1．实验目的

（1）掌握Pandas Dataframe的使用方法。

（2）掌握数据探索基本方法。

（3）掌握数据可视化基本方法。

（4）能够处理缺失值、对数据进行恰当的转码。

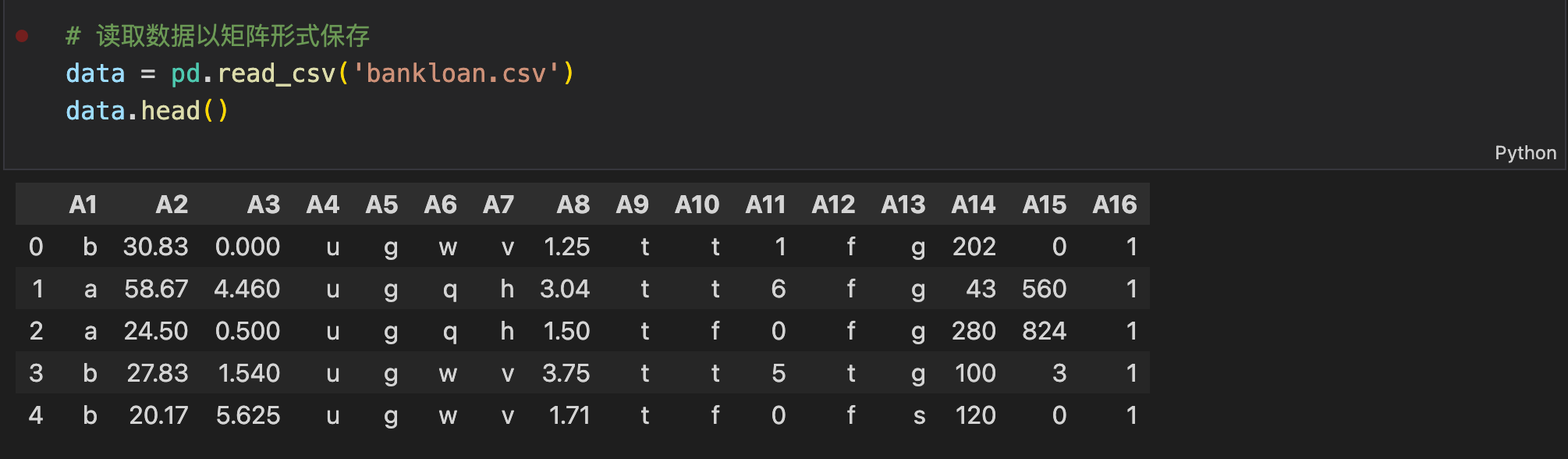
（5）能够选择合适的数值规范化方法

2．实验数据

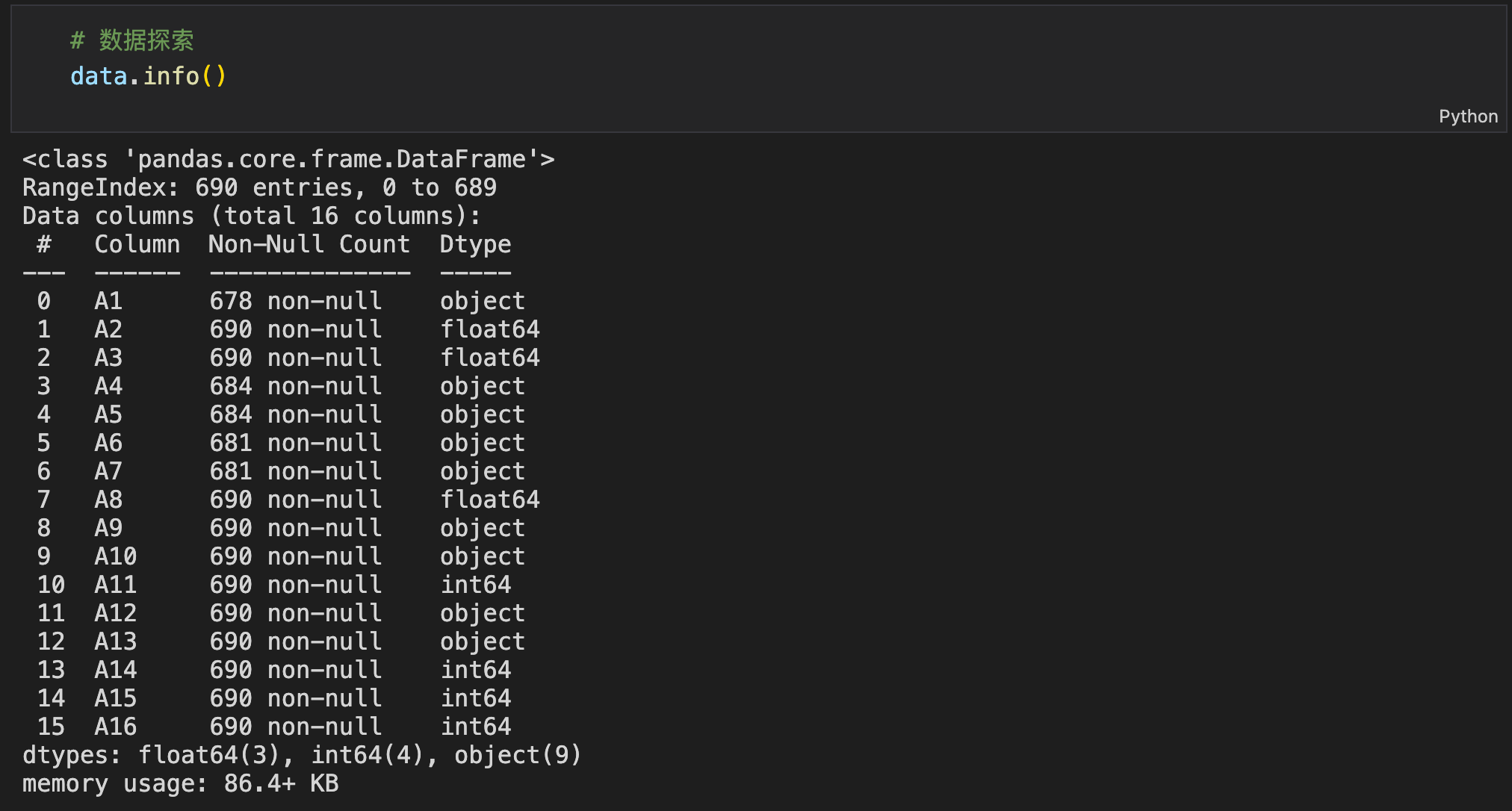
数据文件bankloan.csv。银行贷款数据，一共包含690个样本，16个属性。基于隐私保护的考虑数据已经脱敏，属性名称用A1，A2，…，A16来表示，其中A16是类标号属性，有两个取值：“+”和“﹣”，“+”表示信贷申请被通过，“﹣”表示信贷申请被否决。需要注意的是，一些属性的值为“NA”，表示值缺失。

