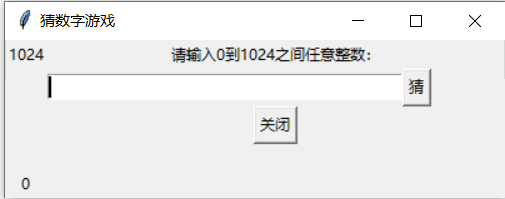
**《大数据开发语言课程设计》选题参考**

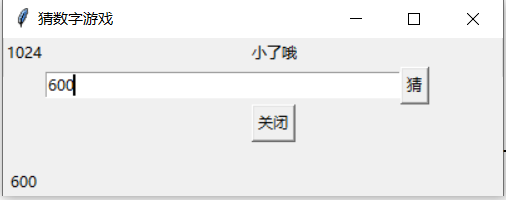
**提示：**请在代码中写好注释

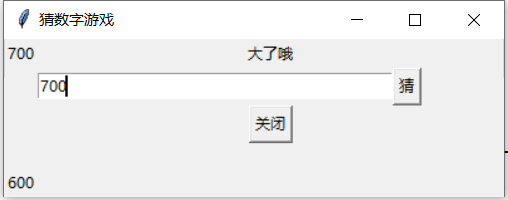
1. 猜数字游戏设计（难度系数：3）

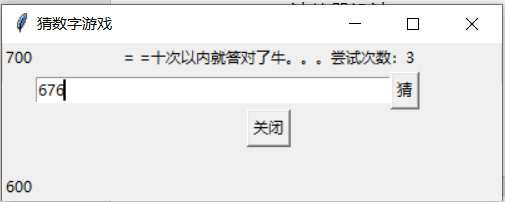
要求：程序随机生成1024以内的数字，让玩家去猜，如果猜的数字过大或过小都会进行提示，程序还会统计玩家猜的次数。使用python内置的Tkinter模块开发游戏界面。（备注：程序存储为.py文件）

游戏参考界面如下：





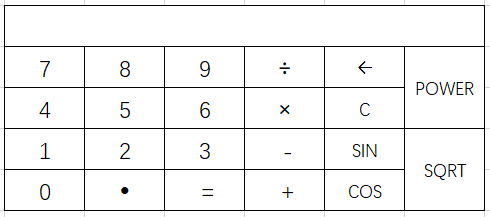




1. 计算器设计（难度系数：4）

要求：程序可以进行加减乘除四则运算，并可以计算幂次方、余数，正、余弦和自然对数。使用python内置的Tkinter模块开发游戏界面。（备注：程序存储为.py文件）

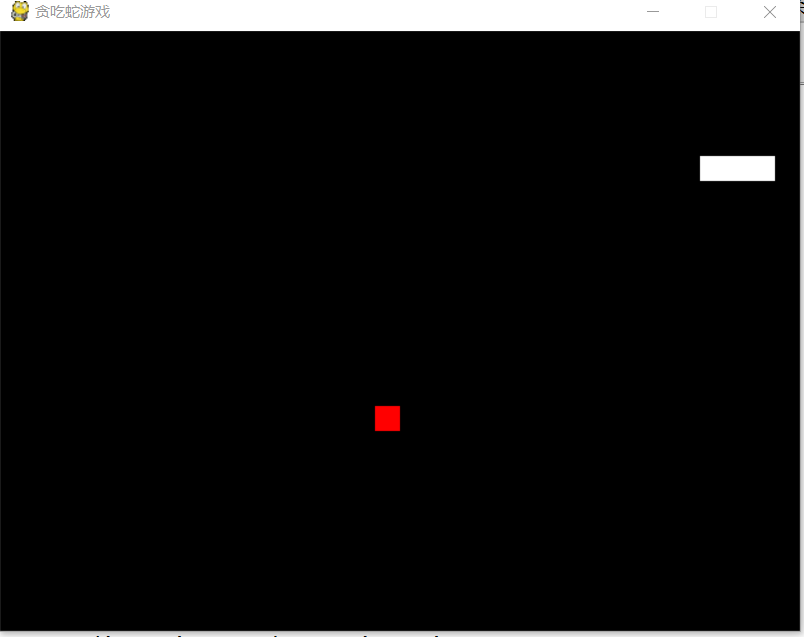
游戏参考界面如下：



1. 贪吃蛇游戏设计（难度系数：6）

要求：通过玩家控制蛇移动，不断吃掉食物（如草莓，豆子等）后，蛇身增长，直到蛇身碰到边界游戏结束。使用python的Pygame模块进行游戏开发。（备注：程序存储为.py文件）

