连锁便利店商品滞销分析项目报告

项目报告作者：刘付雄

时间：2023年10月11日

目录

[1. 绪论 2](#_Toc147918101)

[1.1 项目背景 2](#_Toc147918102)

[1.2 问题描述 3](#_Toc147918103)

[2. 数据预处理 3](#_Toc147918104)

[2.1 数据预处理流程 3](#_Toc147918105)

[2.2 缺失值检测 3](#_Toc147918106)

[2.3 异常值检测 3](#_Toc147918107)

[2.3 特征处理 4](#_Toc147918108)

[3. 数据分析 5](#_Toc147918109)

[3.1 不同商品及各类商品销售情况 5](#_Toc147918110)

[3.2 不同月份的商品销售情况 6](#_Toc147918111)

[3.3 不同季度的商品销售情况 7](#_Toc147918112)

[3.4 不同销售点的销售情况 8](#_Toc147918113)

[4. 营销策略与建议 9](#_Toc147918114)

# 绪论

## 1.1 项目背景

连锁便利店作为零售行业的重要一环，面临着商品滞销的问题。商品滞销不仅会导致货物积压和库存成本增加，还会影响店铺的销售额和利润率。因此，连锁便利店需要通过数据分析，深入挖掘商品销售数据，找出原因并提出解决方案，进而调整商品结构和销售策略，更好地节约成本、提升销售额和市场竞争力。这对于提升连锁便利店的经营效益具有重要意义。

## 1.2 问题描述

根据某连锁便利店商品销售数据，解决以下问题：

1. 通过对数据进行分析，得到商品的销售情况、影响商品销售量的各因素，包括商品种类、销售的月份和季节、销售点类型等。
2. 根据数据分析结果，制定相应的销售策略优化方案以及建议。

# 数据预处理

## 2.1 数据预处理流程

**图1 数据预处理流程图**

## 2.2 缺失值检测

对某连锁便利店的商品销售的数据集作整体的缺失值检测与处理：

对于整体数据集中的所有数据框的缺失情况，我们用pandas库下DataFrame类中的isnull()和sum()的两个方法，调用查看各指标缺失值的总和，结果显示数据集的所有列均无缺失值，具体缺失情况如表1所示。

**表1 各特征缺失情况记录**

|  |  |
| --- | --- |
| **数据列** | **是否缺失** |
| 商品代号 | False |
| 销售月份 | False |
| 数据点类型 | False |
| 销售额（万元） | False |

上述缺失值检测所涉及代码：

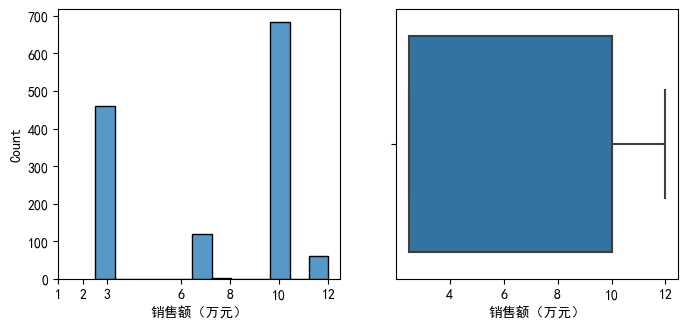
|  |
| --- |
| #检测缺失值  df.isnull().sum() |

输出结果如下：

|  |
| --- |
| 商品代号 0  销售月份 0  销售点类型 0  销售额（万元） 0  dtype: int64 |

## 2.3 异常值检测

本文考虑将数据中个别数值明显偏离其余值的数值划分为异常值视为异常值。如图2销售额的分布以及箱线图所示，销售额基本分布在3、7、10、12万元，从箱线图上可以看出，并没有异常值出现，因此无需处理异常值。



**图2 销售额分布直方图及箱线图**

上述异常值检测所涉及代码：

|  |
| --- |
| #销售额分布  plt.figure(figsize=(8,3.5))  plt.subplot(121)  g = sns.histplot(data=df,x="销售额（万元）")  g.set\_xticks([1,2,3,6,8,10,12])  plt.subplot(122)  sns.boxplot(data=df,x="销售额（万元）") |

