|  |  |
| --- | --- |
| 学号 | 10214518414 |



课 程 设 计

课程名称 python程序设计课程设计

|  |  |
| --- | --- |
| 题 目 | 英文名字的数据分析 |
| 专 业 | 软件工程 |
| 班 级 | \_\_\_1182 |
| 姓 名 | \_\_\_胡思凡 |
| 成 绩 | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 指 导 老 师 | \_\_\_\_\_\_\_\_\_\_\_徐勇\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

2021 年 1 月 11 日至 2021 年 1 月 22 日

武汉华夏理工学院信息工程学院

**课 程 设 计 任 务 书**

课程名称：python程序设计课程设计 指导教师：徐勇

班级名称： 软件1182 开课院、系：计算机与网络工程系

**一、课程设计目的与任务**

“python程序设计课程设计”是一个综合性的学习实践型实验教学环节，将在“python程序设计”课程的授课基础上，对python的基础语法、python 对文件的操作、python 对数据的操作、数据处理基础、数据可视化等若干个知识点进行综合运用。

python是一种具有天然开源基因的编程语言，了解开源社区和广泛的使用开源工具，也是Python学习实践的重要环节。因此在本次课程设计中，要求掌握jupyter notebook、Git等常用工具，以及github等重要开源社区的使用。

**二、课程设计的内容与基本要求**

数据分析的基本技术和方法在“python程序设计”课程中已经进行完整的讲授，由于数据分析本身是基于业务场景的，因此本次课程设计环节更加偏重于实际的业务场景的实践。通过对近期互联网热点的调查，准备了15个具体的业务场景，用于本次课程设计的具体任务场景。具体包括如下：

1. 北上广深租房状况分析；
2. 蔡某坤粉丝数及转发数据真假状况分析；
3. 地震的数据分析；
4. 英文名字的数据分析；
5. 外籍英文老师收入虚高情况数据分析；
6. 我国城市空气污染和烟花燃放的关系分析；
7. 针对996工作，程序员群体的看法的分析；
8. 吴某凡微博热点的分析；
9. 节假日长假景点人满为患的数据分析；
10. 针对荔枝的品种、销售地等维度，进行价格数据分析；
11. 分析芒果TV《我是大侦探》的观众评论数据；
12. 针对当前儿科医生的缺乏，对相关数据进行分析；
13. 著名网游《绝地求生》的数据分析；
14. 实习岗位状况的数据分析；
15. 电影《流浪地球》的观众评价的数据分析。

每三个学生组建一个课程设计小组，最后的任务输出包括代码每小组一份、课程设计报告每人一份、答辩ppt每小组一份，并进行课程设计成果答辩。小组成员均参与前述工作，但是每个人的侧重点不同。

每个课程设计小组可以从上述15个场景中选取一个作为课程设计的选题，选题中提供了待分析的数据，和现有的分析方法。各小组，通过学习和实践现有的分析方法，理解实战分析的思维过程并锻炼实际动手能力，再此基础上可以扩展更多维度的分析和数据展现形式。

每个班每个选题最多只能被两个小组选中，先选先得。同时，如果各小组发现更有意思的场景，并能够获取到相关待分析的数据，也可以申请作为选题方向。

本次课程设计的目标是培养学生的团队协作能力、对python知识点的综合运用、对实际场景的理解和适应能力、针对答辩的表达能力等。注重过程，期待成果，但不强求结果的尽善尽美。

**三、学时分配进度安排**

|  |  |  |
| --- | --- | --- |
| 序号 | 设计内容 | 所用时间 |
| 1 | 下发任务书，学生查阅资料 | 1天 |
| 2 | 组建团队，并进行选题和团队匹配 | 1天 |
| 3 | 各小组对自己的选题开始进行研究 | 2天 |
| 4 | 各小组开始准备数据并形成初步处理意见 | 1天 |
| 5 | 代码和实现分析的动手实践 | 2天 |
| 6 | 形成初步的报告书和ppt | 1天 |
| 7 | 答辩并完成报告书 | 2天 |
| 合 计 | | 2周 |

**四、课程设计考核及评分标准**

**1.设计报告要求**

课程设计报告要求逻辑清晰、层次分明、书写整洁。课程设计报告为每人一份，同一个小组的各成员的整体报告内容结构一致，但个人侧重点不同，个人着重撰写自己的工作内容，其他人的内容只要体现文档结构的完整性即可。

