**老猫荐书**

[老猫荐书【第1期】 |《Python项目案例开发从入门到实战——爬虫、游戏和机器学习》](http://mp.weixin.qq.com/s?__biz=MzI5MzIwNDI1MQ==&mid=2650126235&idx=2&sn=0ad94614369cc4378576abee07b4c73e&chksm=f474afefc30326f9ea3d24cc33db1a2a2b349c115f22fbc2800891ef58d64f72d9c113a0a339&scene=21#wechat_redirect)

[老猫荐书【第2期】 |《微信小程序开发零基础入门》](http://mp.weixin.qq.com/s?__biz=MzI5MzIwNDI1MQ==&mid=2650126341&idx=2&sn=d18b99ecc5975c9d731ec013502c44ce&chksm=f474a071c303296757aa270440365007a6ca18fa12ca7acb9a109e237793e59d17c2f92993b5&scene=21#wechat_redirect)

[老猫荐书【第3期】 |《Java EE框架整合开发入门到实战——Spring+Spring MVC+MyBatis》](http://mp.weixin.qq.com/s?__biz=MzI5MzIwNDI1MQ==&mid=2650126433&idx=2&sn=426b1edc6150e87b76a3f1c11882866b&chksm=f474a015c30329036967241f69ee75c1e6b3e0b2c73f9b2d6722b94a6269f34e5d1acf98bd26&scene=21#wechat_redirect)

**摘要**：**用 Python 分析过往 36 年春晚节目数据，发现一些趣事。**

马上就要过年了，距离 2019 己亥猪年的除夕已不足一个礼拜，提起除夕，多数人马上想到「春节联欢晚会」这道丰盛的「年夜大餐」。看过那么多春晚，哪一年、哪些节目、哪些人你还留有深刻印象呢。

记忆中，只完整地看过 2005 年到 2015 年十年春晚。**05 年之前，还很小，看不懂；15 年之后，长大了，也看不懂。**

如今，距离第一届春晚 1983 年，整整过去了 36 年，3 轮的「十二生肖」年。趁今年春晚还没到，来回顾一下过往 36 届春晚的一些有趣数据。

**分析内容**

接下来，通过 Python 数据分析，会回答下面这些问题，在知道答案之前，你可以先猜猜看：

* **谁导演春晚次数最多？**
* **谁主持春晚次数最多？**
* **哪两年的除夕刚好是同一天？**
* **谁上春晚次数最多，堪称「钉子户」？**
* **港台明星上春晚次数对比**
* **歌曲、小品、相声类节目数量对比**