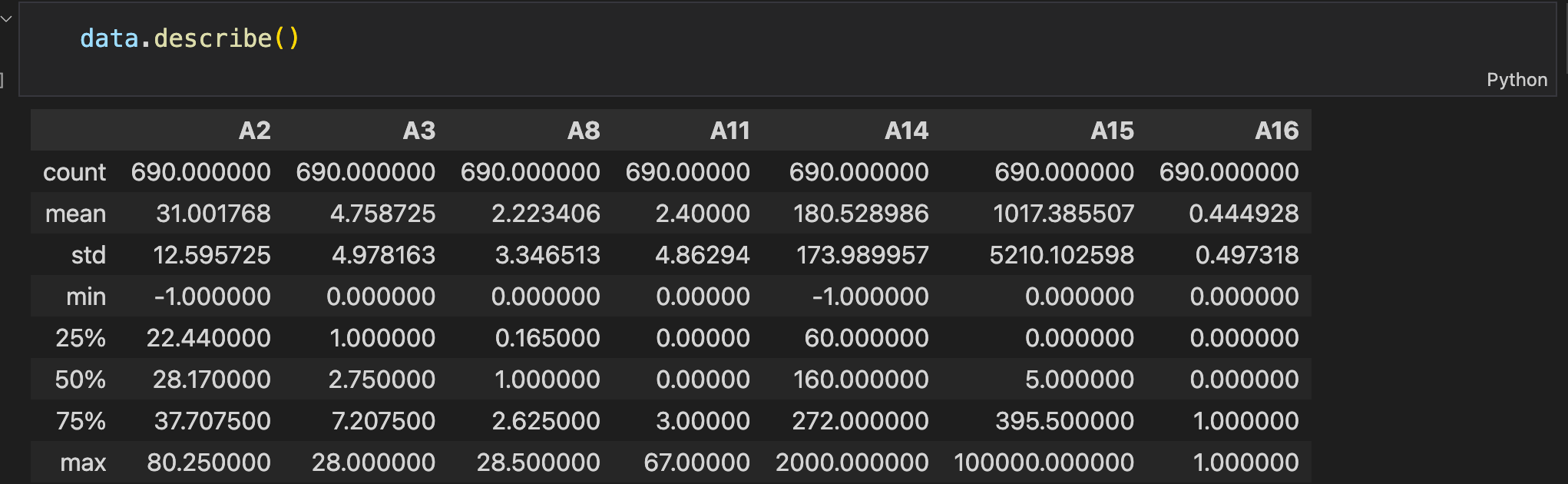
3．实验步骤和实验结果

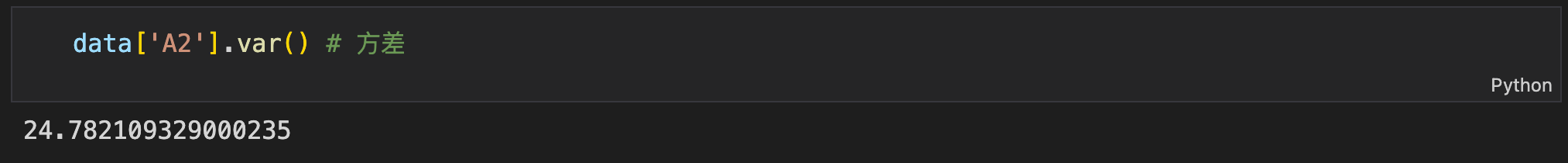
（1）读取数据，以矩阵存储数据

