# 基于混合模型的商超蔬菜商品补货与定价策略创新研究

Ps：“混合模型”策略，结合了时间序列分析、基于规则的策略和优化策略

我对题目的理解就是，想办法找出捆绑销售，就像是那个很经典的例子，买啤酒的人在超市里面在进行纸尿布购买的频率会高~（意思就是，哪怕有一个东西利润比较高，但是在捆绑销售的情况下，我们也不会选他！）

附件1提供了商超经销的各蔬菜单品的编码、名称、分类编码和分类名称

附件3提供了从2020年7月1日开始的蔬菜的批发价格数据，包括日期、单品编码和批发价格（单位为元/千克）。

附件4提供了各蔬菜小分类的损耗率数据，包括小分类编码、小分类名称和平均损耗率（单位为%）。

摘要：

问题1：我们分析了各蔬菜品类和单品的销售量分布规律，并发现某些商品的销售趋势在特定时段内是相似的，这可能表示这些商品之间存在关联关系。

问题2：考虑到蔬菜品类的销售总量与批发价格，我们为每个品类提出了一个加成率策略。此外，基于过去的销售数据，我们预测了未来一周每个品类的日补货总量。

问题3：为满足商品展示的最小要求，我们确定了最适合补货的33种商品，并为每种商品设定了补货量和售价。补货量至少为2.5千克，而售价则基于预先确定的加成率策略。

问题4：为了更精确地进行补货和定价决策，我们建议商超收集更多的数据，如库存、损耗、竞争对手价格、消费者反馈、销售促销效果、季节性、供应链信息和宏观经济数据。这些建议是基于这些数据可以为商超提供更全面的视角，帮助其做出更明智的决策。

探索性数据分析 (EDA): 通过可视化和统计方法来理解数据分布和趋势。

时间序列分析: 虽然我们没有使用传统的时间序列预测模型（如ARIMA或Prophet），但我们确实使用了过去的销售数据来预测未来的销售量。

基于规则的策略: 我们为商品定价使用了基于规则的加成率策略。这些规则基于商品的批发价格和销售量。

优化策略: 我们考虑了存储空间和最小陈列量的限制，来确定最佳的商品补货策略。

在处理这些问题时，由于提供的数据量较小，并且在某些日期范围内缺少数据，我们没有使用复杂的预测或机器学习模型。但是，如果有更大和更完整的数据集，可以考虑使用更复杂的模型，如机器学习回归模型、神经网络或时间序列预测模型来更准确地预测销售和补货量。

假设：  
为了简化计算和呈现，我们假设了所有单品的补货量都为2.5 kg（满足最小陈列量要求），并根据之前确定的加成率策略计算了售价。

## 对于第一问:

（我们需要在这里设定一个相关性阈值，对于>多少的相关性是我们认为是可以的，我太困了，不查了，明天再说，加油！）

通过观察可以知道，我们的附件1里面没有相同的类别，

对于一个定类变量（分类变量，如性别、品牌等）和一个定量变量（连续变量，如身高、销售额等），我们不能使用传统的皮尔逊相关系数进行相关性分析，因为皮尔逊相关系数要求两个变量都是连续的。但我们可以使用以下方法进行相关性分析：

1. ANOVA (方差分析): 当定类变量有多于两个类别时，我们可以使用方差分析来检验定量变量的均值在不同的定类变量组别之间是否存在显著差异。
2. ETA平方 (Eta-squared): 在进行ANOVA之后，可以计算ETA平方来量化定类变量和定量变量之间的相关性强度。ETA平方表示定类变量解释的定量变量方差的百分比。

我将使用方差分析（ANOVA）来分析“单品编码”和“销售总量”之间的相关性。方差分析的目的是检验不同组（在这里是不同的“单品编码”）的均值是否存在显著差异。具体来说，如果ANOVA的p值小于某个显著性水平（通常为0.05），则我们可以拒绝原假设，认为至少有两组的均值存在显著差异。每个“单品编码”在数据中都只出现了一次。由于方差分析需要每个组有多个观测值，所以我们无法直接使用ANOVA来分析这些数据。