**数据获取**

网上搜了挺久都没有找到齐全的春晚节目数据，连春晚官网也没有，结果在维基百科上找到了。



右侧信息表有导演、主持人、除夕当天日期这几项数据。



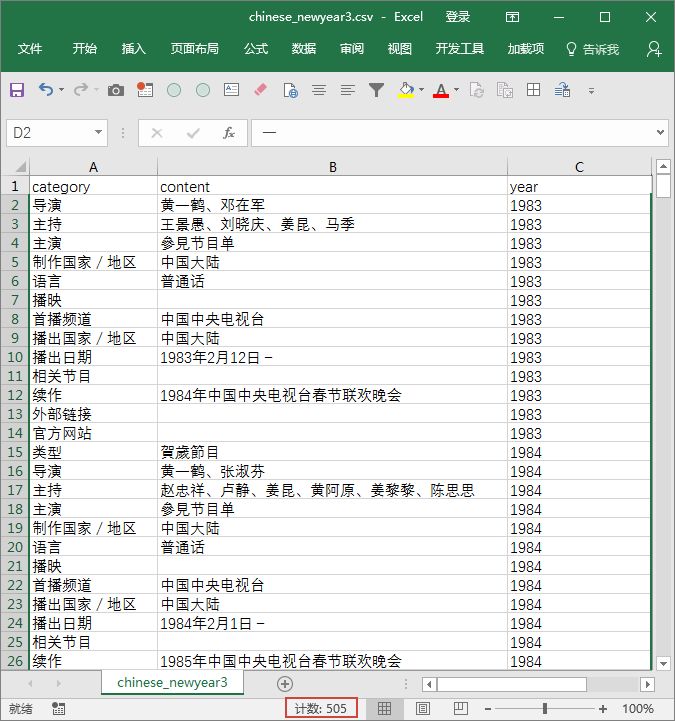
节目单表是每一年春晚上表演的节目，包括：节目类型、节目名、演员名这几项数据。

Python 抓取这类表格数据，方法简单，几行代码就能搞定，修改 URL 的 page 参数，可以循环遍历抓取 1983 到 2018 年所有的节目数据。

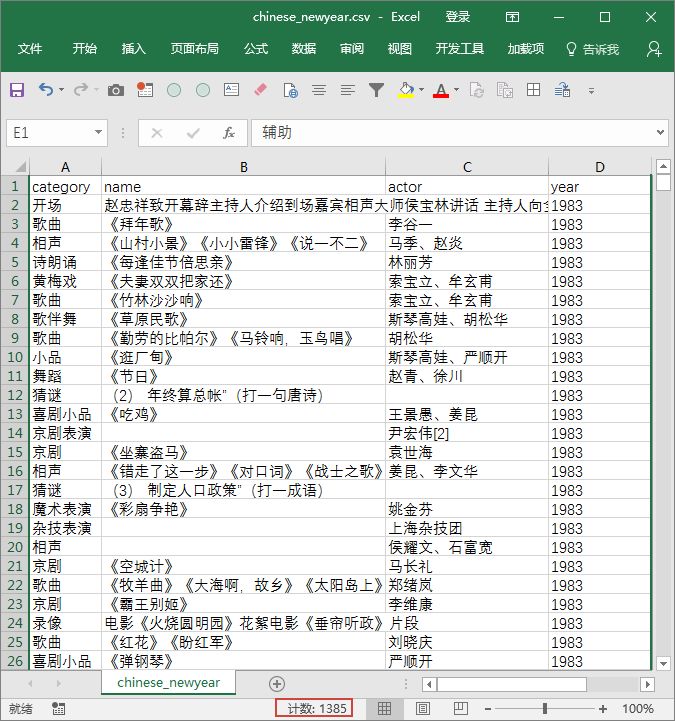
数据抓取代码如下：

1import requests  
 2import pandas as pd  
 3from urllib.parse import quote  
 4  
 5def get\_content(year):  
 6    keywords = quote('年中国中央电视台春节联欢晚会')  
 7    url = 'https://zh.wikipedia.org/wiki/{}{}'.format(year,keywords)  
 8    # 1 节目单； 0 节目信息  
 9    if year != 2014:  
10        response = pd.read\_html(url)[1]  
11    else:  
12        response = pd.read\_html(url)[3]  
13    response['year'] = year  
14    response.drop([0],inplace=True) #删除首行  
15    response.to\_csv('chinese\_newyear.csv',mode='a',encoding='utf\_8\_sig',index=0,header=0)  
16  
17if \_\_name\_\_ == '\_\_main\_\_':  
18    for year in range(1983,2018):  
19        get\_content(year)

抓取下来的节目信息：



抓取下来的节目数据：

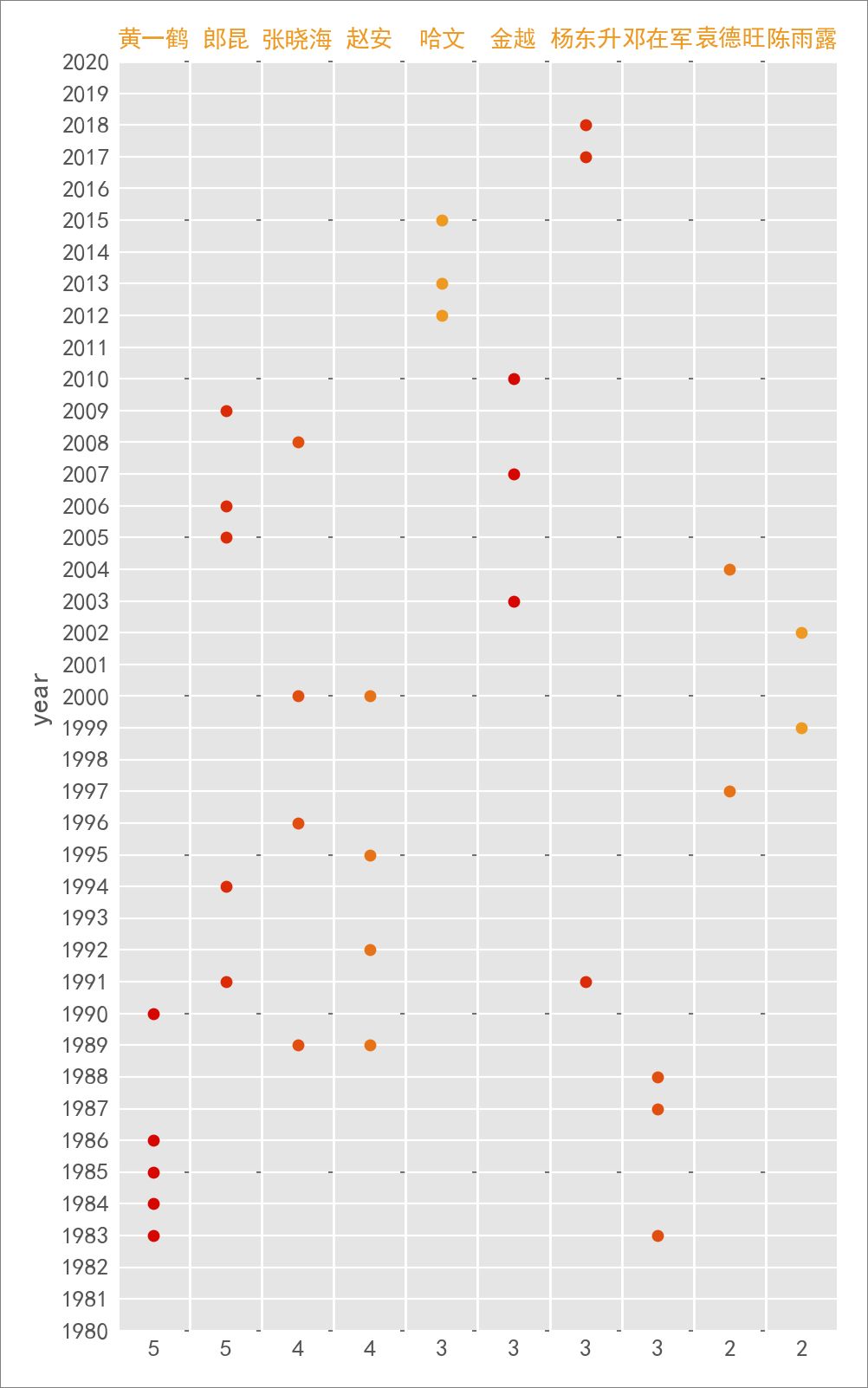


抓取下来的数据是脏数据，用 Python 清洗处理一下就可以分析，这些不是重点，所以下面直接进入分析环节，来一探究竟。

**数据分析**

**▌谁导演春晚次数最多？**

导演是春晚的总负责人，好比厨师，厨师决定了春晚大餐好不好吃。36 年间，有很多导演负责过春晚，比如近些年的哈文、朗昆，你可能想知道他们是不是导演次数最多的，下面就来看看导演次数最多的十大导演：



**导演次数最多的是黄一鹤和朗昆导演，两个人都导演了 5 次。**

黄一鹤导演对于 80 后之后的人来说，不算熟悉，因为他导演春晚的时候是在 80 年代，很多人都没有出生。朗昆则相对熟悉些，最近一次导演是 2009 年。那一年的春晚，是印象最深刻的一届，因为诞生了赵本山最棒的小品《不差钱》（个人之见）。

