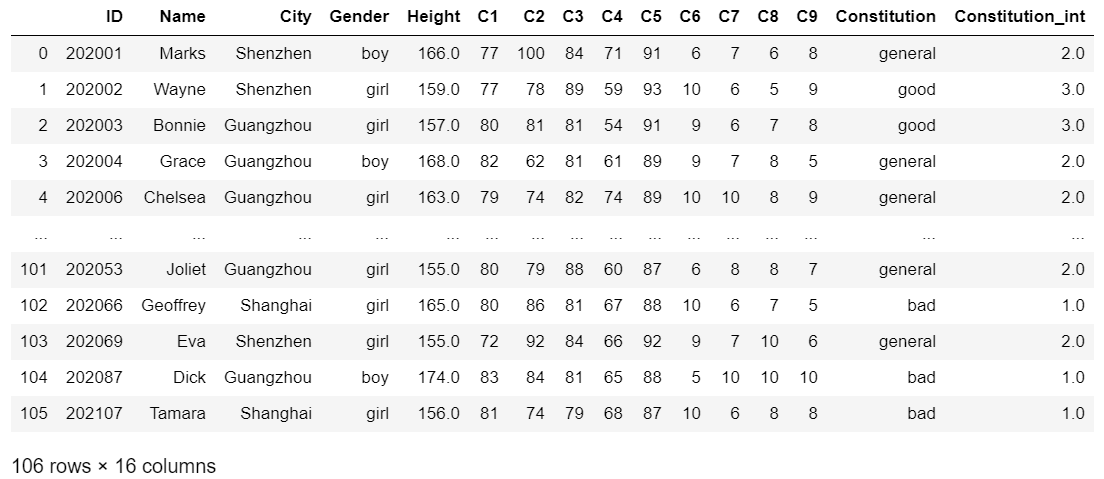
实验二 《数据统计和可视化》

### 实验题目

基于实验一中清洗后的数据练习统计和视化操作，100个同学（样本），每个同学有11门课程的成绩（11维的向量）；那么构成了一个100x11的数据矩阵。以你擅长的语言C/C++/Java/Python/Matlab，编程计算：

1. 请以课程1成绩为x轴，体能成绩为y轴，画出散点图。
2. 以5分为间隔，画出课程1的成绩直方图。
3. 对每门成绩进行z-score归一化，得到归一化的数据矩阵。
4. 计算协相关矩阵，并画出混淆矩阵。
5. 根据协相关矩阵，找到距离每个样本最近的三个样本，得到100x3的矩阵（每一行为对应三个样本的ID）输出到txt文件中，以\t,\n间隔。