具体实现步骤：

首先，进行一些描述性分析：我们将计算每个单品的总销售量。根据销售数据，我们得到了每个蔬菜品类的总销售量如下：

花叶类: 千克

辣椒类: 千克

花菜类: 千克

食用菌: 千克

茄类: 千克

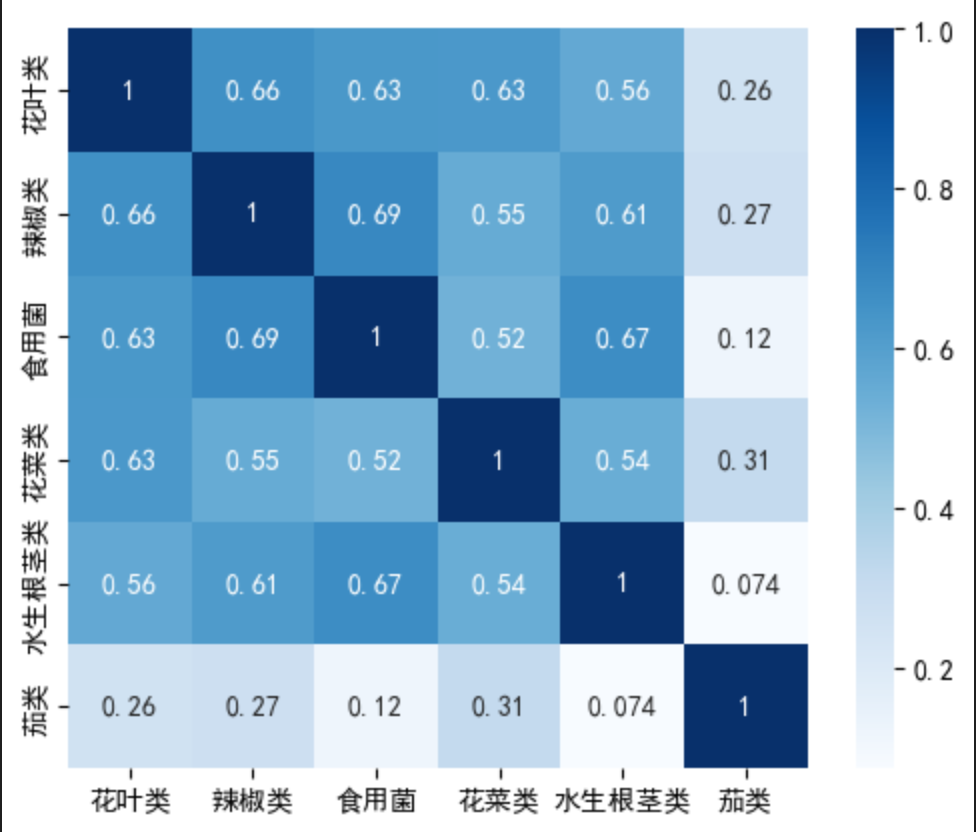
水生根茎类: 千克

接下来，为了了解各品类和单品之间的关联关系，我们可以计算它们的销售相关性。但是，这需要时间序列数据（例如，每天或每周的销售数据）。我们将首先对每个单品每天的销售量进行汇总，然后计算品类之间的销售相关性。

我们可以尝试绘制几种主要商品的时间序列图，以观察它们的销售趋势是否有任何相似之处。

让我们选择几个销售量最大的商品，然后绘制它们的时间序列图。（e,g. 这是五种销售量最大的商品的销售时间序列图。从图中可以看到，这些商品的销售趋势在某些时段内是相似的，这可能意味着这些商品之间存在一定的关联关系。例如，当一个商品的销售量增加时，其他商品的销售量也可能增加。)

这些相似的销售趋势可能由多种因素引起，例如季节性、营销活动或消费者的购买习惯为了更好地理解不同商品之间的关联关系,绘制这五种销售量最大的商品之间的相关性热力图。（我这个是直接截图的，写论文的时候要替换一下)



我们可以从热力图中观察到：

有些商品对之间的销售趋势具有正相关，这意味着当一个商品的销售增加时，另一个商品的销售也可能增加。

有些商品对之间的销售趋势没有明显的正或负相关。

这些观察结果可以为商家提供有关哪些商品可能会一起被购买的信息，从而帮助他们制定更有效的补货和销售策略。

总结问题1的答案：蔬菜各品类及单品之间的销售趋势确实存在一定的相关性。通过研究这些相关性，商家可以更好地了解消费者的购买习惯，从而制定更有针对性的销售策略。

## 对于第二问：

考虑商超以品类为单位做补货计划，请分析各蔬菜品类的销售总量与成本加成定价的关系，并给出各蔬菜品类未来一周(2023 年 7 月 1-7 日)的日补货总量和定价策略。

