

文章编号: 2095-2163(2022)12-0221-05

中图分类号: TP311.52

文献标志码: A

# 基于 Python 的人口普查数据可视化分析

邓慈云, 余国清

(湖南信息职业技术学院, 长沙 410200)

**摘要:** 本文以第七次人口普查数据作为研究对象, 采用 Pandas 对普查数据进行读取和处理, 选择 Pyecharts 直观分析展示人口问题的现状和发展趋势。

**关键词:** Python; 人口普查; Pyecharts; 数据可视化

## Visual analysis of the national population census data based on Python

DENG Ciyun, YU Guoqing

(Hunan College of Information, Changsha 410200, China)

**[Abstract]** This paper takes the data of the Seventh National Census as the research object, the Pandas library is used to read and process the data, the Pyecharts is used to analyze the current situation and development trend of census data.

**[Key words]** Python; national census; Pyecharts; data visualization

## 0 引言

时下, 大数据已逐渐成为技术热词, 对数据进行统计分析和可视化的工具也陆续问世。Python 语言由于其简洁性、易读性、可扩展性, 以及具有丰富的标准库, 备受学术界众多科研机构的关注与青睐。Pandas 是 Python 的一个数据分析包, 提供了大量快速便捷地处理数据的函数和方法。Pyecharts 是一款用于生成 Echarts 图表的类库, 可以通过图形参数配置轻松绘制精美的图表。而在社会经济和信息技术飞速发展的同时, 2021年7月20日, 中共中央、国务院印发《关于优化生育政策促进人口长期均衡发展的决定》, 宣布实施三孩政策并配套实施积极生育支持措施。国内人口在持续增长的同时, 却也表现出人口老龄化速度变快、出生率断崖式下降和男女性别比失衡等问题, 一直引起社会各方的高度重视。

目前研究人口问题的相关文献很多, 但是对于人口普查数据进行可视化展示和分析探讨的较少。鉴于此, 本文以第七次人口普查数据为研究对象, 选用 Python 语言、Pandas 库和 Pyecharts 可视化工具对总人口数及增速、男女人口数及性别比例和城镇乡村人口分布等情况进行了直观分析展示, 以期对未来有关研究工作提供参考基础。

## 1 整体设计

### 1.1 数据来源

本文主要选取国家第七次人口普查结果数据中的总人口、人口性别结构、人口年龄结构和人口出生率等作为主要研究对象。采用的数据来源于国家统计局的开源数据。在国家统计局官网(<https://data.stats.gov.cn/index.htm>)年度数据中人口类别获取上述普查数据结果, 并以 csv 格式下载保存数据到本地。

### 1.2 流程设计

采用 Python 技术可视化分析的过程为: 首先, 利用 Pandas 库读取 csv 文件; 然后, 筛选和处理数据; 最后, 使用 Pyecharts 库对数据进行可视化。整个项目流程如图 1 所示。



图1 项目流程图

Fig. 1 Flow chart of the project

## 2 可视化过程实现

### 2.1 数据读取

本文使用总人口数、人口出生率、死亡率和自然增长率及人口年龄结构等数据信息, 包含的数据字段有: 年末总人口、男性和女性人口、城镇和乡村、人

**作者简介:** 邓慈云(1983-), 女, 硕士, 讲师, 主要研究方向: 人工智能和可视化技术; 余国清(1971-), 男, 硕士, 副教授, 主要研究方向: 数据分析、模式识别。

收稿日期: 2022-03-25

口出生率和0~14岁人口等。数据完整性和数据质量均良好,只需删除每个文件中表头内容和多余的补充说明文字即可。利用Pandas库读取csv数据文件,因篇幅有限,仅介绍其中一个数据文件的读取、筛选和处理。首先,利用方法`read_csv()`读取csv文件。相关代码如下:

```
dfpeo = pd.read_csv('D:/file/总人口.csv',
encoding='GBK')
#查看相关数据文件的前5行数据
print(dfpeo.head())
```

## 2.2 数据筛选和处理

从读到的文件数据中,通过筛选及相关处理后才能获得可以用于绘图的数据。如筛选出年末总人口数后,将其转换成列表,并去掉字段说明。因考虑到绘制的图表便于用户阅读,故采取年份按从低到高排列。为配合这一设计,调用方法`reverse()`将年末总人口数列表中的元素反向排序,并将列表元素类型转换成float。相关代码如下:

```
all = list(dfpeo.iloc[0])
all = all[1:]
all.reverse()
alldata = []
for i in range(len(all)):
    alldata.append(float(all[i]))
```