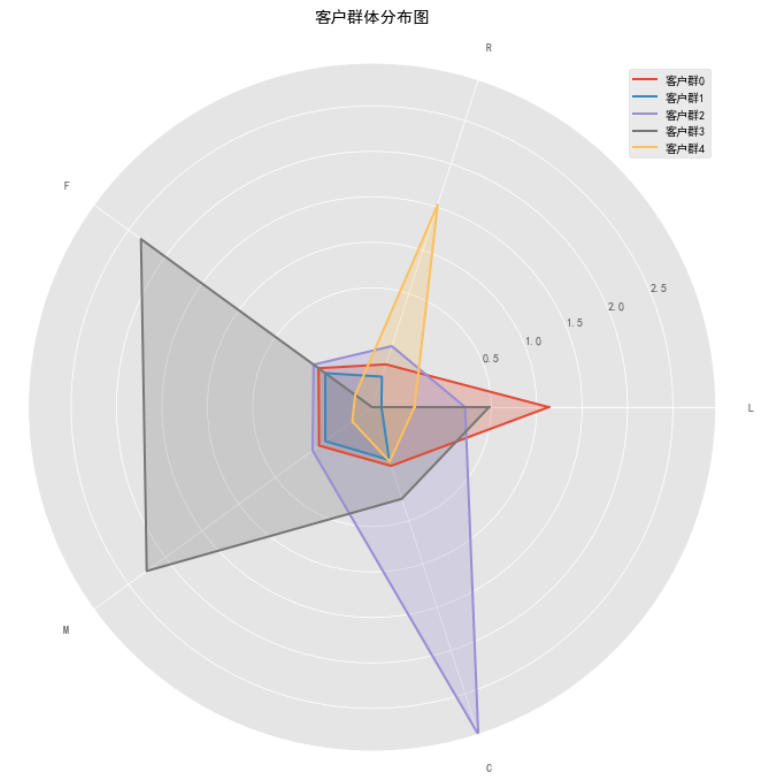
游戏参考界面如下：



1. 航空公司客户价值分析（难度系数：8）

要求：利用pandas、sklearn、matplotlib对air\_data.csv中的数据进行航空公司客户价值分析。主要包括：

1. 利用pandas读取air\_data.csv文件中的数据。
2. 利用pandas进行数据探索，如利用shape属性，head( )和describe( )函数等。
3. 利用pandas进行数据预处理：丢弃票价为空的记录，保留票价非0，或者平均折扣率不为0且总飞行公里数大于0的记录。
4. 利用pandas构建LRFMC模型，以2014-03-31为结束时间。其中：L为客户关系长度（会员入会时间距观测窗口结束的月数），R为消费时间间隔（客户最近一次乘坐公司飞机距观测窗口结束的月数），F为消费频率（客户在观测窗口内乘坐公司飞机的次数），M为飞行里程（客户在观测窗口内累计的飞行里程），C为折扣系数的平均值（客户在观测窗口内乘坐舱位所对应的折扣系数的 平均值）。
5. 利用sklearn进行数据标准化处理
6. 利用sklearn中的KMeans聚类算法，对客户进行分群处理，并找出聚类中心和聚类类别。
7. 利用matplotlib绘制雷达图，进行客户群体分析。



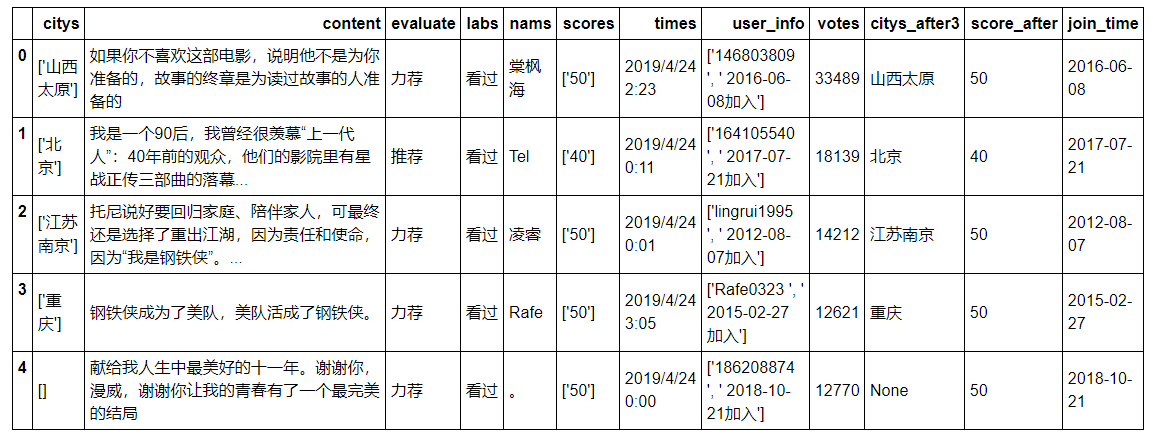
1. 结合雷达图，给出分析结论。并针对不同价值的客户类别给出个性化服务建议，及相应的营销策略。