### 实验数据

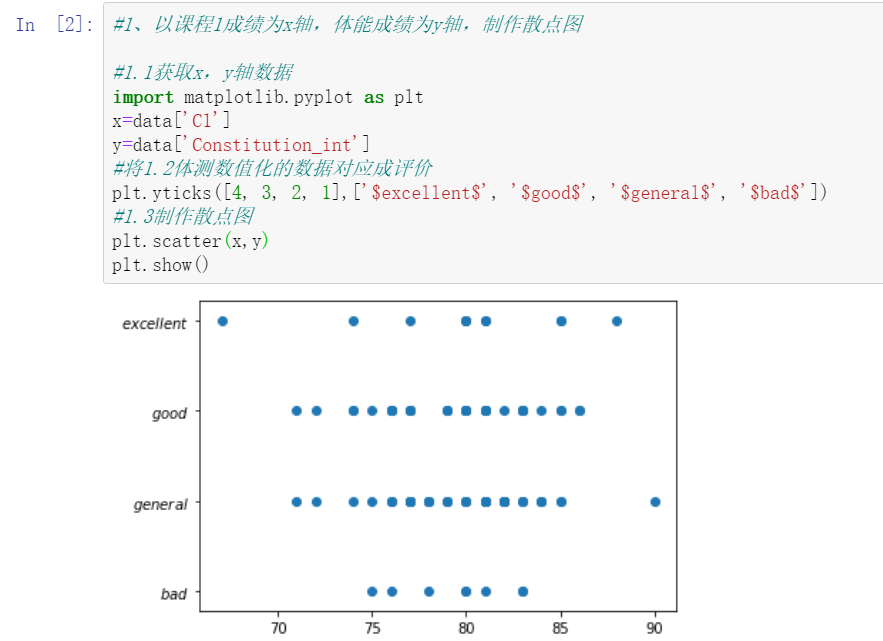


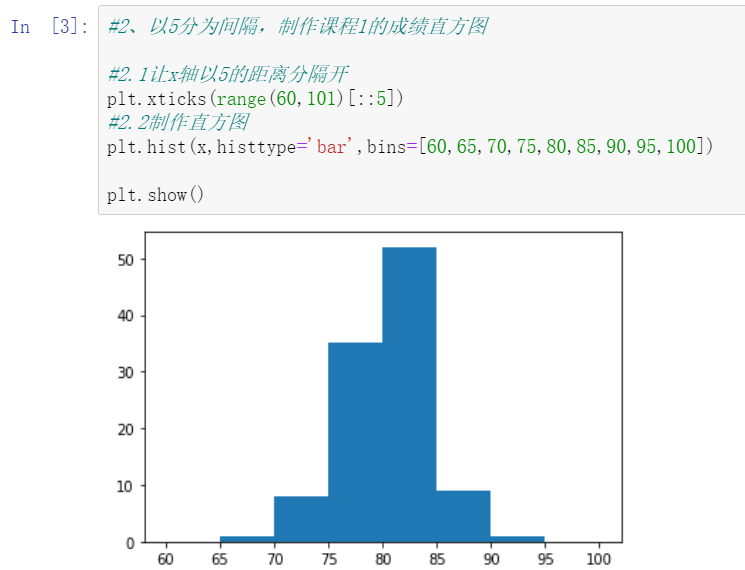
### 实验环境

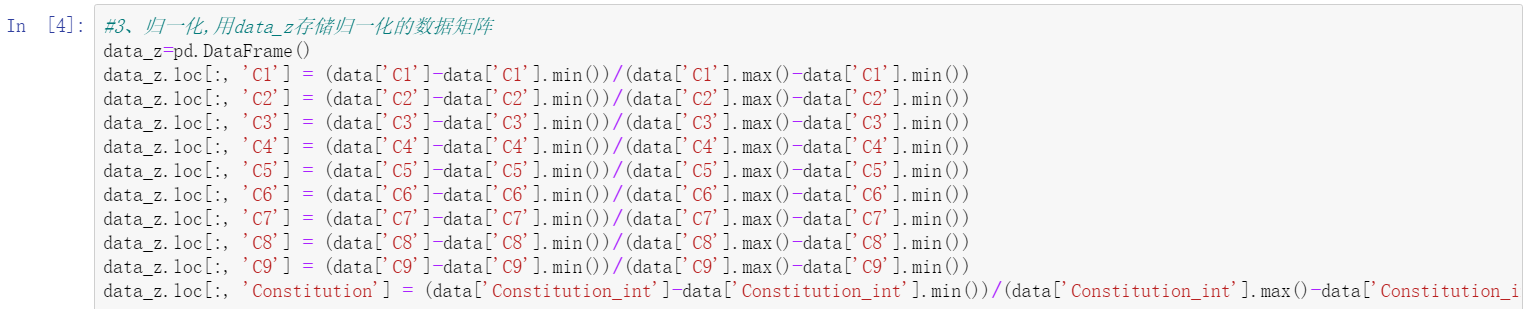
Windows 10、Jupyter

### 实验代码和结果截图



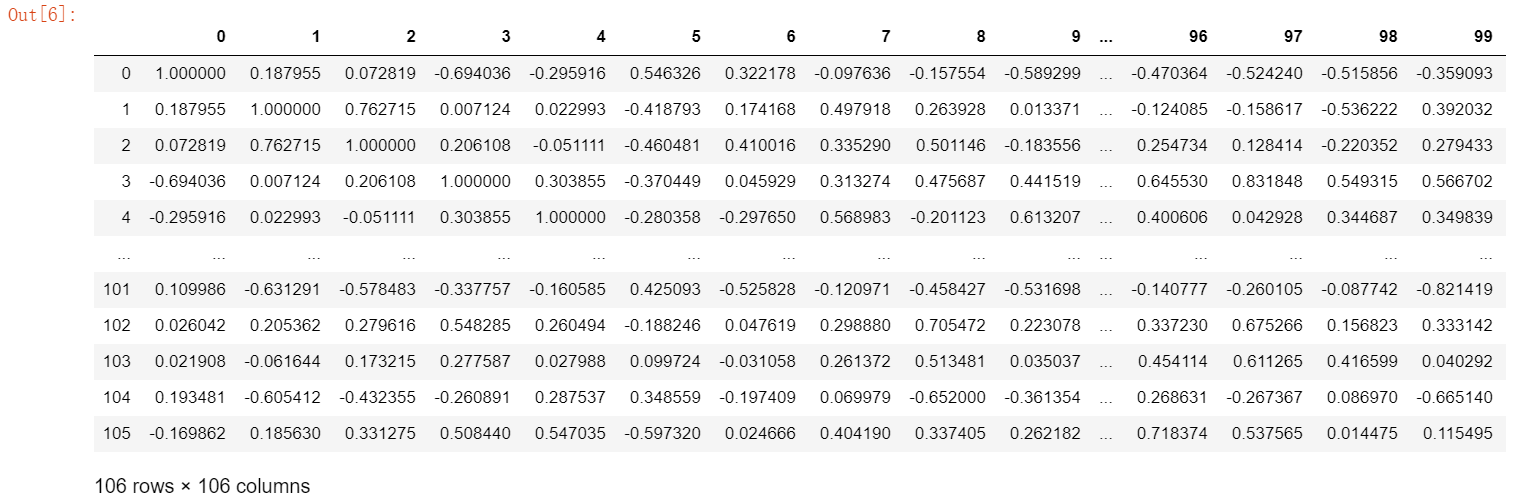


