冬季奥运会数据分析

--“助力冬奥小分队”社会实践团前期数据准备

作者：周殷喆 2018211129班 2018210365号

摘要

关键词：python 可视化 数据分析与深度学习

Data Analysis of Winter Olympic Games

Author:linhaiyuyin(Yinzhe Zhou)

Abstract

Key words: python, Visual Data Analysis and Deep Learning

**引言**

2022年，北京张家口两市联办的第24届冬季奥林匹克运动会将正式召开。为了响应国家号召：青年服务国家，同时也为了进一步锻炼个人实践能力，我们信通院129,107班几位同学申请组建了“助力冬奥小分队”2019暑期社会实践团。这次社会实践旨在宣传推广北京冬奥会，让更多的人了解冬奥会，参与冬奥项目，助理北京冬奥。为了达到更好的实践效果，我决定借助本次python大作业，对历届冬奥会的数据进行分析，先获取更多更广的信息，知己知彼。

冬季奥运会和夏季奥运会在1992年及之前是同年举办，每四年一届。从1992年之后，冬季、夏季奥运会拆分，冬季奥运会从1994年开始，夏季奥运会从1996年开始各自每四年举办一次。每年的冬季奥运会项目由主办方决定，根据东道主的具体情况会有所增减。这对我们接下来的数据分析会产生一些影响。此外，每届冬季奥运会都仅接收以团队为单位的代表参赛，而且正式奥运会比赛之前会在各个国家，各个大洲举行预选赛，只有成绩在录取线之内的选手，才会获得奥运会的参赛资格。因此，奥运会成果可以被看作是各人以及团队，也就是国家的整体竞技水平、国民精神状态的某种表现。

中国第一次参加东家奥运会是在1980年的第十三届冬奥会上。该赛事在美国普莱西德湖举行。这也是中国奥委会在1979年于国际奥委会获得合法席位之后参加的首届冬季奥运会。中国共派出28名男女运动员，参加了滑冰、滑雪、现代冬季两项等十八个单项的比赛。但是由于首次参赛，我国男女选手与世界先进水平有较大差距，无一人进入前六名。直到第十六届冬季奥运会，中国选手一举夺得三枚银牌，实现了奖牌“零的突破”。

1. **基本情况介绍**

本项目github链接：<https://github.com/linhaiyuyin/A-social-practice-about-2022-Beijing-winter-olympic>

下面介绍本次项目的基本逻辑：

1. 寻找或爬取冬季奥林匹克运动会相关的数据集，包括参赛选手基本信息：姓名，代号，团队（国家），年龄，性别，参与项目，参与年份，获得的奖项等等。
2. 检验获取的数据集，筛选目的信息并测试导入情况
3. 第一项：统计冬季奥运会奖牌榜前20（本来是要统计前10，但是前十里边没有中国）并制作奖牌榜。
4. 制作冬季奥林匹克运动会词云，准备进行具体的项目分析
5. 分析哪个国家在哪一届奥运会获得奖牌最多，根据概数据，探索与奖牌数相关的影响因子
6. 探究参加冬奥会的男女选手的数量变化趋势
7. 分析男女选手参与人数最多的项目
8. 分析中国冬奥会相关数据：
9. 中国冬奥会奖牌榜
10. 中国男女选手在冬奥会上广泛参与的项目以及获奖较多的项目
11. 中国奖牌和老牌冰雪项目强国的对比
12. **具体实施**

需要的python库：

import plotly

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from plotly.graph\_objs import \*

