**市场分析04**

**寻找新的消费者**

**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**

寻找新的消费者

1.案例简介：

一家西班牙银行推出了一项新的定期存款业务，为了寻找新的用户，银行在2008年的5月至2010年的11月期间，开展了17次电话营销活动，并详细记录了与每一位消费者的通话信息，其中包括年龄、职业、贷款情况等信息。旨在从这些信息中找到未与自身建立联系的消费者，并通过对这部分消费者进行分析，发现这部分消费者的特征。

1. **解题思路**

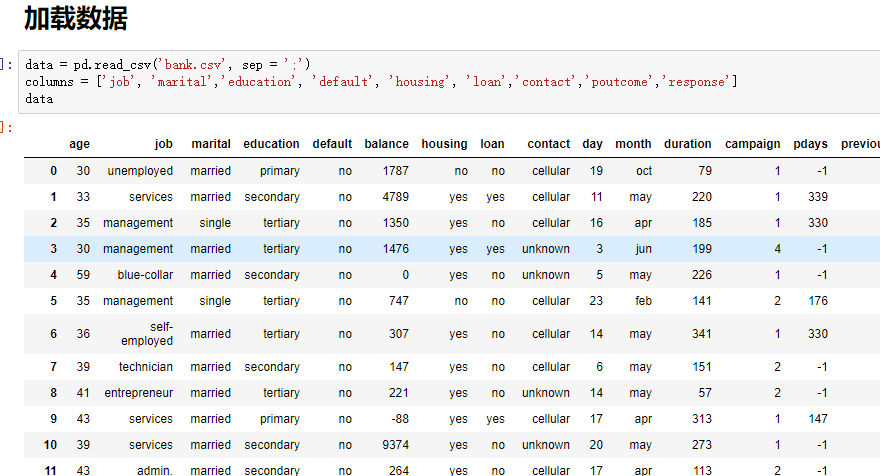
通过K-means算法聚类，对消费者的特征进行划分，努力划分出不同类型的消费者。

**K-means算法**是最简单的分类算法，K-means算法是硬[聚类算法](https://baike.baidu.com/item/%E8%81%9A%E7%B1%BB%E7%AE%97%E6%B3%95/1252197" \t "https://baike.baidu.com/item/K-means/_blank)，是典型的基于原型的[目标函数](https://baike.baidu.com/item/%E7%9B%AE%E6%A0%87%E5%87%BD%E6%95%B0/10829077" \t "https://baike.baidu.com/item/K-means/_blank)聚类方法的代表，它是数据点到原型的某种距离作为优化的目标函数，利用函数求极值的方法得到迭代运算的调整规则。K-means算法以[欧式距离](https://baike.baidu.com/item/%E6%AC%A7%E5%BC%8F%E8%B7%9D%E7%A6%BB/2809635" \t "https://baike.baidu.com/item/K-means/_blank)作为相似度测度，它是求对应某一初始聚类中心向量V最优分类，使得评价指标J最小。算法采用[误差平方和](https://baike.baidu.com/item/%E8%AF%AF%E5%B7%AE%E5%B9%B3%E6%96%B9%E5%92%8C" \t "https://baike.baidu.com/item/K-means/_blank)准则函数作为聚类准则函数。

## 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. **统计样本数据，处理数据转化成虚拟变量**

处理Job marital education,三个属性的取值生成虚拟变量，使之可以通过K-means聚类分析：

1. Job将属性分为三类虚拟变量White Collar,Blue Collar, other
2. marital 属性生成虚拟变量divorced，married
3. education属性生成虚拟变量primary，secondary，tertiary

在python可以通过pandas库的get\_dummies方法来实现

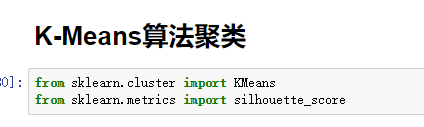


1. **将数据转换成可以聚类的数据类型：**

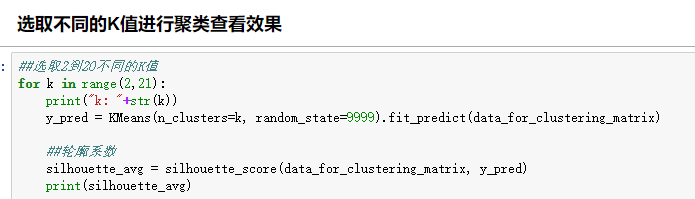
Scikit-learn中K-means 算法需要numpy矩阵的格式，所以就选择聚类属性age, job marital education 四个属性进行聚类。