## 2.3 绘制和保存图表

本文绘制了男性、女性人口对比和城镇、乡村人口对比等6张图。为了呈现总人口数的变化和增长

率的差异,绘制了双坐标轴图,将柱状图和折线图组合在一起。研发代码具体如下:

```
bar = Bar()
bar.add_xaxis(year)
bar.add_yaxis("总人口数", alldata, category_
gap = "50%", z = 0)
bar.set_global_opts(yaxis_opts = opts.AxisOpts
(name = "万人", interval = 2 000, min_ =
'134 000'))
bar.extend_axis(yaxis = opts.AxisOpts (axislabel_opts =
opts.LabelOpts(formatter = '{value}%o'), interval =
2, max_ = '8'))
line = Line().add_xaxis(year).add_yaxis("比
比上年增长", zzl, yaxis_index = 1)
bar.overlap(line)
line.overlap(bar)
bar.render('fig01.html')
```

## 3 实验结果与分析

### 3.1 男性、女性人口及性别比分析

从性别结构看,男性人口高于女性人口,性别比都近似维持在105左右(以女性为100),但这个数据已经逐渐在往好的方向转变。其中,2012年男性人口69 660万人,女性人口66 262万人,性别比为105.13;2021年男性人口72 311万人,女性人口68 949万人,性别比为104.88,如图2所示。

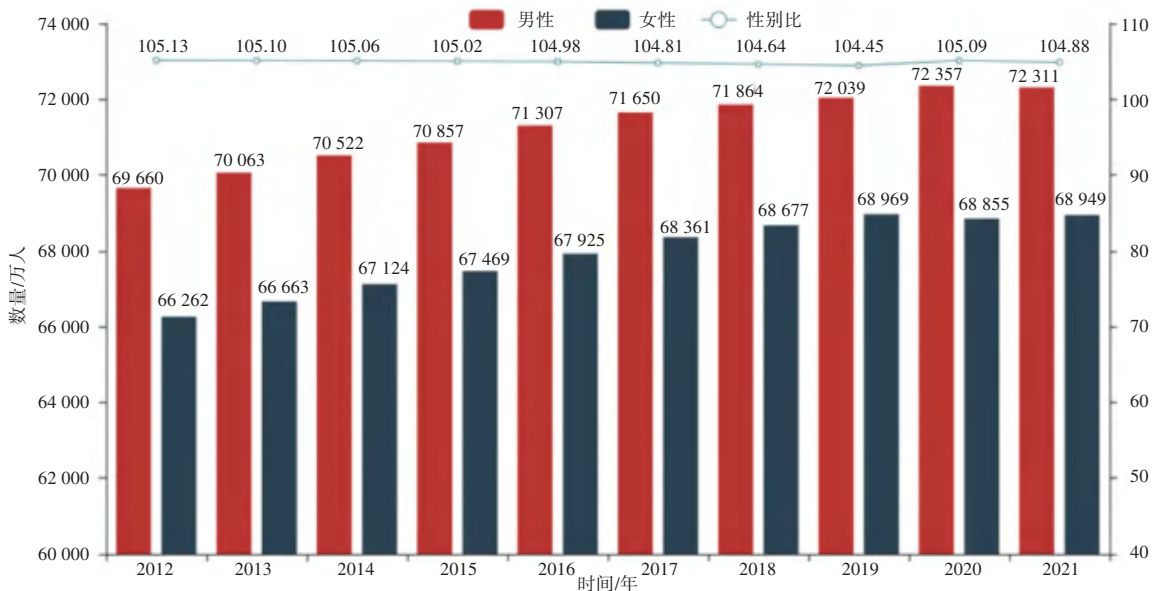


图2 男性、女性人口对比

Fig. 2 Proportion of male and female

### 3.2 城镇、乡村人口结构分析

从城乡结构看,中国城镇人口数量长期占据中国人口总数 50% 以上的比例,尤其自 2017 年起占比达到 60% 以上,且逐年攀升。2021 年城镇人口达到

91 425 万人,比上年末增加 1 205 万人,城镇人口占总人口比重为 64.72%,比上年末提高 0.83 个百分点;乡村人口数量 49 835 万人,减少 157 万人,比上年末下降了 0.83 个百分点,如图 3、图 4 所示。