from wordcloud import WordCloud, STOPWORDS

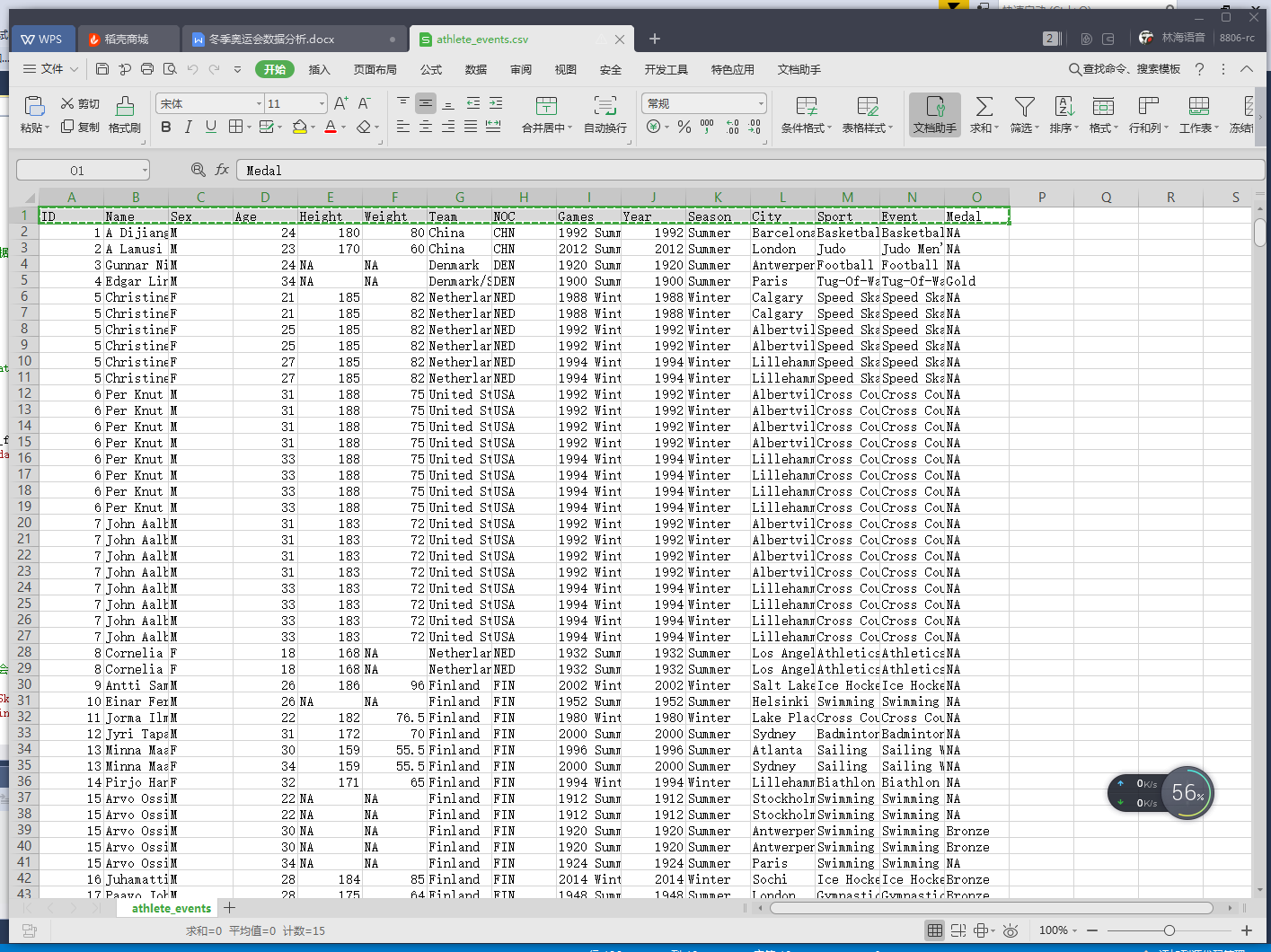
import colorlover as cl

**二．1.数据集搜集**

经过不懈努力，终于发现了开源的数据集。链接如下：

<https://www.kesci.com/home/project/5b6912bc9889570010c244f3/dataset>

这份数据集是从第一届奥运会开始的，包含了选手的各项信息：



包括选手的ID，NAME，Sex，Age,Height,Weight,Team,Year,Games,City,Sport,Event,Medal等全部信息。但是并没有区分冬季还是夏季奥运会。所以我们首先对数据进行了导入和筛选：

#导入数据文件

f\_p='athlete\_events.csv'#文件默认为是已经下载好的本地文件

athlete\_events=pd.read\_csv(f\_p)

'''

#检验数据导入情况

print(athlete\_events.head())

'''

athlete=athlete\_events#从今往后，athlete就是数据集

athlete = athlete[athlete['Season'] == 'Winter']#从今往后，athlete就是只含冬季奥运会项目的数据集

**二．2.制作奖牌榜**

#统计历届冬奥会奖牌榜前10

def countmedalcountrys():

athlete\_winners = athlete[athlete.Medal.notnull()]

winner\_team = athlete\_winners.groupby('Team').size().to\_frame('medal\_count')#采用升序排列

winner\_team = winner\_team.reset\_index().sort\_values('medal\_count',ascending=False)

return winner\_team.head(20)

#将奖牌榜绘制成图像显示

def Drawmedalcountrys(data):

x = data['Team'].astype(str)

y = data['medal\_count'].astype(str)

ax = data.plot.bar(x='Team',y='medal\_count')

ax.set\_xlabel('Top 20')

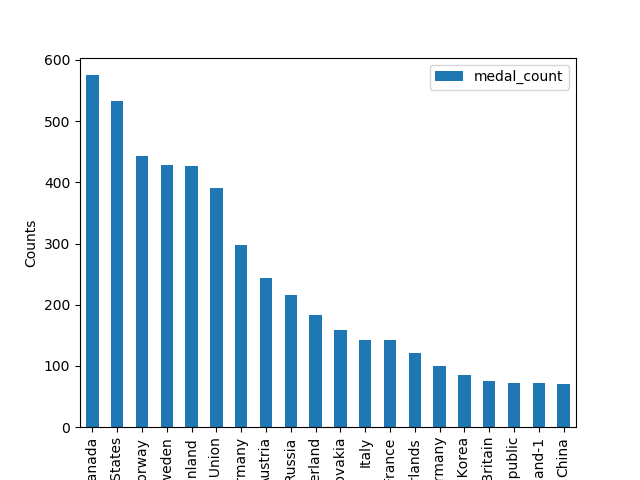
ax.set\_ylabel('Counts')

plt.show()

print (data)

top20data = countmedalcountrys()

