**市场分析13**

**预测消费规模**

**实验报告**

****

**学号：201600301079**

**姓名：崔玉峰**

**班级：2016级**

## 问题描述

1. **问题题目**

一个连锁餐馆的经营者想要发现一个新的餐馆的位置。利用已有的餐馆的销售总额和两公里半径范围内竞争者的数目以及可以收集到这些餐馆附近的人口、收入数据，来建立一个预测销售额的精确模型，从而利用该模型来发现未来餐馆的位置。

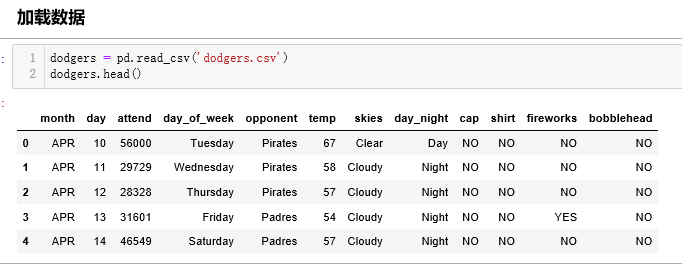
1. **解题思路**

对于这种问题经过多次的实验已经有很多数据科学的基础，此类问题的大体解决思路都是：首先观察数据，可以通过数据可视化的方式，查看数据的分布特点，变化趋势；然后对数据处理，为之后喂给模型做训练打好基础；选取合适的模型进行训练；分析评价模型的情况；通过训练好的模型进行预测。

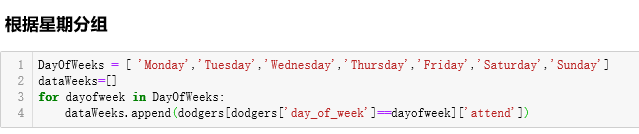
## 具体实现：

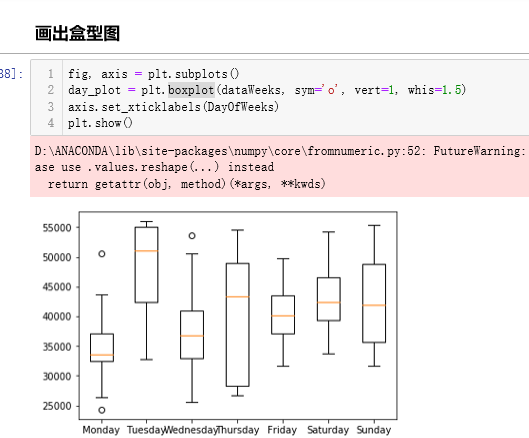
1. **加载数据**

根据所给的.csv数据，通过pandas库读进来



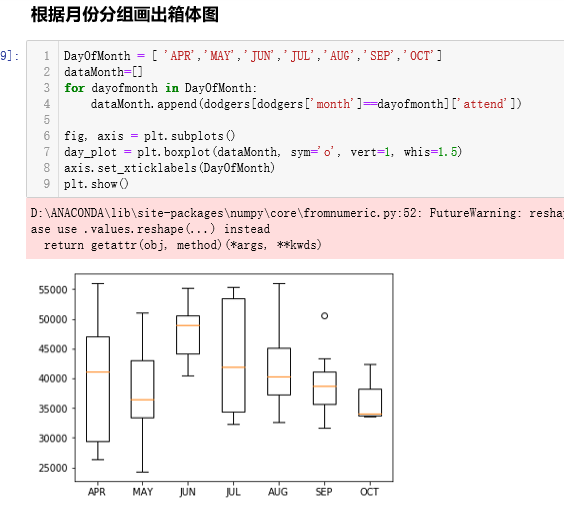
1. **以星期分组画出盒型图**

以比赛星期（dayofweek）属性为例，取值只在星期一到星期日，根据星期取值统计数据并且，调用matplotlib的boxplot方法画出盒型图。 



1. **根据月份分组画出盒型图**

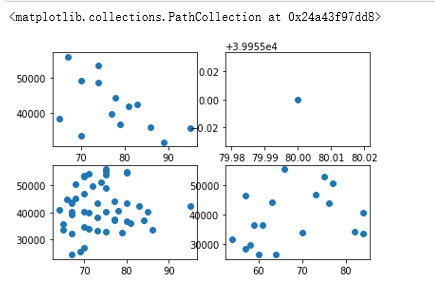
同理根据月份（**month**）属性的取值对数据进行分组并画出盒型图



1. **根据比赛时间和天气画出散点图**

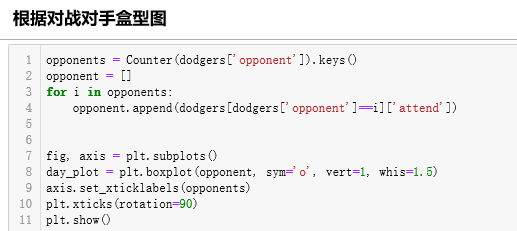
根据比赛在白天或者晚上；天气是晴天还是阴天对数据进行划分，根据情况画出散点图：

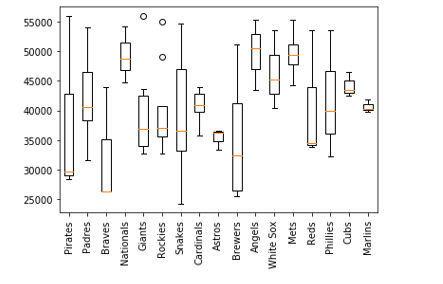




1. **根据对战对手画出盒型图**

统计不同的对战对手，对数据进行分组，然后画出盒型图：





## 结果分析：

## 实验结论和感悟：

本次实验的数据和代码会通过.ipynb的形式上传可以随时运行查看。

本次实验通过建立K-Means聚类对样本数据进行了简单的聚类，分析出两类不同的消费者。