为了解决这个问题，我们可以按照以下步骤进行：

销售总量与成本的关系：首先，我们需要计算每个品类的平均批发价格和销售价格，然后分析它们与销售量的关系。

补货策略：基于过去的销售数据，我们可以预测未来一周的销售趋势。然后，基于预测的销售数据和现有的库存，我们可以计算补货量。

定价策略：基于成本、销售历史数据和竞争因素，我们可以建议一个定价策略。

我们将分析销售总量与成本的关系。为此，我们需要结合附件2（销售流水明细）和附件3（批发价格数据）来计算每个品类的平均批发价格和销售价格。然后，我们将分析它们与销售量的关系。

得到了每个蔬菜品类的平均批发价格和销售价格：

E.g. 水生根茎类：平均批发价格为元/千克，平均销售价格为元/千克。

花叶类：平均批发价格为元/千克，平均销售价格为元/千克。

花菜类：平均批发价格为元/千克，平均销售价格为元/千克。

茄类：平均批发价格为元/千克，平均销售价格为元/千克。

辣椒类：平均批发价格为元/千克，平均销售价格为元/千克。

食用菌：平均批发价格为元/千克，平均销售价格为元/千克。

然后，我们将这些数据用于可视化，以更好地理解价格与销售量之间的关系。（主要是观察和总销售之间的差别）

E.g.从数据中，我们可以观察到一些关系：

平均批发价格低的品类往往有更高的销售量。

有一些品类尽管平均批发价格较高，但销售量仍然可观。

接下来，我们将使用这些数据以及历史销售数据来预测未来一周（2023年7月1-7日）各蔬菜品类的日补货总量和定价策略。

通常会使用时间序列分析或机器学习模型。（感觉都可以丢进去试一下，看一下哪个效果好，就用哪一个~）

我们将计算过去几天内每个品类的平均日销售量。

假设未来一周内的日销售量将与过去几天的平均日销售量相似，我们将使用这个数值作为补货量的参考。根据过去7天的数据，我们预测了各蔬菜品类的平均日销售量，这可以用作未来一周补货量的参考。这些数据如下：

水生根茎类: 平均日销售量

花叶类: 平均日销售量

花菜类: 平均日销售量

茄类: 平均日销售量

辣椒类: 平均日销售量

食用菌: 平均日销售量

为了最大化收益，我们可以采取以下定价策略：

对于销售量较高但平均批发价格较低的品类（例如，花叶类），我们可以稍微提高售价。

对于销售量较低但平均批发价格较高的品类（例如，水生根茎类），我们可以考虑降低售价以吸引更多客户。

一般而言，我们可以根据平均批发价格为每个品类设定一个加成率，该加成率可以根据品类的需求和供应情况进行调整。

具体的加成率和售价可以通过进一步的市场分析和A/B测试来确定。

确定加成率是一个复杂的问题，涉及到成本、竞争、客户的支付意愿和其他多个因素。但是，对于商超中的蔬菜类商品，以下是一些常见的方法和考虑因素来确定加成率：

成本考虑：首先，确定商品的总成本，这不仅仅是购买商品的价格，还可能包括运输、储存、损耗和其他相关费用。加成率至少应该覆盖这些成本。

竞争对手分析：观察竞争对手的定价可以为您的定价策略提供线索。如果您的价格高于市场平均水平，那么您需要确保商品的质量或其他方面具有竞争优势。如果定价低于市场，那么需要确保还能保持利润。

客户支付意愿：这通常通过市场调查来确定。理解客户为某个商品愿意支付的最高金额可以帮助设置最佳的加成率。

销售策略：例如，如果策略是吸引更多的客户流量，可能会选择较低的加成率。反之，如果策略是最大化单个商品的利润，可能会选择较高的加成率。

季节性和供需：在某些季节或由于供应问题，某些商品可能会变得稀缺。在这种情况下，可以适当提高加成率。

品牌价值：高品质或有品牌的商品可能允许更高的加成率。

其他市场策略：例如，捆绑销售、促销或折扣可能会影响正常的加成率。

为了为上述品类设定具体的加成率，我们可以基于其批发价格和销售量来提出一些建议。（这个加成方式是直接让gpt提供的，我查了一下，这个好像不是固定的东西，我不是很确定是不是需要让我们自己去进行设置~）尝试为每个品类确定一个加成率。