Drawmedalcountrys(top20data)

奖牌榜如图：

文字显示如下：

Team medal\_count

8 Canada 575

69 United States 533

47 Norway 443

60 Sweden 428

23 Finland 426

56 Soviet Union 390

26 Germany 297

1 Austria 244

50 Russia 216

61 Switzerland 183

17 Czechoslovakia 158

34 Italy 143

24 France 142

44 Netherlands 122

19 East Germany 100

55 South Korea 86

29 Great Britain 75

16 Czech Republic 73

62 Switzerland-1 72

11 China 70

中国刚好第20位。大多数不怎么关注冬季奥运会的人肯定会很诧异，为什么奖牌榜第一名是加拿大，中国排名为什么这么靠后？是不是数据有问题？当然不是，首先，这份数据是所有冬季奥运会的历史数据，是从1896年开始的。当时，北洋水师刚刚遭遇大败，中国还处在水深火热之中，一代代有志青年长期战斗在民族解放与富强的第一线，当然无暇顾及象征和平与进步的体育竞技。其次，冬季奥运会大致分为两个板块：冰上项目和雪上项目。中国短道速滑、花样滑冰队是世界著名强队，但是中国在雪上项目几乎一直处于疲软的状态。而反观加拿大、挪威、美利坚、这三个典型的多雪山国家，他们包揽奖牌榜前三位就在情理之中了。

**二．3.制作冬季奥运会词云**

词云这个概念是我在搜索数据集时意外发现的，简单的说它就是按照词汇出现的频率给词汇群设置大小，然后显示在一张图片上。这个概念可以很直观地体现选手参与各个项目的相对人数。可以作为之后分析的一个前导。词云制作如下：

#制作冬奥会项目的词云

print(athlete\_events.Sport.unique())#网上找来的历届冬季奥运会项目词库

[ 'Speed Skating'

'Cross Country Skiing' 'Ice Hockey' 'Biathlon' 'Alpine Skiing'

'Luge' 'Hockey''Bobsleigh' 'Figure Skating' 'Nordic Combined'

'Freestyle Skiing' 'Rugby Sevens' 'Trampolining'

'Triathlon' 'Ski Jumping' 'Curling' 'Snowboarding'

'Rugby' 'Short Track Speed Skating' 'Skeleton' 'Lacrosse' 'Polo'

'Cricket' 'Racquets' 'Motorboating' 'Military Ski Patrol' 'Croquet'

'Jeu De Paume' 'Roque' 'Alpinism' 'Basque Pelota' 'Aeronautics']

stopwords = set(STOPWORDS)

def show\_wordcloud(data, title = None):

wordcloud = WordCloud(

stopwords=stopwords,

max\_words=200,

max\_font\_size=40,

scale=3,

random\_state=1 #选取一个随机位置开头

).generate(str(data))

fig = plt.figure(1, figsize=(15, 15))

plt.axis('off')

if title:

fig.suptitle(title, fontsize=20)

fig.subplots\_adjust(top=2.3)

plt.imshow(wordcloud)

plt.show()

show\_wordcloud(athlete['Sport'], title = "WordCloud")

词云如下：



词云显示如下：

['Basketball' 'Judo' 'Football' 'Tug-Of-War' 'Speed Skating'

'Cross Country Skiing' 'Athletics' 'Ice Hockey' 'Swimming' 'Badminton'

'Sailing' 'Biathlon' 'Gymnastics' 'Art Competitions' 'Alpine Skiing'

'Handball' 'Weightlifting' 'Wrestling' 'Luge' 'Water Polo' 'Hockey'

'Rowing' 'Bobsleigh' 'Fencing' 'Equestrianism' 'Shooting' 'Boxing'

'Taekwondo' 'Cycling' 'Diving' 'Canoeing' 'Tennis' 'Modern Pentathlon'

'Figure Skating' 'Golf' 'Softball' 'Archery' 'Volleyball'

'Synchronized Swimming' 'Table Tennis' 'Nordic Combined' 'Baseball'

'Rhythmic Gymnastics' 'Freestyle Skiing' 'Rugby Sevens' 'Trampolining'

'Beach Volleyball' 'Triathlon' 'Ski Jumping' 'Curling' 'Snowboarding'

'Rugby' 'Short Track Speed Skating' 'Skeleton' 'Lacrosse' 'Polo'

'Cricket' 'Racquets' 'Motorboating' 'Military Ski Patrol' 'Croquet'

'Jeu De Paume' 'Roque' 'Alpinism' 'Basque Pelota' 'Aeronautics']

简单看词云，可以发现，乡村滑雪、越野滑雪、短道速滑、大道速滑是最热门的几类项目。

在制作词云的过程中，出现了两个主要的问题：

1. 尽管选取的数据集是已经处理过的，只有冬季奥运会项目的数据，但是最终的词云里还是出现了夏季的项目。经过查看数据集我们发现：在南非等南半球国家举办的奥运会，北方冬季举办的是夏季奥运会！真是让人心酸。苦思良久，最终决定忽略这个问题，因为南半球的冬季奥运会次数太少。
2. 在制作词云图片的时候，初始的图片总是挤在Figure界面的最上方，需要手动在图片界面调整一次之后才能正常显示。这个问题目前只好通过手动调整一次来解决。但是在其他的部分都没有出现类似的问题。

