# Part1设计思路

本题目决赛阶段需要预测的是500家酒店在2016年2月份的每日间夜量。

## 数据分析

基于对2015年数据的分析，我们发现不同的日期属性对于间夜量影响较大，其中日期属性可分为节假日、工作日以及周末。

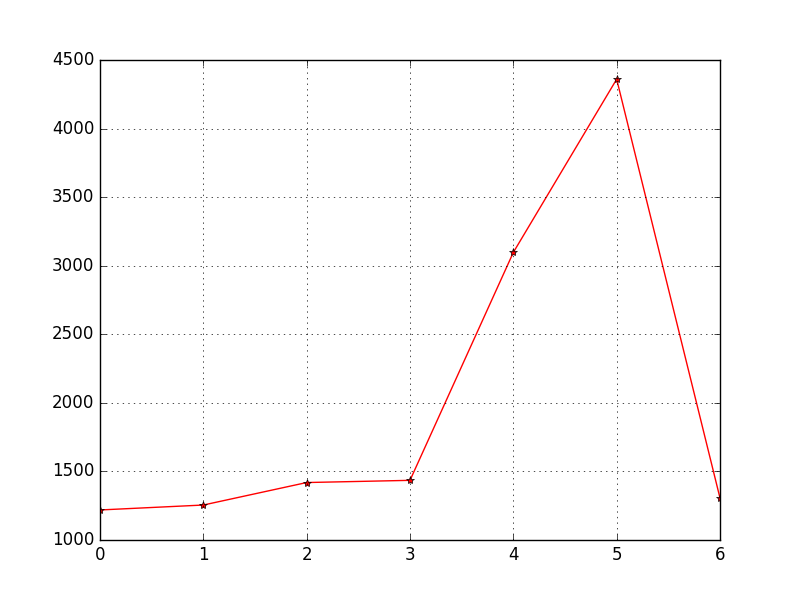


图1 酒店351661在2015年每个工作日的间夜量总和

如图2，可以看出酒店351661在一周当中的周一至周四以及周日的间夜量基本保持稳定，周五和周六的间夜量则有不同程度的提高。经过分析，我们认为周一至周四由于是工作日，住酒店的人数不多，可以理解。周日虽然算是周末，但是因为周一要上班，所以住酒店的人数也和平日相近。同理，因为周六和周日不用工作，所以在周五和周六晚上住酒店的人会比较多。经过对其他500家酒店进行画图分析，发现绝大部分都和酒店351661相近。

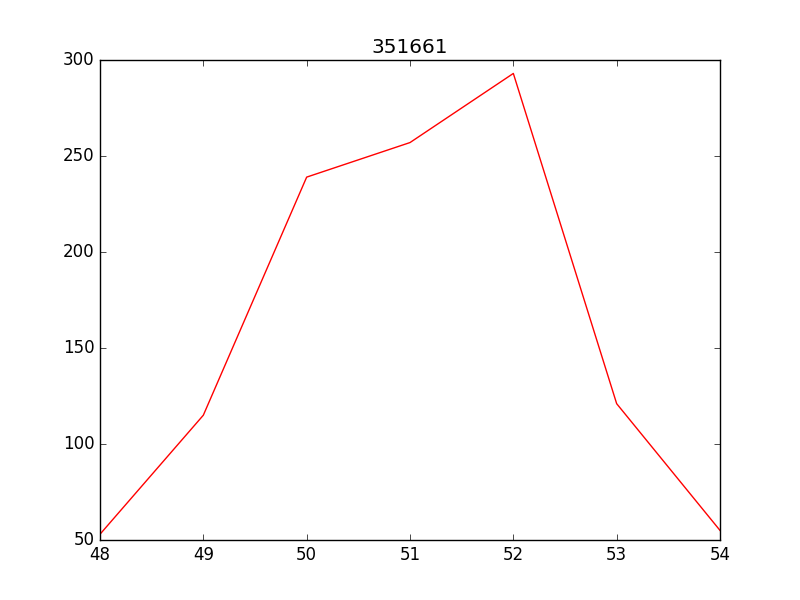


图2 酒店351661在2015年春节期间每日间夜量

如图2，可以看出在春节期间，除夕和春节最后一天的酒店间夜量很低，春节假期的第3至第5天酒店间夜量很高。经推测分析，我认为除夕间夜量低是因为除夕当天大家都习惯与家人一起在家中过年，因此在酒店居住的人很少。而春节最后一天大家基本都会回家准备第二天的上班，在酒店居住的人因此不多。而春节假期的中间几天则是大家最放松，最适宜出行游玩的几天，因此酒店间夜量很高。