哈文一共导演了 3 次，都在 2010 年之后，其他的导演就不那么熟悉了，相比于主持人、演员，他们是幕后工作者。

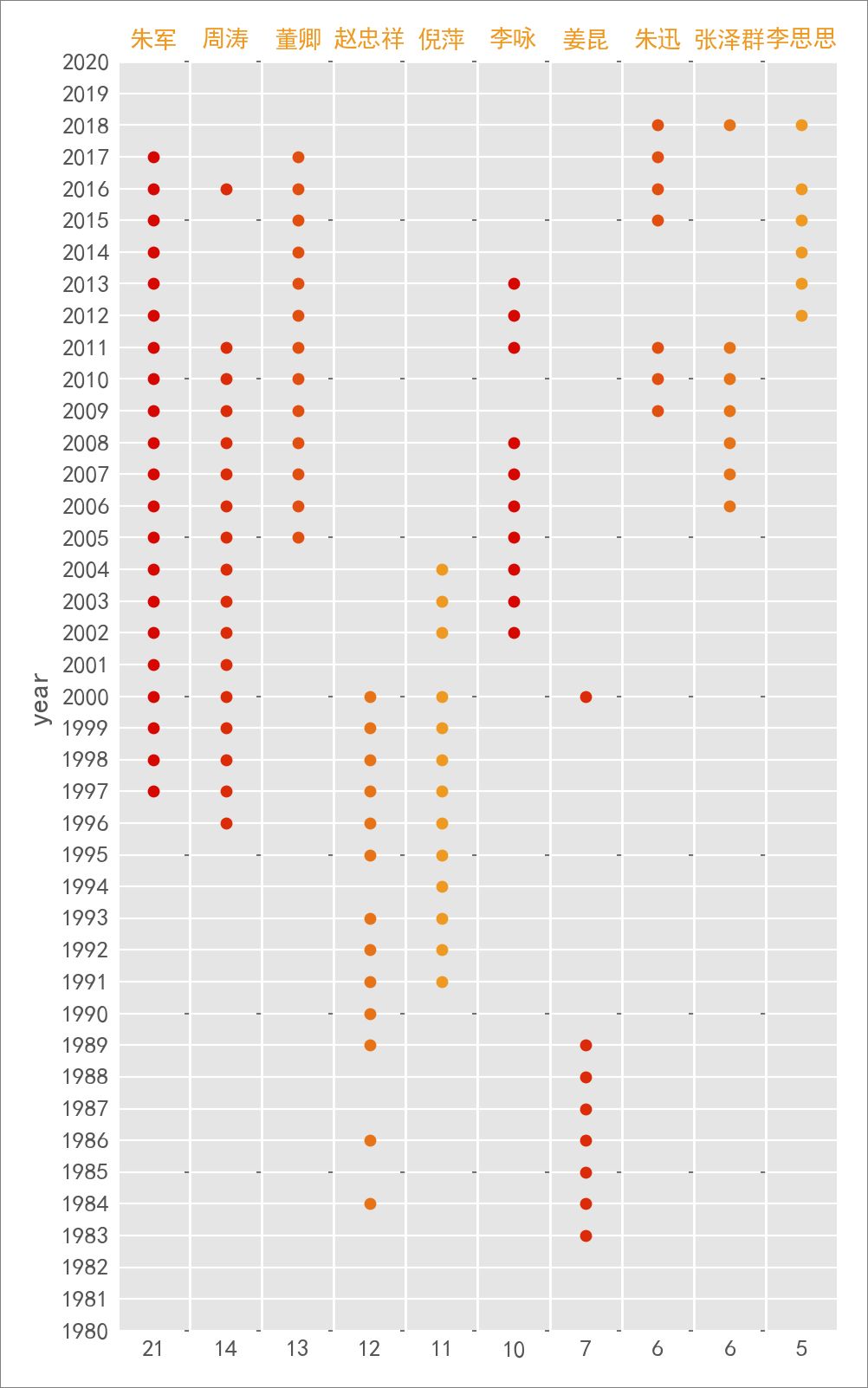
代码实现如下，关键点在于 For 循环绘制子图：

|  |  |
| --- | --- |
| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 | def get\_director():     data = pd.read\_csv('chinese\_newyear3.csv',encoding='utf\_8\_sig')     # 筛选导演主持人     data = data[data['category'] == '导演']     # data = data[data['category'] == '主持']     data2 = data['content'].str.split('、',expand=True)      # 统计出现次数     data2 = data2.apply(pd.value\_counts)     data2['col\_num'] = data2.sum(axis=1)     data2.sort\_values(by='col\_num',ascending=False,inplace=True)     data2 = data2['col\_num'][:10][::-1]     return data,data2  def analysis6(data,data2):     data = data.set\_index('year')     data2.sort\_values(ascending=False,inplace=True)     lst = list(data2.index)[:10]     lst\_num = list(data2)[:10]      colorsall = [color1,color2,color3,color4,color5,color1,color2,color3,color4,color5]     for i,name in enumerate(lst):         data3 = data['content'].str.contains(name,na=False).astype('int')         data3 = pd.DataFrame(data3[data3.values == 1])         data3['year'] = data3.index          axs = fig.add\_subplot(1, 10, 1+i)         data3.plot(             ax=axs,             x='content',             y = 'year',             kind = 'scatter',             subplots=True,             sharey=True,             color=colorsall[i],             )         new\_ticks = np.linspace(1980,2020,41)         plt.yticks(new\_ticks)         plt.tick\_params(direction='in')  #标签朝里         plt.tick\_params(which='major',length=0)  # 不显示刻度标签长度         plt.xticks([]) #去掉坐标标签         plt.xlabel('%i' %(lst\_num[i]),fontsize=10)         plt.xlim(0,2)          plt.title(name,color=color5,fontsize=10)         plt.tight\_layout()         fig.subplots\_adjust(hspace=0,wspace=0) # 调整子图间距为0      plt.savefig('导演次数最多的主持 TOP 10.png',dpi=200)     plt.show() |

**▌谁主持春晚次数最多？**

导演过后就是主持人了，他们堪称春晚的门面，大家也更熟悉些，不管是早年的赵忠祥、倪萍，中生代的朱军、董卿，还是近年新晋的康辉、李思思以及我们新疆大帅锅尼格买提，只要一提名字，你马上就能对号入座。