**二．4.分析那些国家在哪届冬奥会上夺得奖牌最多**

实现代码如下：

#下面分析哪个国家在哪一届获得的奖牌最多

def countCountryMedal():

athlete\_winners=athlete[athlete.Medal.notnull()]

winner\_teams=athlete\_winners.groupby(['Team','Year']).size().to\_frame('medal\_count').reset\_index().sort\_values('medal\_count',ascending=False)

return winner\_teams.head(20)

print(countCountryMedal())

展示效果如下：

Team Year medal\_count

493 United States 2010 89

77 Canada 2014 82

76 Canada 2010 80

74 Canada 2002 72

491 United States 2002 70

75 Canada 2006 67

420 Sweden 2006 64

160 Finland 1998 58

353 Russia 2014 56

494 United States 2014 52

385 Soviet Union 1988 52

422 Sweden 2014 51

**470 Unified Team 1992 51**

382 Soviet Union 1976 50

199 Germany 2002 49

492 United States 2006 48

73 Canada 1998 47

163 Finland 2010 47

383 Soviet Union 1980 46

379 Soviet Union 1964 45

在这项数据当中存在一项：470 Unified Team，这是1991年12月前苏联解体，各联盟国没有完全自立之前，他们以独联体的身份临时参加的一次比赛。虽然苏联解体，运动员们却丝毫没有泄气，他们仍旧获得了那届冬奥会的最佳成绩。在1994年冬奥会独立开展之后，加拿大，美国，几乎包揽了奖牌榜，偶尔也有芬兰，瑞典，俄罗斯和德意志的身影。通过这项数据，我们基本认定了当今世界的冰雪体育项目强国为：加拿大，美利坚，瑞典，芬兰，俄罗斯和挪威。他们将是我们中国代表队最好的学习对象和最强的竞争对手。

**二．5.额外话题**

现代的冬季奥林匹克运动会是完全男女平等，自由竞争的，但是男女选手在冬奥会赛场上的参与度是从什么时候开始接近的呢？下面就来简单分析下冬奥会男女选手数量的变化趋势，以及男女选手参赛人数最多的项目：

#参加冬奥会男女选手数量变化

def SexTrendD():

sexathlete=athlete.groupby(["Year","Sex"]).size()

sexathlete=sexathlete.unstack()

ax=sexathlete.plot()

ax.set\_title("Sex")

ax.set\_xlabel("Year")

ax.set\_ylabel("Counts")

plt.show()

SexTrendD()

#分析男女选手参与人数最多的项目

def MostEventsSex():

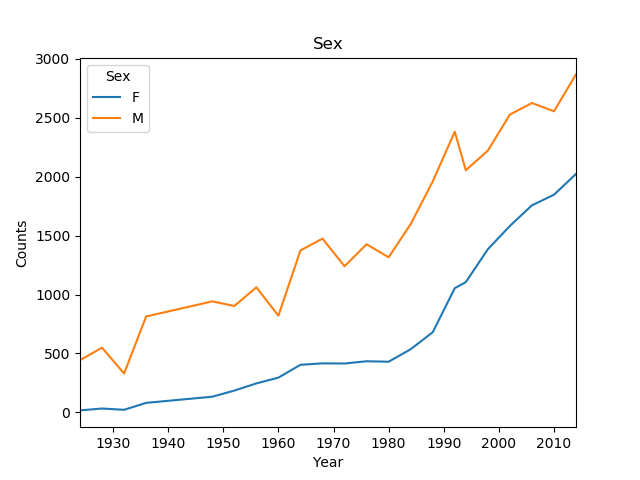
events=athlete.groupby(["Event","Sex"]).size().sort\_values(ascending=False).reset\_index()

print(events[events.Sex=='M'].head(5))

print(events[events.Sex=='F'].head(5))

MostEventsSex()

下面是男女选手参与人数的变化：



下面是男女选手参与人数变化趋势和参与项目的排名：

Sex F M

Year

1924 17 443

1928 33 549

1932 22 330

1936 81 814

1948 133 942

1952 185 903

1956 246 1061

1960 295 821

1964 404 1374

1968 416 1475

1972 415 1240

1976 434 1427

1980 430 1316

1984 536 1598

1988 680 1959

1992 1054 2382

1994 1105 2055

1998 1384 2221

2002 1582 2527

2006 1757 2625

2010 1847 2555

2014 2023 2868

Event Sex 0

0 Ice Hockey Men's Ice Hockey M 4702

1 Bobsleigh Men's Four M 1715

2 Alpine Skiing Men's Slalom M 1593

3 Alpine Skiing Men's Giant Slalom M 1535

4 Alpine Skiing Men's Downhill M 1164

Event Sex 0

11 Alpine Skiing Women's Slalom F 966

12 Alpine Skiing Women's Giant Slalom F 961

20 Ice Hockey Women's Ice Hockey F 754

22 Alpine Skiing Women's Downhill F 708

23 Cross Country Skiing Women's 10 kilometres F 667

从第一项数据以及图像中我们惊奇地发现，男女人数始终相差在800-1000名左右。根据奥组委官方给出的通报，造成男女运动员人数差异的直接原因是报名人数而不是比赛项目或者举办国家。也就是说，当前冬季奥运会男女人数相差较大的原因主要是女同胞参与度不足。

