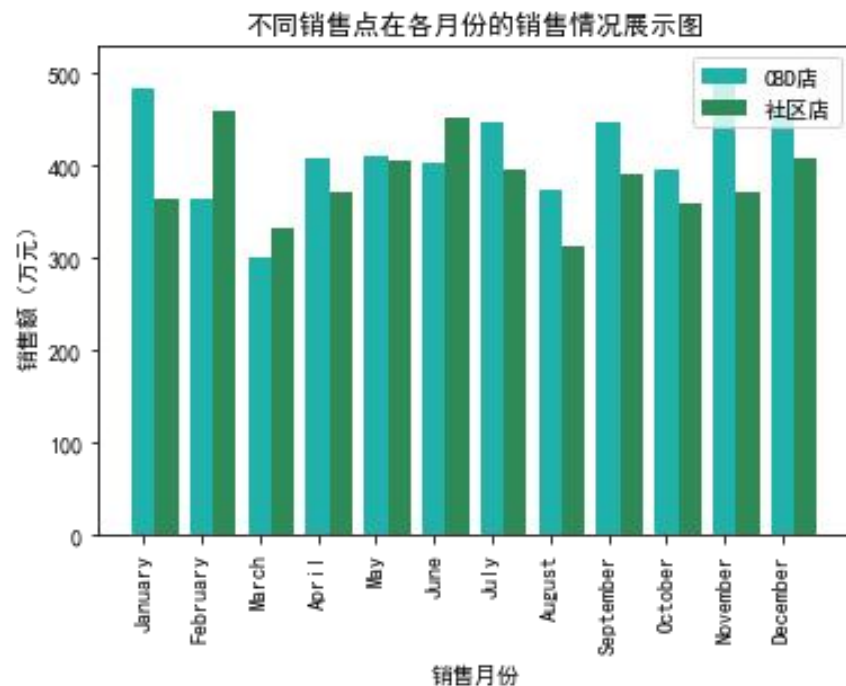


不同店铺类型的销售额汇总情况



## CBD店与社区店月销售额比较

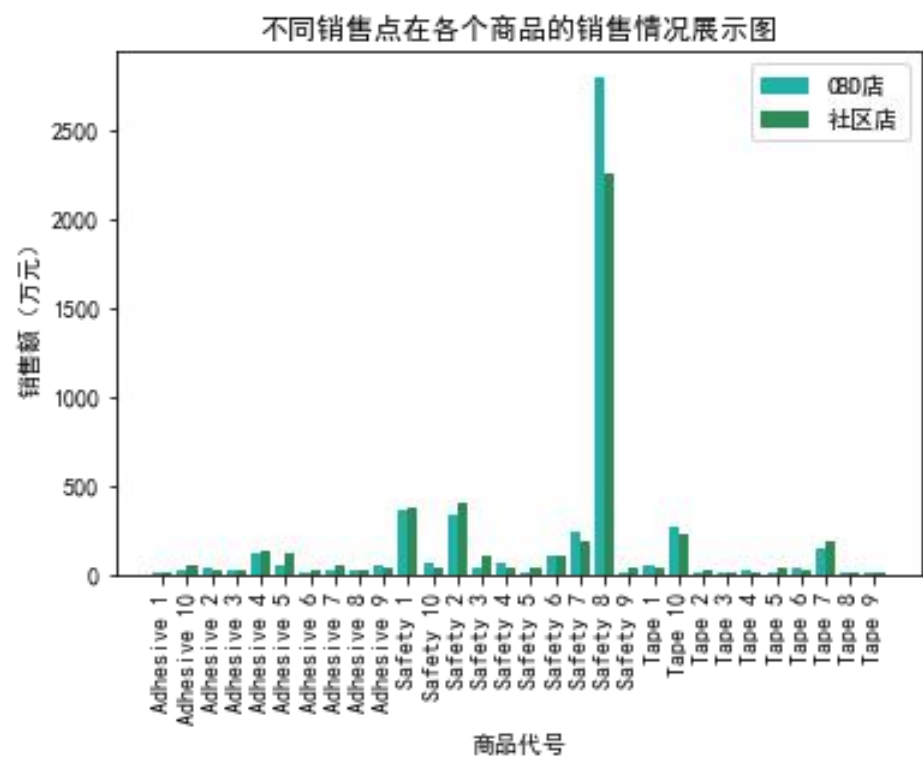
- 通过对CBD店与社区店月销售额进行统计分析，一方面可以精确得到其在各个月份总的销售情况和变化趋势，另一方面则是通过各个月得销售额从而计算每个店在不同月份得销售额的偏差大小，从而体现该店每个月销售额稳定性。
- 由CBD店与社区店的月销售额柱状图可知，CBD店总体上每个月的销售额均比社区店销售额高。