可你知道主持界的「钉子户」是谁？「常青树」又是哪些人？ 来看看春晚主持次数最多的 TOP 10 名单：



一眼望去，十个人每一个都很熟悉。

排第一的是 **主持了 21 年的朱军**，称得上劳模了，**从 1997 年连续不间断地住持到 2017 年**，远超其他主持人。

排第二的是周涛，主持了 14 年，2011 年之后退居幕后，2016 年又复出了一次。对她的印象，莫过于 2003 年和冯巩合作过的一个小品《马路情歌》：



排第三的是董卿，从 05 年开始主持，只在 2018 年缺席了一年，大家都喜欢她，所以她的缺席也成为去年春晚的一大讨论话题。好消息是，今年的春晚她会继续主持。

董卿春晚上的亮点太多了，印象最深的是和刘谦三度搭档过的魔术节目。刘谦技术够神，董卿配合地也神。另一个好消息是，自 13 年阔别春晚舞台之后，刘谦今年也会再度登台，二人是否还会继续搭档呢，拭目以待。



再往后是一对著名搭档，赵忠祥和倪萍老师，可惜印象不深，找了找他们曾主持过的春晚照片，满满的历史感。



（1992 年春晚）



（1996 年春晚）

再之后是李咏，刚刚离开不久，喜欢他的风格，希望天堂也有话筒，也有春晚。

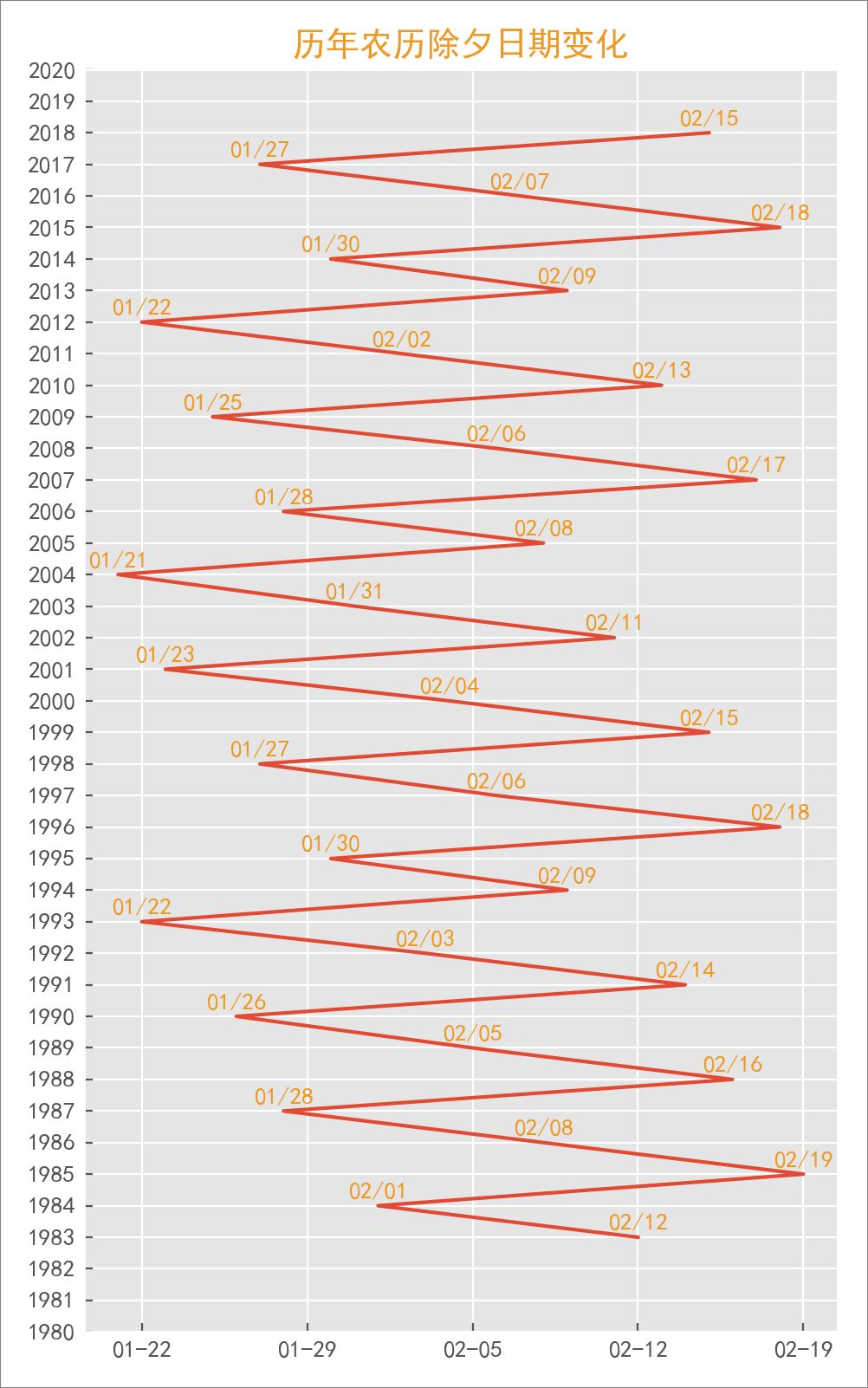


说完主持人，下面说一说除夕的日期。

**▌哪两年的除夕刚好是同一天？**

小时候一直没有搞懂中国的农历，想不明白为什么每年除夕的日子都不一样，为什么过年不是在元旦。后来才知道这是我们老祖宗的智慧，因为农历比阳历准确。

每年除夕的日期都在变化，那你是否好奇过「某两年的除夕是在同一天」这个问题？来看看：



以从下往上，从左往右的顺序看上面这张图，会发现几件有意思的事：

* **除夕最早的一年是 2004 年， 1 月 21 日。**
* **除夕最晚的一年是 1985 年， 2 月 19 日，最早最晚差了近一个月。**
* **2019 年除夕是 2 月 4 日，和 2000 年是一样的。**
* **有不少年份的除夕在同一天。**比如：1993 和 2012 年都在 1 月 22 日，1998 和 2017 年的 1 月 27 日，1987 和 2006 年的 1 月 28 日。很有意思对吧，但至今最多只有两年的除夕是在同一年，没有出现过「三年除夕都在同一天」这种现象，以后也许会。

