**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**

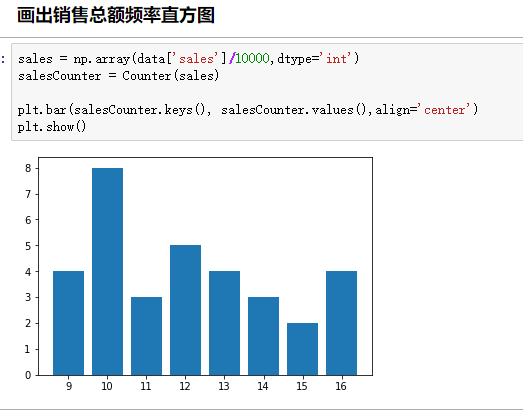
根据所给的.csv数据，通过pandas库读进来



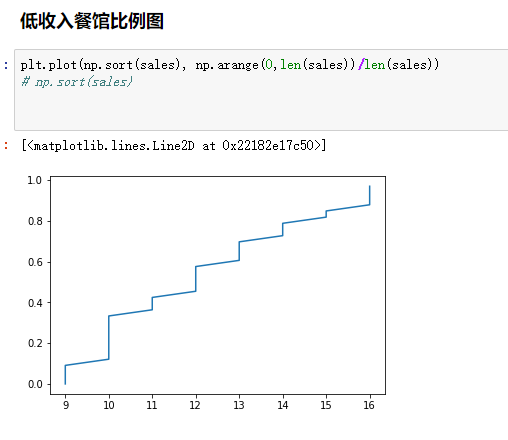
1. **对数据进行分析**

1. **画出销售总额频率直方图**

**销售数量频率直方图，横坐标为销售总额（万），纵坐标为次数。**

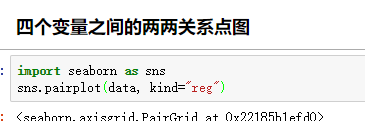


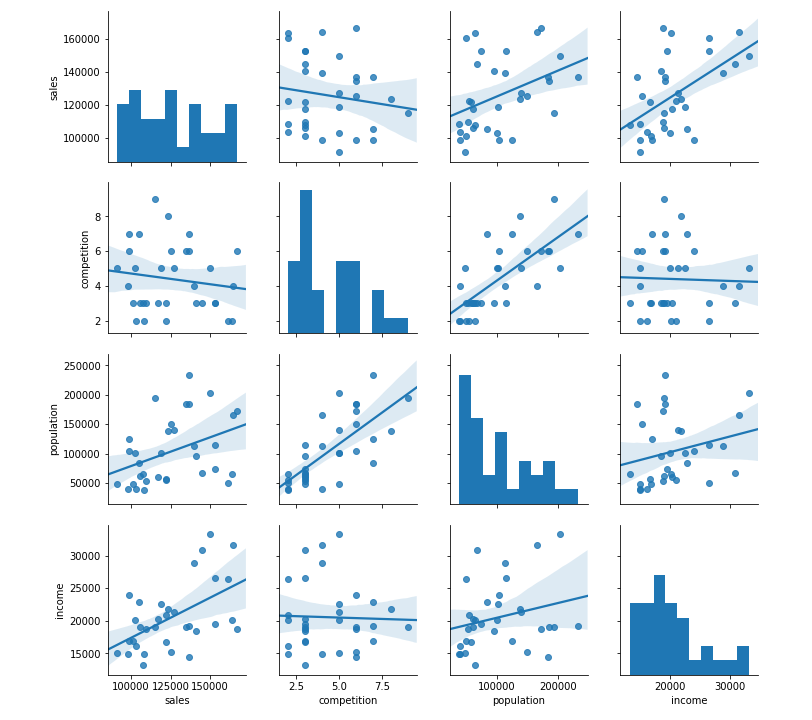
1. **画出低收入餐馆比例图：**



1. 画出四个变量之间两两关系点图：

因为数据每个样本共有四个维度，很难直接展现在同一个坐标系中，所以通过，绘制两两变量点图的方式，观察变量之间的关系：

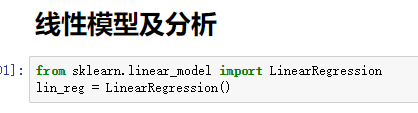




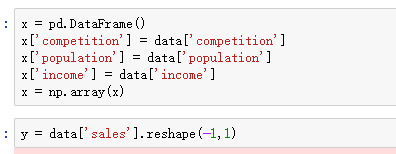


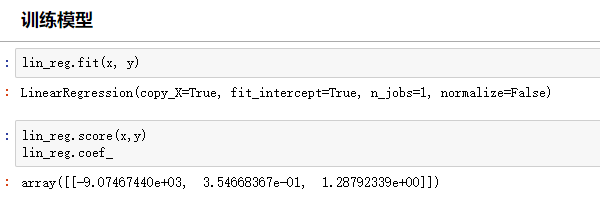
1. **建立线性回归模型**

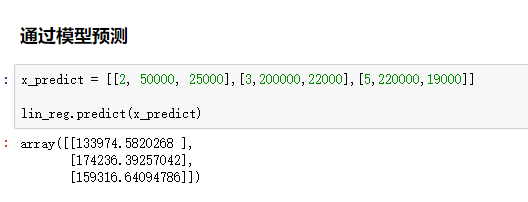
通过scikit-learn机器学习库的,建立线性模型，并对数据进行分析：



生成训练数据集:







## 结果分析：

通过结果观察结果分析数据，训练模型得到了一个可以信赖的线性模型可以根据餐馆提供的相关数据对其的销售额进行预测，并且这个模型相对简单只要几个参数，可以很方便的存储和使用。

## 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。

具体的实现并未完全按照老师给定的代码进行实现，因此在结果上会与老师给定的代码的输出有所不同，但是在大体结果上还是一致的。

通过几次的数据科学题目也可以总结出一些共性的东西了，处理类似的数据分析题目的流程大致为：

获取数据

对数据进行预处理

对数据进行可视化操作观察数据的特点

根据数据特点和需求选择模型

筛选和处理数据，并喂给模型做训练

优化模型参数

得出结论分析结果。