课程设计考核将综合考虑学生考勤和参与度、团队协作能力，过程管理能力、成果达成情况等。

**2.过程要求**

整个过程要求通过github来进行过程化发布，即阶段性地在github上提交过程结果。

**3.评分标准**

|  |  |
| --- | --- |
| **评分依据** | **评分成绩** |
| 1．团队协作能力 | 25分 |
| 2．python综合运用能力 | 25分 |
| 3．态度认真、刻苦钻研、遵守纪律 | 10分 |
| 4．过程完成、对工具的使用、对github的运用 | 20分 |
| 5．课程设计答辩逻辑清晰，内容正确 | 10分 |
| 6. 课程设计期间的课堂考勤、创新能力 | 10分 |
| 总分 | 100分 |

注：按上述六项分别记分后求和，根据小组成员贡献率综合评定，记载个人最后成绩。

成绩等级：优（90分—100分）、良（80分—89分）、中（70分—79分）、及格（60分—69分）、60分以下为不及格。

**五、指导地点与时间**

本课程设计将安排在第20-21周，采用腾讯会议和QQ群的方式，以在线形式进行。具体安排如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 周次 | 星期一 | 星期二 | 星期三 | 星期四 | 星期五 |
| 第20周 | 第5-8节 | 第5-8节 |  | 第5-8节 |  |
| 第21周 |  | 第5-8节 |  | 第5-8节 |  |

执笔:徐 勇 日期：2021-1-8

审阅:钱小红 日期：2021-1-8

**目录**

[一．项目背景 1](#_Toc16921)

[二．软件开发环境 1](#_Toc14435)

[2.1工作平台 1](#_Toc3944)

[2.2开发环境 1](#_Toc12444)

[2.3开发需要安装的包 1](#_Toc7783)

[2.4进程汇报 1](#_Toc21570)

[2.5实现手段 1](#_Toc6824)

[三．数据分析及处理 1](#_Toc7811)

[3.1数据分析目标 1](#_Toc8448)

[3.2数据处理 2](#_Toc30155)

[四．数据分析 2](#_Toc3920)

[4.1 2010-2017年,Top15的男生、女生英文名 2](#_Toc13568)

[4.1.1 2010-2017年,Top15的男生英文名 2](#_Toc32613)

[4.1.2 2010-2017年,Top15的女生英文名 3](#_Toc30540)

[4.2 每个年代，最流行的英文名字 5](#_Toc11782)

[4.2.1每个年代，最流行的男性英文名字 5](#_Toc887)

[4.2.2每个年代，最流行的女性英文名字 5](#_Toc22392)

[4.3 2000年后越来越流行的名字 6](#_Toc23271)

[4.3.1 2000年后越来越流行的男性英文名字 6](#_Toc7411)

[4.3.2 2000年后越来越流行的女性英文名字 7](#_Toc20921)

[4.4 影响美国人取名的因素 8](#_Toc25305)

[4.5 top100名的英文名 9](#_Toc148)

[五．数据分析 11](#_Toc21948)

[课程设计成绩评定表 12](#_Toc9767)

**一．项目背景**

项目主要分析美国婴儿取名数据，为我们取英文名提供一些建议。

**二．软件开发环境**

**2.1工作平台**

jupyter notebook

**2.2开发环境**

python3.8

**2.3开发需要安装的包**

pandas、numpy、scip、matplotlib、seaborn、pyecharts

**2.4进程汇报**

Github

**2.5实现手段**

主要使用matplotlib来进行信息的图形化显示。

包的安装是在cmd中通过输入” pip install <包名> (==版本号)/\*下载指定版本\*/” 来进行下载。

注意: pyecharts包需要下载0.5.11版本的，下载的1.0版本的一些用法与0.5版本的不同。

**三．数据分析及处理**

**3.1数据分析目标**

1. 2010-2017年，Top15的男生、女生英文名
2. 每个年代，最流行的英文名字

3. 以前很流行，现在不流行的名字

4. 2000年后越来越流行的名字

5. 影响美国人取名的因素(体育明星、电视明星)

