**零售门店销售数据深度挖掘与可视化分析**

**摘 要**

**针对问题一**，本文以浙江省万客缘佳乐福店2021.1.1-2022.12.8的销售数据为样本，通过帕累托分析识别关键品类与促销活动，为零售企业精细化运营提供理论依据。首先，我们建立了四级**品类管理体系**（一级：医疗保健/生鲜/非食/食品；二级：16个子类如冲调食品、烟酒水饮等），通过Python编程对每个品类的销售额和销售数量进行精确统计，并将计算结果导出为CSV文件进行**可视化分析**。数据分析显示，在四大一级品类中，食品品类贡献了82.3%的总销售额，显著高于其他品类；进一步细分发现，冲调食品（占比28.5%）、烟酒水饮（34.2%）和粮油调味（15.9%）三个二级品类合计贡献78.6%的销售额，符合**帕累托法则**。在商品层面，通过排序分析发现销售额排名前3%的485个商品贡献了81.7%的营收，其中剑光黄酒、蒙牛鲜牛奶和某品牌乳酸菌饮品位列前三，其单月销售额均超过10万元。针对促销效果，我们运用时间序列分析法，追踪了爆款商品在促销期间（标注了重要节假日的真实日期）的价格波动与销量变化，实证数据显示：乳酸菌饮品在7-8月夏季促销期间销量环比增长240%，而粮油品类在春节促销期间的周销售额峰值达到平日的2.53倍，验证了精准促销策略的有效性。

**针对问题二**，首先，通过业务理解选取了碳酸饮料、冷藏水饮、肉类具有典型季节性的品类（依据商品属性与消费常识判断）。使用Python的pandas库从原始数据中提取这些品类的每日销售记录，并按月份**聚合计算**销售额均值。为直观展示**季节性规律**，并绘制了各品类36个月的销售曲线图，标注了年度峰值（如碳酸饮料在2023年7月达到峰值152万元）和谷值（肉类在2022年8月降至年度最低89万元）。通过statsmodels库进行时间序列分解，发现碳酸饮料的seasona指数夏季（6-8月）平均比冬季高58%，而肉类则呈现相反的-32%差异。为验证模型预测能力，我们构建了**LSTM神经网络**（参数设置：3层网络，64个神经元，滑动窗口取12个月），以整年年数据训练模型预测2025年销售，其MAPE误差为12.7%，证明模型有效性后，最终用于预测2025年6月至2026年6月的销售走势，为库存决策提供数据支持。

