数据分析：

1总览

ID:  
•Survived(存活与否)  
•Pclass（客舱等级，较为重要）  
•Name（姓名，可提取出更多信息）  
•Sex（性别，较为重要）  
•Age（年龄，较为重要）  
•Parch（直系亲友）  
•SibSp（旁系）  
•Ticket（票编号）  
•Fare（票价）  
•Cabin（客舱编号）  
•Embarked（上船的港口编号）

查看前几项 查看是否存在null值

#### 2使用统计学与绘图对数据初步分析

1. 乘客等级更高 获救的几率越大
2. 女生获救概率更高
3. 有旁系亲友存活率高
4. 有一个兄弟姐妹的一般有更高的生存率
5. 婴儿的生存率奇高
6. 有cabin存活率高
7. 途径港口上船的人生存几率较大

#### 3数据清洗

1. 查看缺失值
2. 使用随机森林模型补齐age特征
3. 对cabin实行二分类—yes/no
4. 对cabin embarked sex pclass实行编码
5. 并对age fare实行归一化

#### 4建模和优化

对比多个模型，选择效果最好的，网格搜索调参。

#### 5进行预测

使用调参好的模型进行预测，将预测值放入logic.csv文件中。