6. 为什么同一发音的名字，有很多不同的拼写变体？

7. 一些具有特殊含义的名字，有多少人取？

8. 一些特殊的名字

9.把所有州的姓名合在一起，查看所有州一起英文名使用情况和使用最高的前一百个名字。

**3.2数据处理**

1.通过import导入相关的包

2.通过data =pd.read\_csv(‘NationalNames.csv’),来读取csv文件

3.通过data .head()读取前5行的数据,观察数据是否干净(干净则直接使用,不干净则需要进行数据清洗)

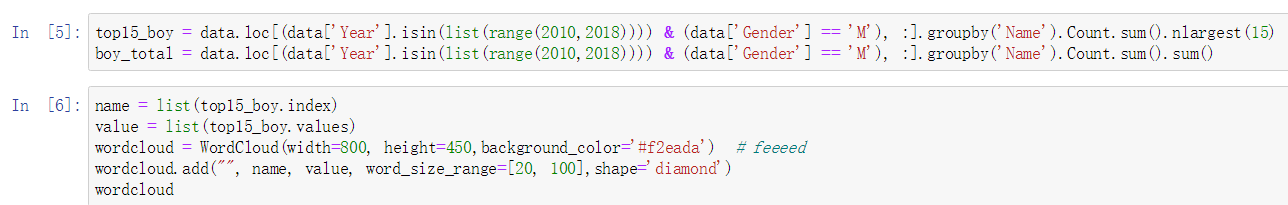
4.通过data .info()获知样本数据的相关信息概览:行数，列数，列索引，列非空值个数，列类型，内存占用。

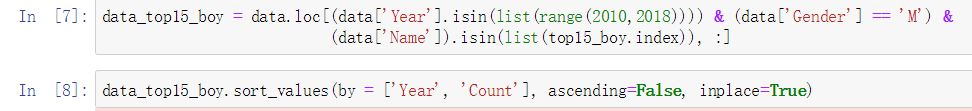
**四．数据分析**

**4.1 2010-2017年,Top15的男生、女生英文名**

4.1.1 2010-2017年,Top15的男生英文名

代码如图4.1.1.1所示：





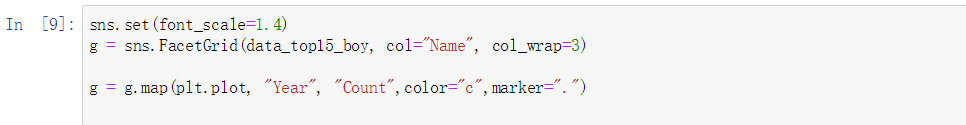


图4.1.1.1 Top15的男生英文名

词条图如图4.1.1.2所示：



图4.1.1.2 Top15的男生英文名词条图

Top15的男生英文名在2010-2017年的变化如图4.1.1.3所示：

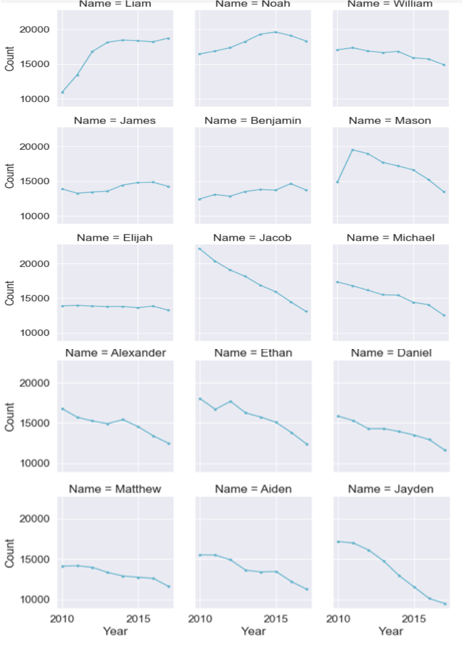
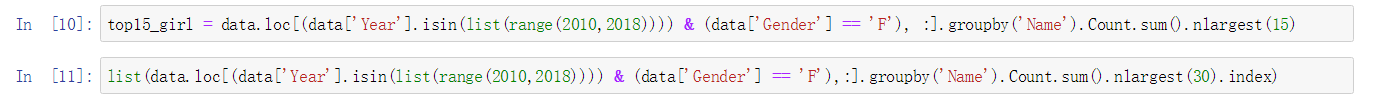
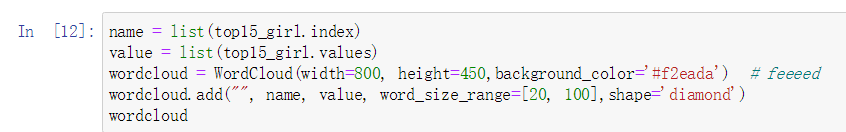


