数据挖掘互评作业3：分类与预测

姓名：曹健 学号：3120190978

1. **数据**
2. 数据集选择：**Hotel booking demand**
3. 数据集描述：

该数据集包含城市酒店和度假酒店的预订信息，包括预订时间、停留时间，成人/儿童/婴儿人数以及可用停车位数量等信息

1. **问题**

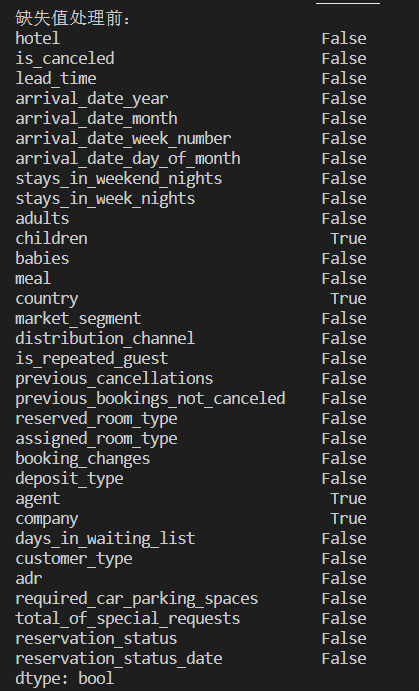
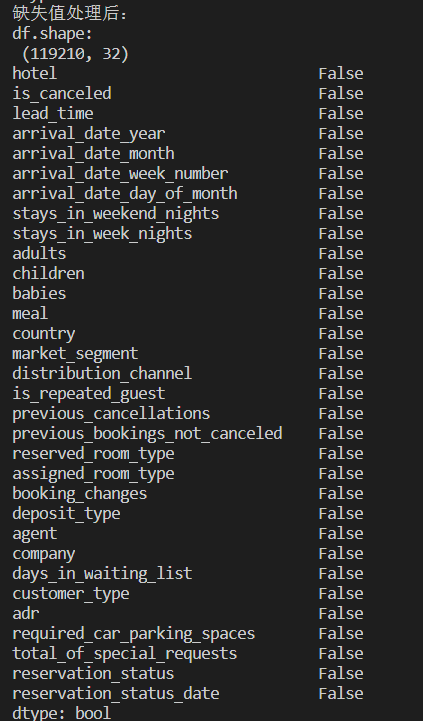
基于这个数据集，进行以下问题的探索：

* 基本情况：城市酒店和假日酒店预订需求和入住率比较；
* 用户行为：提前预订时间、入住时长、预订间隔、餐食预订情况；
* 一年中最佳预订酒店时间；
* 利用Logistic预测酒店预订。