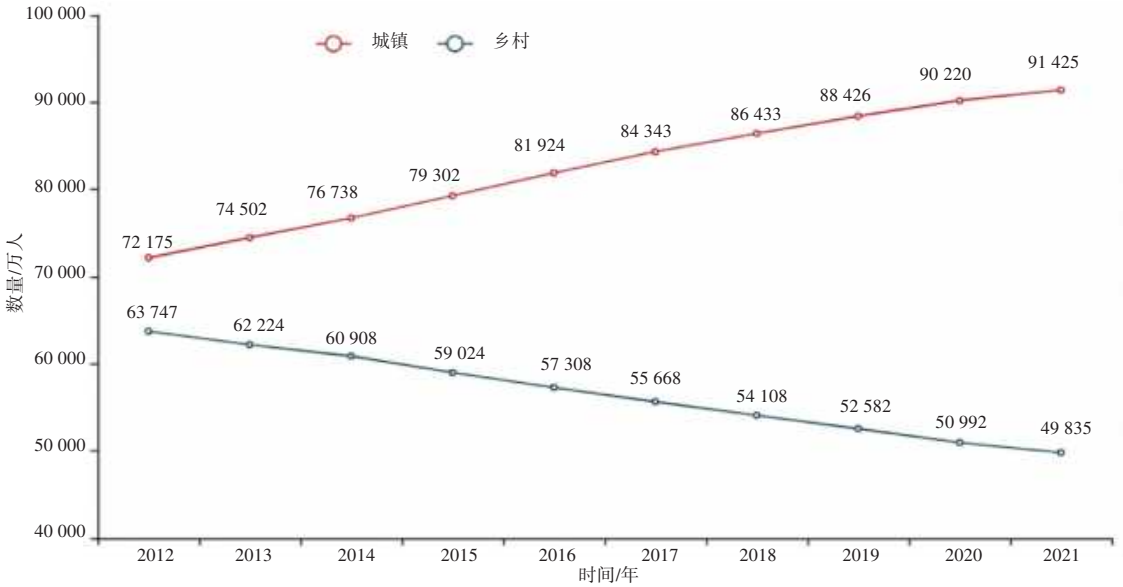


图 3 城镇、乡村人口结构

Fig. 3 Composition of urban and rural populations

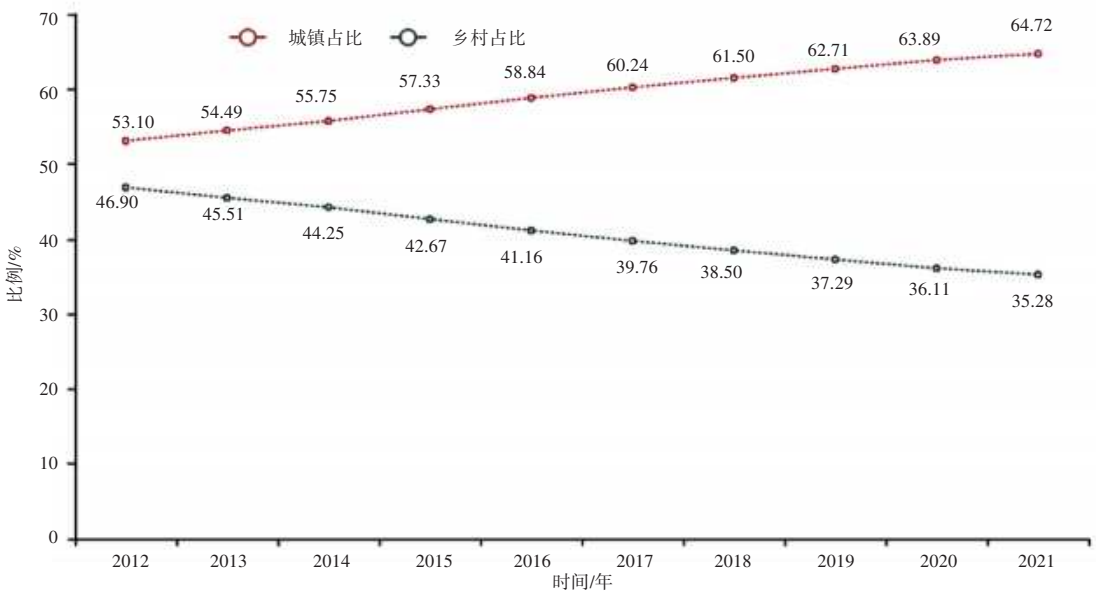


图 4 城镇、乡村人口占比

Fig. 4 Proportion of urban and rural populations

### 3.3 人口年龄结构分析

从年龄构成看,2020 年中国 0~14 周岁的人口 25 277 万人,占总人口的 17.9%;随着老龄化进程的加剧,15~64 周岁劳动年龄人口数量和比重呈双降态势。劳动年龄人口数量从 2012 年的 100 718 万人下降到 96 871 万人,占比从 2012 年的 74.1%持续