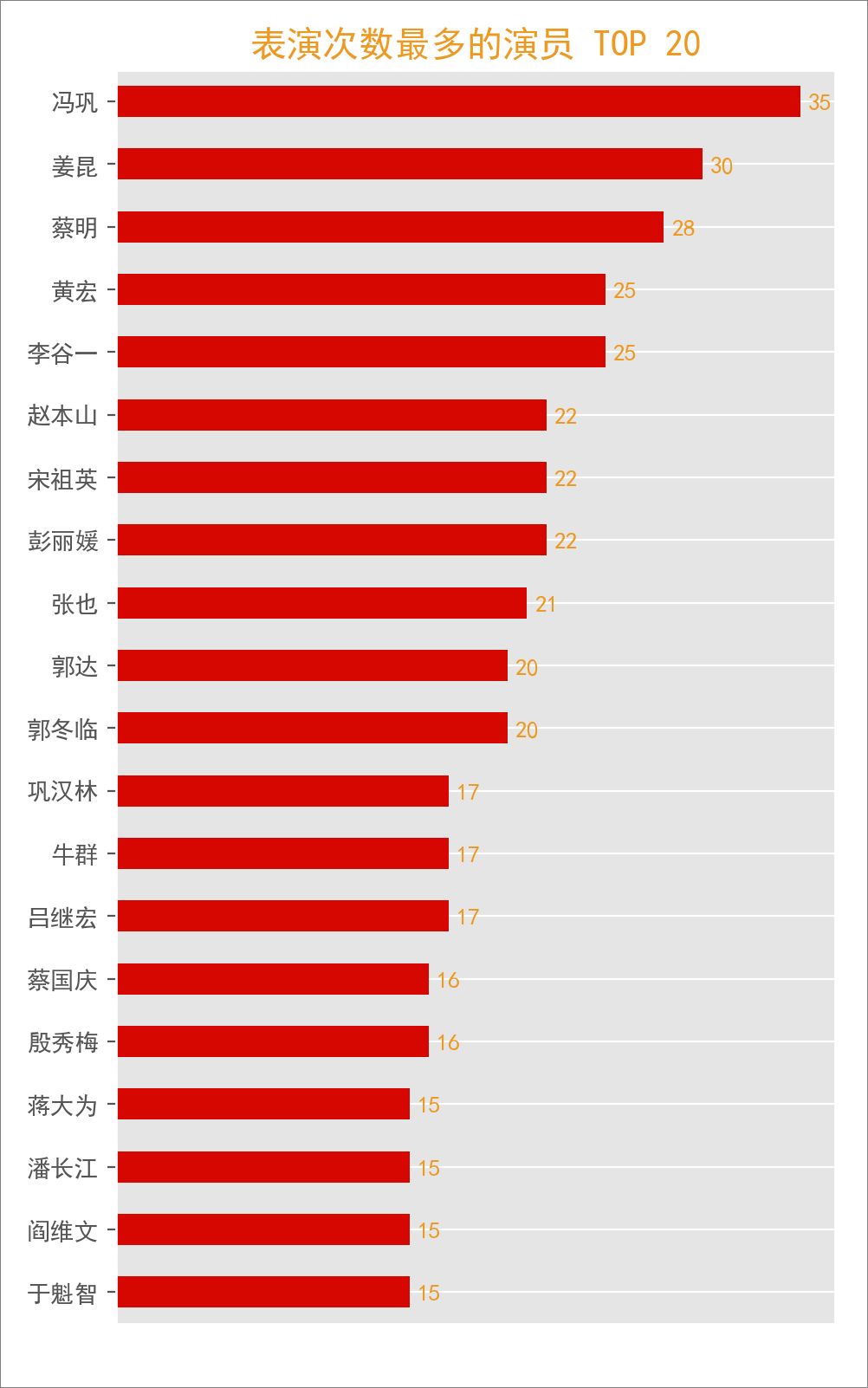
代码实现如下：

|  |  |
| --- | --- |
| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 | # 分析除夕日期 def get\_date():     fig = plt.figure(figsize=(5,8))     ax = fig.add\_subplot(111)     data = pd.read\_csv('chinese\_newyear3.csv',encoding='utf\_8\_sig')     data = data[data['category'] == '播出日期']     data = data['content'].str.extract(r'.\*?年(.\*?)月(.\*?)日.\*?')     data = data.reset\_index(drop=True)     data.columns = ['month','day']     data['year'] = np.arange(1983,2019)      # int 转 string     data = data.applymap(str)     data['year2'] = '1900'     data['date'] = data['year2'].str.cat([data['month'],data['day']],sep='/')     data = data.apply(pd.to\_numeric,errors='ignore')     data['date'] = pd.to\_datetime(data['date'])     ax.plot(         data['date'],         data['year'],         )     new\_yticks = np.linspace(1980,2020,41)     date\_format = mpl.dates.DateFormatter("%m-%d")     ax.xaxis.set\_major\_formatter(date\_format)     plt.yticks(new\_yticks)     plt.tick\_params(direction='in')  #标签朝里     content = list(zip(data['date'],data['year']))     for x, y in content:         x2 = '%s' %x.strftime('%m/%d') # 只显示月日格式         # print(x,'\n',y)         plt.text(x, y+0.2,x2, ha='center', color=color4)      plt.title('历年农历除夕日期变化',color=color4,fontsize=14)     plt.tight\_layout()     plt.savefig('除夕日期变化.png',dpi=200,)     plt.show() |

 下面，来看看更有意思的。

**▌谁上春晚次数最多，堪称「钉子户」？**

很多人梦想着这辈子能上一次春晚，但绝大多数人都只能坐在电视跟前，而有些人则天生是上春晚的，上春晚就像家常便饭，一上就是几十年。你能猜到这**36 年谁上春晚次数最多么？**来看看：