- 与此同时，在计算的CBD店与社区店月销售额的基础上，对CBD店与社区店的销售额数据进行统计分析，其中包括月销售额均值、标准差、最大销售额等。
- 在月份最高销售额中，CBD店高达503.50（万元），而社区店为457.00（万元），单在一个月的销售额上CBD店高出46.50（万元），从这一方面上看，CBD店的可选择性更强；在月销售额均值上，CBD店月销售额均值为415.46（万元），社区店为383.88（万元），其中月销售额均值CBD店比社区店大31.58（万元），从这一方面上看，CBD店的盈利效果更好。

## CBD店与社区店不同商品的销售额对比

- 不同店的各个商品的销售量各有不同，对不同商品的销售额进行对比，有利于进一步地分析。
- 大部分商品在CBD店与社区店的销售额差别不是很大，但是商品“Safety 8”在CBD店的销售额相对于在社区店的销售额相差较大。



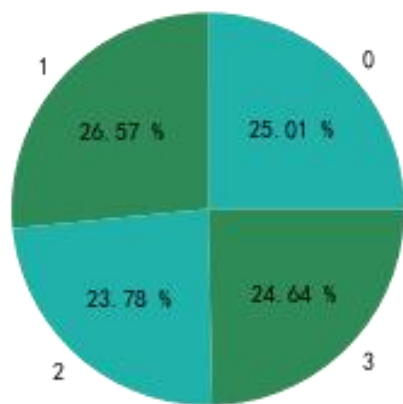
综合上面的分析，CBD店无论是在年销售总额、月均销售总额还是最高销售额以及各个商品的销售额上，其盈利程度都比社区店略胜一筹，因此，对于新开拓销售店地址的选取，选择CBD店将会更有利。

- 年休假是国家根据劳动者工作年限和劳动繁重紧张程度每年给予的一定期间的带薪连续休假。员工培训是组织人力资源资产增值的重要途径，也是企业组织效益提高的重要途径。只有合理的安排员工年休假和培训时间才能最大程度的调动员工的积极性与最大程度的发挥员工的能力进行工作。另外合理的订货量能最大程度的获取利润和减少货物的消极堆积。
- 根据已有的数据，通过分析各季节销售额的情况，安排员工休假，根据各个季节各个月份销售额的情况安排店员培训，最后按照所有商品的季度销售额情况，合理的安排订货量。