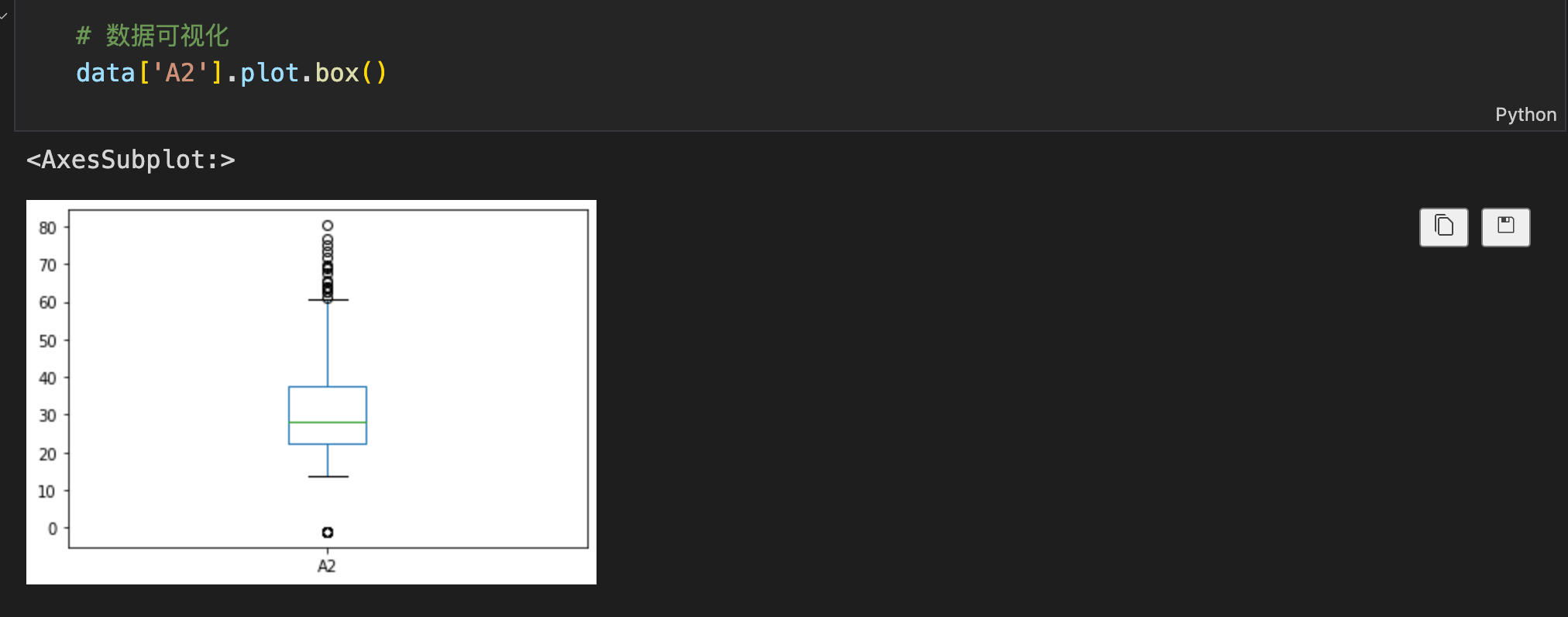


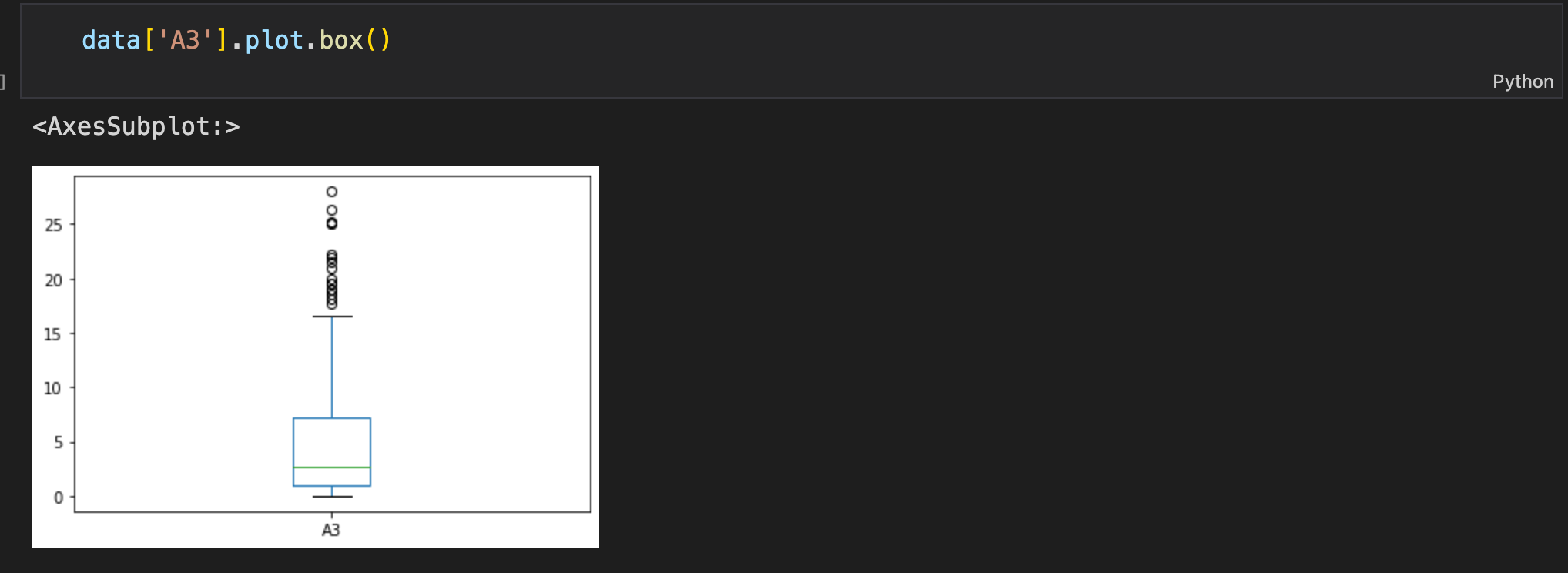
1. 探索数据，并适当可视化