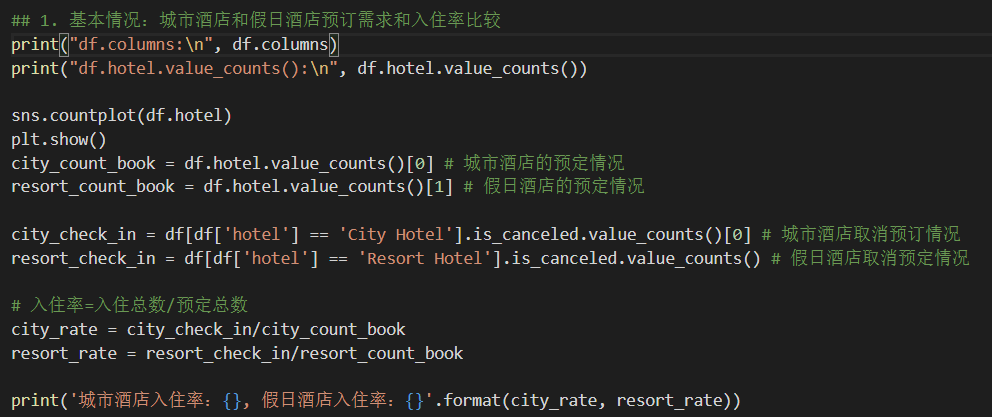
**三、数据挖掘**

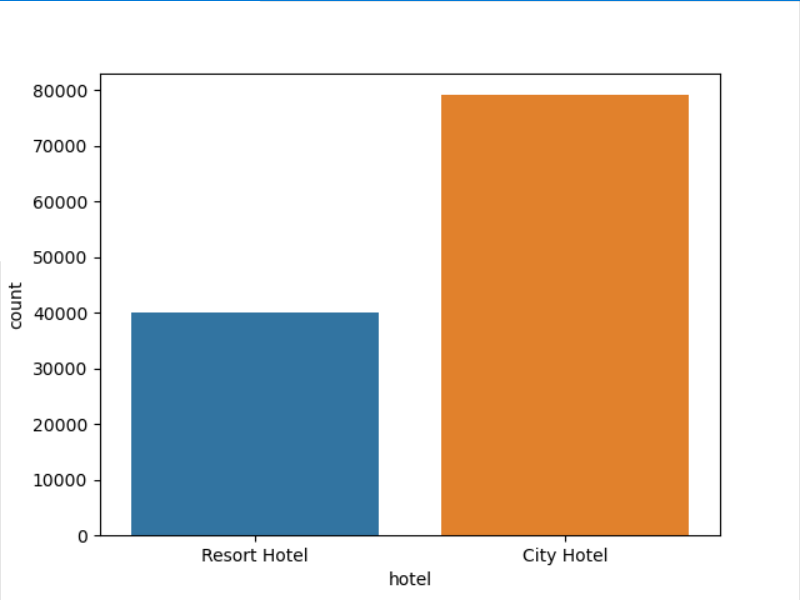
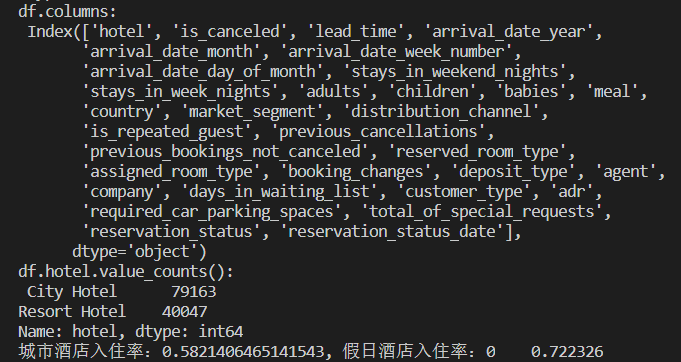
**1. 缺失值处理**



**2. 基本情况：城市酒店和假日酒店预订需求和入住率比较**

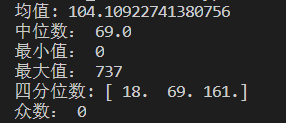


分析可知，就酒店预定来说城市酒店比假日酒店更受欢迎，人们更喜欢预定城市酒店，但是假日酒店的酒店入住率更高。

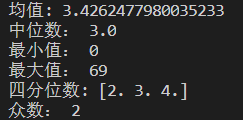
**3. 用户行为：提前预订时间、入住时长、餐食预订情况**

3.1 提前预订时间



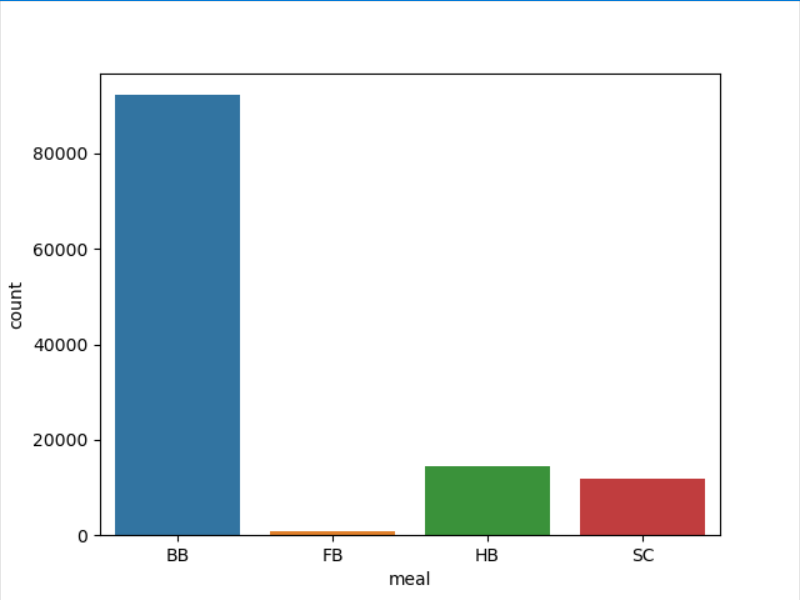
顾客平均提前预定时间为104天左右，预定最久的天数为737天，将近两年多。大部分顾客都是当天预定当天入住