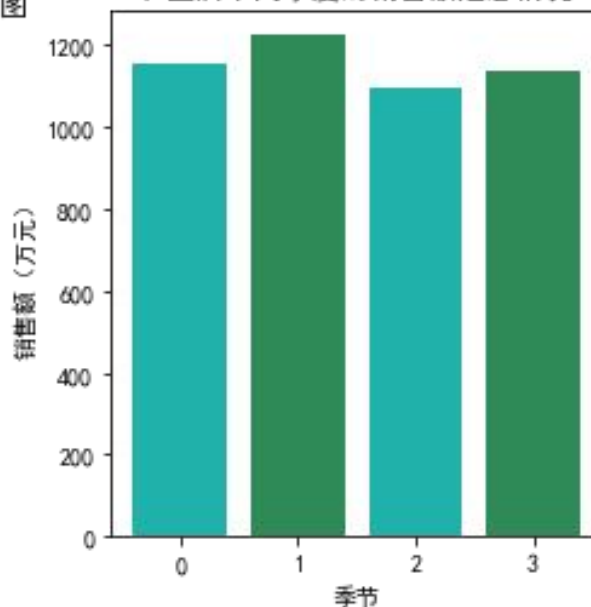
## 店员年休假的安排

- 分析各季度销售的情况，在忙碌的季节安排少人年休假，在不忙碌的季节安排多人年休假。
- CBD店在上半年相对空闲，社区店则在下半年相对空闲。

社区店不同季度的销售额占比情况展示图



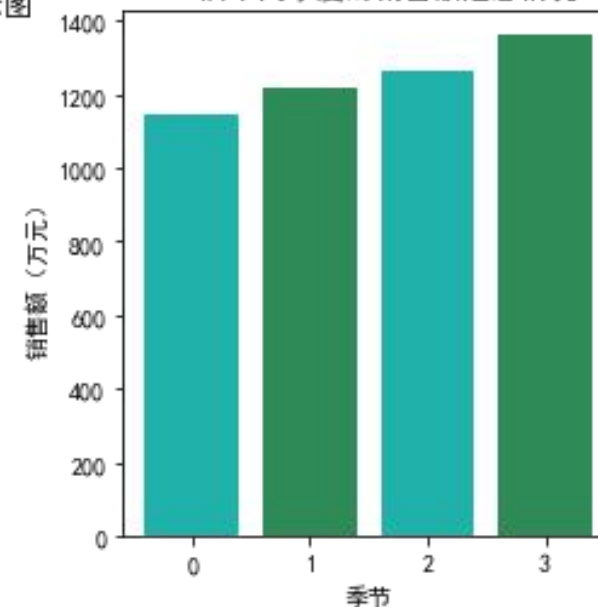
社区店不同季度的销售额汇总情况



CBD店不同季度的销售额占比情况展示图



CBD店不同季度的销售额汇总情况

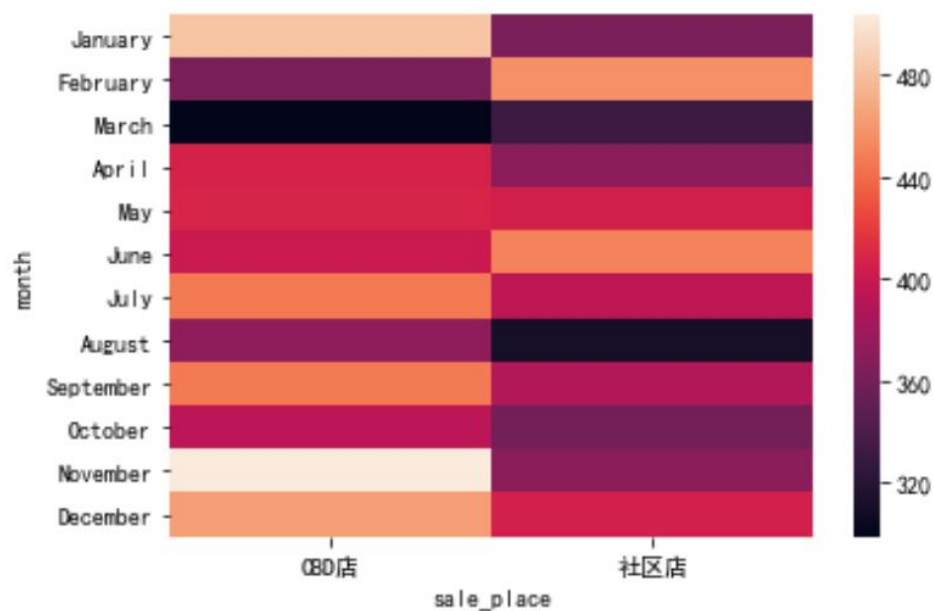




## 店员培训时间的安排

- 绘制各个月份销售额热力图，分析各月份销售额的情况以及变化趋势，以此来安排培训时间。

- 由绘制出CBD各个月份销售额热力图可知，第一季度的二三月份、第三季度的八月份销售额较低，且在上述月份之后的月份销售额相较于本月都有明显的提升。
- 由绘制出社区店各个月份销售额热力图可知，第一季度的三月份、第三季度的八月份销售额较低，且在上述月份之后的月份销售额相较于本月都有明显的提升。
- 而员工培训的目的是为了更好地销售，因此在上述月份进行员工培训，既能避开员工人手不足的问题，也能为后面月份销售额的大幅度提升做好准备。