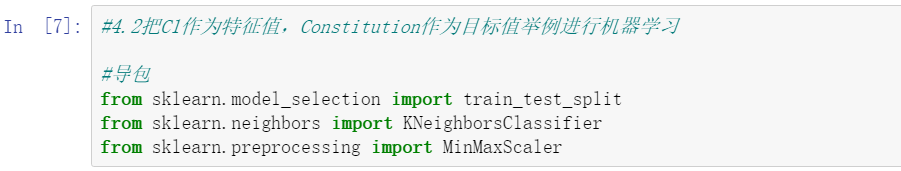


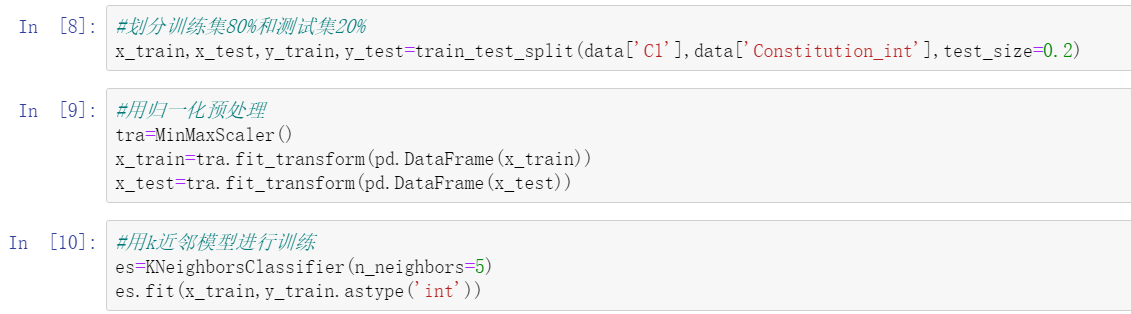




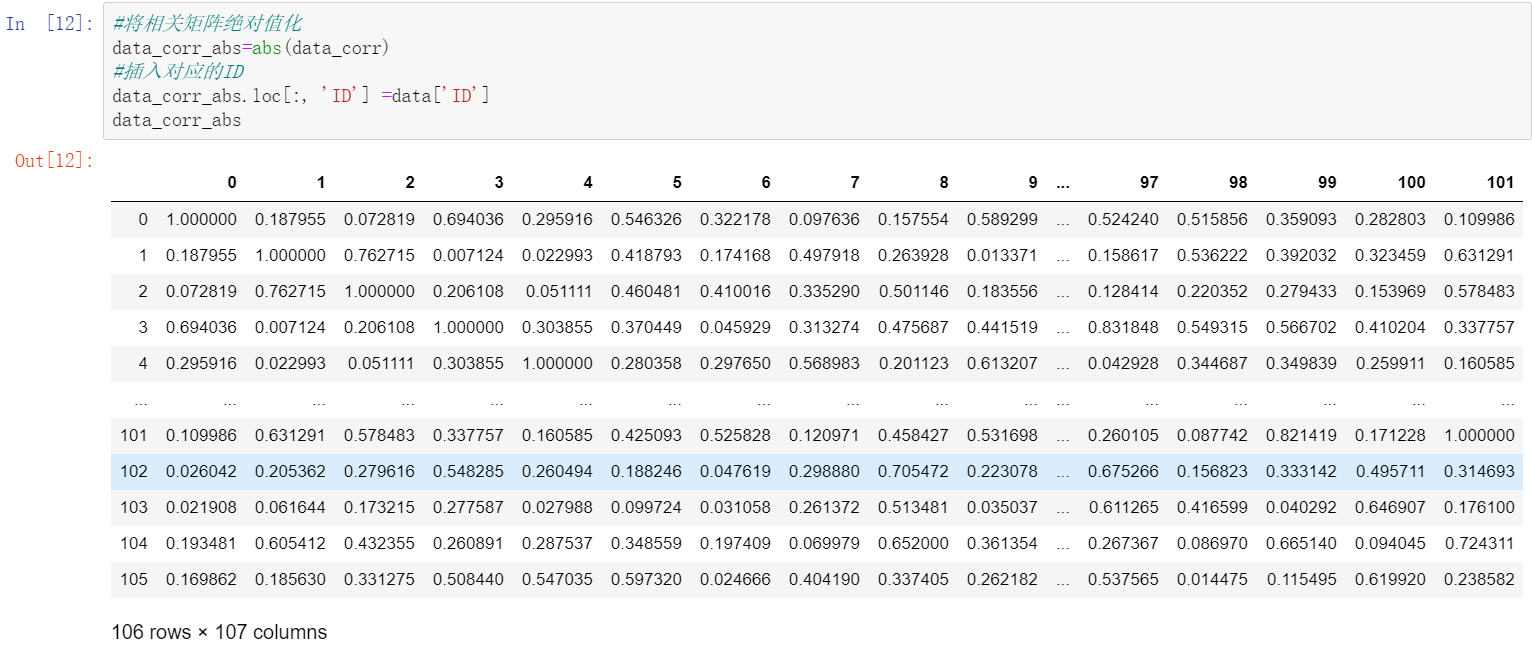


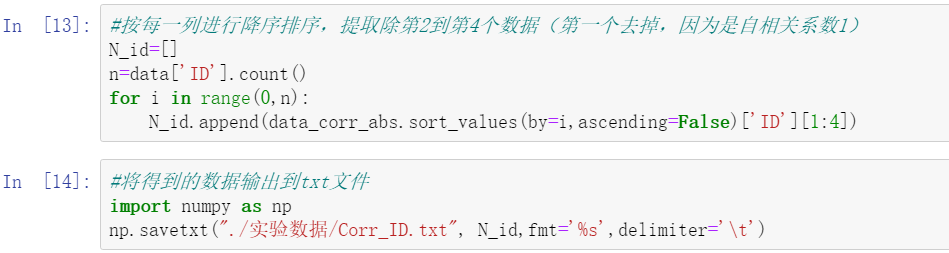


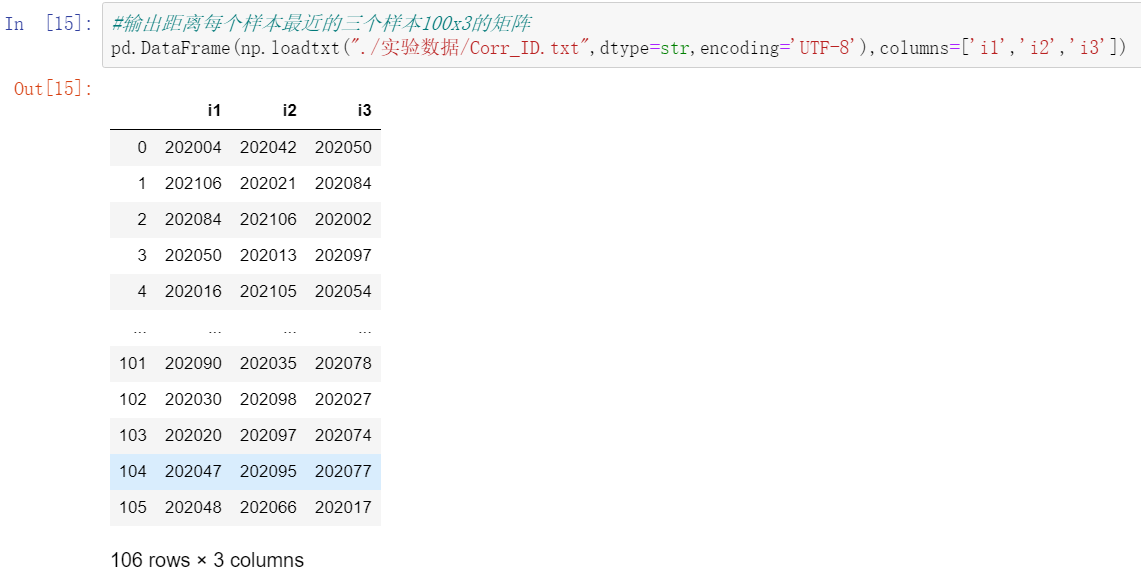












### 实验心得

通过对收集到的数据进行统计和分析，我们能提取有用信息形成模型，从而能对未知数据进行预测。若实际数据存在，则我们可以通过比较实际与预测的数据来评估该模型的好坏。统计和分析数据的目的跟数据挖掘差不多，目的就是把隐藏在大批杂乱无章的数据中进行信息的集中、萃取和提炼出来，以便找出所研究对象的内在规律。类似“淘金”。而数据可视化可以使数据变得更容易理解。通过数据可视化我们可以从复杂数据中理出头绪,化繁为简,变成看得见的财富,从而实现更有效的决策过程。