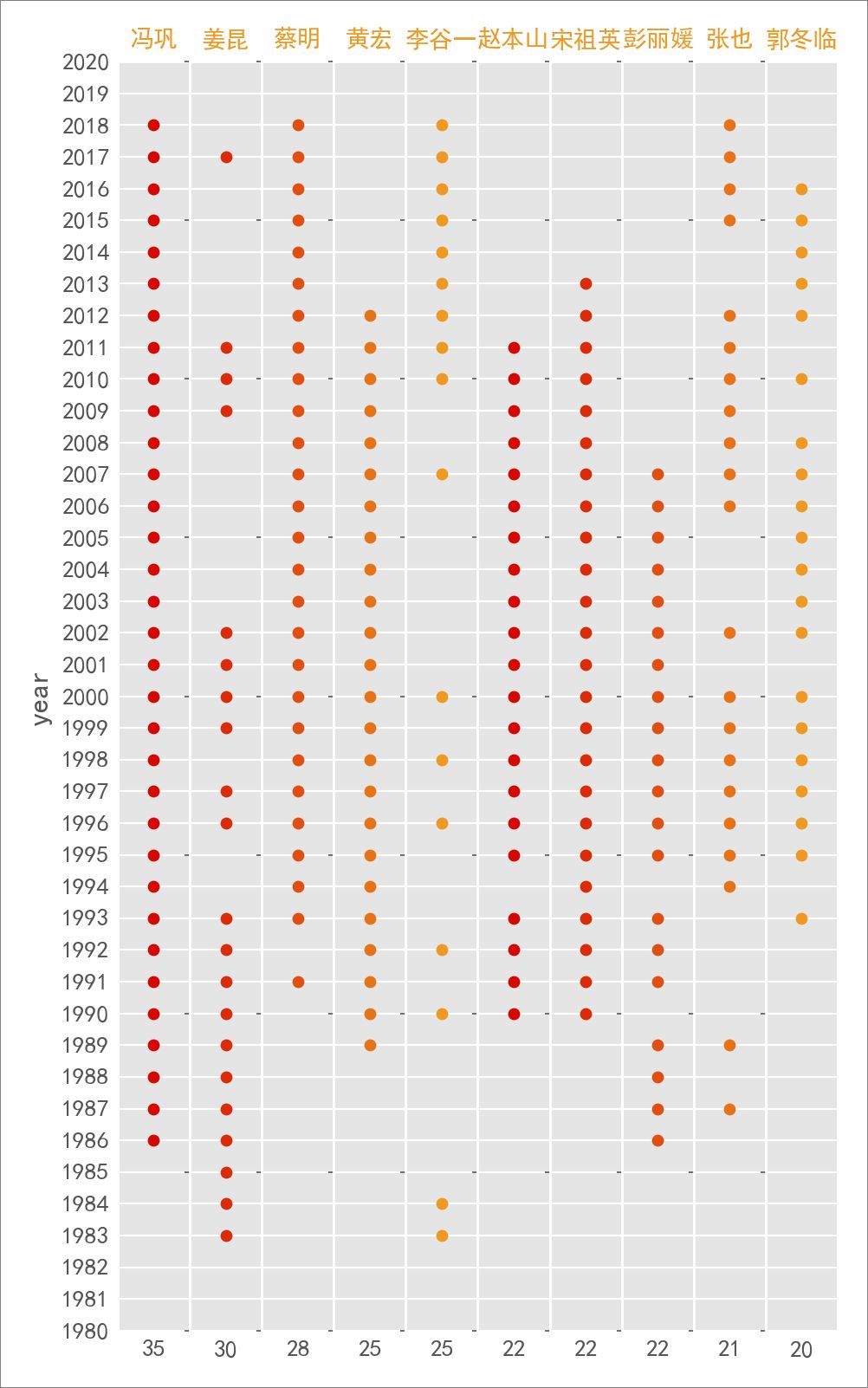
春晚表演次数最多的是相声演员冯巩大叔，你猜对了么。**春晚总共 36 年历史，他就登台了 35 次，十足的「钉子户」。**