## 2.3 特征处理

为了更好对商品销售数据进行数据分析，我们对一些数据特征作如下处理：

（1）针对商品代号特征，我们发现这些商品的前缀英文名称有三类，因此将商品分为三大品类，具体如下表2：

**表2 各商品的商品大类**

|  |  |
| --- | --- |
| **商品代号** | **商品大类** |
| Tape 1、Tape 2、… 、Tape 10 | Tape |
| Safety 1、Safety 2、… 、Safety 10 | Safety |
| Adhesive 1、Adhesive 2、… 、Adhesive 10 | Adhesive |

（2）季度：通过特征“销售月份”我们可以添加销售季度特征，1~3月为一季度，4~6月为第二季度，7~9月为第三季度，10~12月为第四季度，集体如下表3所示：

**表3 月份转换季度形式**

|  |  |
| --- | --- |
| **销售月份** | **季度** |
| January、February、March | 一季度 |
| April、May、June | 二季度 |
| July、August、September | 三季度 |
| October、November、December | 四季度 |

最后，为了更加直观得分析数据，我们将销售月份英文名称转换为数值1~12的形式。

上述预处理所涉及代码：

|  |
| --- |
| #添加商品类别列  def date\_to\_season(good):  if good.startswith('Adhesive'):  return 'Adhesive'  elif good.startswith('Safety'):  return 'Safety'  else:  return 'Tape'  df['商品类别'] = df['商品代号'].apply(date\_to\_season)  #增加季季度列  def date\_to\_season(month):      if month in('February','January','March'):          return '一季度'      elif month in('June','April','May'):          return '二季度'      elif month in('September','July','August'):          return '三季度'      else:          return '四季度'  df['季度'] = df['销售月份'].apply(date\_to\_season)  import calendar  list(calendar.month\_name).index('December')  df['月份num'] = pd.to\_datetime(df['销售月份'],format='%B').dt.strftime('%m')  df.head() |

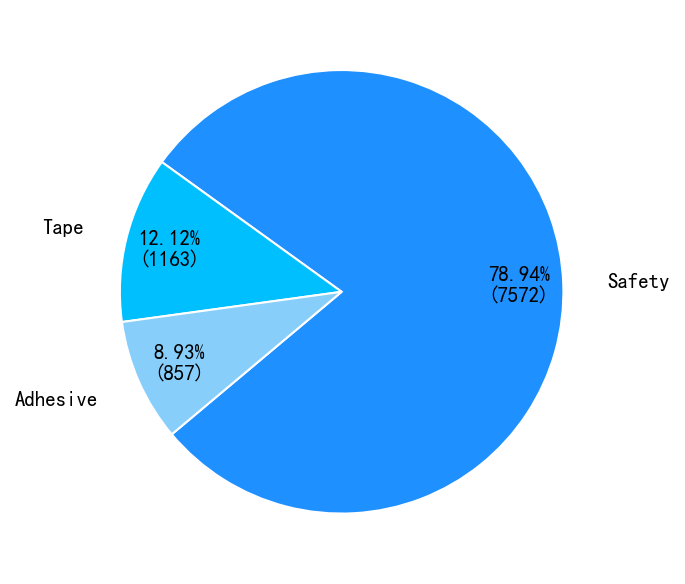
输出结果如下：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 商品代号 | 销售月份 | 销售点类型 | 销售额（万元） | 季度 | 月份num | 商品类别 |
| 0 | Tape 10 | April | CBD店 | 2.5 | 二季度 | 04 | Tape |
| 1 | Safety 8 | August | 社区店 | 10.0 | 三季度 | 08 | Safety |
| 2 | Safety 2 | February | 社区店 | 10.0 | 一季度 | 02 | Safety |
| 3 | Safety 8 | November | 社区店 | 10.0 | 四季度 | 11 | Safety |
| 4 | Tape 10 | October | 社区店 | 2.5 | 四季度 | 10 | Tape |

