

关于中国人口，你需要关心的问题

原创 闲欢 Python技术 3月11日

中国的人口总数已经突破14亿了，你知道吗？中国人口的出生率、死亡率和自然增长率你了解吗？中国人口的男女比例你清楚吗？不要着急，跟随着我来一起了解。

获取人口数据

我们的目标是

获取新中国成立后70年的总人口数据，以及人口的出生率、死亡率和自然增长率数据。

我们从中国最权威的地方获取人口数据——国家数据。网址是：<http://data.stats.gov.cn/easyquery.htm?cn=C01>。打开网页后，如下图：

左边是菜单树，包含各种各样的统计数据。右边是数据表格，表格右上方有选择项，我们可以看到总人口数据可以选择时间段。

要想获取数据，我们先要分析请求，于是我们打开了网页开发者工具，我们可以很容易地找到获取数据的请求：

通过分析请求参数，我们可以得出几个变动参数的含义：

- m: 固定为 QueryData。
- dbcode: 固定为 hgnd。
- rowcode: 固定为 sj。
- colcode: 固定为 zb。
- wds: 固定为 []。
- dfwds: 数组。当 wdcode 为 zb 时，valuecode 表示指标，A0301表示总人口，A0302表示人口出生率、增长率和死亡率。当 wdcode 为 sj 时，valuecode 表示时间，LAST5 表示5年，LAST70 表示70年。
- k1: 时间戳，可以不用管。

弄清楚了这些，我们就可以开始发送请求获取数据了：

```
1 # 总人口
2 dfwds1 = '[{"wdcode": "sj", "valuecode": "LAST70"}, {"wdcode": "zb", "valuecode": "A0301"}]'
3 # 人口出生率、死亡率、自然增长率
4 dfwds2 = '[{"wdcode": "sj", "valuecode": "LAST70"}, {"wdcode": "zb", "valuecode": "A0302"}]'
```

```

5 url = 'http://data.stats.gov.cn/easyquery.htm?m=QueryData&dbcode=hgnd&rowcode=sj&colcode=zb&wds=
6 # 将所有数据放这里，年份为key，值为各个指标值组成的list
7 # 因为 2019 年数据还没有列入到年度数据表里，所以根据统计局2019年经济报告中给出的人口数据计算得出
8 # 数据顺序为历年数据
9 population_dict = {
10
11 }
12
13 response1 = requests.get(url.format(dfws1))
14 get_population_info(population_dict, response1.json())
15
16 response2 = requests.get(url.format(dfws2))
17 get_population_info(population_dict, response2.json())
18
19 population_dict['2019'] = [2019, 140005, 71527, 68478, 84843, 55162, 10.48, 7.14, 3.34]
20 save_excel(population_dict)

```

考虑到我们可能需要调试程序，不宜多次频繁请求网站，所以我们这里将获取到的数据存入 Excel 表格中，并且将两份数据合成一个表格。

```

1 # 人口数据生成excel文件
2 def save_excel(population_dict):
3     # .T 是行列转换
4     df = pd.DataFrame(population_dict).T[:::-1]
5     df.columns = ['年份', '年末总人口(万人)', '男性人口(万人)', '女性人口(万人)', '城镇人口(万人)', '乡村人
6                 '口自然增长率(%)']
7     writer = pd.ExcelWriter(POPULATION_EXCEL_PATH)
8     # columns参数用于指定生成的excel中列的顺序
9     df.to_excel(excel_writer=writer, index=False, encoding='utf-8', sheet_name='中国70年人口数据')
10    writer.save()
11    writer.close()

```

大家注意一点，国家数据网站上最新的人口数据是2018年的，所以2019年我得从另外渠道（国家统计局网站）获取，并且加入表格中。我获取数据的地址是：http://www.stats.gov.cn/tjsj/zxfb/202001/t20200117_1723383.html。