从第二项数据当中，我们发现男女选手最喜爱的项目有明显的差别。由于缺乏相关的知识，这项数据我们仅作了解。

1. **中国冬奥会数据分析**

**三．1中国冬奥会数据**

def CountryC():

athlete\_winners=athlete[athlete.Medal.notnull()]

winner\_team=athlete\_winners.groupby(['Team','Year']).size().to\_frame('medal\_count').reset\_index().sort\_values('Year',ascending=False)

winner\_China=winner\_team[winner\_team.Team=="China"]

return winner\_China

def DrawCountryDe(data):

x = data['Year'].astype(str)

y = data['medal\_count'].astype(str)

ax=data.plot(x='Year',y='medal\_count')

C=str(data.Team)

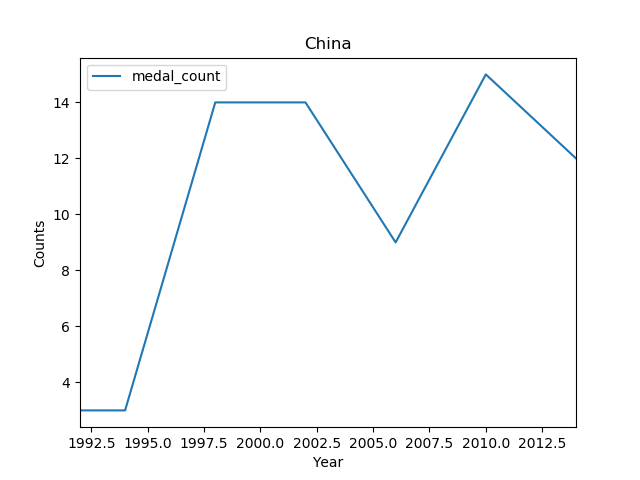
ax.set\_title(C)

ax.set\_xlabel('Year')

ax.set\_ylabel('Counts')

plt.show()

print(data)

DrawCountryDe(CountryC())

Team Year medal\_count

96 China 2014 12

95 China 2010 15

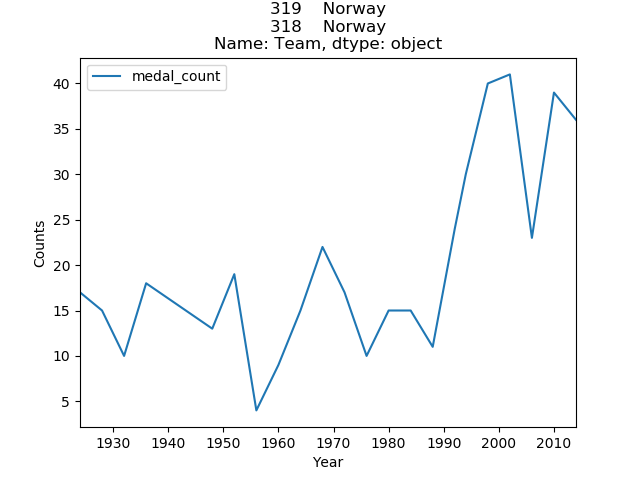
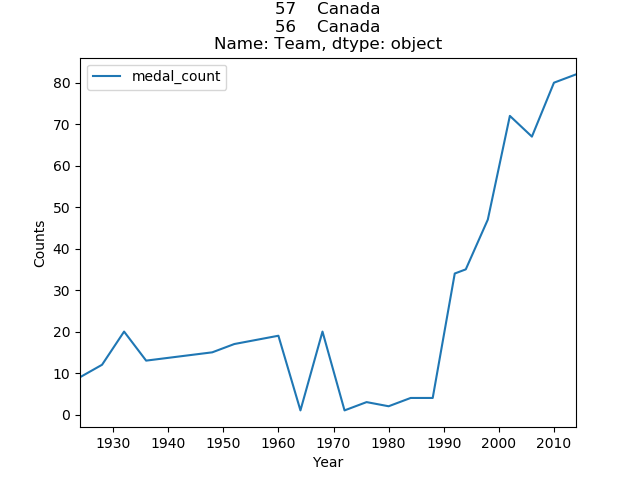
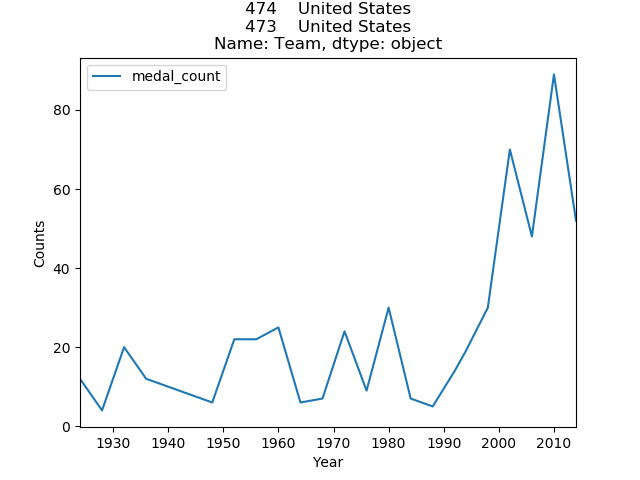
94 China 2006 9