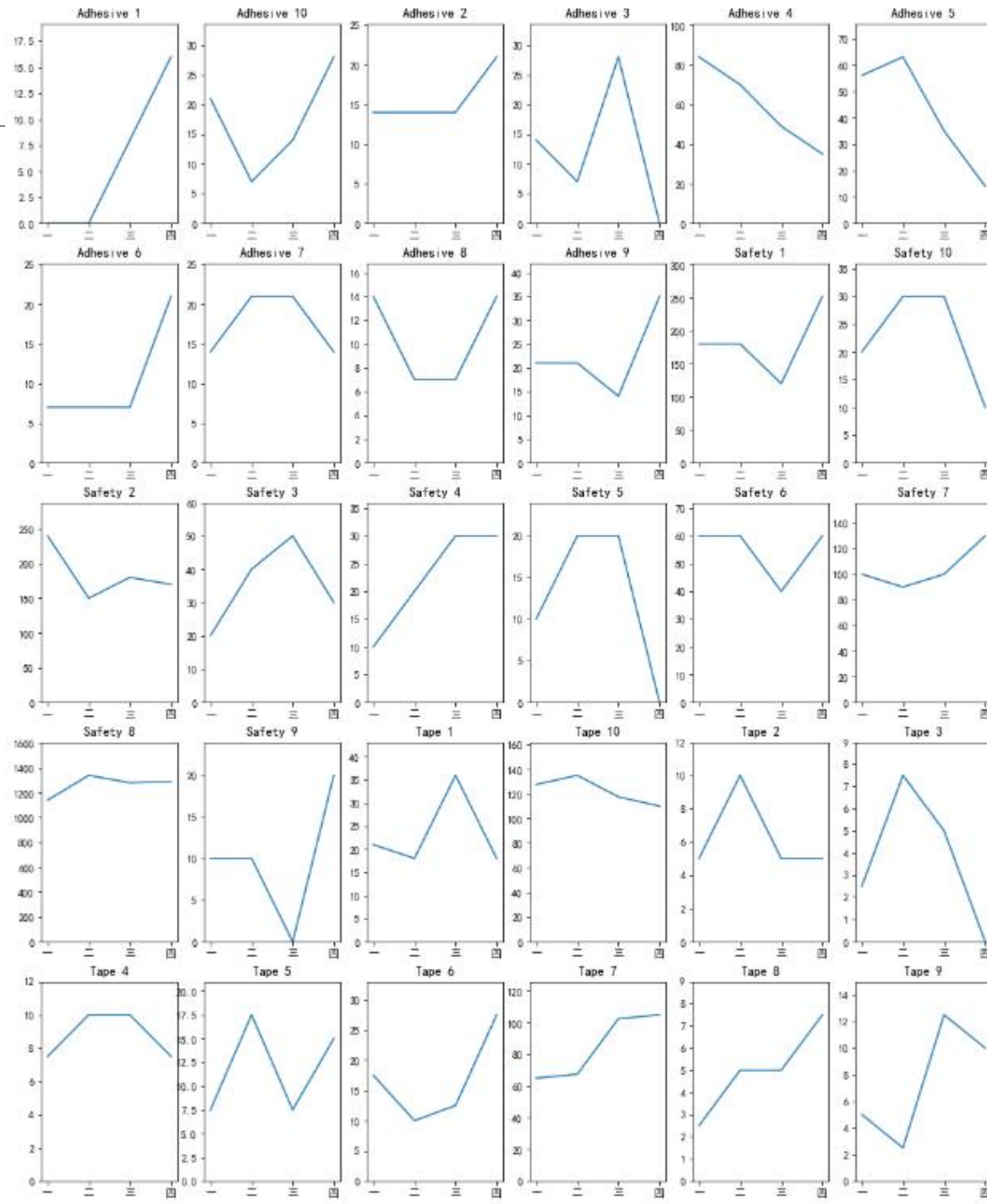


## 订货量的安排

- 绘制各个商品的季度销售额的折线图，分析每个商品各个季度具体的销售额的情况以及销售额的变化趋势，从而适当的增加或减少订货量。
- 销售额变高之前提高订货量避免缺货，销售额变低之前减少订货量避免货物冗余。