前 5 名除了李谷一老师是歌手以外，其他人都是相声或小品演员，也都是你熟悉的面孔：姜昆、蔡明和黄宏。相声和小品为春晚贡献了很多人才。

5 - 10 名歌唱家占多，有宋祖英，还有一位重磅人物，没想到还上了这么多次吧。另外还有两名小品演员本山大叔和黄宏。可惜，不见二人表演已多年。

10 - 20 名还有很多熟悉的人，比如郭冬临、巩汉林、潘长江等。

来具体看一下他们都上了哪些年的春晚：



（注：部分演员表演次数和图中点数不一致，是因为某些年 TA 不只表演了一个节目）

**冯巩老师从 86 年至今，雷打不动地从未缺席过春晚。**他上的最多，每年的开场白都是那句「我想死你们了」，对春晚和观众的热爱之情可见一斑。

他的作品太多了，相声小品通吃，多的让人记不住。



**姜昆老师，登台最早，第一年春晚就上台了。** 这些年，消失几年后又出现在春晚舞台，也许是为了以免大家太过于想他。

最喜欢他 1987 年的《虎口遐想》相声，30 年后的 2017 年又再度带来了《新虎口遐想》。岁月在他的脸上仿佛没有留下痕迹。



（1987 年《虎口遐想》）



（2017 年《新虎口遐想》）

**蔡明老师，是女性中上春晚最多的。**她和许多人合作过，和谁合作效果都好。



（1991 年《陌生人》）

黄宏老师一连登台 25 年，小品中扮演过很多角色，可惜 2012 年之后就没有上台了。

**李谷一老师，春晚演出时间跨度最大**，1983 年有她，2018 年还有她。对她的印象莫过于每年春晚尾声的那曲《难忘今宵》。实际上，李谷一老师曾一人撑起了早期的春晚。1983 年的春晚，她一人连唱 7 首歌，前无古人，也后无来者。



赵本山老师，对他的评价只有一句：**赵本山之后再无春晚。**他在就是压轴，每一部作品都看过很多遍，现在也只有靠怀念，众多作品中，最喜欢两部：



（2005 年《功夫》）

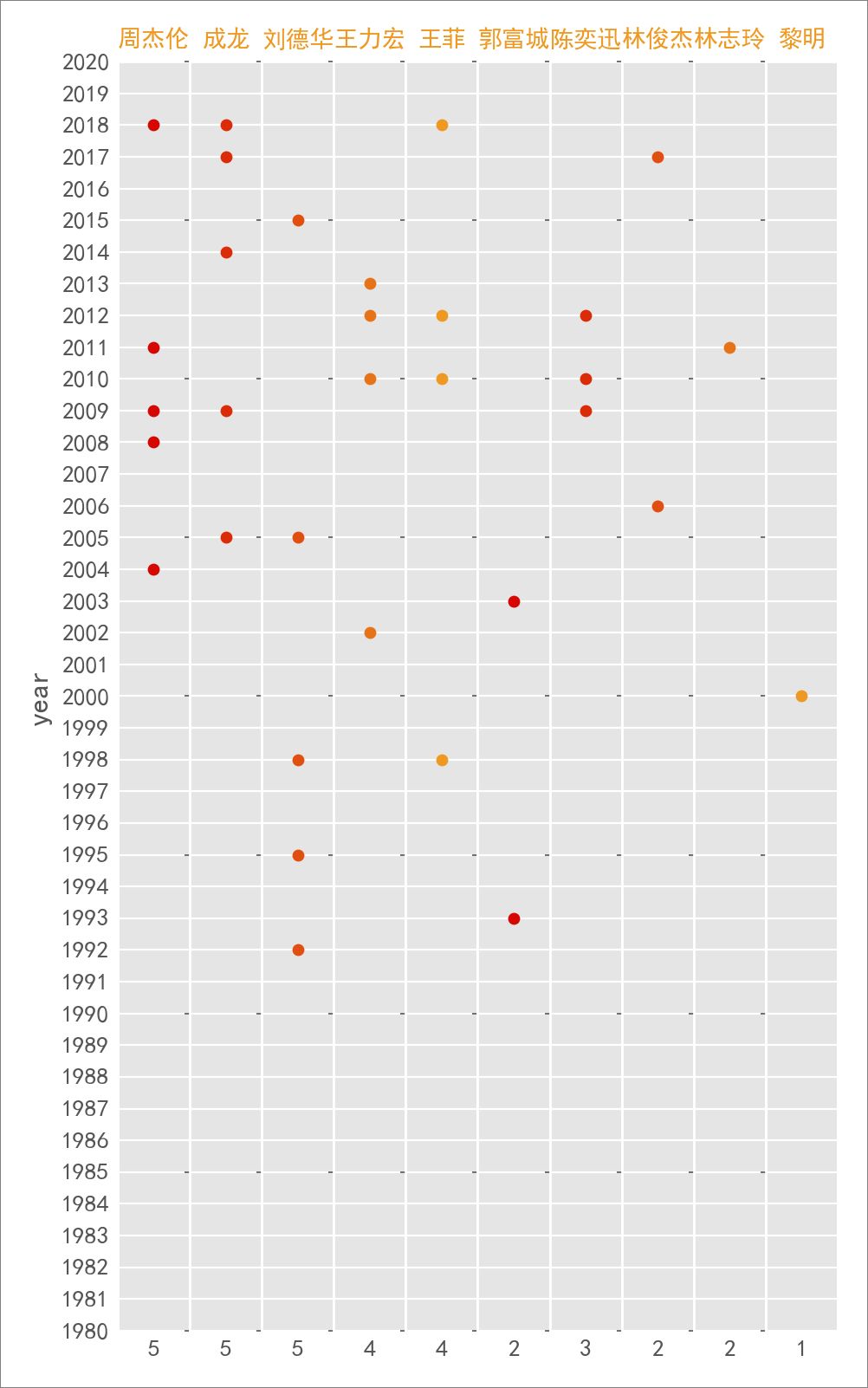


（2009 年《不差钱》）

**▌港台明星上春晚次数对比**

不知你发现没有，前面这些人员都来自内地，港台及海外明星并没有出现，但不可忽略的是，他们的演出让春晚更精彩。

来看看都有哪些港台明星上过春晚，上了几次：



（注：这里选的 10 位明星是第一时间想到的，全凭个人主观印象）

**周董上过 5 次春晚，你猜对了么？**最近一次是去年的《告白气球》，当时朋友圈都刷爆了。除此之外，他其他几次登台，都惊艳无比，04 年的《龙拳》，08 年的《青花瓷》，09 年的《本草纲目》以及 11 年的《兰亭集序》。



15 年，偶像换了一波又一波，他却依旧是那个光芒四射的周杰伦。

成龙大哥同样上了 5 次，一身正气、中国功夫。

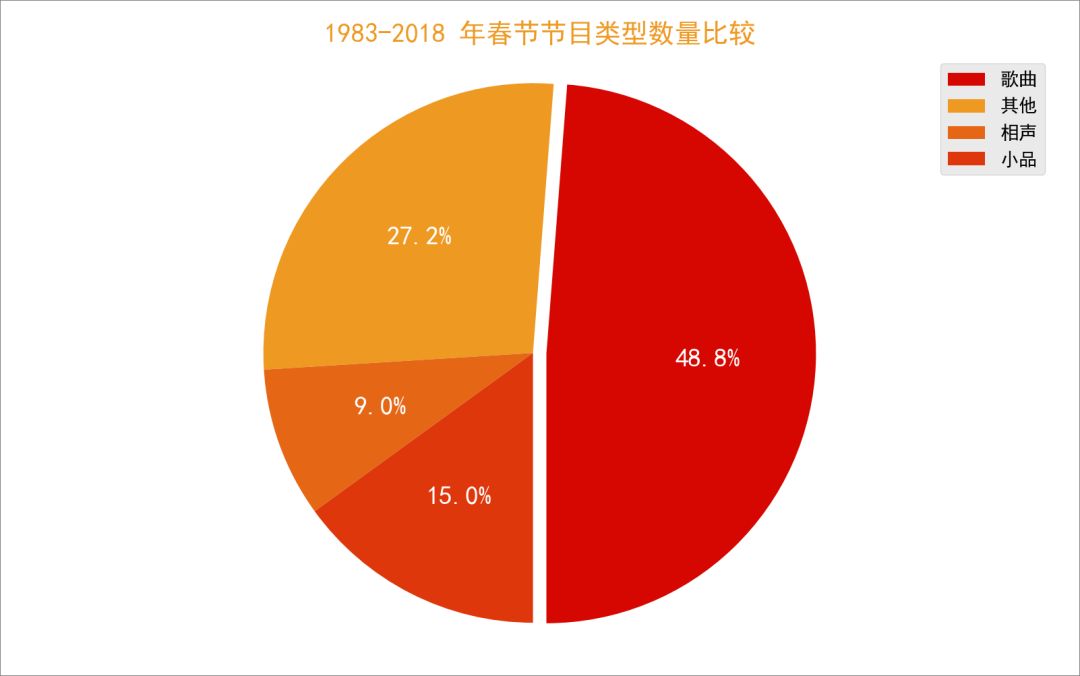
刘德华是四大天王中上春晚最多、最早的，2005 年唱了那首《恭喜发财》之后，每年过年在大街小巷你都能听到。

