该项目的目标是帮助某餐饮企业有效的提高利润，推进企业持续发展，项目的主要内容有如下几部分：

(1) 从系统数据库中迁移与分析相关的数据到分析数据库中，包括客户信息，菜品详情和订单详情等数据。

(2) 对数据进行预处理，包括数据清洗、特征选取和数据变换等，并统计菜品数据中的用餐人数、销售额、热销度和毛利润等。

(3) 构建ARIMA模型，预测销售额，并对模型结果进行分析。

(4) 分别使用基于物品的协同过滤算法和基于用户的协同过滤算法对菜品进行智能推荐，并对推荐结果进行评价。

(5) 根据订单表，使用Apriori算法对菜品进行关联分析，并对结果进行分析。

(6) 使用K-Means算法对客户进行价值分析，识别出有价值的客户

(7) 使用决策树算法对客户流失情况进行预测，并对结果进行分析

**一 统计餐饮菜品数据**

1. 统计每日用餐人数与销售额

统计每日用餐人数与销售额有两种方法，即分组聚合法和透视表法。分组聚合是对二维结构化数据中的某个字段（或多个字段）分组，并对分组内字段进行汇总的算法。

1.1. 统计结果分析

根据订单表的数据，统计2016年8月的每日用餐人数与每日营业额，如图1和图2所示：

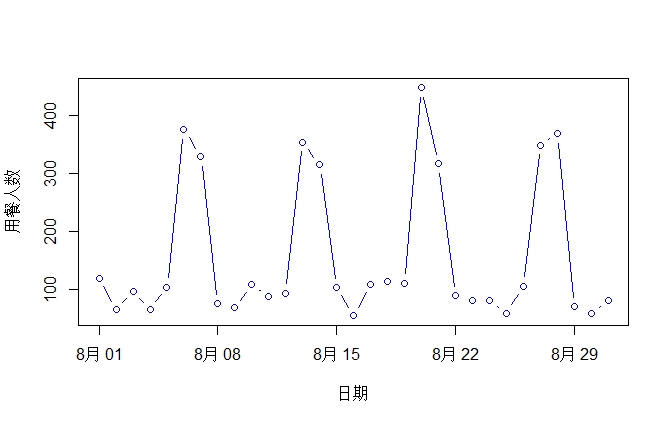


图1

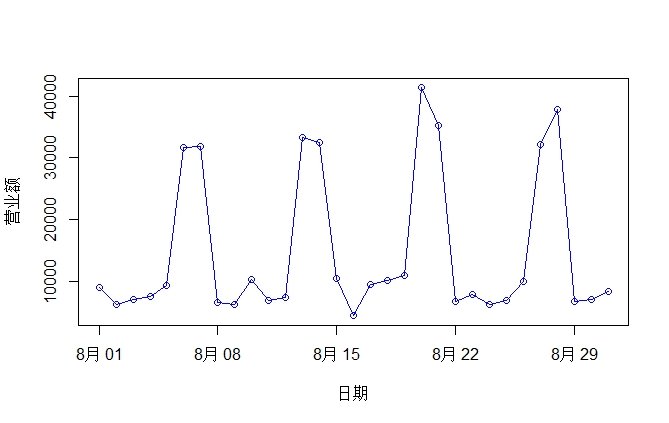


图2

从图中可以观察到，随着日期的改变，用餐人数呈现周期性的变化，观察可知每7天为一个周期。此外，营业额也随着用餐人数的变化而发生增减变化。

2. 统计菜品热销度

热销度即在一定时间内产品销量的程度。根据餐饮企业最近31天的菜品销售统计每个菜品的热销度，其公式如下所示：经过Min-Max标准化后计算得分，最高分为1分，最低分为0分

其中，为某项菜品的热销度评分，其值范围为 0~1； 为某项菜品的销售份数；为该餐饮企业最近31天内有销售记录的菜品中最大的销售份数；为该餐饮企业最近31天内有销售记录的菜品中最小的销售份数；