下降到 68.6%。劳动年龄人口规模及比重下降的趋势十分明显。65 周岁及以上人口从 2012 年的 12 777 万人增加至 19 064 万人,占比从 2012 年的 9.4% 上升到 13.5%。老年人口比重增幅均呈现不断上升的趋势,如图 5、图 6 所示。

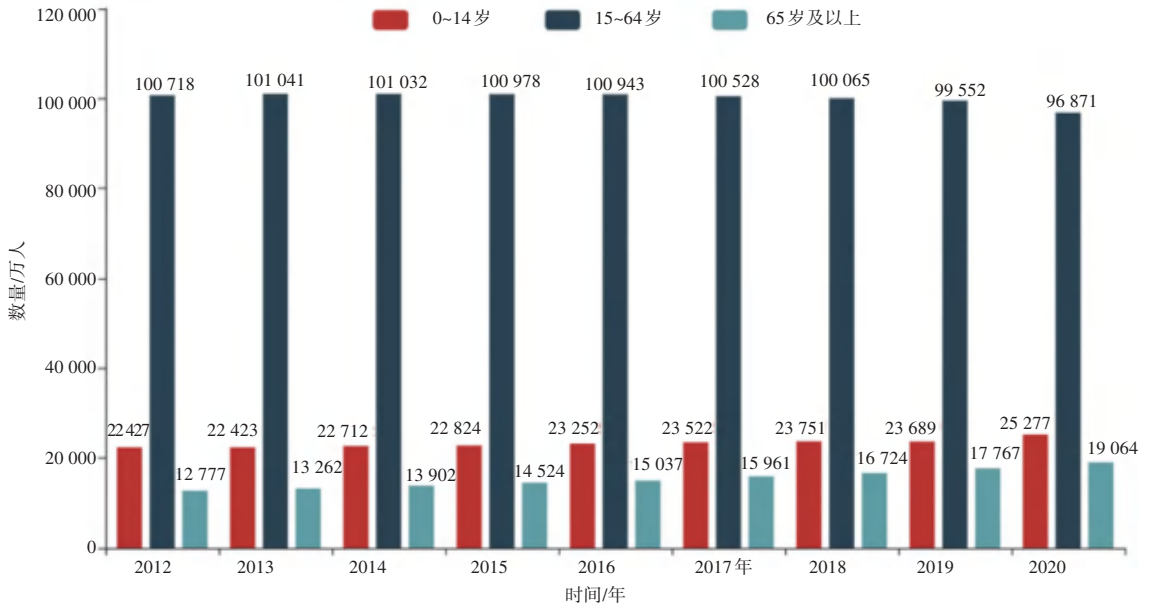


图5 人口年龄结构

Fig. 5 Composition of populations age

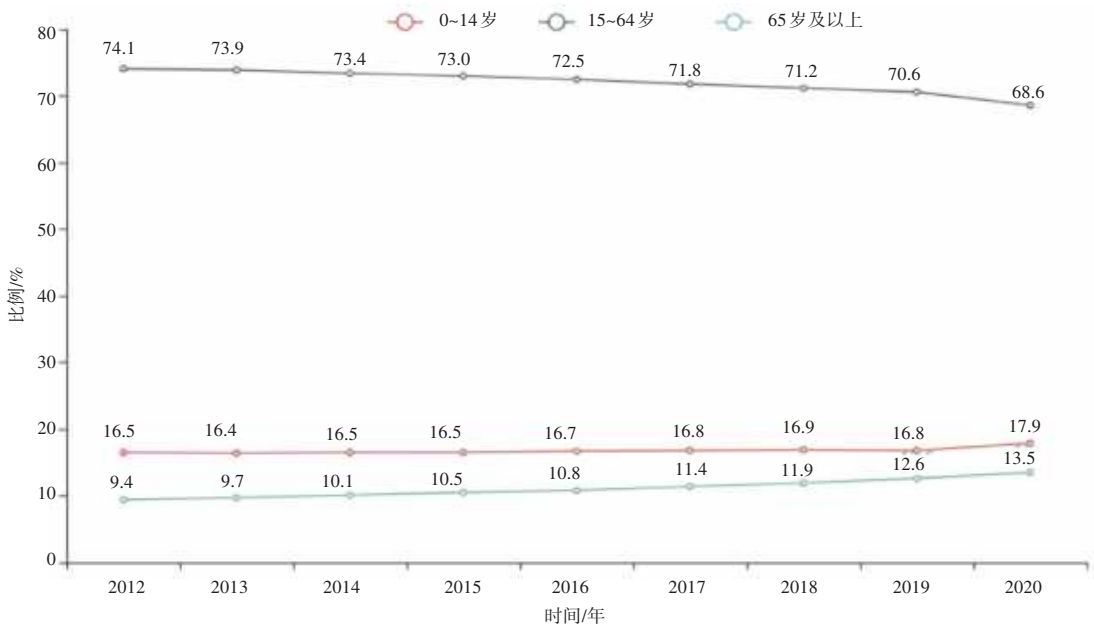


图6 人口年龄占比

Fig. 6 Proportion of populations age

### 3.4 人口出生率和死亡率分析

2012年中国人口出生率为14.57‰,是近10年来出生率最高的一年,然后逐年下降,至2021年人口出生率跌至7.52‰。2014年和2016年人口出生率相比上一年有小幅的增加,其中,2014年和2016年人口出生率分别为13.83‰和13.57‰。2012~

2020年期间,人口死亡率整体呈缓慢下降态势。2012年人口死亡率7.13‰,2020年人口死亡率7.07‰,较2012年下降0.06‰,但2021年人口死亡率有明显上升,达到7.18‰,与这一年的出生率仅相差0.34‰,如图7所示。



图 7 人口出生率和死亡率

Fig. 7 Birth rate and mortality

### 4 结束语

本文基于第七次全国人口普查数据,利用 Python 中的 Pandas 对男性、女性人口等数据进行读取、筛选和处理后,使用 Pyecharts 对男性、女性人口对比和城镇、乡村人口对比、人口出生率和死亡率等方面总体情况进行分析。通过可视化,有助于直观地了解国内人口实情,后续还将对其做进一步的完善。

### 参考文献

[1] 侯瑾菲,梁艺多. 基于 Python 的政府开放数据可视化应用研究

[J]. 科学技术创新, 2021 ( 34 ): 160-162.

[2] 韦依洋,吴一凡,李永远. Python 技术在数据可视化中的应用研究[J]. 福建电脑,2022,38(01):27-31.