93 China 2002 14

92 China 1998 14

91 China 1994 3

90 China 1992 3

通过分析这组图片及数据，我们发现我国的确是在1992年底十六届冬奥会上实现了奖牌0的突破，但是在这之后，我国在每届冬奥会上所能取得的成就都十分有限。可见我们国家在冰雪项目上还有很长的路要走，为什么这么说呢，为了让大家有更加直观的认识，我们决定将之前分析出的老牌强国也画出数据来：

这下，差距立显！

Team Year medal\_count

77 Canada 2014 82

76 Canada 2010 80

75 Canada 2006 67

74 Canada 2002 72

73 Canada 1998 47

72 Canada 1994 35

71 Canada 1992 34

70 Canada 1988 4

69 Canada 1984 4

68 Canada 1980 2

67 Canada 1976 3

66 Canada 1972 1

65 Canada 1968 20

64 Canada 1964 1

63 Canada 1960 19

62 Canada 1956 18

61 Canada 1952 17

60 Canada 1948 15

59 Canada 1936 13

58 Canada 1932 20

57 Canada 1928 12

56 Canada 1924 9

Team Year medal\_count

494 United States 2014 52

493 United States 2010 89

492 United States 2006 48

491 United States 2002 70

490 United States 1998 30

489 United States 1994 19

488 United States 1992 14

487 United States 1988 5

486 United States 1984 7

485 United States 1980 30

484 United States 1976 9

483 United States 1972 24

482 United States 1968 7

481 United States 1964 6

480 United States 1960 25

479 United States 1956 22

478 United States 1952 22

477 United States 1948 6

476 United States 1936 12

475 United States 1932 20

474 United States 1928 4

473 United States 1924 12

Team Year medal\_count

339 Norway 2014 36

338 Norway 2010 39

337 Norway 2006 23

336 Norway 2002 41

335 Norway 1998 40

334 Norway 1994 30

333 Norway 1992 24

332 Norway 1988 11

331 Norway 1984 15

330 Norway 1980 15

329 Norway 1976 10

328 Norway 1972 17

327 Norway 1968 22

326 Norway 1964 15

325 Norway 1960 9

324 Norway 1956 4

323 Norway 1952 19

322 Norway 1948 13

321 Norway 1936 18

320 Norway 1932 10

319 Norway 1928 15

318 Norway 1924 17

**三．2究竟是哪些项目一直无法突破呢？**

为了进一步找到我们中国体育健儿擅长以及不擅长的项目领域，我们进一步分析中国运动员参赛项目主要有哪些：

#分析中国男女选手在冬奥会上广泛参与的项目

def MostEventsChina():

Cevents=athlete.groupby(["Event","Team"]).size().sort\_values(ascending=False).reset\_index()

print(Cevents[Cevents.Team=="China"].head(10))

MostEventsChina()

反馈回来的结果是：

Event Team 0

101 Ice Hockey Women's Ice Hockey China 58

308 Speed Skating Women's 500 metres China 34

335 Speed Skating Women's 1,000 metres China 32

400 Short Track Speed Skating Women's 3,000 metres... China 28

490 Short Track Speed Skating Men's 5,000 metres R... China 25

500 Speed Skating Men's 500 metres China 25

526 Biathlon Women's 7.5 kilometres Sprint China 24

583 Biathlon Women's 15 kilometres China 23

615 Freestyle Skiing Women's Aerials China 22

625 Speed Skating Men's 1,000 metres China 22

通过分析这项数据，我们很清楚地认识到：除了Biathlon（冬季两项）和自由式滑雪空中技巧Freestyle Skiing Women's Aerials，中国队员参加的其他项目全部都是冰上项目skating或者曲棍球Hockey！那么我们专注于冰上项目，成绩如何呢？

#分析中国男女选手在冬奥会上广泛参与的项目

def MostEventsChina():

Cevents=athlete.groupby(["Event","Team"]).size().sort\_values(ascending=False).reset\_index()

CeventsM=athlete.groupby(["Event","Team","Medal"]).size().sort\_values(ascending=False).reset\_index()

print(Cevents[Cevents.Team=="China"].head(10))

print(CeventsM[CeventsM.Team=="China"].head(10))

MostEventsChina()