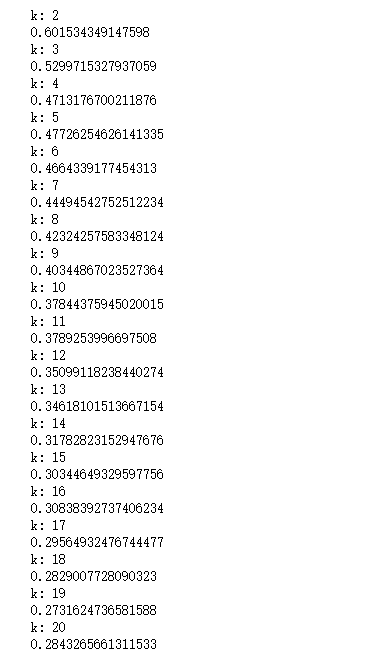


1. **建立K-means模型进行聚类**
2. 通过调用Scikit-learn 机器学习库的K-Means 算法对数据进行分析：



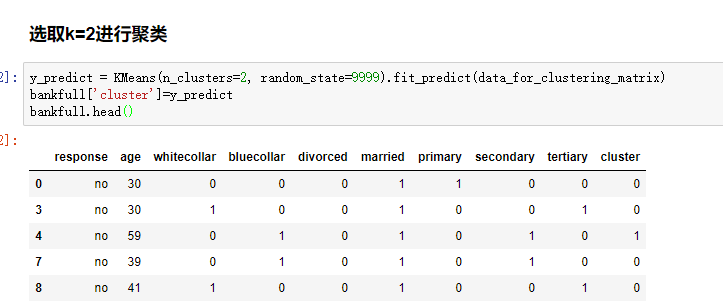
1. 可以选取不同的K值聚类，并通过轮廓系数判断聚类效果：



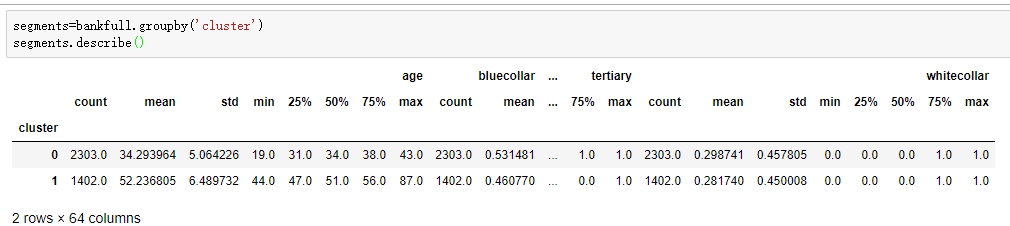


通过查看结果发现K=2时轮廓系数最大，聚类效果最好.

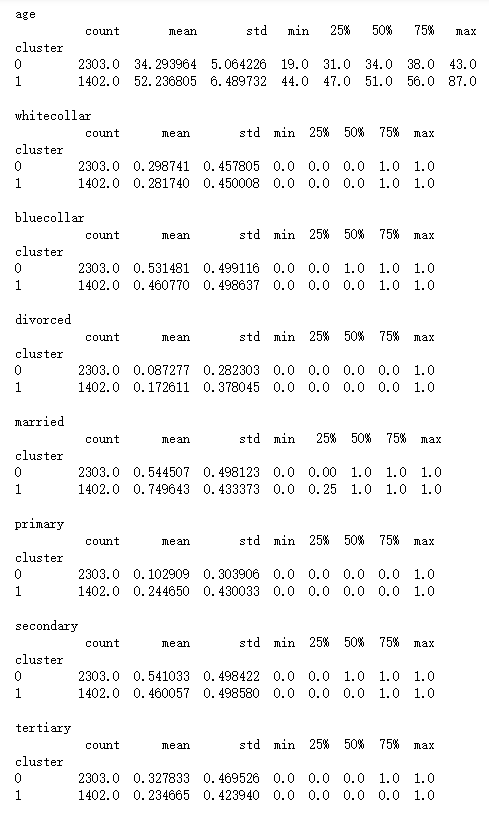
1. 选取K=2对数据进行聚类并为数据添加聚类结果



1. 对聚类结果进行分析



## 结果分析：



通过K-Means聚类分析,聚类出两类不同的消费者，这两类消费者的区别在于年龄上，第一类普遍年龄偏小；第二类年龄偏大；其次是在婚姻状况上，第二类消费者大多数为已婚。在受教育程度上，第一类的教育水平普遍高于第二类，但是两类的都主要分布在secondary中；最后在工作上两类的差别不大

## 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。