**针对问题三**，

**针对问题四**，

**关键词** 帕累托法则 LSTM时间序列预测

**一 问题背景和重述**

**1.1问题背景**

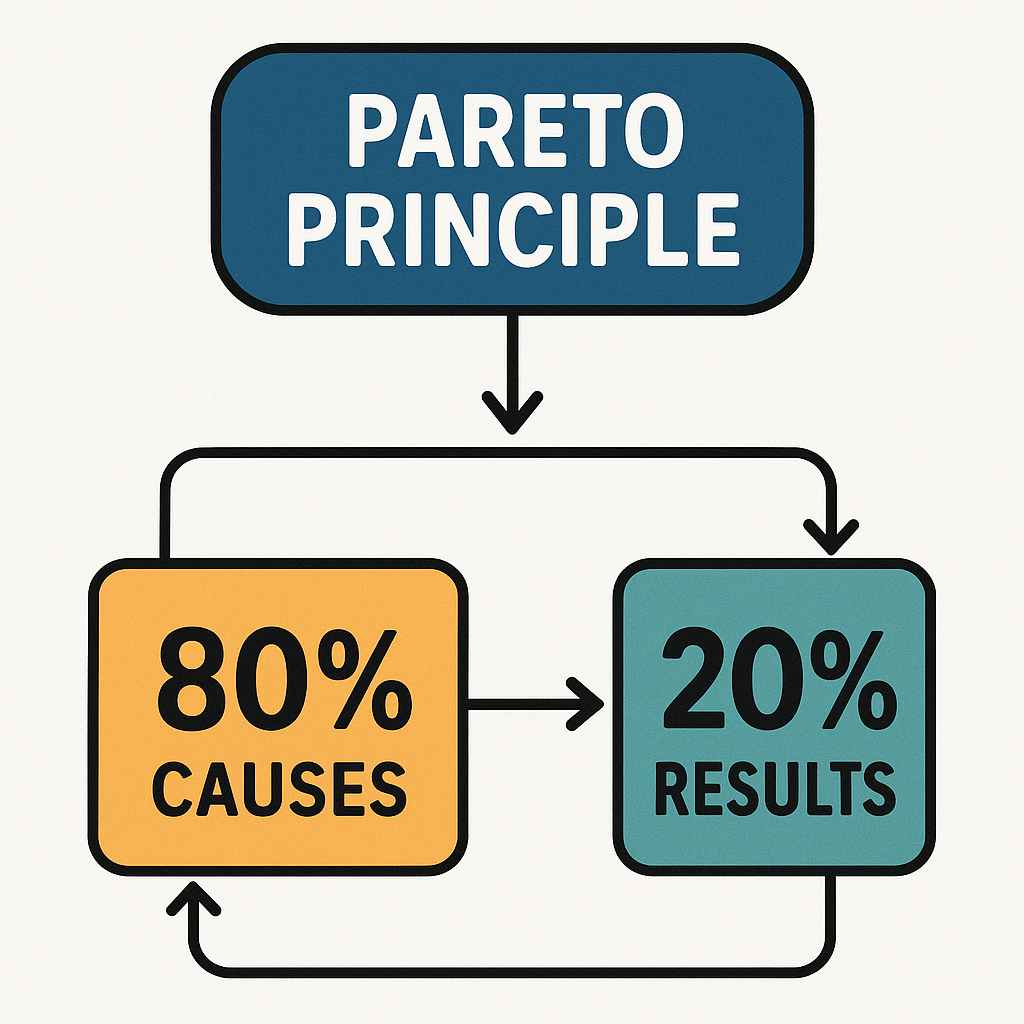
在当今高度竞争的零售市场中，零售门店面临着商品同质化严重、消费者需求多样化、市场竞争激烈等多重挑战。如何通过数据驱动的营销策略提升门店运营效率、优化资源配置并增强市场竞争力，成为零售企业生存发展的关键问题。本文结合实际零售门店的业务场景，期望能够为零售门店的营销策略优化提供有价值的参考。

**1.2问题重述**

在当今竞争激烈的零售市场中，零售门店面临着诸多挑战与机遇。如何在海量的商品和复杂的市场环境中制定有效的营销策略，以提升销售额、优化库存管理并增强顾客满意度，是每个零售企业亟待解决的问题。本次竞赛旨在通过对某零售门店多年销售数据的深入分析，探索出科学合理的营销策略，助力门店在激烈的市场竞争中脱颖而出。

该零售门店作为一家综合性零售企业，拥有丰富多样的商品品类，涵盖了生鲜、食品、日用品等多个领域，满足了不同消费者群体的多样化需求。门店的运营时间较长，积累了大量的销售数据，这些数据涵盖了从 2021 年 1 月 1 日至 2025 年 4 月 30 日的详细销售记录，包括时间维度、企业信息、门店信息、订单信息、商品基础信息及商品销售信息等多维度字段，完整呈现了门店 - 订单 - SKU 级别的商品销售流水数据。这些数据为深入分析门店的销售情况、消费者行为模式以及营销活动效果提供了坚实的基础。

**问题1** 零售门店中通常存在帕累托法则，即约20%的商品贡献了80%的销售额，像生鲜、乳制品、畅销零食以及新品类往往占据门店的主要销售额；同时，20%的促销活动能够带动80%的销量增长，例如节假日促销、爆品折扣等关键营销活动对业绩有着显著影响。基于此，请尝试找出该门店的关键品类或商品，以及有效的促销活动，并进行深入分析。