我们生成的数据表格是这样的：

分析人口数据

获取到我们需要的数据后，我们就可以利用数据来分析我们关心的话题了，这里主要以图表方式来呈现。

人口总数分析

我将70年的总人口数制作成一个柱状图，来展现新中国成立以来人口数的变化情况：

```
1 # 处理数据
2 x_data = pdata['年份'].tolist()
3 # 将人口单位转换为亿
4 y_data1 = pdata['年末总人口(万人)'].map(lambda x: "%.2f" % (x / 10000)).tolist()
5 y_data2 = pdata['人口自然增长率(%)'].tolist()
6 y_data3 = pdata['人口出生率(%)'].tolist()
7 y_data4 = pdata['人口死亡率(%)'].tolist()
8
9 # 总人口柱状图
10 bar = Bar(init_opts=opts.InitOpts(width="1200px", height="500px"))
11 bar.add_xaxis(x_data)
12 bar.add_yaxis("年末总人口（亿）", y_data1, category_gap="10%", label_opts=opts.LabelOpts(rotate=90))
13 bar.set_global_opts(
14     title_opts=opts.TitleOpts(title="年末总人口变化情况", pos_bottom="bottom", pos_left="center"),
15     xaxis_opts=opts.AxisOpts(
16         type_="category",
17         name='年份',
18         # 坐标轴名称显示位置
19         name_location='end',
20         # x轴数值与坐标点的偏移量
21         # boundary_gap=False,
22         axislabel_opts=opts.LabelOpts(is_show=True, margin=10, color="#000", interval=1, rotate=90),
23         # axisline_opts=opts.AxisLineOpts(is_show=True, symbol="arrow"),
24         axistick_opts=opts.AxisTickOpts(is_show=True, is_align_with_label=True),
25         axispointer_opts=opts.AxisPointerOpts(type_="line", label=opts.LabelOpts(is_show=True))
26     ),
27     # y轴相关选项设置
28     yaxis_opts=opts.AxisOpts(
29         type_="value",
30         position="left",
31     ),
32     legend_opts=opts.LegendOpts(is_show=True)
33 )
```

画图的代码我就不详细讲解了，大家感兴趣可以去看一下 [pyecharts](#) 作图。

总人口数的柱状图效果如下：

从图中可以看出，我们的总人口数除了1960年和1961年两年是减少的，其他年份都是增加的。而那两年是由于严重自然灾害导致的大饥荒，饿死了好多人。我们也可以看到在2019年，我国的人口总数正式突破14亿人，以后需要更新数据了，再也不要再在别人面前说我们国家有13亿人了。

接下来，我们将人口出生率、死亡率和自然增长率用折线图来展现，代码如下：

```

1 # 自然增长率、出生率、死亡率折线图
2 line = Line(init_opts=opts.InitOpts(width="1400px", height="500px"))
3 line.add_xaxis(x_data)
4 line.add_yaxis(
5     series_name="自然增长率(%)",
6     y_axis=y_data2,
7     label_opts=opts.LabelOpts(
8         is_show=False
9     )
10 )
11 line.add_yaxis('出生率(%)', y_data3, label_opts=opts.LabelOpts(is_show=False))
12 line.add_yaxis('死亡率(%)', y_data4, label_opts=opts.LabelOpts(is_show=False))
13 line.set_global_opts(
14     title_opts=opts.TitleOpts(title="人口自然增长率、出生率、死亡率", pos_bottom="bottom", pos_left="left"),
15     xaxis_opts=opts.AxisOpts(
16         name='年份',
17         name_location='end',
18         type_="value",
19         min_="1949",
20         max_interval=1,
21         # 设置x轴不必与y轴的0对齐
22         axisline_opts=opts.AxisLineOpts(is_on_zero=False),
23         axislabel_opts=opts.LabelOpts(is_show=True, color="#000", interval=0, rotate=90),
24         axistick_opts=opts.AxisTickOpts(is_show=True, is_align_with_label=True),
25         axispointer_opts=opts.AxisPointerOpts(type_="shadow", label=opts.LabelOpts(is_show=True))
26     ),
27     # y轴相关选项设置
28     yaxis_opts=opts.AxisOpts(
29         name='比例',
30         type_="value",
31         position="left",
32         min_=-10,
33         axislabel_opts=opts.LabelOpts(is_show=True)
34     ),
35     legend_opts=opts.LegendOpts(is_show=True)
36 )