# 数据分析

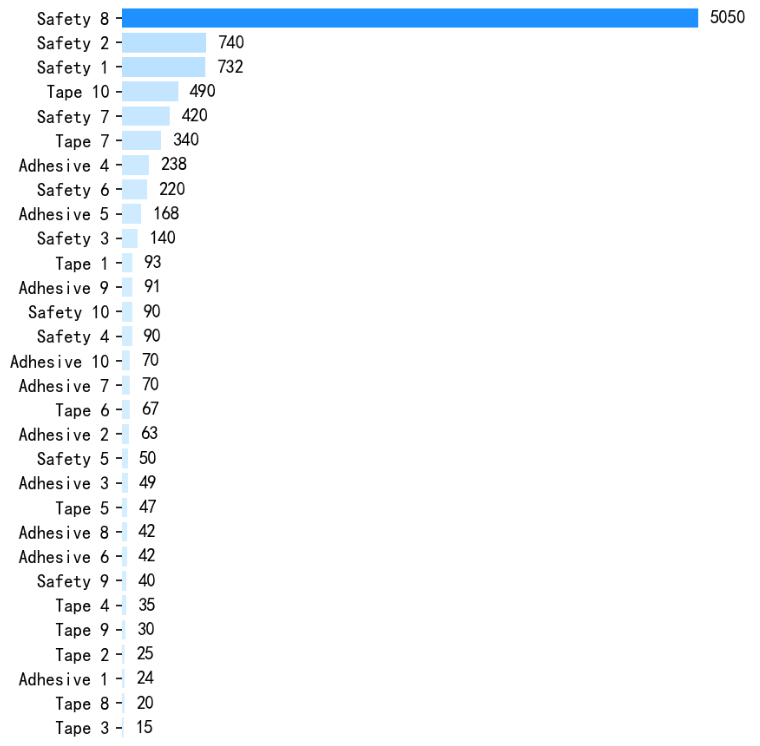
## 3.1 不同商品及各类商品销售情况

绘制出各类商品的销售占比，如图2所示，销售占比最高的是Safety类商品，占78.94%，销售额高达7572万元。



**图3 各类商品的销售占比**

进一步绘制各商品的销售额条形图，如图3所示，在三十种商品中，销售额最高是Safety品类的Safety 8，远超其它商品。由此，我们可以得出一些结论：该商品在顾客中非常受欢迎，并且有很大的市场需求；这款商品可能具有较高的利润空间，因为它的销售额高，意味着商家可以通过销售这款商品获得更多的利润。那么商家保持这款商品供应充足，确保市场上的需求得到满足。