图4.1.1.3 Top15的男生英文名在2010-2017年的变化

4.1.2 2010-2017年,Top15的女生英文名

代码如图4.1.2.1所示：





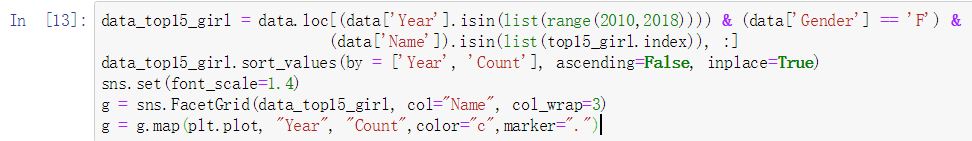


图4.1.2.1 Top15的女生英文名

词条图如图4.1.2.2所示：



图4.1.2.2 Top15的女生英文名词条图

Top15的女生英文名在2010-2017年的变化如图4.1.2.3所示：

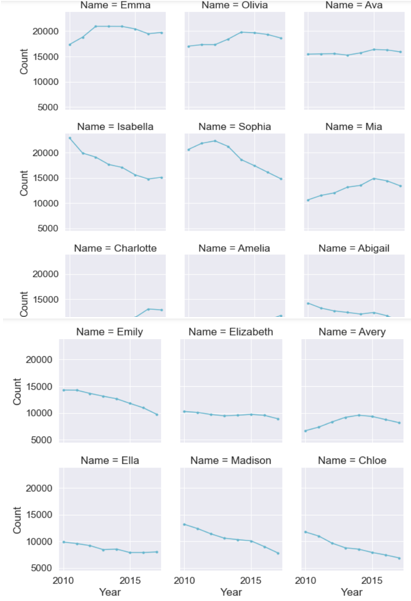
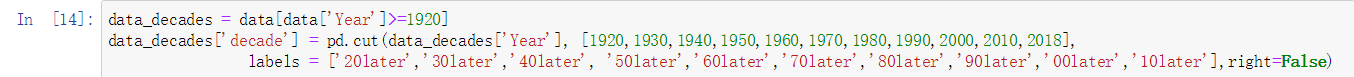


