**项目报告**

**所选项目（同学填写）**

| **数据集** | **数据集文件** | **实践者** |
| --- | --- | --- |
| （例子）项目一 | 02\_movies.csv |  |

**项目过程（同学填写）**

**数据处理的详细过程**

| **内容** | **子项** | **分析结果（结果或者代码）** |
| --- | --- | --- |
| 数据概览 | 数据行数 | 8569  print(len(array)) |
| - | 数据列数 | 6  print('column len=',array.shape[1]) |
| 数据概览 | 数据行数 |  |
| 索引器 | 获取100行3列数据 | print('100行3列\n',array[0:100,0:3]) |
| 索引器 | 获取某行数据 | print('获取第1行\n',array[1]) |
| 索引器 | 获取某列数据 | print('获取第2列\n',array[:,2]) |
| 索引器 | 获取前两行前两列数据 | print('2行2列\n',array[0:2,0:2]) |
| 缺失值 | 统计缺失值列 | col\_null = df.isnull().sum(axis=0)  print(col\_null) |
| 缺失值 | 删除某个缺失值列语言 | # 可以通过axis参数来删除含有空数据的全部列  df = df.dropna(axis=1)  print(df) |
| 缺失值 | 选取2列，填充缺失值 | df.fillna(df.mean()['title':'overview']) |
| 分组统计 | 尝试分组统计某个维度数据 | print('分组统计',df['rating'].value\_counts()) |
| 数据可视化 | 计数图(countplot) | 首先ijuypter中外面：ipython notebook --matplotlib=inline  ax = sns.countplot(x="rating", data=df)  plt.show(sns) |
| 数据可视化 | 直方图(distplot) | l = array[:,4]  ax = sns.distplot(l)  plt.show(sns) |
| 数据可视化 | 散点图(scatterplot) | sns.scatterplot(data=df, x="rating", y="popularity")  plt.show(sns) |
| 建模（加分项） | 模型建立与评估 | # create and plot models  使用多项式模型拟合populority和rating的关系  fp1, res1, rank1, sv1, rcond1 = sp.polyfit(x, y, 1, full=True)  f1 = sp.poly1d(fp1)  fx = sp.linspace(0,x[-1], 1000)  # generate X-values for plotting  plt.plot(fx, f1(fx), linewidth=4)  plt.legend(["d=%i" % f1.order], loc="upper left") |

**评测结果(老师评估)**

**评级等级**

* 优秀(S): 优秀:代码可运行/代码符合规范/统计数据准确/数据多维度显示/有分析结果说明/模型初步建立
* 良好(A): 代码可运行/统计数据准确/数据可视化
* 合格(B): 代码运行有 2 次以下错误/统计数据 2 处以内不准确/数据能可视化
* 待合格(C): 代码运行出 3 次以上错误/统计数据部分不准确

**评级**

| **评级** | **评语** |
| --- | --- |
|  |  |