**图4 各商品的销售额条形图**

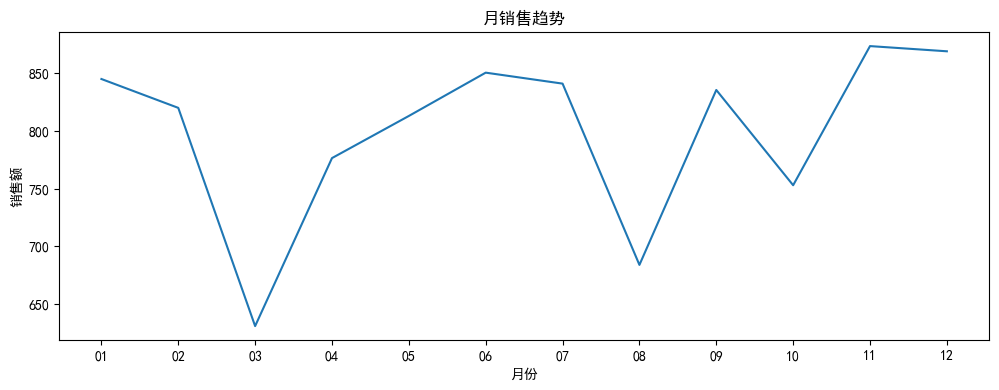
上述分析及可视化所涉及代码：

|  |
| --- |
| df\_good = df.groupby('商品类别')['销售额（万元）'].sum().reset\_index()  df\_good.sort\_values('销售额（万元）',ascending=False,inplace=True)  df\_goodId = df.groupby('商品代号')['销售额（万元）'].sum().reset\_index()  df\_goodId.sort\_values('销售额（万元）',ascending=True,inplace=True)  # plt.bar(x=0,height=0.5,bottom=df\_goodId['商品代号'],width=df\_goodId['销售额（万元）'])  plt.figure(dpi=150)  plt.pie(df\_good['销售额（万元）'],          labels=df\_good['商品类别'],          startangle=-140,          # explode=(0.01,0.01),          # autopct='%.1f%%',          # startangle=90,#旋转90度,          autopct=make\_autopct(df\_good['销售额（万元）']),          wedgeprops={'edgecolor': 'w'},          colors=['#1E90FF','#00BFFF','#87CEFA'],          labeldistance=1.2,#标签放在里面          pctdistance=0.8,#百分比标签往里缩          # textprops={'fontsize':12,'color':'w'},          )  plt.show()  from matplotlib import cm  from matplotlib.colors import ListedColormap, LinearSegmentedColormap  blue\_colors = ['#D4EEFF', '#1E90FF'] # 淡蓝色到深蓝色  cmap = LinearSegmentedColormap.from\_list('my\_cmap', blue\_colors)  norm = plt.Normalize(df\_goodId['销售额（万元）'].min(), df\_goodId['销售额（万元）'].max())  norm\_y = norm(df\_goodId['销售额（万元）'])  color = cmap(norm\_y) #  plt.figure(figsize=(6,8),dpi=150)  ax=plt.gca()  #设置图片的右边框和上边框为不显示  ax.spines['right'].set\_color('none')  ax.spines['top'].set\_color('none')  ax.spines['left'].set\_color('none')  ax.spines['bottom'].set\_color('none')  plt.barh(df\_goodId['商品代号'], df\_goodId['销售额（万元）'],color=color)  for i, value in enumerate(df\_goodId['销售额（万元）']):  plt.text(value+100, i, str(int(value)), ha='left', va='center')  plt.xlabel('销售额')  plt.show() |

输出结果如图3、4。

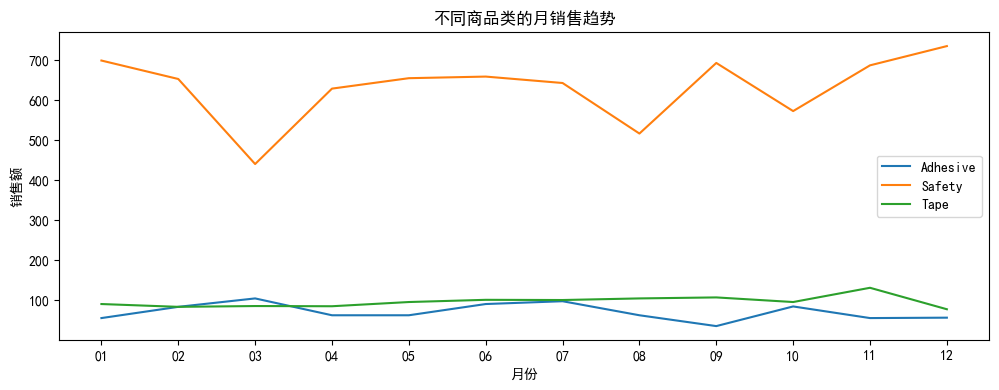
## 3.2 不同月份的商品销售情况

本文通过python的matplotlib库来绘制出各个月份的总体销售变化情况，如图2各月份总体销售折线图所示，在4~7月、11~2月销售情况较好，且比较稳定，在3月份时销售额最低，8~10月之间销售情况不稳定。这种趋势可能是3月份受某些节假日影响，如3月份刚好是在春节结束不久，一些人还并未复工复产复学。那么商家就此情况应该在3、8、10月份减少商品进货，在较为稳定的4~7月和11~2月之间加大进货量。