因此，我决定按照工作日（周一至周四以及周日）、周末（周五和周六）以及节假日（2月份春节7天假期）分别对数据进行分析和预测。

## 模型建立

对于2016年2月份的工作日、周末以及节假日，我一共建立了两个模型。第一个模型用来预测工作日和周末。第二个模型则是预测春节7天假期中每天的间夜量。

### 模型一 均值模型

对于2月份的每个工作日（即周一至周四以及周日），统计过去两个月（即2015年12月和2016年1月）中的所有工作日（不包含处于节假日期间的工作日）的平均间夜量，当做每个工作日的预测值。对于2月份的每个周五和周六，则分别取前两个月中所有周五和周六的平均间夜量作为周五和周六的预测值。

### 模型二 XGBoost

对于春节七天假期，我利用XGBoost训练了一个模型，模型的训练集是所有酒店在2015年2月过年期间的数据。

模型的特征包括：

1. 当天是节假日中的第几天
2. 酒店基础特征staticinfo（排除了具有null字段的记录）
3. 当天起价
4. 前一个月的入住总量，平均每日入住量，入住量方差
5. 历史最高间夜量
6. 春节前29天至春节前8天非工作日平均值

模型的主要参数设置为：树深度4,5,6；学习率0.03,0.05,0.07；树的个数100,150,200做笛卡尔积，因此一共有27个模型。

将训练好的模型用于对2016年2月份春节7天假期进行预测，并将最终结果取平均值，作为最终的预测结果。

## 结果生成

将模型一预测的平日结果和模型二预测的春节7天假期的结果进行拼接，生成最后的part1部分的预测结果。

# part2 部分

## I 均值策略预测春节以外的平时产量：

1. 每家酒店分别按时间顺序提取每天产量数据

1. 平滑节假日，将节假日的取值改为该酒店此时节假日所在的周几的平均值

2. 建立大小为7天的滑动窗口

3. 选取每个时间窗内的7天的工作日做均值，每个时间窗做一个样本，建立y=ax+b，y是周末的值，x是工作日均值，

每次滑动一天，这样样本数量就有365+31-7+1个，求出a和b，再选取预测2016年2月前21天工作日均值为x，求出y，此时所求的x

和y分别作为预测的29天的工作日和周末的值

## II 多元线性回归模型预测春节产量：

**1. 提取训练特征：**

按照需要预测的500个酒店为单位，分别提取：

2015年元旦两天的和

2015年春节七天前（也就是2月11日）的前21天时间时间内的工作日的平均夜量

2015年春节七天前（也就是2月11日）的前21天时间时间内的周末的平均夜量

2015年春节除夕前一天夜量

2015年春节除夕夜量

2015年春节大年初一夜量

2015年春节大年初二夜量

2015年春节大年初三夜量

2015年春节大年初四夜量

2015年春节大年初五夜量

2015年春节大年初六夜量

正常情况下，训练数据为500\*11，但是有的酒店2015年春节附近没有数据，所以实际情况少于500条

**2. 提取预测特征：**

按照需要预测的500个酒店为单位，分别提取：

2016年元旦两天的和

2016年春节七天前（也就是1月的最后21天）的前21天时间时间内的工作日的平均夜量

2016年春节七天前（也就是1月的最后21天）的前21天时间时间内的周末的平均夜量

这里的数据规模为500\*3

**3. 训练数据:**

建立多元回归模型yi = a\*x1 + b\*x2 + c\*x3 + d训练模型。i为春节的第几天，0代表除夕，6代表初六

y为夜量，x1代表元旦和特征，x2代表工作日特征，x3代表周末特征。训练得出a,b,c,d

**4. 预测结果**

根据训练的参数，通过酒店2016年的特征，可以预测分别出春节7天的酒店夜量，替换掉I中对应的数据

# part3部分

**I. 春节分布拟合：**

1.对春节七天的分布进行了处理，求出500家每家酒店，春节7天每天产量占7天总和的比例，再算出全局500家酒店每天比例的均值。

2.对part1和part2融合之后的数据，七天的值进行求和，再按照训练出的比例求出七天每天的值，再按照权重2:1的分配方式与part1&part2求出的原有值融合（原有值权重大一些为2）

**II. 波动拟合：**

1. 与part2部分的I一样，将工作日也与周末一样用y=ax+b拟合，x是时间窗口内工作日均值，yi分别代表周一到周五分别拟合，再按照权重2:1的分配方式与part1&part2求出的原有值融合（原有值权重大一些为2）