## 分析销售季节性特点

- 每个商品的季度销售额都不一致，需要根据各个商品的销售额的情况，进行相应的增加或减少该商品的订货量。



## 分析滞销商品和畅销商品

- 根据销售情况，进行相应的订货量的控制。因为有些商品在某些季节是热销商品，在某些季节是滞销商品，所以需要根据销售情况进行合理的安排订货，如在第四季度商品总销售额明显大于其他季度，因此对应上在第四季度上销售量明显增多的商品，为此可以增大该商品的订货量以供购物者的需求。

	滞销商品	畅销商品
season		
0	[Adhesive 1, Tape 3, Tape 8, Tape 2, Tape 9]	[Safety 7, Tape 10, Safety 1, Safety 2, Safety 8]
1	[Adhesive 1, Tape 9, Tape 8, Adhesive 3, Adhes...]	[Safety 7, Tape 10, Safety 2, Safety 1, Safety 8]
2	[Safety 9, Tape 3, Tape 2, Tape 8, Adhesive 6]	[Tape 7, Tape 10, Safety 1, Safety 2, Safety 8]
3	[Adhesive 3, Tape 3, Safety 5, Tape 2, Tape 4]	[Tape 10, Safety 7, Safety 2, Safety 1, Safety 8]

# Part3 个性化定制方案

- 根据分析结果，可以从选址、主要铺货产品、进货、补货、促销和员工休假与培训等方面制定个性化方案，具体介绍如下。

## 选址

- 建议客户选择开设CBD店，CBD店的销售总额比社区店整整多出了300多万元，利润可观。

## 主要铺货产品

- “Adhesive 2” “Safety 4” “Safety 7” “Safety 8” “Safety 10” “Tape 10” 等商品在CBD店的销售额明显高于在社区店的销售额，在不同地区不同店铺中销售这类商品比较有利，因此可以选择这些商品作为主要铺货产品。

## 进货

- 按照销售额的大小决定每种商品的进货量，如CBD店销售额最高的“Safety 8”商品，销售额远远超过了其他商品，那么订货的时候即可适当的增加该商品的进货量。
- 对于其他商品，可以通过逐级递减方式调整进货量，这样可以有效地最大化盈利。
- 对于季度进货量的安排，根据CBD店各个商品的季度销售数量折线图可知销售额的季度变化。假设在每个季度初进货，那么根据折线图表示的销售额的大小，结合上个季度的库存，相应地改变进货量的多少。
- 销售额大，进货量就增多；销售额小，进货量就减少。



## 补货

- 在销售过程中，难免会发生商品库存不足的情况，这个时候就需要进行补货。
- 进货是在每个季度初进货，若是在第一个月结束后发现当月销售额比数据显示的要大，那么就可以进行补货，补货的数量不用多，但是要确保充足；第二个月结束后，统计第一、二季度销售额的总和与数据进行比较，以此来确定是否需要补货。

## 促销

- 在销售过程中，若是商品销售情况不好，且商品季节性强，就有可能导致商品冗余，那么需要进行促销活动来减少因商品过期而导致的损失。
- 对于促销时点的安排，建议安排在商品销售额最高的季度与商品销售额最低的季度之间，这样既能避免影响热销季度的盈利，又能避免由于淡季导致滞销。
- 若商品销售额最高与最低的季度是连续的，则在最高的季度末到最低的季度初这段时间安排促销；若不是连续的，则在最高、最低两个季度之间的季度安排促销。

## 员工休假与培训

- 对于休假，可以按照季度销售额倒数之比来安排相应比例的员工的休假，这样可以避免出现人手不足的情况。
- 对于培训，为了确保人手充足，可以选择销售额较低的月份进行培训，而培训的目的是为了提高员工的能力来更好地进行销售，因此可以在销售额较高的月份前选择销售额较低的月份进行培训。



# Thanks

广东泰迪智能科技股份有限公司

[www.tipdm.com](http://www.tipdm.com)

热线：40068-40020

电话：020-22205718

网址：[www.tipdm.com](http://www.tipdm.com)

邮箱：[services@tipdm.com](mailto:services@tipdm.com)

地址：广州市黄埔区开泰大道36号