**图5 各月份总体销售折线图**

我们进一步对各大类商品作各月份的销售变化分析，如图3所示，发现Safety类商品各个月份的销售额远大于其它两类商品，而Adhesive类和Tape类商品的变化波动不大。由于原始数据只有商品的销售金额，并没有销量或者商品单价，所以我们并不能因为Safety类商品的销售金额高而得出其销量高的结论，那么我们只能推测：（1）Safety类商品的价格相较其它两类商品较为昂贵；（2）Safety类商品的销量远高于其它两类商品。若第二种假设成立，那么我们建议商家应加大Safety类商品的进货量，下文都以高销售金额即高销量这种假设成立为前提。通过各类商品曲线变化可以看到，Safety类与Adhesive类商品的变化趋势在许多月份之间是相反的，那么这两类商品很可能是负关联的。



**图6 不同类商品的月销售趋势**

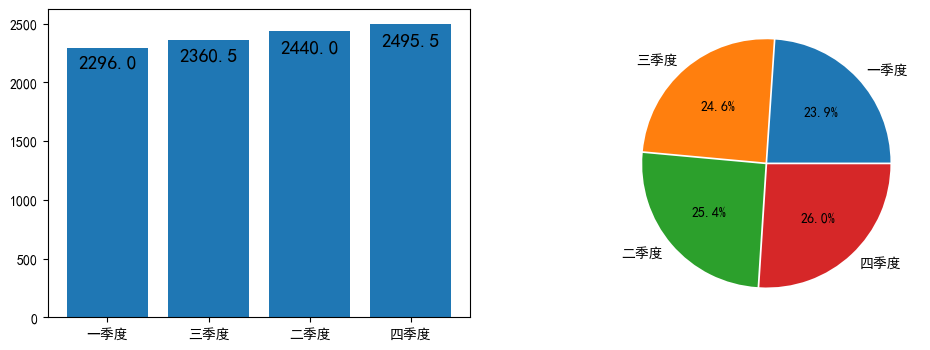
上述分析及可视化所涉及代码：

|  |
| --- |
| df\_month = df.groupby(['月份num'])['销售额（万元）'].sum().reset\_index()  df\_month\_group = df.groupby(['商品类别','月份num'])['销售额（万元）'].sum().reset\_index()  df\_month\_group = df\_month\_group.pivot(index='月份num', columns='商品类别', values='销售额（万元）')  df\_month\_group.reset\_index(inplace=True)  plt.figure(figsize=(12,4))  plt.plot(df\_month['月份num'],df\_month['销售额（万元）'],)  plt.title('月销售趋势')  plt.xlabel('月份')  plt.ylabel('销售额')  plt.show()  plt.figure(figsize=(12,4))  plt.plot(df\_month\_group['月份num'],df\_month\_group['Adhesive'],label = 'Adhesive')  plt.plot(df\_month\_group['月份num'],df\_month\_group['Safety'],label = 'Safety')  plt.plot(df\_month\_group['月份num'],df\_month\_group['Tape'],label = 'Tape')  # 设置图表标题和轴标签  plt.title('不同商品类的月销售趋势')  plt.xlabel('月份')  plt.ylabel('销售额')  plt.legend(loc=0)  plt.show() |

输出图表如图5、6。

## 3.3 不同季度的商品销售情况

绘制出各季度的总体销售额和各季度的销售额占比，如图4所示，一季度的销售额占比最低，为2296万元，四季度的销售额占比最高，为2495.5万元，第一季度到第二季度和第三季度到第四季度都是销售额上升趋势，那么第一季度可视为淡季，第四季度视为旺季，而这两个季度又相邻，那么我们建议商家应该在第一季度末和第三季度末加大商品的进货量。



**图7 各季度总体销售情况**

进一步，我们统计出各个季度的畅销商品和滞销商品，以帮助商家更加清晰明确地去采购货物。我们规定每个季节的销售额前五的商品为畅销商品，销售额倒数前五的商品为滞销商品，具体统计如下表4，畅销大部分都是Safety类，而滞销商品中基本是Adhesive类和Tape商品。