王力宏上了 4 次，2012 年龙年，再次唱了那首家喻户晓的《龙的传人》。

王菲是港台明星中，上春晚次数最多的女星之一，98 年和那英的《相约九八 》影响力空前。其他几次《传奇》、《因为爱》和《岁月》也空灵好听。

**▌歌曲、小品、相声等节目数量对比**

最后，再来看看春晚各类节目构成比例。印象中，最多的节目就是歌曲，其次是小品、相声。可以看到，歌曲类节目基本占了所有节目的一半，小品占了 15%，相声是 9%。其他类型则是一些杂技、戏剧、舞蹈这些，未作统计。



代码实现如下：

|  |  |
| --- | --- |
| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 | # 各节目数量 def analysis3(data):     fig = plt.figure(figsize=(8,5))     num\_all = data.shape[0]     # 歌曲节目数量     num\_song = data[data['category'].str.contains('歌|尾|开场')].shape[0]     # 小品数量     num\_sketch = data[data['category'].str.contains('小品')].shape[0]     # 相声数量     num\_crosstalk = data[data['category'].str.contains('相声')].shape[0]      # 其他节目数量     other = num\_all - sum([num\_song,num\_sketch,num\_crosstalk])     lst = [num\_song,num\_sketch,num\_crosstalk,other]      sizes = [num\_song,other,num\_crosstalk,num\_sketch]     labels = ['歌曲','其他','相声','小品']     colors\_pie = [color1,color4,color3,color2]     explode = [0.05,0,0,0]     plt.pie(         sizes,         autopct='%.1f%%',         labels=labels,         colors=colors\_pie,         shadow=False,         startangle=270,         explode=explode,         textprops={'fontsize':14,'color':colors}         )     plt.title('1983-2018 共 36 年春晚节目类型数量比较',color=colorstitle,fontsize=fontsize\_title)      plt.tight\_layout()     plt.axis('equal')     plt.axis('off')     plt.legend(loc='upper right')     plt.savefig('1983-2018 共 36 年间各类节目类型数量比较.png',dpi=200)     plt.show() |

**你对春晚有什么想说的嘛？欢迎留言~**

**-END-**

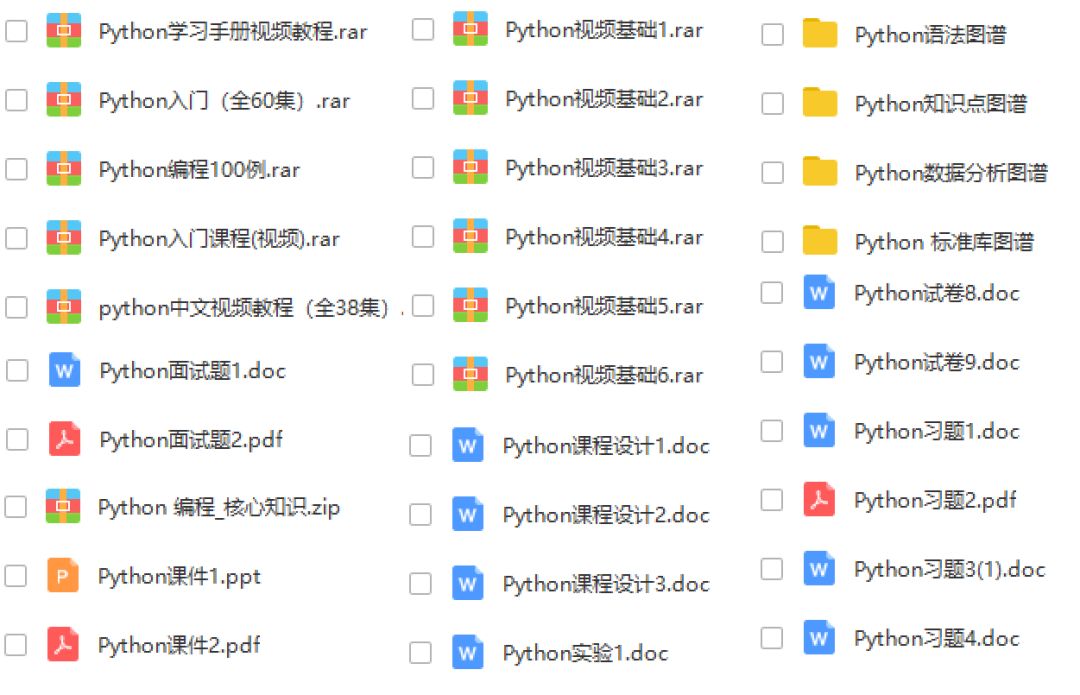


**京东首发热卖中**

* 
* **600分钟超长视频讲解！**本书案例具有实用性，例如校园网搜索引擎、小小翻译器、抓取百度图片这些爬虫案例略加修改可以应用到实际项目中；还有通过微信通信协议开发微信机器人、机器学习的文本分类、基于卷积神经网络的手写体识别等案例；另外是一些大家耳熟能详的游戏案例，例如连连看、推箱子、中国象棋、网络五子棋、两人麻将、人物拼图和飞机大战等游戏。
* 
* **扫描，京东优惠购书中...**

**凡是在京东购买用户，将订单和评价截图发到邮箱itbook8@163.com**

**将获得额外Python大礼包（资源丰富）**



**更多精品图书**

* 
* 
* 