[3] 赵宝筑,王彪,吴和俐,等. 基于 Python 的乌当区气温数据可视化应用[J]. 数字技术与应用,2021,39(11):103-105.

[4] 贾艳平,翟晋刚. 基于 Python 爬虫技术的游客评论数据可视化分析[J]. 安阳师范学院学报,2021(05):51-54.

[5] 李佳佳. 基于 Python 的股票数据可视化分析—以新能源汽车行业为例[J]. 河北软件职业技术学院学报,2021,23(03):15-18.

[6] 蔡文乐,周晴晴,刘玉婷,等. 基于 Python 爬虫的豆瓣电影影评数据可视化分析[J]. 现代信息科技,2021,5(18):86-89,93.

(上接第 220 页)

### 4 酒精浓度在线检测仪实验验证

对设计好的酒精浓度在线检测仪进行实验验证,观察其是否能满足正常使用要求。购买所需要的元器件,利用焊接技术焊制出的酒精浓度在线检测仪,如图 5 所示。设置的阈值为 25 mg/100 mL,检测到空气中的酒精浓度超过设置的阈值,则红灯闪烁,蜂鸣器响起。



图 5 酒精浓度在线检测仪实验结果

Fig. 5 Alcohol concentration online detector experimental results

通过实验可以看出,当酒精浓度超过设置的阈值,则发出报警提示,低于阈值、报警停止。实验说明

本文设计的酒精浓度在线检测仪能够满足使用要求。

### 5 结束语

设计了基于 STC89C51 单片机的酒精浓度在线检测仪,MQ-3 酒精传感器采集信息,LCD1602 液晶显示模块输出信息。经实验表明:该检测仪能够较为准确快速地检测出当前环境下的酒精浓度,并通过液晶显示屏显示出数值,超过设定值将开启声光报警,另外还可以通过按键设定不同的阈值,具有较好的实际应用价值。

### 参考文献

[1] 刘彤,崔悦,李书臣,等. 汽车语音—酒检防酒驾自动闭锁控制系统设计[J]. 电子世界,2018(20):161-162.

[2] 刘艳红. 基于多传感器信息融合的汽车酒驾测控系统研究[D]. 长春:长春理工大学,2016.

[3] 李居尚. 基于单片机的防酒驾系统设计[D]. 长春:吉林大学,2013.

[4] 刘杨,张健,黄建芬,等. 基于 STC89C51 单片机酒精浓度报警系统[J]. 电子测试,2021(24):23-25,56.