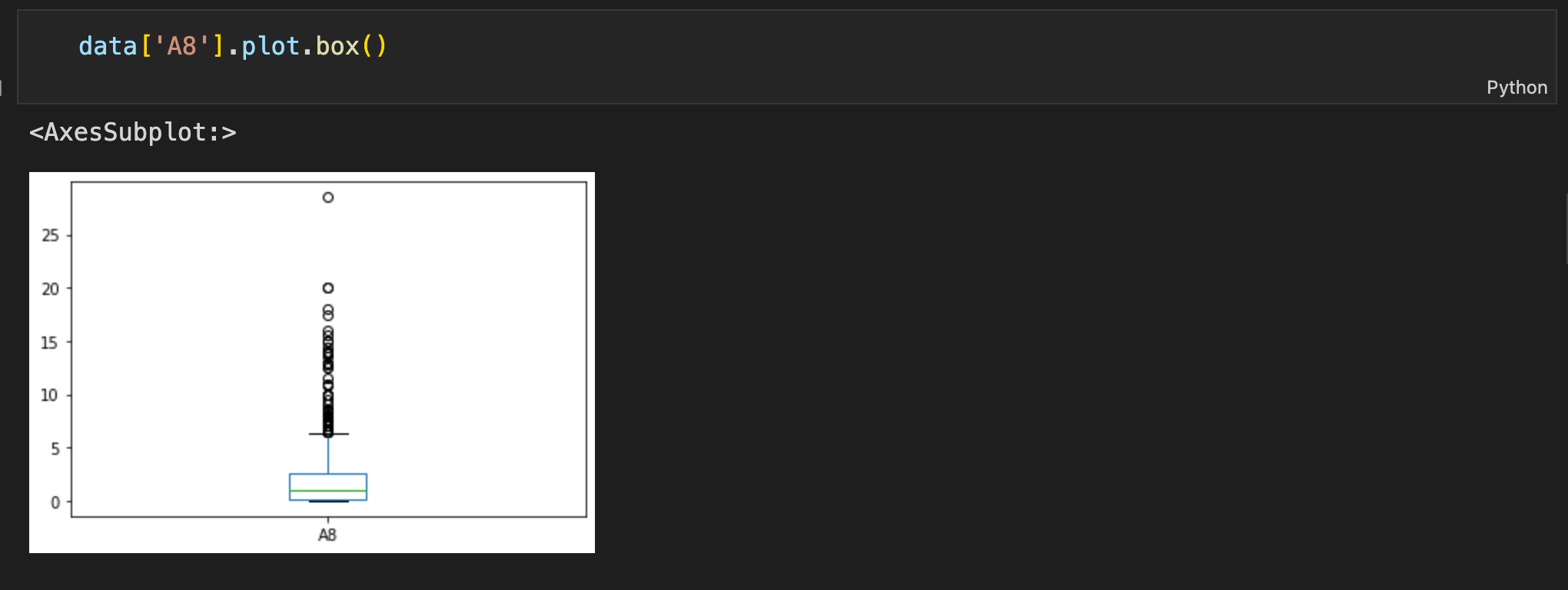


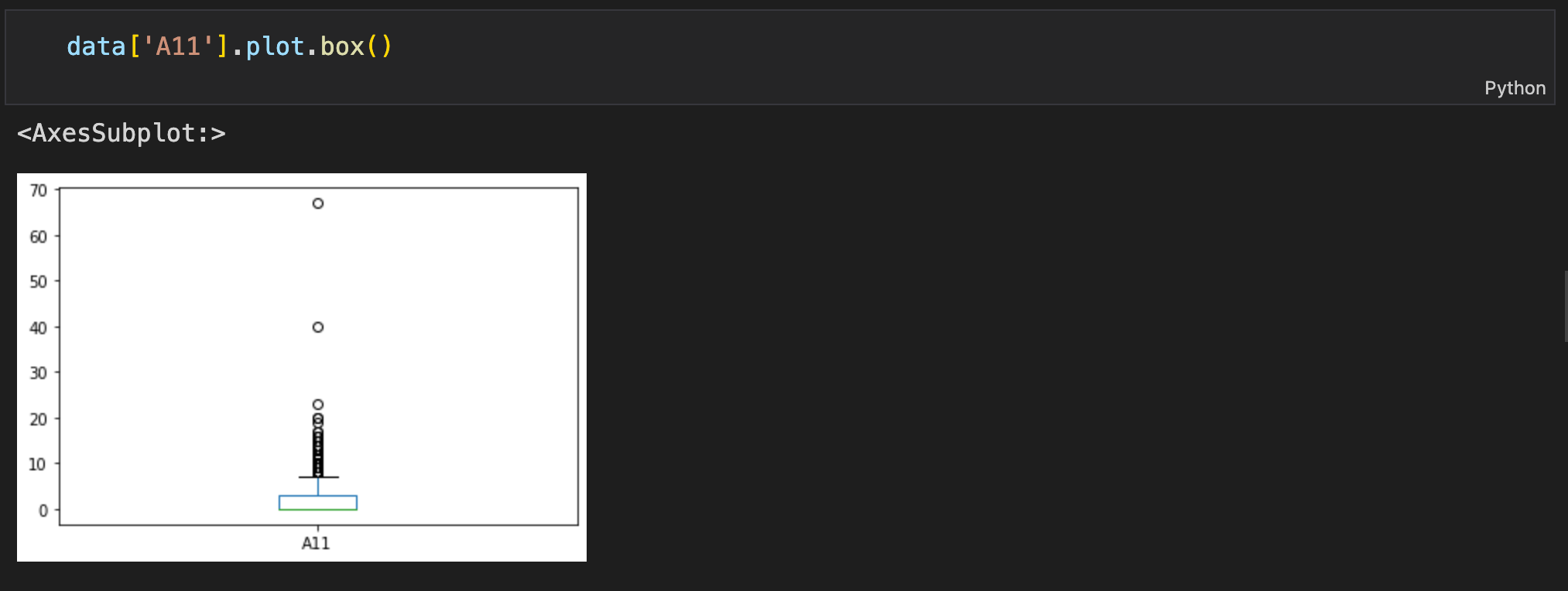


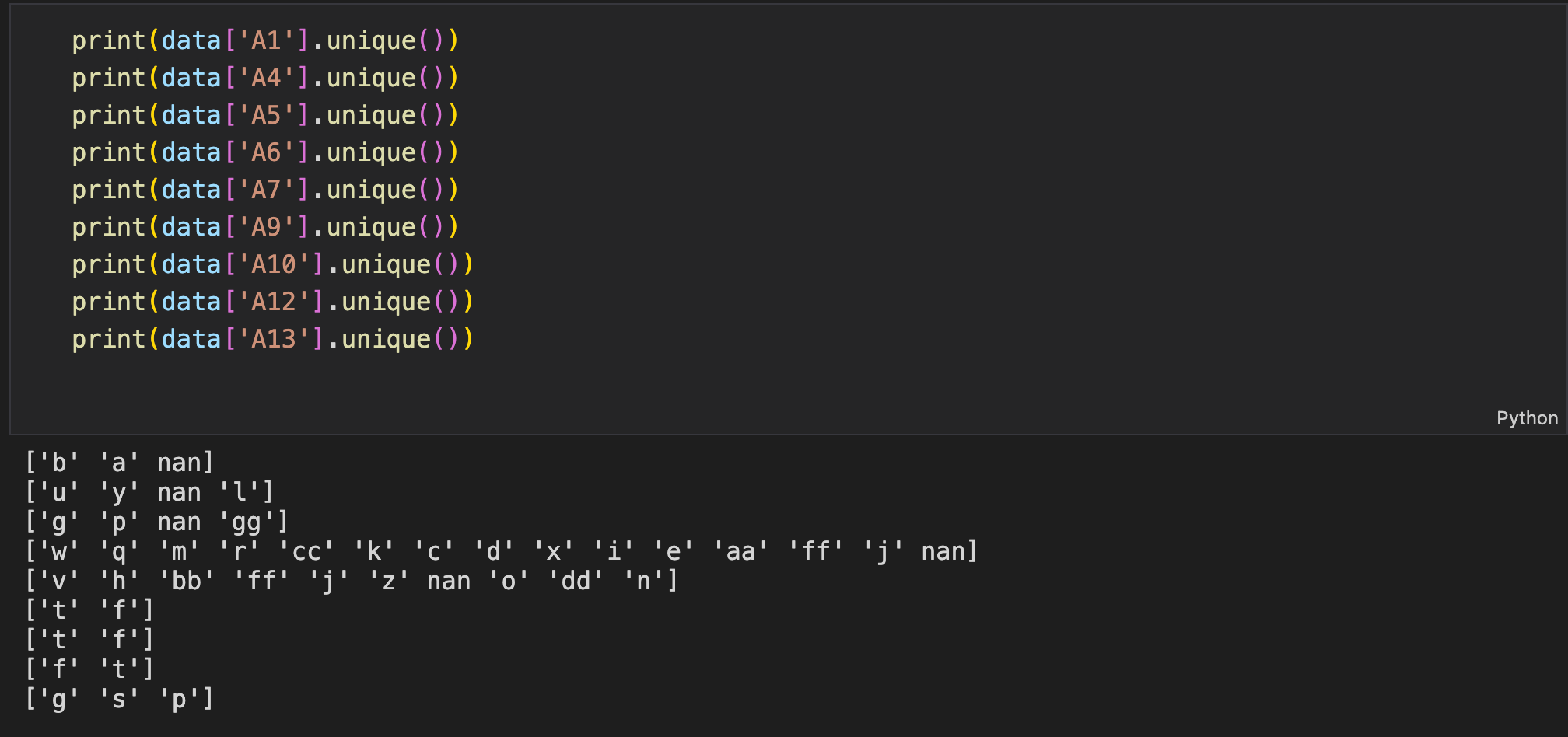




