具体的步骤是在做什么在code当中由批注展示，本文件大致描述一下步骤以及code中一些变量是什么。

Code的步骤：

1. 对train和test进行了data cleaning，把NaN转成了0.
2. 将string的feature通过pd.factorize转成了计算机可识别的数值，并为了保证train和test对应的编码一致进行了一定的调整。
3. 将无用的数据去除将数据转成了矩阵的形式，并把test文件分为了test和predict
4. 引入sklearn中的logistic regression model进行训练和预测。

相关变量：

1. dfc：dataframe of train.csv
2. dfct: dataframe of test.csv
3. dfc=dfc.values后的dfc：train的feature矩阵
4. test：test的feature矩阵
5. Predict：predict的数据矩阵

上述矩阵前7列是feature，最后一列是label即Correct First Attempt