（上边这个函数其实和刚才统计总数的是一个函数）

反馈的结果是这样的：

Event Team Medal 0

49 Short Track Speed Skating Men's 5,000 metres R... China Bronze 13

101 Short Track Speed Skating Women's 3,000 metres... China Silver 8

283 Short Track Speed Skating Women's 3,000 metres... China Gold 4

344 Short Track Speed Skating Women's 500 metres China Gold 4

355 Curling Women's Curling China Bronze 4

365 Freestyle Skiing Women's Aerials China Silver 4

528 Short Track Speed Skating Women's 500 metres China Silver 3

553 Short Track Speed Skating Women's 1,000 metres China Silver 3

735 Speed Skating Women's 500 metres China Bronze 2

739 Speed Skating Women's 500 metres China Silver 2

分析上边的数据，我们专注于参加冰上项目，当然也只能在冰上项目有所突破啦，但是其实顶尖水平运动员（获得过奖牌的），最高比例也只有13/25，而平均比例就直接惨不忍睹了。

**三．3.其他强国对比**

为了进一步探究我们的竞争对手实力有多么恐怖，我们对Canada、USA、Norway三个国家做与中国相同的分析如下：

def compare():

Oevents=athlete.groupby(["Event","Team"]).size().sort\_values(ascending=False).reset \_index() OeventsM=athlete.groupby(["Event","Team","Medal"]).size().sort\_values(ascending=False). reset\_index()

print(Oevents[Oevents.Team=="Canada"].head(10))

print(Oevents[Oevents.Team=="Norway"].head(10))

print(Oevents[Oevents.Team=="United States"].head(10))

print(OeventsM[OeventsM.Team=="Canada"].head(10))

print(OeventsM[OeventsM.Team=="Norway"].head(10))

print(OeventsM[OeventsM.Team=="United States"].head(10))

compare()