（备注：请使用Jupyter Notebook）

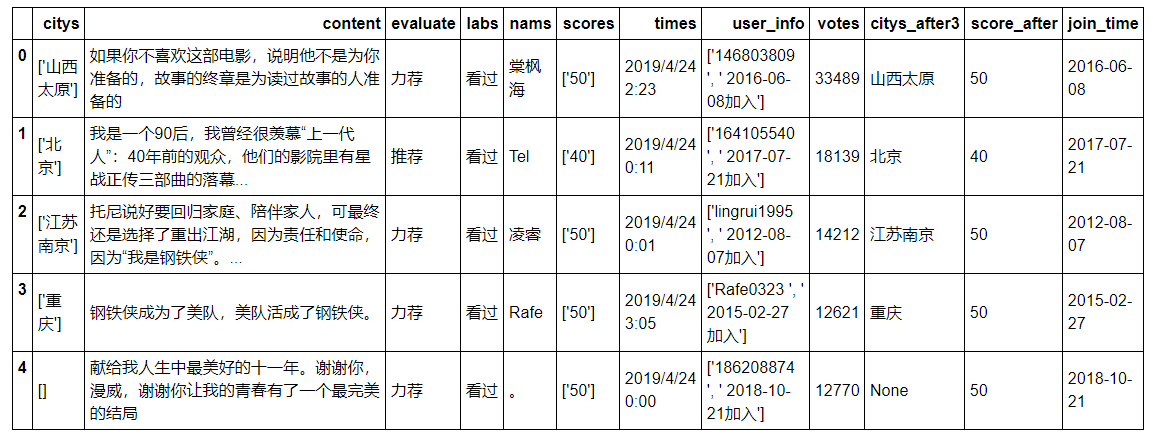
1. 豆瓣影评数据分析（难度系数：8+2）

要求：利用requests或其他爬虫库爬取豆瓣影评数据，使用xpath提取所需信息，并pandas进行数据分析、利用matplotlib和wordcloud进行可视化展示（爬虫属于选做内容）。主要包括：

1. 爬取影评数据，《复仇者联盟4》的网址为：<https://movie.douban.com/subject/26100958/comments?status=P>
2. 使用xpath提取用户名、发表时间、评分、投票数量、短评正文、常居城市和入会时间的内容。
3. 使用pd.DataFrame将提取的数据转换为DataFrame类型的数据。
4. 使用data.to\_csv将数据存储到douban.csv文件中。
5. 利用pandas读取douban.csv文件中的数据。
6. 利用pandas进行数据探索，如利用shape属性，head( )和describe( )函数等.
7. 使用re.findall进行数据预处理。



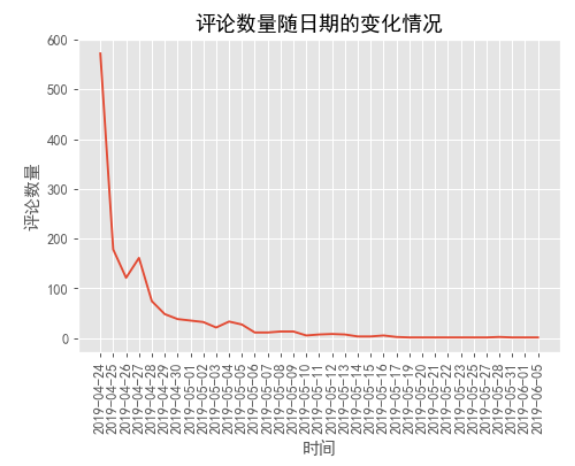
1. 使用pandas进行特征提取。

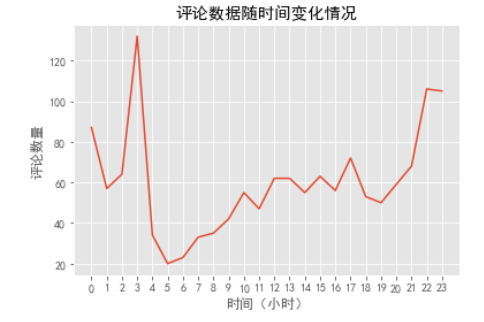


1. 使用jieba.load\_userdict添加用户自定义词典（dict.txt）
2. 对content内容进行分词处理。
3. 依据stoplist.txt文件去除停用词。（若是stoplist.txt中没有包含你要去除的停用词，请自己添加）
4. 对content分词后的内容进行词频统计。
5. 利用matplotlib、wordcloud绘制词云，并显示。依据评分分为好评、中评、差评，分别做词云展示，其中，10-20:差评 30：中评 40-50：好评。

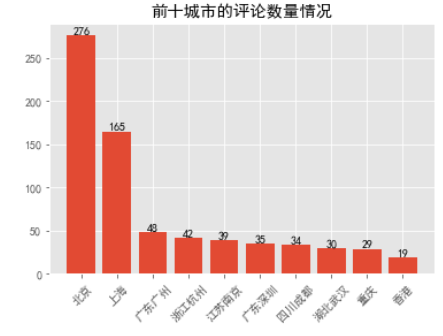


1. 依据times中的数据进行数据分析，利用pandas绘制评论数量随日期的变化情况，以及评论数量随时辰的变化情况。





1. 依据城市信息对常居城市进行数据分析，利用pandas绘制常居城市分布情况。



（备注：请使用Jupyter Notebook）