```

运行代码，我们得到的图像如下：

同样的，我们可以看到如下几点：

- 1960年和1961年两年因为大饥荒导致死亡率激增。
- 可能由于大饥荒过后国家鼓励生育，人口增长率在1962年出现大幅反弹，并于1963年达到峰值后开始下跌。
- 图中除了特殊年份外，死亡率趋于平稳，所以自然增长率和出生率呈现一定的正相关性。
- 在2016年出生率出现小幅反弹，没错，那年是放开二胎的元年。

- 虽然放开了二胎，但是大家好像不热衷于生小孩了，人口出生率在近两年下降明显。

人口性别分析

对于人口性别分析，我们从两个方面出发，一个是目前我国总人口中的男女比例，另一个方面是我国男女人口差额的变化情况。

我们首先用一个饼图来看看2019年的男女比例：

```
1 pie = Pie()
2     pie.add("", [list(z) for z in zip(['男', '女'], np.ravel(sex_2019.values))])
3     pie.set_global_opts(title_opts=opts.TitleOpts(title="2019中国男女比", pos_bottom="bottom", pos_left="center"))
4     pie.set_series_opts(label_opts=opts.LabelOpts(formatter="{b}: {d}%"))
```

运行代码得到的效果图如下：

很显然，目前中国男性比女性多了2个多百分点。青年男士依然面临僧多粥少的局面啊。

下面我们来计算一下新中国成立以来男女的差额，然后通过折线图来看看变化情况，代码如下：

```
1 line = Line(init_opts=opts.InitOpts(width="1400px", height="500px"))
2     line.add_xaxis(x_data)
3     line.add_yaxis(
4         series_name="男女差值",
5         y_axis=y_data_man_woman.values,
6         # 标出关键点的数据
7         markpoint_opts=opts.MarkPointOpts(
8             data=[
9                 opts.MarkPointItem(type_="min"),
10                opts.MarkPointItem(type_="max"),
11                opts.MarkPointItem(type_="average")
12            ]
13        ),
14        label_opts=opts.LabelOpts(
15            is_show=False
16        ),
17        markline_opts=opts.MarkLineOpts(data=[opts.MarkLineItem(type_="average")])
18    )
19    line.set_global_opts(
20        title_opts=opts.TitleOpts(title="中国70年(1949-2019)男女差值（万人）", pos_left="center", pos_top="top"),
21        legend_opts=opts.LegendOpts(is_show=False),
22        xaxis_opts=opts.AxisOpts(
23            name='年份',
24            name_location='end',
25            type_="value",
26            min_="1949"
```

```

26         max_interval=1,
27         # 设置x轴不必与y轴的0对齐
28         axisline_opts=opts.AxisLineOpts(is_on_zero=False),
29         axislabel_opts=opts.LabelOpts(is_show=True, color="#000", interval=0, rotate=90),
30         axistick_opts=opts.AxisTickOpts(is_show=True, is_align_with_label=True),
31         axispointer_opts=opts.AxisPointerOpts(type_="shadow", label=opts.LabelOpts(is_show=True)
32     ),
33     yaxis_opts=opts.AxisOpts(
34         name='差值（万人）',
35         type_="value",
36         position="left",
37         axislabel_opts=opts.LabelOpts(is_show=True)
38     ),
39 )
40

```

运行代码，得到的图像如下：

从图中可以看出：

- 在60年代，男女的差额是最小的一段时期，只有不到2000万。
- 在80年代末到90年代末这期间差值缩小明显，而后又迅速攀升到2000年的峰值4131万。这个变化我不知道是什么政策或事件导致的，你们知道吗？
- 近些年，男女的差值是缓慢减少的，这可能是对于单身男士的唯一利好消息了吧。

总结

本文从国家数据官方网站获取到中国人口的数据，然后通过图标将总人口、人口变化率以及人口性别等几个方面展现出来，从而可以直观地看到我国人口的总体情况。通过阅读本文，对于中国人口的陈旧记忆，你是否更新了呢？

文中示例代码：python-100-days

PS： 公号内回复「Python」即可进入 Python 新手学习交流群，一起 **100天计划**！

-END-

Python 技术
关于 Python 都在这里