**项目概述**

在本项目中，你将分析一个电影数据集，然后传达你的发现。你将使用 Python 库 NumPy、Pandas 、 Matplotlib 和 Seaborn 来使你的分析工作更加轻松。

**我需要安装什么？**

你需要安装 Python 和以下库：

* pandas
* numpy
* matplotlib

**为何要进行此项目？**

此项目将向你介绍数据分析过程，以及所有工作的相互配合。后期的纳米学位将重点关注数据分析过程的各个环节。

你将使用 Python 库 NumPy、Pandas 和 Matplotlib，这些库使你能够更加轻松地在 Python 中编写数据分析代码！不仅如此，它们也是雇主们青睐的技能！

**我将学到什么？**

完成此项目后，你将：

* 了解典型数据分析过程中所涉及到的所有步骤
* 轻松提出可用给定数据集回答的问题，并解答这些问题。
* 了解如何调查数据集中的问题，以及将数据整理成你可以使用的格式
* 练习传达你的分析结果
* 能够在 NumPy 和 Pandas 中使用向量化运算，以加快数据分析代码的运行速度
* 熟悉 Pandas 的 Series 和 DataFrame 对象，它们能使你访问数据更方便
* 了解如何使用 Matplotlib 生成图形，展示你的发现
* 请在这里查看P2\_Explore\_Movie\_Dataset： [**https://github.com/udacity/AIPND-cn**](https://github.com/udacity/AIPND-cn)

## 我要如何完成该项目？

* 该项目与数据分析入门课程相关，但根据你的知识背景，你可能无需学习全部课程即可完成该项目。

### 介绍

* 在此项目中，你需要进行数据分析，并创建一个文档分享你的发现。你应该首先了解数据集，想想它可以用来回答哪些问题。然后使用 Pandas 和 NumPy 回答你最感兴趣的问题，并编写一份报告来分享你的结论。你不需要使用推论统计或机器学习知识来完成此项目，但是你需要在报告中声明，你的结论是暂时的，可能需要进一步改进。

### 第一步 - 选择数据集

* 可以通过以下链接进行数据集下载： **[TMDb电影数据](https://s3.cn-north-1.amazonaws.com.cn/static-documents/nd101/explore+dataset/tmdb-movies.csv" \t "_blank)**

### 第二步 - 组织

* 你可以使用 Jupyter Notebook，这样，你可以在同一份文档中提交编写的代码和报告你的发现。

### 第三步 - 分析数据

* 思考你可以用数据集回答哪些问题，然后开始着手回答这些问题。 试着完成项目中的任务，你应该在你的调查里分析至少一个因变量和三个自变量。在恰当的地方使用 NumPy 和 Pandas。

### 第四步 - 分享发现

* 分析完数据以后，编写一份报告，分享你觉得最有意思的发现。如果你使用 Jupyter Notebook，则与你用于执行分析的代码一起分享发现。确保你的报告文本包含在 Markdown 单元中，以清楚地将你的注释和发现与代码分开。当然你也可以使用其他工具和软件来制作最终报告，但请确保你能以 HTML 或 PDF 文件的形式提交报告，以便轻松打开。

## 项目提交

请在这里查看P2\_Explore\_Movie\_Dataset： [**https://github.com/udacity/AIPND-cn**](https://github.com/udacity/AIPND-cn)

# 准备好提交了吗？

## 评估

我们将使用[**此评估准则**](https://review.udacity.com/#!/rubrics/1861/view)来评估你的项目。这个项目的目的是帮助你熟悉项目提交流程，让你之后的课程学习和项目提交更加顺畅。

## 提交项目

点击下方的“提交项目”按钮提交你的项目。由于各地区的网络差异，第一次打开页面的时间可能会比较长。

**请提交 ipynb 格式、html 格式、py格式的文件**。

## 问题

无法提交项目？别担心！有任何问题，都可以求助我们的班主任，或者发邮件给我们的学习助手 [**support@youdaxue.com**](mailto:support@youdaxue.com)。我们会尽快帮助你解决。

## 接下来？

我们专业的老师评审完你的项目后，你会马上收到一封邮件。在等待的这段时间里，你可以学习接下来的课程。

## 评审标准

## 数据导入与处理

1. 导入库以及数据

给分点：学生成功导入所需要的库，并通过Pandas读取tmdb-movies.csv 中的数据。

2. 了解数据

给分点：学生成功完成 5 个 tasks。

3. 清理数据

给分点：学员对数据表中的空值进行了处理。

## 根据指定要求读取数据

1. 简单读取

给分点：完成3个简单读取的tasks

2. 逻辑读取

给分点：学员完成逻辑读取中的2个任务

3. 分组读取

给分点：学员使用 Groupby 命令完成2个分组读取任务

## 绘图与可视化

1. 对 popularity 最高的20名电影绘制其 popularity 值。

给分点：对 popularity 最高的20名电影绘制其 popularity 值。

2. 分析电影净利润（票房-成本）随着年份变化的情况，并简单进行分析。

给分点：分析电影净利润（票房-成本）随着年份变化的情况，并简单进行分析。

3. 选做题

给分点：选择最多产的10位导演（电影数量最多的），绘制他们排行前3的三部电影的票房情况，并简要进行分析。

4. 选做题

给分点：分析1968年~2015年六月电影的数量的变化。

5. 选做题

给分点：分析1968年~2015年六月电影 Comedy 和 Drama 两类电影的数量的变化。