数据集选用Building Permits。采用标称属性进行挖掘，丢弃数值属性。首先按实验1中操作，用‘N’填补Structural Notification，Voluntary Soft-Story Retrofit，Fire Only Permit，TIDF Compliance，Site Permit的缺失项。这些字段除TIDF Compliance外均为布尔型，Y表示是，空表示否，因此可以用N填充。TIDF Compliance字段只有两条记录不为空，一条值是Y，一条是P，或许空也表示否，可用N填充。此部分代码由python实现。

频繁项集挖掘代码由C++实现，采用apriori算法，但是由于数据集过大，程序在支持度2%，置信度50%的配置下执行数小时依然没有结果，只得减小数据集规模进行测试，然而就算只读入2000条数据，执行时间也十分长，而提高支持度、置信度又会导致无关联规则产生。考虑到算法会对1-项集进行全连接，所以执行效率较低，而我又并不掌握多线程编程的技术，难以在短时间内速成，所以就我个人的编程能力而言，此实验在此数据集上恐难以完成。

提交文档为python和C++代码，以及python程序生成的文件，此文件首行的数据条数需要人工录入，此值讲控制C++程序读入的行数。