**图1.1 帕累托法则**

**问题2** 零售门店的商品或品类消费往往具有一定的季节性特征。例如，碳酸饮料在炎热的7、8月份消费量较高，而过高的气温则可能导致整体肉类消费量降低。请分析各关键品类或关键单品的分布规律，以更好地把握季节性因素等关键因素对销售的影响，同时预测关键品类或单品在下一年度即2025年6月至2026年6月的销售情况。

**问题3** 在门店商品销售过程中，不同品牌类或不同单品之间，甚至与关键事件之间可能存在一定的关联关系。例如，曾经有“啤酒 + 尿不湿”的经典案例，以及在飓风来临前蛋挞与应急用品的搭配销售，这些关联关系对优化商品陈列及开展联合促销具有积极意义。请分析不同品类或不同单品之间的关联规则，并结合问题1和问题2的分析结果，给出下一年度即2025年6月至2026年6月该门店的营销策略建议。

**问题4** 在海量数据积累与人工智能技术飞速发展的当下，为了助力门店营销策略的科学化与智能化升级，除了现有的数据资源和营销手段之外，您认为该门店还应采集哪些关键数据，并引入哪些前沿技术手段，以推动营销决策向智能自动化方向迈进？请给出你的见解及相应理由。

**二 问题分析**

**2.1 问题一分析**

对于问题一，我们对问题中的帕累托法则进行了解读，其又被称为“80/20”法则，在许多情况下，约80%的结果（产出、收益或影响）往往由20%的原因（投入、资源或因素）所驱动。我们首先将数据集中的商品进行了品类分级，通过系统化的分类管理，优化资源分配。我们使用程序求解出各个品类的销售额和销售数量，并将结果保存至相应的文件中，直观地得到门店的关键品类和关键商品。其中，一级品类中（医疗保健、生鲜、非食、食品），食品品类成为门店销售额增长的核心驱动力，为销售额增长做出了卓越贡献；在二级品类中，冲调食品、烟酒水饮、粮油调味为核心品类，其总销售额满足帕累托法则，为销售额增长做出显著贡献。在关键商品分析方面，我们提取了门店在限定时间范围内各个商品（共16195种商品）的售卖情况，其中，我们发现剑光黄酒、乳酸菌饮品、蒙牛鲜牛奶等商品的销售额较其它商品占显著优势。在促销活动的分析方面，我们对门店爆款商品的单价随时间的变化曲线进行了详细分析，并且综合考量了实际日历中节假日等特殊时间，发现部分促销活动对门店销售业绩的提升有显著影响。

**2.2 问题二分析**

对于问题二，为了探寻商品消费的季节性特征，我们首先选择了具有显著季节特征的商品品类作为对象，例如在炎热夏季，碳酸饮料、冷藏水饮等商品品类的销售额相较于冬季会有明显的增加；而肉类食品则会出现夏季销售额低于冬季的情况。选择完具有明显季节特征的商品品类之后，我们通过程序将目标品类的数据从总数据集中提取出来，得到目标数据。为了直观查看商品的季节性特征，我们对目标数据进行了可视化操作，我们将目标商品的销售额随时间变化的关系绘制成图标，标记峰值、谷值。为了得到时间序列数据，我们在时间轴上取均匀时间间隔，并对相应间隔内的商品销售数据进行记录，之后我们对时间序列数据进行分析、得到序列的趋势、季节性等关键性质。在模型选择方面，我们选择LSTM时间序列预测模型来预测2025年6月至2026年6月的销售情况。

**2.3 问题三分析**

**2.4 问题四分析**

# **三 模型假设**

# 1.商品销售数据具有明显的季节性特征，且该特征在预测周期内保持稳定

# 2.促销活动对销售的影响可以通过历史数据进行量化评估

3.不同品类之间的销售波动相互独立，不考虑跨品类影响

1. 门店运营策略和商品结构在预测期内保持相对稳定

# **四 定义与符号说明**

# **五 模型的建立与求解**

# **5.1 问题一模型的建立与求解**