反馈回来的结果是这样的：

Event Team 0

2 Ice Hockey Men's Ice Hockey Canada 360

18 Ice Hockey Women's Ice Hockey Canada 101

73 Speed Skating Men's 500 metres Canada 68

81 Alpine Skiing Men's Downhill Canada 65

85 Speed Skating Men's 1,500 metres Canada 64

110 Ski Jumping Men's Normal Hill, Individual Canada 56

113 Alpine Skiing Women's Downhill Canada 56

134 Alpine Skiing Men's Slalom Canada 53

167 Alpine Skiing Women's Giant Slalom Canada 49

174 Speed Skating Women's 1,000 metres Canada 48

这是加拿大人的数据，他们参赛的主要人数在曲棍球，短道速滑，高山滑雪上

Event Team 0

8 Ice Hockey Men's Ice Hockey Norway 223

24 Ski Jumping Men's Normal Hill, Individual Norway 88

29 Cross Country Skiing Men's 50 kilometres Norway 84

32 Speed Skating Men's 1,500 metres Norway 81

33 Nordic Combined Men's Individual Norway 80

48 Speed Skating Men's 5,000 metres Norway 73

49 Cross Country Skiing Men's 4 x 10 kilometres R... Norway 72

79 Speed Skating Men's 10,000 metres Norway 66

92 Speed Skating Men's 500 metres Norway 62

104 Alpine Skiing Men's Slalom Norway 57

这是挪威人的数据，他们参赛的主要人数在曲棍球，跳台式滑雪、越野滑雪、短道速滑上

Event Team 0

0 Ice Hockey Men's Ice Hockey United States 386

19 Ice Hockey Women's Ice Hockey United States 101

25 Speed Skating Men's 500 metres United States 85

26 Speed Skating Men's 1,500 metres United States 85

28 Ski Jumping Men's Normal Hill, Individual United States 85

43 Nordic Combined Men's Individual United States 75

45 Cross Country Skiing Men's 4 x 10 kilometres R... United States 74

51 Alpine Skiing Men's Downhill United States 72

56 Alpine Skiing Women's Slalom United States 70

58 Alpine Skiing Men's Slalom United States 70

这是美利坚人的数据，他们参赛的主要人数分布在曲棍球、短道速滑、跳台式滑雪上。

**再看这三个国家的奖牌分布在哪些项目上：**

Event Team Medal 0

0 Ice Hockey Men's Ice Hockey Canada Gold 131

4 Ice Hockey Women's Ice Hockey Canada Gold 81

7 Ice Hockey Men's Ice Hockey Canada Silver 73

15 Ice Hockey Men's Ice Hockey Canada Bronze 35

32 Ice Hockey Women's Ice Hockey Canada Silver 20

38 Short Track Speed Skating Women's 3,000 metres... Canada Silver 17

45 Curling Men's Curling Canada Gold 14

46 Short Track Speed Skating Men's 5,000 metres R... Canada Gold 14

69 Curling Men's Curling Canada Silver 10

70 Curling Women's Curling Canada Bronze 10

加拿大人的奖牌主要分布在曲棍球上，前五位居然都是曲棍球！然后是Track Speed，这个是大道速滑

Event Team Medal 0

16 Cross Country Skiing Men's 4 x 10 kilometres R... Norway Silver 30

30 Cross Country Skiing Women's 4 x 5 kilometres ... Norway Silver 20

47 Cross Country Skiing Men's 4 x 10 kilometres R... Norway Gold 14

50 Biathlon Men's 4 x 7.5 kilometres Relay Norway Silver 12

57 Ski Jumping Men's Large Hill, Team Norway Bronze 12

75 Nordic Combined Men's Individual Norway Gold 10

86 Speed Skating Men's 5,000 metres Norway Gold 9

93 Speed Skating Men's 1,500 metres Norway Bronze 8

104 Speed Skating Men's 1,500 metres Norway Gold 8

105 Nordic Combined Men's Team Norway Gold 8

挪威人的奖牌主要分布在越野滑雪上，这是一个纯粹的雪上强国

Event ... 0

2 Ice Hockey Men's Ice Hockey ... 117

8 Ice Hockey Women's Ice Hockey ... 61

14 Ice Hockey Men's Ice Hockey ... 36

27 Ice Hockey Women's Ice Hockey ... 20

31 Ice Hockey Women's Ice Hockey ... 20

67 Ice Hockey Men's Ice Hockey ... 11

78 Short Track Speed Skating Men's 5,000 metres R... ... 9

88 Short Track Speed Skating Women's 3,000 metres... ... 9

110 Short Track Speed Skating Men's 5,000 metres R... ... 8

123 Figure Skating Women's Singles ... 8

美利坚人的奖牌首先分布在曲棍球上，然后在短道速滑项目上也有一席之地。

这样一来，我们中国的问题就很明显了：首先，曲棍球项目就像夏季运动会当中的足球，几乎没有我们中国队的席位。其次在滑雪、跳台滑雪等雪上项目中，中国队表现不佳，参与度不高。最后，值得一提的是，我们在短道速滑项目上有比较强的优势。但是细心的人们可以发现，上边中国队奖牌数据中，短道速滑是以铜牌为主，银牌为辅，几乎没有金牌的。这个冬季项目爱好者都明白，金牌主要在韩国队手上。

综上所述，简要阐明中国冬季项目的现状：冰上项目为主，雪上项目较劣势。团队球类项目无缘名次，冰上项目经常屈居第二。

1. **小结**

经过这次简单的数据分析，我们总结找出了冬季奥林匹克运动会历史上的强国：加拿大，美国，挪威等。通过这些强国和中国的比较，我们发现了中国在雪上项目处于巨大劣势，在冰上项目一直屈居第二的现状。此外，我们还意识到，女同胞对体育运动项目的关注度，参与度不足，也是导致男女参赛人数有较大差距的原因之一。这对我们接下来的社会实践起到了很关键的指导作用。我们将主要聚焦在雪上项目的宣传推广，同时鼓舞更多的女同胞积极参与到奥林匹克体育运动当中来。争取在2022年冬季奥运会开始时，展现出一个更好、更强大的中国风貌。

1. **存在的问题与未来拓展**

首先，本次实验区分冬季和夏季奥运会的方式，是Season，但是没有意识到在南半球举办的冬季奥运会是7-8月，也就是北半球的夏季。这导致了大概5%左右的数据错乱（因为南半球举办的冬季奥运会本身就非常少）

然后是对数据集的处理不是很成熟。在这份数据集中，有很多信息是统计不完全的，即有空的格子，但是在做分析的时候，我们忽略了这部分信息，而不是按照惯例使用平均数来代替。这可能待本次实验带来一定的误差。之后将进一步处理相关的数据。

其次是绘图的时候存在较大的问题，由于我队绘图语言掌握的不充分，导致图像上其实有诸多问题暂时没有解决。之后将完善图像的设计。

最后是深度问题。本次实验的分析基本停留在大团队、大项目类的总结性分析上，没有深入到具体的数据当中去，这一点是本次项目后续的发展方向。

1. **参考资料**
2. **百度百科：[www.baike.baidu.com](http://www.baike.baidu.com)**
3. **百度翻译：[www.fanyi.baidu.com](http://www.fanyi.baidu.com)**
4. **科赛：[www.kesci.com](http://www.kesci.com),**
5. **数据集：www.kesci.com/home/project/5b6912bc9889570010c244f3/dataset**
6. **制作词云：<https://blog.csdn.net/xc_zhou/article/details/80637754>**
7. **SQL分析奥运会：<https://blog.csdn.net/weixin_42215007/article/details/81984930>**
8. **Python数据文件处理：<https://blog.csdn.net/heavenmark/article/details/73518853>**
9. **大数据文摘：https://mp.weixin.qq.com/s?\_\_biz=MjM5MTQzNzU2NA==&mid=2651671119&idx=2&sn=f90221e8351ed7a9433385782fa59a40&chksm=bd4c6ddc8a3be4ca2df7279eb3e38858c718bf05867cf52300e74ab5de000b6a86faa2907be1&mpshare=1&scene=1&srcid=&key=6ff676ea8e4091fdecddc95655f0c0e5d34eebdcbcd69a12a8fc227af7eff18bd2dfcc9ceb5b85f6dcaa7e25d95fb23ce10e04a62729136152ec46becde24b07972511b9d7519ab1c99bf51ead804e17&ascene=1&uin=MjM0OTc5NTc1&devicetype=Windows+10&version=62060833&lang=zh\_CN&pass\_ticket=uCDHFJpF8ZbwFOrHXGBw0In8UKF0B8saAYwomoUxcxA%3D**
10. **Github相关：www.github.com**