**表4 各季度的滞销商品和畅销商品**

|  |  |  |
| --- | --- | --- |
| **季度** | **滞销商品** | **畅销商品** |
| 一季度 | Adhesive 1, Tape 3, Tape 8, Tape 2, Tape 9 | Safety 7, Tape 10, Safety 1, Safety 2, Safety 8 |
| 二季度 | Adhesive 1, Tape 9, Tape 8, Adhesive 3, Adhesive 6, Adhesive 8, Adhesive 10 | Safety 7, Tape 10, Safety 2, Safety 1, Safety 8 |
| 三季度 | Safety 9, Tape 3, Tape 2, Tape 8, Adhesive 6, Adhesive 8 | Tape 7, Tape 10, Safety 1, Safety 2, Safety 8 |
| 四季度 | Adhesive 3, Tape 3, Safety 5, Tape 2, Tape 4, Tape 8 | Tape 10, Safety 7, Safety 2, Safety 1, Safety 8 |

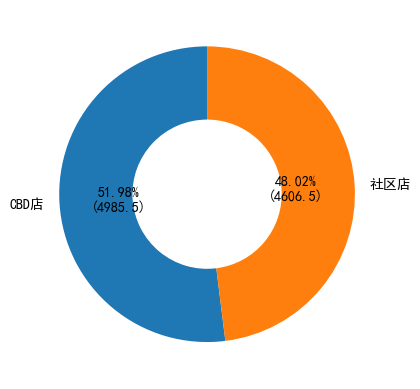
上述分析及可视化所涉及代码：

|  |
| --- |
| df\_season = df.groupby('季度')['销售额（万元）'].sum().reset\_index()  plt.figure(figsize=(12,4))  plt.subplot(121)  plt.bar(df\_season['季度'],df\_season['销售额（万元）'])  for x,y1 in enumerate(df\_season['销售额（万元）']):      plt.text(x,y1-y1/12,str(y1),fontsize=14,ha='center',va='bottom')  plt.subplot(122)  plt.pie(df\_season['销售额（万元）'],          labels=df\_season['季度'],          explode=(0.01,0.01,0.01,0.01),          autopct='%.1f%%',)  plt.show() |
| df\_season\_good = pd.pivot\_table(df,index='商品代号',columns='季度',values='销售额（万元）',aggfunc=sum).reset\_index()  df\_season\_good.fillna(0,inplace=True)  df\_season\_good.set\_index('商品代号',inplace=True)  data = pd.DataFrame([],index=df\_season\_good.columns,columns=['滞销商品','畅销商品'])  for i in df\_season\_good.columns:      data.loc[i,'滞销商品'] = list(df\_season\_good[i].sort\_values().index[:5])      data.loc[i,'畅销商品'] = list(df\_season\_good[i].sort\_values().index[-5:]) |

