题目十八：2000-2015 年电影评分数据分析

【数据说明】

数据集“tmdb_5000_movies.csv”包含 20 个字段，4803 行，每一行代表的是一个电影，具体信息如下：

序号	字段名	数据类型	字段描述
1	Budget	Numeric	预算
2	Genres	String	流派
3	Homepage	String	主页
4	Id	Numeric	电影编号
5	Keywords	String	关键词
6	original_language	String	初始语言
7	original_title	String	初始标题
8	overview	String	概述
9	popularity	Numeric	受欢迎程度
10	production_companies	Numeric	发行公司
11	production_countries	String	发行国家
12	release_date	Date	发行日期
13	revenue	Numeric	收入
14	Runtime	Numeric	上映时间
15	spoken_languages	String	语言
16	Status	String	状态
17	Tagline	String	标语
18	Title	String	标题
19	vote_average	Numeric	平均评分
20	vote_count	Numeric	评分人数

【任务】

1、用 pandas 库读取“tmdb_5000_movies.csv”文件，查看前三行、后两行。

2、用 pandas 数据预处理模块将缺失值丢弃处理，选择列 Id、release_date、title、vote_average、vote_count，并导出到新的 csv 文件“tmdb_5000_movies_vote.csv”。

3、利用 pandas 库重新读取新的数据集“tmdb_5000_movies_vote.csv”，按照字段 vote_average 降序排列所有数据集，导出为文本文件“tmdb_5000_movies_vote_descending.txt”，要求数据之间用逗号分隔，每行末尾包含换行符。

4、重新读取新的数据集“tmdb_5000_movies_vote_descending.txt”，选择 vote_average 字段，统计最大值 maxValue、最小值 minValue、平均值 meanValue。

5、重新读取文件“tmdb_5000_movies_vote_descending.txt”，利用上一步统计结果最大值 maxValue、最小值 minValue，利用 category = [minValue, 5, 7,9，maxValue]和 labels = ['bad', 'ok', 'good', 'excellent']将 vote_average 进行离散化；并将离散化结果作为一个新的列 Label 添加到原始数据集，并保存为“tmdb_5000_movies_vote_descending_result.csv”文件；根据离散化结果画出饼状图，保存为“tmdb_5000_movies_vote_descending_result_ pie.png”，要求分辨率不低于 300dpi。

【要求】

1、根据以上数据处理任务，设计并编程实现“数据分析与可视化系统”，要求

① 各个任务选择用菜单实现（菜单可用字符串输出模拟，或者 Tkinter 形式实现）。

② 各个任务名称自己定义，须由独立的函数实现，且每个任务执行成功与否须给出必要的文字提示。

③ 数据输入和结果输出的文件名须由人工输入，且输出结果都要以文件形式保存。

④ 为保持程序的健壮性，各个任务执行过程中需要进行必要的判断（如文件是否存在、输入是否合法等）、程序异常控制等。

2、根据以上统计结果，书写不少于 300 字的结果分析。