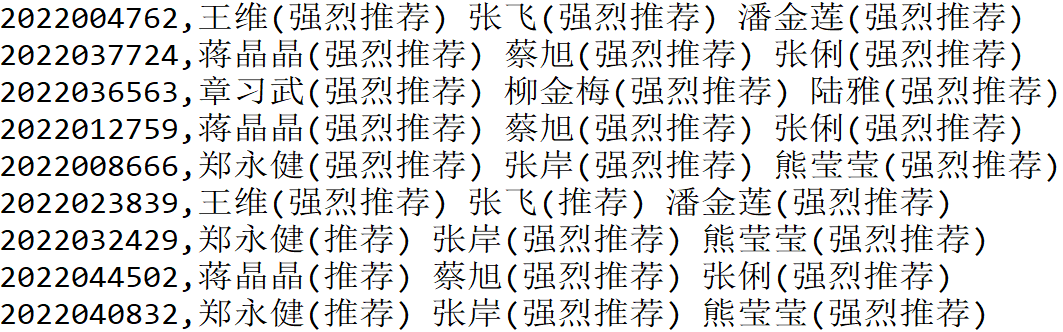
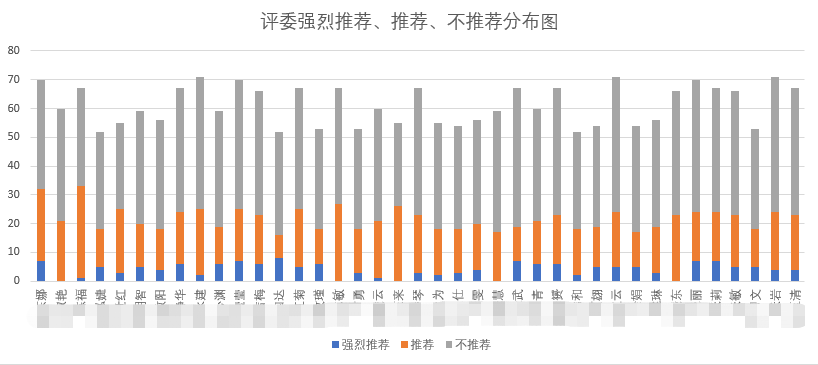
《大数据计算机基础》2022秋季学期作业1

**以Python为主，Excel为辅，利用压缩包提供的contestData.txt和work\_cat.csv两个数据文件(编码均为utf-8)，完成如下作业。如果Excel不会，请自学。**

1. contestData.txt是一次竞赛产生的数据，其内容如下。第1列是作品编号，后续为评委姓名以及该评委给该件作品的推荐等级，每件作品都有三位评委。假定“强烈推荐”为2，推荐为1，不推荐为0。
2. 基于两份给定的数据，产生两个数据文件，分别是runResult00.txt，第一列为标题，分别是作品编号、得分构型、作品类别；runResult01.txt,第一列为标题，分别是作品编号，评委姓名，推荐等级。
3. 基于runResult01.txt，在Excel中，通过“数据透视表”功能，生成每个评委给出的强烈推荐、推荐及不推荐的数量。即第一列为姓名，以后各列分别是强烈推荐、推荐、不推荐，相对应单元格为对应次数，并以柱状图形式分别显示出每位评委推荐等级的分布，如下图所示，能很明显看出评委风格。



1. 以runResult00.txt生成如下统计图及其相关数据。注意构型中，210表示一位评委强烈推荐，一位评委推荐，一位评委不推荐。210这样的三位数中，等级最高在前，即不能出现201,021这样的构型，要做数据规范。因为数量不大，可以穷尽列出所有组合可能，然后生成字典对应的标准构型。



1. 程序以hw00\_序号.py命名，如果有多个程序文件，分别命名为hw00\_00.py、hw00\_01.py、……等；所有Excel操作集中到一个文件，命名为hw00.xlsx。Excel操作可以用WPS完成。上述文件打包压缩成hw00.zip，在教学网上提交。提交截止时间：2022年10月6日23:59前。
2. 通过上述操作，可以总体把握每位评委的评审风格，也能总体了解作品质量以及各种分布等，利于决策，评估风险，比如：200要重点监督，一位强烈推荐，两位不推荐，这件作品是争议很大还是一位评委被公关等等。