图4.1.2.3 Top15的女生英文名在2010-2017年的变化图

**4.2 每个年代，最流行的英文名字**

4.2.1每个年代，最流行的男性英文名字

代码如图4.2.1.1所示：



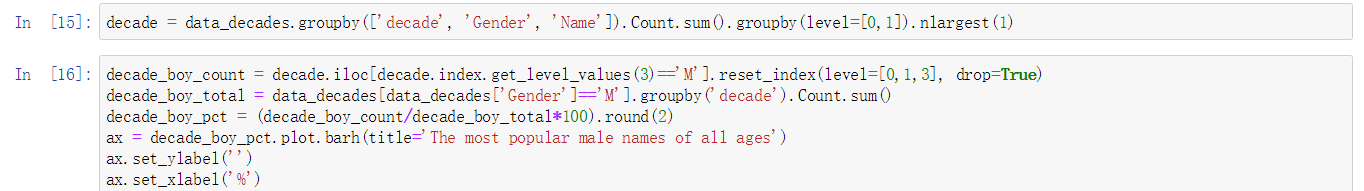


图4.2.1.1 流行的男性英文名

流行的男性英文名柱状图如图4.2.1.2所示：

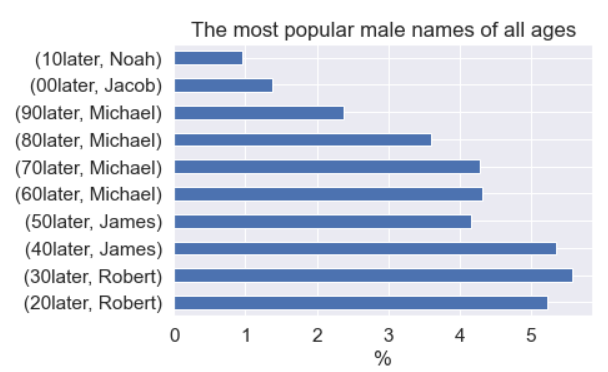


图4.2.1.2 流行的男性英文名柱状图

4.2.2每个年代，最流行的女性英文名字

代码如图4.2.2.1所示：

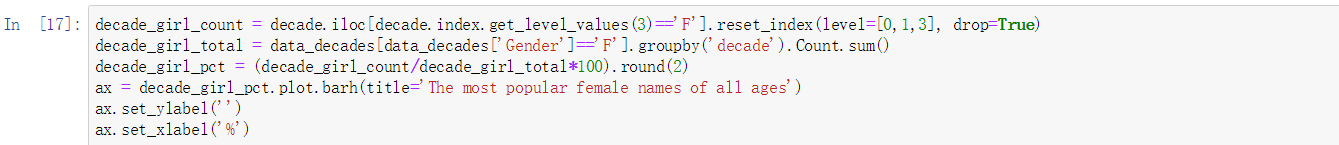


图4.2.2.1 流行的女性英文名

流行的女性英文名柱状图如图4.2.1.2所示：

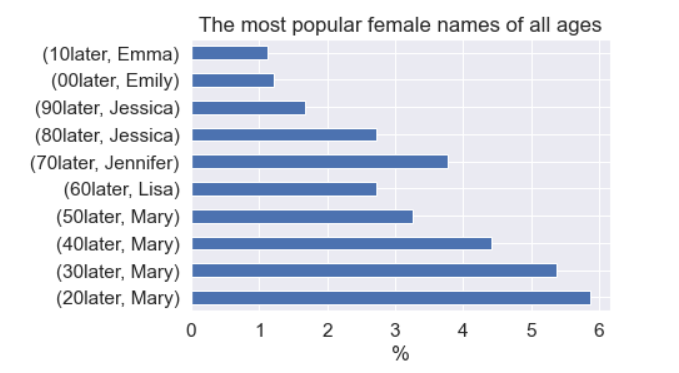


图4.2.1.2 流行的女性英文名柱状图

**4.3 2000年后越来越流行的名字**

4.3.1 2000年后越来越流行的男性英文名字

