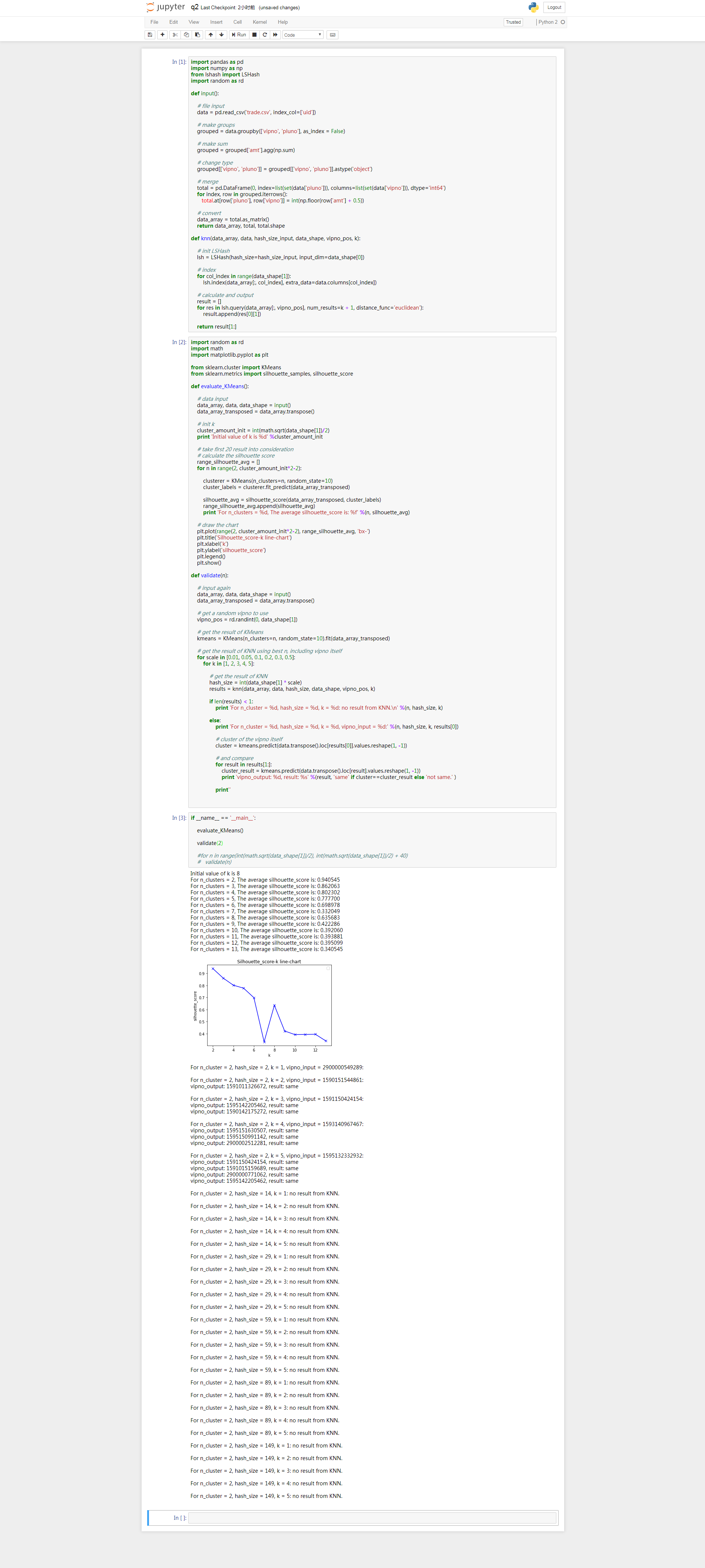
## 代码运行

[点击查看原图](q2.png)



### 讨论分析

1. K-Means算法的基本原理为，事先确定常数K，常数K意味着最终的聚类类别数，首先随机选定初始点为质心，并通过计算每一个样本与质心之间的相似度(这里为欧式距离)，将样本点归到最相似的类中，接着，重新计算每个类的质心(即为类中心)，重复这样的过程，知道质心不再改变，最终就确定了每个样本所属的类别以及每个类的质心。
2. 本题目中直接使用从q1获得的训练集矩阵以及sklearn的KMeans方法进行聚类。
3. 在选择最佳k值时，在2-10的取值区间内进行实验，观察K-Means聚类的结果标签集。发现对该训练集进行聚类，实际产生的结果是一个大的簇以及多个零星的、由单个数据构成的簇组成的簇集。即该聚类行为是将离群数据（或噪声数据）分离的聚类操作，本质是一次数据过滤。
4. 因此在选择最佳k值时，可见k=2的silhouette\_score最高。所以将k=2作为最佳簇数。观察k=2时的聚类结果，其实际意义为，从样本集中去除了1个最离群的数据，并且使得其他数据聚合度最高。
5. 使用KNN的结果进行验证，可见k=2时的分类结果可靠。

### 性能比较

查询相关资料，可知K-Means的时间复杂度为O(tkmn)，其中，t为迭代次数，k为簇的数目，m为记录数，n为维数。

进行全部K-Means查询的总时间为8.33s，减去进行KNN查询所花费的2.30s，实际耗时为6.03s。