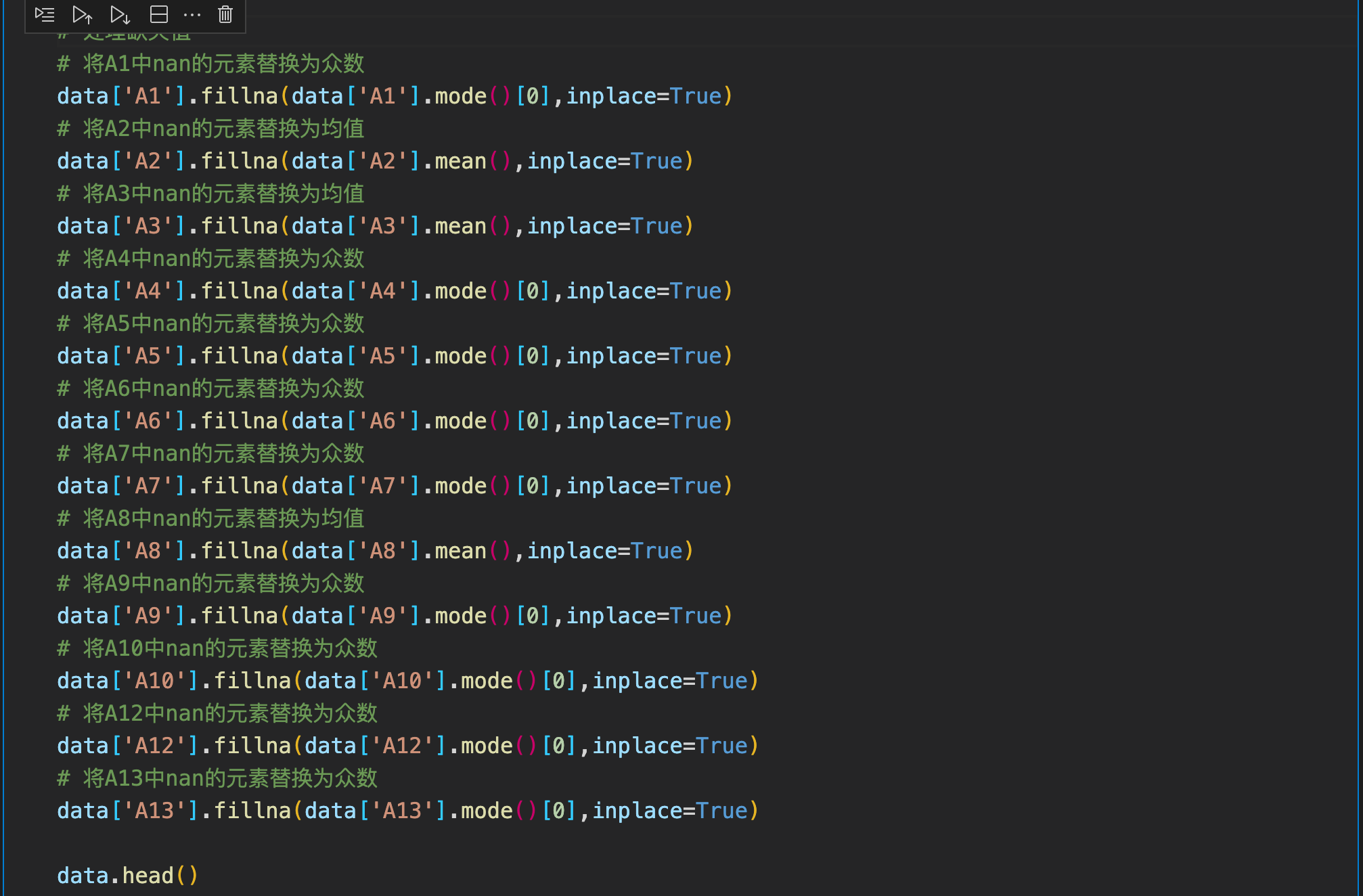


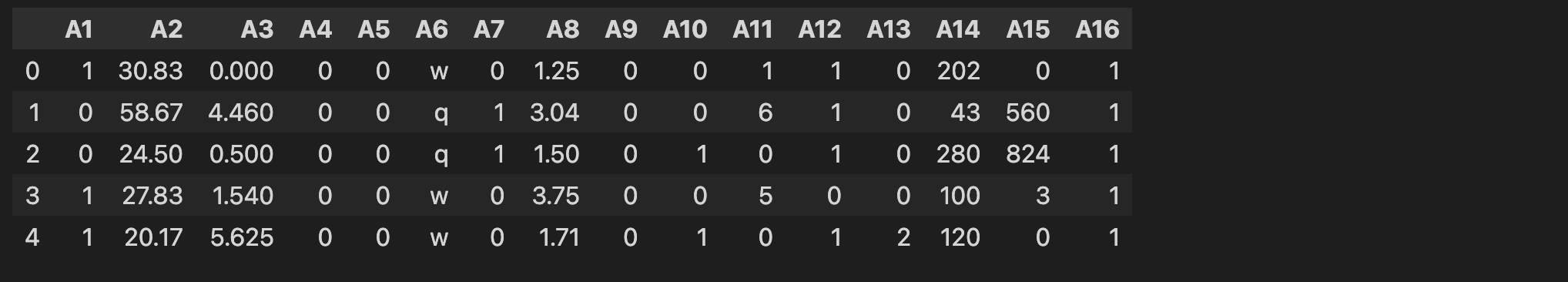




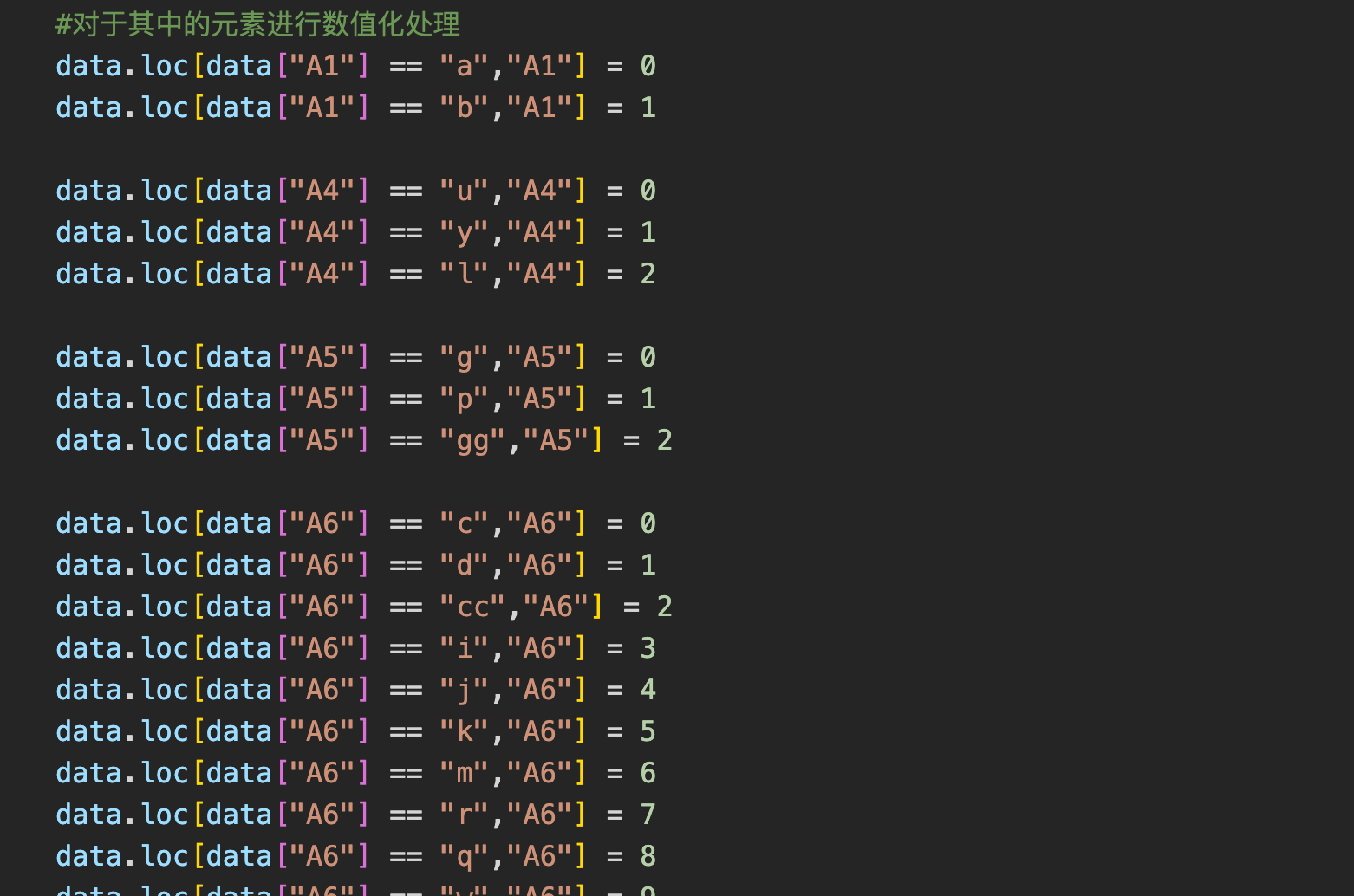