代码如图4.3.1.1所示：

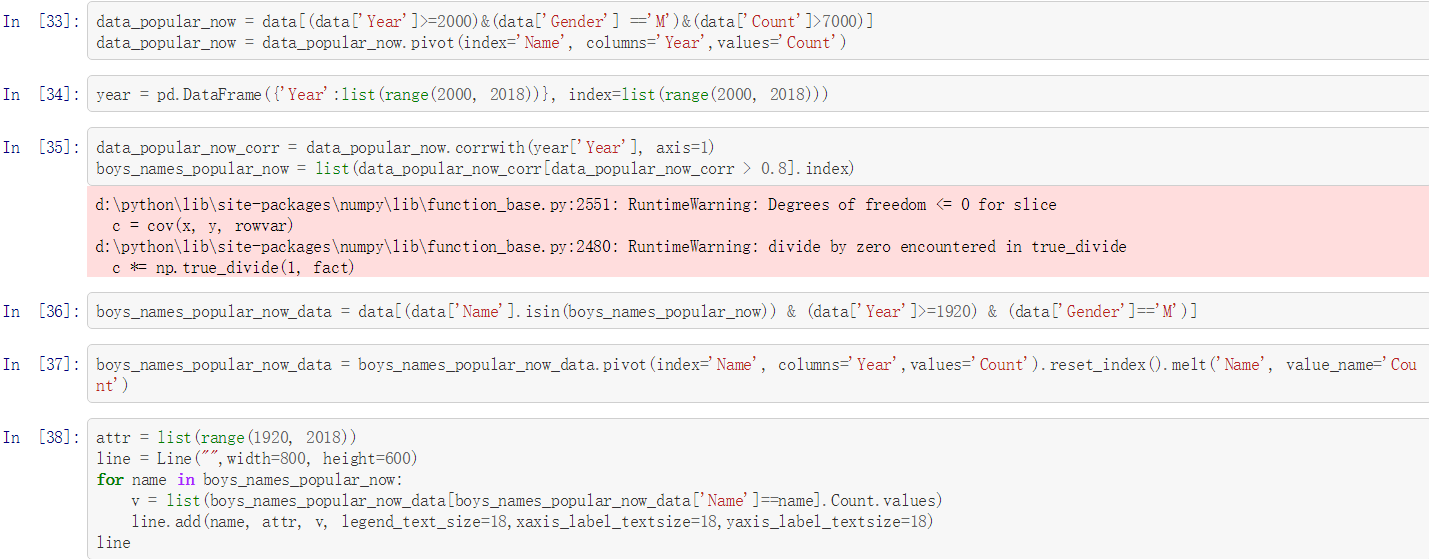


图4.3.1.1 2000年后越来越流行的男性英文名字

2000年后越来越流行的男性英文名字线性图如图4.2.1.2所示：

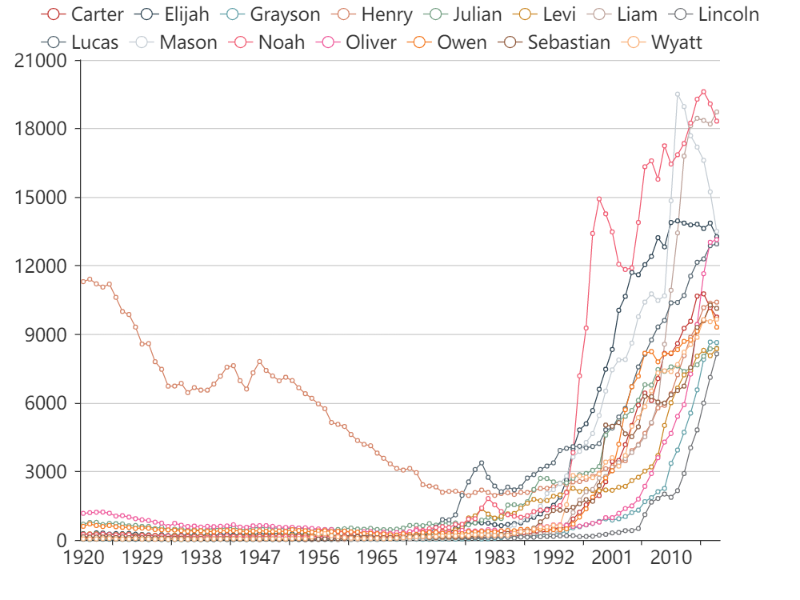


图4.3.1.2 2000年后越来越流行的男性英文名字线性图

4.3.2 2000年后越来越流行的女性英文名字

代码如图4.3.1.1所示：



图4.3.1.1 2000年后越来越流行的女性英文名字

2000年后越来越流行的女性英文名字线性图如图4.2.1.2所示：

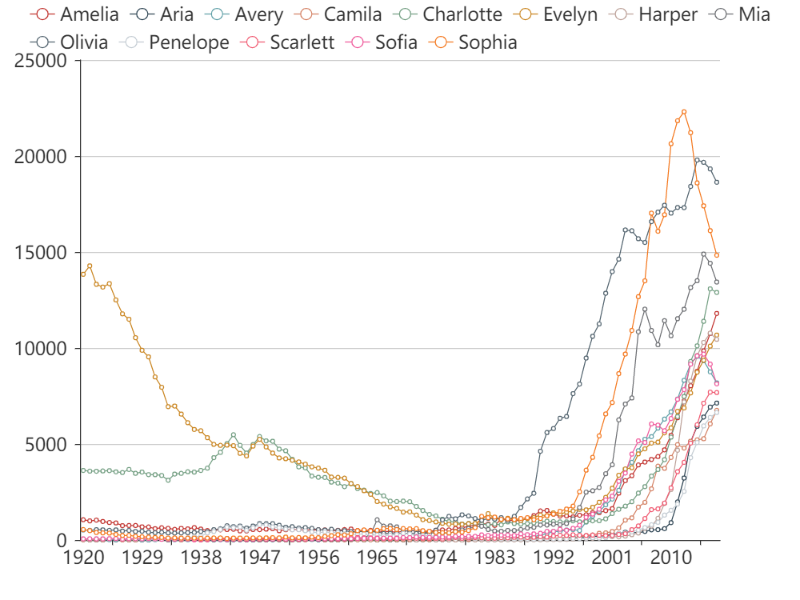


图4.3.1.2 2000年后越来越流行的女性英文名字线性图

**4.4 影响美国人取名的因素**

主要从明星这个方面判断

代码如图4.4.1所示：



38

图4.4.1 影响美国人取名的因素

(想要看不同明星对美国人的取名影响，只需改变上述代码中的明星名)

