**任务1 数据预处理（100分）**

在做数据统计之前，需要对数据进行预处理，具体要求如下：

读取movie\_data.csv文件，以元组的形式输出行数和列数；

删除所有有缺失值的数据行，并输出被删除的行数；

删除所有重复行数据，并输出当前剩余的行数。

【答案】（每行30分，全正确100分）

(4924, 20)

1485

3429

**任务2 数据统计（100分）**

对处理好的数据进行统计，具体要求如下：

筛选出总收入大于50000000并且支出小于100000000的数据行；

输出筛选出的电影中总收入gross字段最高的数值；

输出筛选出的电影中支出budget字段最低的数值；

对筛选出的数据行按照imbd\_score值的大小进行降序排序；

将排序好的前五行数据中的gross值添加到列表中，并输出列表。

【答案】（每行30分，全正确100分）

623279547.0

60000.0

[84185387.0, 214948780.0, 50800000.0, 100614858.0, 52200504.0]

**任务3 数据更新（100分）**

将已经完成预处理和筛选更新后的数据，保存为新的文件。具体要求如下：

* 找出所有以actor开头的字段名，并添加到列表中；
* 删除所有以actor开头的数据列；
* 将最新的数据保存为new\_movie\_data.csv文件；
* 查看new\_movie\_data.csv文件的大小。

【答案】（100分）

116kb 说明：值在114 kb到118kb范围内都正确