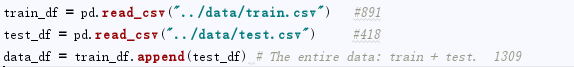
1. 数据预处理

**此处注意数据的处理方式：  
将训练集和测试集合并成数据data\_df，方便进行统一处理  
但是在处理完成后，再将训练集和测试集按照891的长度分开**



年龄(age)属性：

**从Name属性中提取出姓氏Title,  
按照Title分组，并使用各组Age的中位数替换该组Age的空值，  
年龄的并未有很大的影响**

**Mlle,Major,Col,Sir,Don等Title名都是很少出现，因此将其替换为常出现的title**

增加家庭成员数量： **Adding Family\_Size,That's just Parch + SibSp.**

增加家庭存活人数：

**Adding Family\_Survival**

**Name: Family\_Survival, Length: 1309, dtype: float64**

**Number of passengers with family survival information: 420**

**Number of passenger with family/group survival information: 546**

将船票价格离散化：

**将船票价格Fare离散化，分为五个档次**

将年龄离散化: **注意：若要用中位数/均值等替代，最好是取整个数据样本**

将性别做01映射:

丢弃多余属性：  
此时剩下属性：Survived Pclass Sex Family\_Size Family\_Survival FareBin\_Code AgeBin\_Code

1. 训练模型
2. knn网格搜索的最佳模型（0.83253）
3. 流程图