影响美国人取名的因素(Jordan)的柱状图如图4.4.2所示：

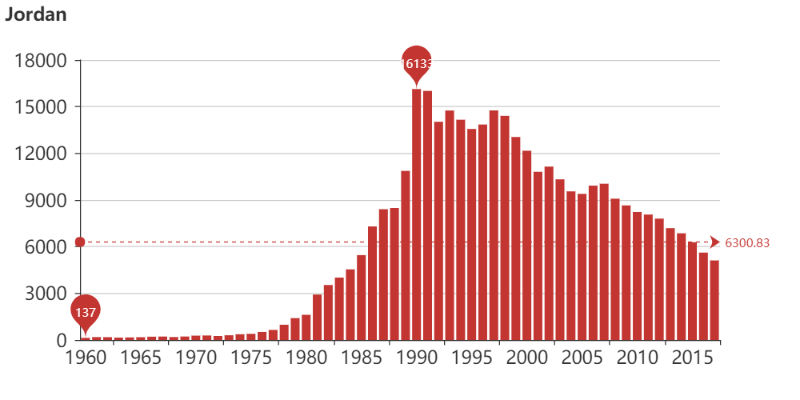


图4.4.2 影响美国人取名的因素(Jordan)的柱状图

影响美国人取名的因素(Kobe)的柱状图如图4.4.3所示：

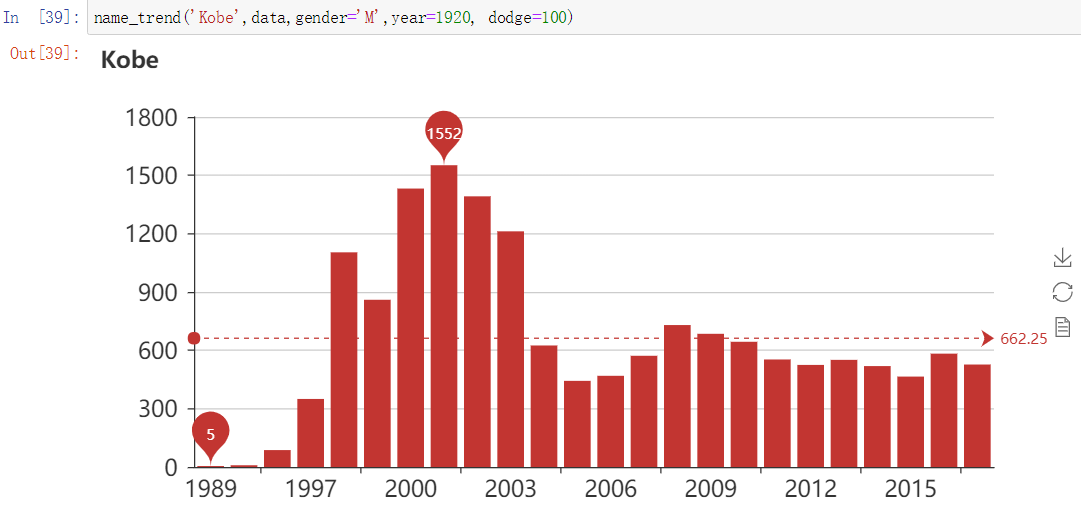


图4.4.3 影响美国人取名的因素(Kobe)的柱状图

**4.5 top100名的英文名**

把所有州的姓名合在一起，查看所有州一起英文名使用情况，查看使用最高的前一百个名字：

所有州英文名使用情况代码及表如图4.5.1所示：

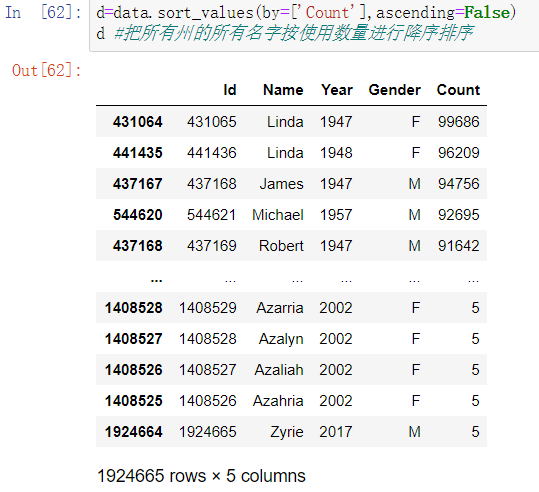


图4.5.1所有州英文名使用情况代码及表

查看使用最高的前一百个名字代码及表如图4.5.2所示：

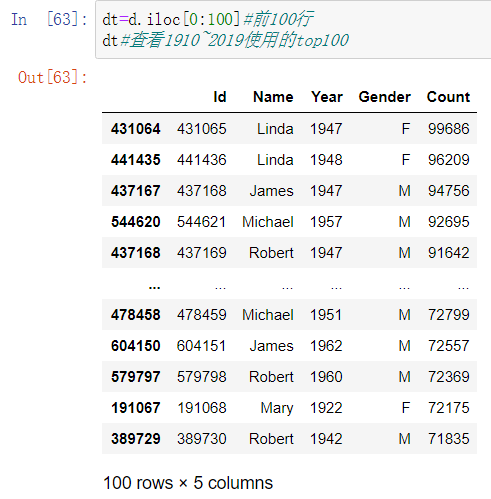


图4.5.2使用最高的前一百个名字代码及表

心得体会

通过本次的课程设计，不仅复习巩固了python语言，也在设计思考中学到了新的知识，也学会了github的基础运用与jupyter notebook的程序设计，通过本次课设，我更深刻的体会到程序设计需要缜密的逻辑思维，编写程序时需要细心谨慎，小心处理好每一个字符，包括每一个缩行、空格都需要注意。

python语言不仅仅是一个设计优秀的程序语言，它能够完成现实生活中的各种任务，包括开发者们日复一日所做的事情，还可以作为编制其他组件、实现独立程序的工具。也体会到一个成功的程序设计离不开一个团队的合作，以及给每个人分配适合的任务是十分重要的。

当然，也有一些不足之处，比如说不能显示出中文字符，最后只能用英文字符代替，虽然这也是一种解决办法，但总归不是完美的。

参考文献