输出图表如图7、表4。

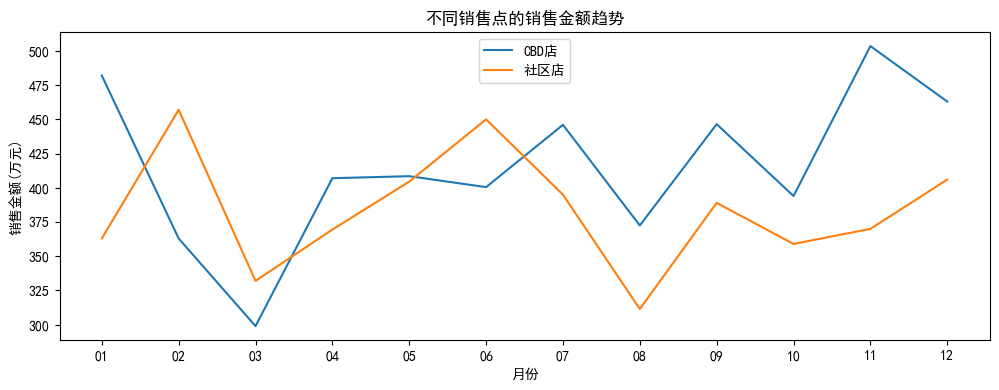
## 3.4 不同销售点的销售情况

绘制出各销售点的销售额占比，如图5所示，CBD店的销售额占比更高，占51.98%，销售额比社区店高了将近400万。那么就此情况是否要向CBD店进更多货？这还是需要进一步考究的，CBD是中心商业区，而社区就是较为偏离商业中心的地区。那么针对地理位置的不同，可能会有两种情况：（1）CBD地区便利店的商品可能会相较社区店更加昂贵，所以其销售金额会更高；（2）CBD店的销售额虽然高，但是利润可能更低，因为除了商品本身，CBD地区的各种成本都会比社区更高。所以对于商家而言，不能只追求高销售额，还要考虑获得的利润是否与进货量成正比。因为此数据集只有销售额这一指标，所以需要商家去衡量向哪个地区进更多商品货物。



**图8 各销售点总体销售情况**

再来做销售点每个月的销售额变化趋势，如图8所示，在2、3、6月份社区店的销售额比CBD店要高，那么商家要考虑在这些月份里向社区店进更多货物。



**图9 各销售点月销售情况**

上述分析及可视化所涉及代码：

|  |
| --- |
| # 不同销售点的销售金额  df\_place = df.groupby('销售点类型')['销售额（万元）'].sum().reset\_index()  #能显示原始数值和百分比的自定义函数  def make\_autopct(values):      def my\_autopct(pct):          total = sum(values)          val = pct\*total/100.0          # 同时显示数值和占比的饼图          return '{p:.2f}%\n({v:.1f})'.format(p=pct,v=val)      return my\_autopct  plt.pie(df\_place['销售额（万元）'],          labels=df\_place['销售点类型'],          startangle=90,          # explode=(0.01,0.01),          # autopct='%.1f%%',          autopct=make\_autopct(df\_place['销售额（万元）'])          )  plt.pie(df\_place['销售额（万元）'],          colors='w',          startangle=90,          radius=0.5,          wedgeprops={'edgecolor': 'w'}          )  plt.show() |
| df\_place\_month = pd.pivot\_table(df,index='月份num',columns='销售点类型',values='销售额（万元）',aggfunc=sum).reset\_index()  plt.figure(figsize=(12,4))  plt.plot(df\_place\_month['月份num'],df\_place\_month['CBD店'],label = 'CBD店')  plt.plot(df\_place\_month['月份num'],df\_place\_month['社区店'],label = '社区店')  # 设置图表标题和轴标签  plt.title('不同销售点的销售金额趋势')  plt.xlabel('月份')  plt.ylabel('销售金额(万元)')  plt.legend(loc=9)  # 显示图表  plt.show() |

输出图表如图8、9。

# 营销策略与建议

根据以上数据分析，我们可以策划一些策略以及针对商家的一些建议：

a. 根据不同商品及各类商品销售情况的分析，商家可以针对热销商品，如Safety 8商品，加大进货量，加强促销活动，吸引更多的消费者。对于滞销商品，即销售额在末尾的商品，则可通过采用降价、组合搭售等方式进行推广。

b. 鉴于不同月份的商品销售情况分析，商家可以根据各个月份的销售情况，在3、8、10月销售较差的月份，制定一些促销计划，以推动销售量的增长。例如，在节假日或者特殊的日子中，可以推出限时促销、打折优惠等活动。而在销售额较高且较稳定的月份，则加大进货量，让销售高峰期更持久。

c. 根据不同季度的商品销售情况分析得知，商品在第二、第四季度的销售较好，，那么商家可以根据这种销售规律，制定相应的季节性采购货物计划和促销计划；此外商家可以根据每个季度的畅销商品和滞销商品，来决定每个季度该将重点放在哪些商品上。

d. 针对不同销售点的销售情况分析，商家可以对销售额高的门店给予奖励或者特别优惠，以激励门店实现更好的销售业绩。同时，也可以向销售额低的门店提供支持、帮助，促进其业绩的提升。在2、3、6月可以考虑给各成本更低的社区店进更多的商品货物。在其它月份时向CBD店进货。