**《Python 数据分析》项目考核**

**一、任务描述**

reviews.csv 是电视剧《安家》的影评数据集，请针对该数据集完成以下数据分析任务：

1. 统计不同Rating 的数量，并进行可视化。
2. 统计评论数量的时间变化趋势，并进行可视化。对变化趋势给出解释。

安家从2月21日开始播放，播放第三天评论数量就一下子冲到40+，说明这部剧在刚开始播出时还是受到很大的关注。随之评论量降低，但是在3-17到3-22之间又来了一拨小高潮，应该是故事到了高潮部分。随着安家结局后，评论数逐渐下降。

1. 分析用户评论的活跃时间段，时间段的长度需结合数据探索自行设定，并给出理由。

用户评论的活跃时间段为2-22-3-22之间。在这段时间内，每天的评论数都保持在20+。

1. 设计指标，度量影片中各个角色受关注的程度。说明度量指标的设计原理，并计算出各个角色的受关注程度。

度量各个角色受关注的程度的指标=角色在评论中出现的次数 / 总评论人数

理由： 角色越被提到次数越多，该角色越受关注

1. 设计指标，度量影片中各个角色受欢迎的程度，说明度量指标的设计理由，并计算出各个角色的受欢迎程度。

欢迎程度 = 角色总得分 / 评论总分

理由：根据不同的评价赋予不同分值，最终的总得分越高，说明该角色越受欢迎

1. 对受关注度最高的 5 个角色，计算他们的受欢迎程度随着时间的变化趋势（需进行可视化），并分析原因。
2. 设计指标，度量影片中的每个角色，与其他角色的关联性。说明度量指标的设计理由，并计算每一对角色之间的关联性指标。

关联性 = 统计两个角色同时出现在同一个评论中的次数然后归一化处理。

理由：能在同一个评论下出现，说明这两个角色必然是存在联系的。统计他们同时的次数，便可以确定关联性的强弱。

1. 自主探索（10 分）
2. rewiews.josn 是以JSON 数据格式存储的影评数据集，请将

其转换为csv 文件格式。

1. 对评论进行分词，统计与“徐文昌”和“房似锦”有关的有意义词语，并绘制词云。