[https://www.liaoxuefeng.com/wiki/896043488029600 (git的的日常使用参考)](https://www.liaoxuefeng.com/wiki/896043488029600(git的的日常使用参考))

https://docs.python.org/3/ (Python官方手册)

**课程设计成绩评定表**

|  |  |  |  |
| --- | --- | --- | --- |
| 课程设计题目 | 英文名字的数据分析 | | |
| 课程设计学生答辩或质疑记录：   1. %matplotlib inline在代码中有什么用？   答：%matplotlib具体作用是当你调用matplotlib.pyplot的绘图函数plot()进行绘图的时候，或者生成一个figure画布的时候，可以直接在你的python console里面生成图像。   1. 怎样导入数据的？   答：通过data = pd.read\_csv('<文件名>')来导入csv文件。   1. 在环境配置中，你遇到过什么问题？如果有，是怎样解决的？   答：遇到过，在包的安装时，我首先安装pyecharts包是1.0版本的，在运行时，发现运行出错，我到网上通过搜索，获知了pyecharts包1.0版本和0.5版本引用包的语法不同，随后我安装了0.5版本的pyecharts包，再运行后，错误没有出现。 | | | |
| **评 分 依 据** | | **分 值** | **评分成绩** |
| 1．团队协作能力 | | 25分 |  |
| 2．python综合运用能力 | | 25分 |  |
| 3．态度认真、刻苦钻研、创新能力 | | 10分 |  |
| 4．过程完成、对工具的使用、对github的运用 | | 20分 |  |
| 5．课程设计答辩逻辑清晰，内容正确 | | 10分 |  |
| 6. 课程设计期间的课堂考勤、遵守纪律 | | 10分 |  |
| 总 分 | | 100分 |  |
| 最终评定等级为：  指导老师签字：  2021 年 1 月 15日 | | | |