3.2 入住时长



顾客平均入住晚数为3晚左右，最大入住晚数为69天，两个多月，其中大部分顾客入住2晚。

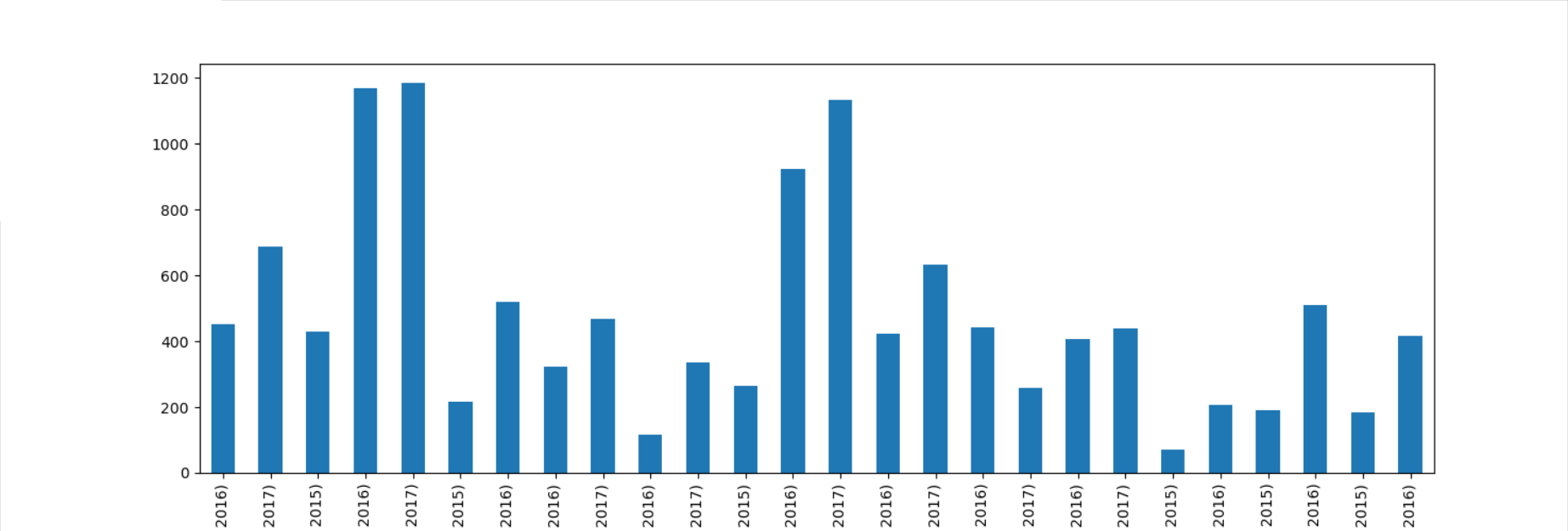
3.3 餐食预订情况



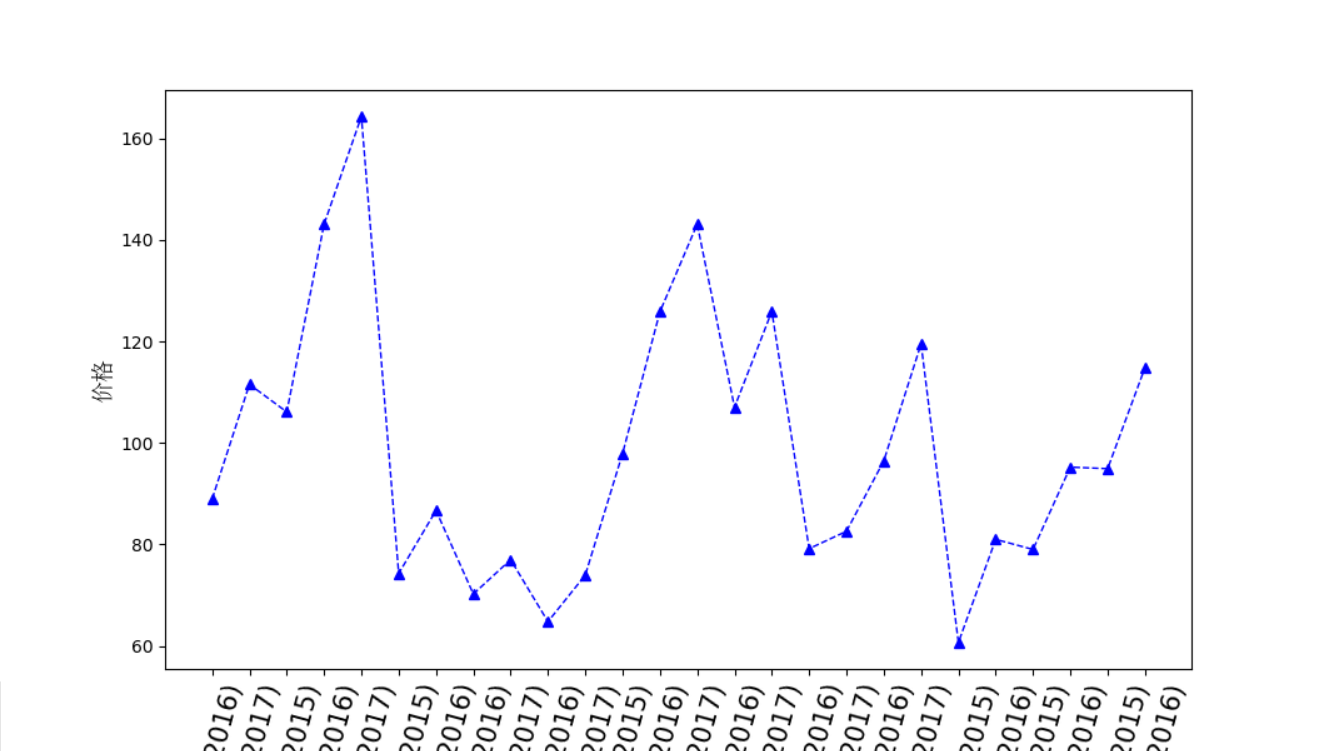
大多数人会在酒店中订餐，其中大部分人预定了BB这个套餐类型，很少人订FB这个套餐类型，少部分人不需要订餐服务（SC）

**4. 一年中最佳预订酒店时间**

4.1 酒店入住情况柱状图

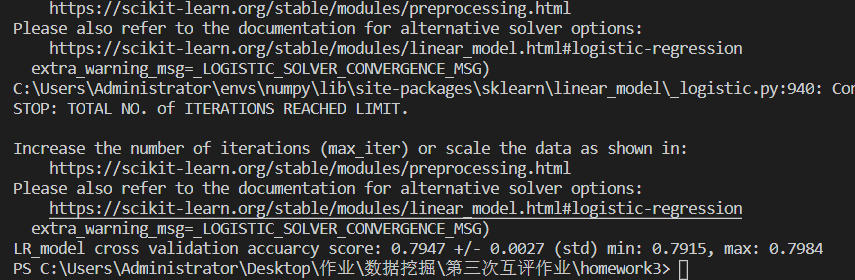


4.2 酒店平均价格-时间段折线图



结合上面两个图来看，对于顾客来说，最佳预定酒店的时间应为每年的1、2月和11、12月，这几个时间段的酒店的入住人数少且价格较低，是最佳的酒店预定入住时间。

**4. 利用Logistic预测酒店预订**



用逻辑回归来预测的准确率结果如上图倒数第二行所示，在79%左右。