1. 恰当处理缺失值





1. 对非数值属性，编程替换为数值





（5）对数值做必要的归一化

按步骤截图，添加必要的说明文字

特别是采取某个处理策略时，应说明“理由和处理的合理性”

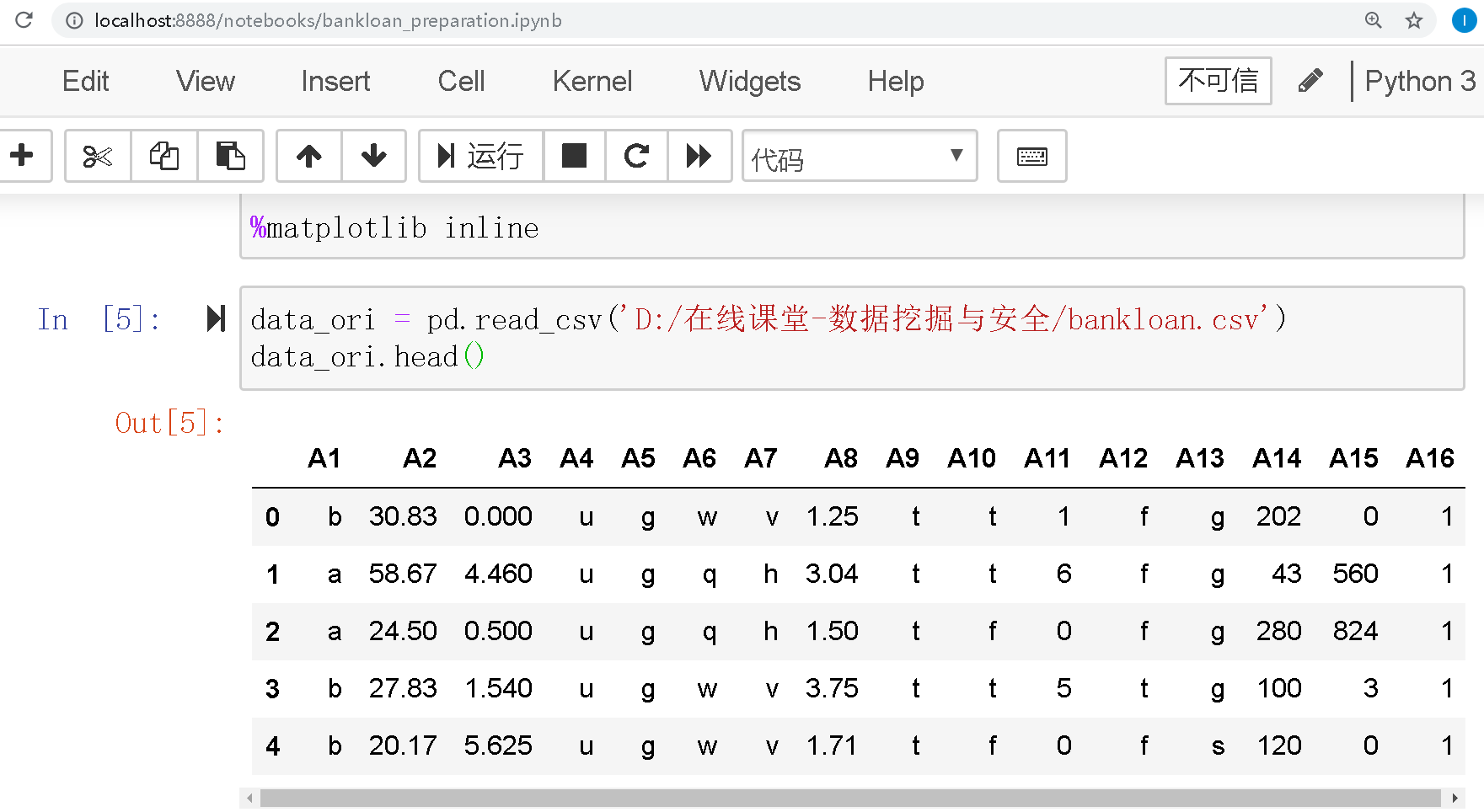
截图示例：

**要求**

**（1）不要截取满屏，不要带有windows桌面背景**

**（2）截图中的关键内容应清晰**

**（3）输出结果，要尽量避免满屏黑色，应包含结果全部内容**



4．分析与讨论

（1）填充缺失值可以采取哪些策略？

（2）何时需要对数据做规范化？本实验数据，适合哪种规范化方法？

回答问题，记录个人知识总结、调试经验等

5．附录

附实现代码