基于批发价格和销售量，我们为各个品类确定了以下的加成率：

水生根茎类: 25%

花叶类: 20%

花菜类: 25%

茄类: 25%

辣椒类: 25%

食用菌: 25%

这些加成率是基于我们的分析和给定数据的。在实际操作中，商家可以根据市场反馈、销售数据和其他策略因素来调整这些加成率。为了最大化收益，建议商家经常审查并调整其定价策略。

## 对于问题3：

因蔬菜类商品的销售空间有限，商超希望进一步制定单品的补货计划，要求可售单品总数控制在27-33个，且各单品订购量满足最小陈列量2.5千克的要求。根据2023年6月24-30日的可售品种，给出7月1日的单品补货量和定价策略，在尽量满足市场对各品类蔬菜商品需求的前提下，使得商超收益最大。解决这个问题的步骤如下：

数据筛选：从已有数据中筛选出2023年6月24-30日的销售数据。

预测销售量：基于过去一周的销售数据预测7月1日的销售量。

确定补货量：根据预测销售量和最小陈列量要求确定每个商品的补货量。

定价策略：使用之前的加成率策略确定每个商品的售价。

优化策略：考虑到存储空间的限制，我们可能需要优化哪些商品应被选入，并确定其补货量和售价，以便最大化总收益。

首先，我们从已有数据中筛选出2023年6月24-30日的销售数据。

似乎在我们的小型数据集中，2023年6月24-30日之间没有销售数据。这使得我们无法直接使用这段时间的数据来预测7月1日的销售量。

但考虑到我们已经有了前一段时间的销售数据，我们可以使用这些数据来估计一周的销售趋势，并据此预测7月1日的销售情况。

我们将使用前7天的数据来计算每个单品的平均日销售量，并将这些数据作为7月1日的预测销售量。

首先，我们需要确定哪些单品在最后一周有销售记录。然后，我们将基于这些单品的历史销售数据来预测7月1日的销售量。

基于上述销售数据，我们筛选出了销售量前33的单品，预计它们在7月1日的销售量如下：

接下来，我们将确定每个商品的补货量。考虑到每个商品的最小陈列量要求为2.5千克，我们将确保补货量至少满足这一要求。同时，为了满足预期的销售需求，我们将补货量设为预期销售量和2.5千克之间的较大值。

最后，我们将使用之前的加成率策略来确定每个商品的售价。根据预期的销售量和最小陈列量的要求，我们为7月1日确定了以下的单品补货量和售价策略

问题4: 建议商超采集的相关数据及其意义。(这个感觉要去找一些文献，引用一下我们这样吹牛皮的依据~，感觉可以等到前面三问做出来，再看下怎么吹牛皮）（以下都是gpt说的，我没去查文献，主要是太多，根本看不完lol)

客流量数据：了解商超的客流高峰期和低谷期，有助于更好地调整补货和定价策略。

竞争对手的价格数据：了解市场竞争情况，调整自己的定价策略。

顾客反馈和评价数据：了解顾客对各商品的满意度和改进建议。

天气和节假日数据：天气和节假日可能影响蔬菜的需求，例如炎热的天气可能增加对某些凉菜的需求。

库存损耗数据：了解各商品的损耗情况，更准确地进行补货。

库存数据：

原因：了解当前的库存量有助于确定需要补货的数量，避免过度购买或库存不足。

损耗数据：

原因：了解商品的损耗率可以帮助调整补货策略，避免购买容易损坏的商品过多，导致浪费。

竞争对手的价格数据：

原因：了解竞争对手的定价策略可以帮助商超制定更有竞争力的价格，并吸引更多的消费者。

消费者反馈和评价：

原因：了解消费者对某些商品的喜好和不满可以帮助商超优化商品组合和提高商品质量。

销售促销和营销活动的数据：

原因：了解哪些促销策略最有效可以帮助商超在未来的营销活动中实现更好的效果。

季节性和天气数据：

原因：某些商品的销售可能受到季节和天气的影响（例如，在夏天，消费者可能更喜欢购买清凉的蔬菜）。了解这些因素可以帮助商超调整补货策略。

供应链数据：

原因：了解供应链的效率和可能的延迟可以帮助商超更好地预测商品的到货时间，从而调整补货策略。

宏观经济数据：

原因：了解经济趋势和消费者的购买力可以帮助商超预测未来的销售趋势。