根据订单详情的数据，计算菜品热销度，求得热销度最高的前10个菜品，其结果在表sales\_volume.csv中，将其结果绘制条形图，如图3所示：

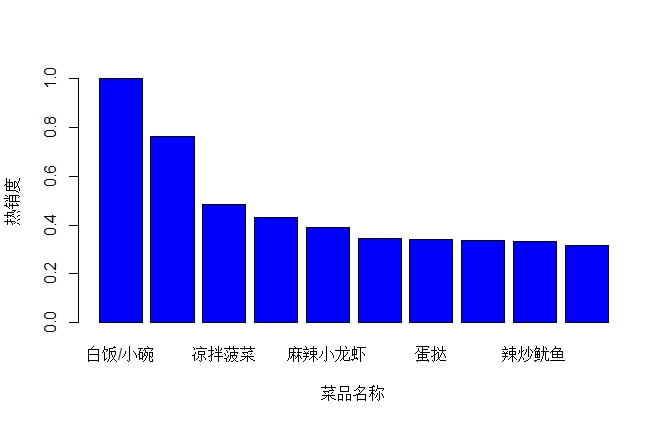


图3

3. 统计菜品的毛利率

毛利率是毛利在销售输入中的占比，其中毛利是菜品单价与菜品相对应的成本之间的差额，计算公式如下所示：

公式中为某项菜品的毛利率修正系数，其值范围为0.1 ~1，当值为负数时设置为0.1；为某项菜品的单价；为某项菜品的估计成本

根据菜品详情的数据计算菜品的毛利率，得出部分菜品的毛利率，结果在表profit.csv中。

二 使用ARIMA算法预测销售额

1. 检验平稳性

选取2016年8月1日-2016年8月28日的数据，针对销售额特征，以天为单位，绘制时序图，判断序列的平稳性，结果如图4所示。

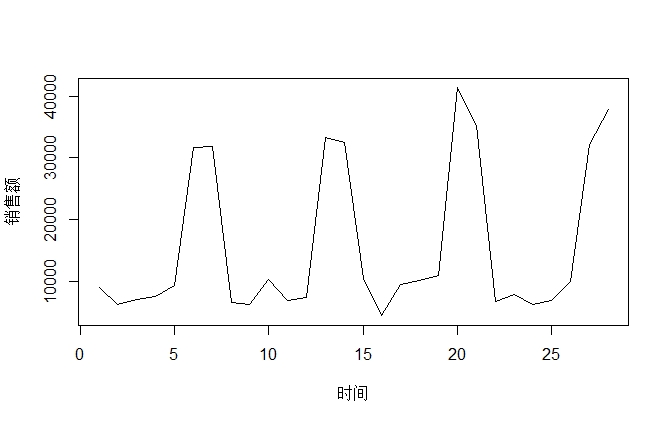


图4

从图中可以看出，序列呈上升趋势，且有一定的周期性，属于非平稳序列。

对每日销售额序列绘制自相关图，如图5所示。

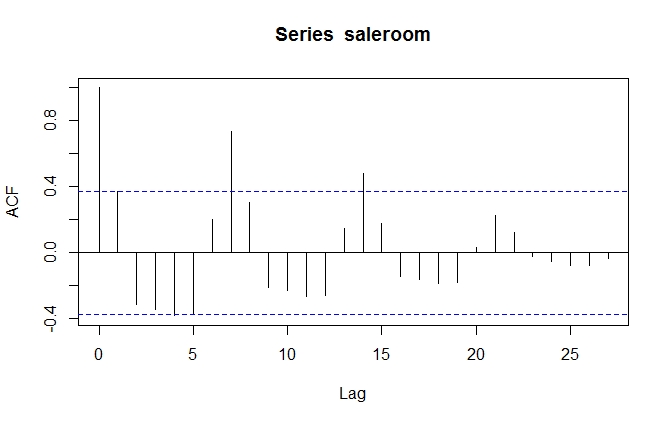


图5

从图中可以看出周期性的波动，以7天为周期，数据稳定的上下波动。

对每日销售额序列进行一次差分，使其变为平稳的时间序列，并对差分后的序列绘制自相关图，如图6所示。

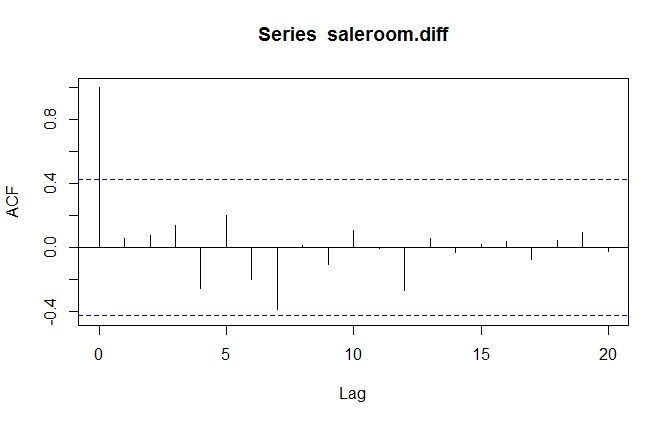


图6

从图中可以看出，数据的周期性已消失，自相关系数大部分都在置信区间内，该数据自始至终都在零轴附近波动，属于平稳序列。

2. 构建ARIMA模型

2.1 模型定阶