

# Table of Contents

- [1 什么是缺失值](#)
- ▼ [2 发现缺失值](#)
  - [2.1 df.isnull\(\)](#)
  - [2.2 df.notnull\(\)](#)
- ▼ [3 剔除缺失值df.dropna\(\)](#)
  - [3.1 参数how](#)
  - [3.2 参数thresh](#)
  - [3.3 参数subset](#)
  - [3.4 练习](#)
- ▼ [4 填充缺失值df.fillna\(\)](#)
  - [4.1 固定值填充](#)
  - [4.2 从前往后填充](#)
  - [4.3 从后往前填充](#)
  - [4.4 df.replace\(\)](#)

```
In [2]: #全部行都能输出
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"

import numpy as np
import pandas as pd
```

## 1 什么是缺失值

在了解缺失值（也叫控制）如何处理之前，首先要知道的就是什么是缺失值？直观上理解，缺失值表示的是“缺失的数据”。

可以思考一个问题：是什么原因造成的缺失值呢？其实有很多原因，实际生活中可能由于有的数据不全所以导致数据缺失，也有可能由于误操作导致数据缺失，又或者人为地造成数据缺失。

在DataFrame中被当作是缺失值来处理的有：

- Python对象中的None
- np中的np.nan,显示标签是"NaN"
  - "NaN"是特殊的浮点数

其实如果DataFrame中存在Python中的None对象，一旦转化为dtype类型数据(object类型除外)，None对象都会转成np.nan。

```
In [3]: df_01=pd.DataFrame([1,2,None]) #在None转化为NaN过程中，object对象转化为float64，因为"NaN"是特殊的
df_01
df_01[0].dtype
```

Out[3]:

	0
0	1.0
1	2.0
2	NaN

Out[3]: dtype('float64')

```
In [4]: type(np.nan)
```

```
Out[4]: float
```

```
In [5]: df_01=pd.DataFrame([1,2,np.nan])
df_01
df_01[0].dtype
```

```
Out[5]:
```

	0
0	1.0
1	2.0
2	NaN

```
Out[5]: dtype('float64')
```

```
In [9]: #在None转化为NaN过程中，如果通过Series来指定数据类型，object对象不会转化为float64
ser_01=pd.Series(range(2),dtype=object)
a=pd.DataFrame(ser_01)

a
a[0].dtype
```

```
Out[9]:
```

	0
0	0
1	1

```
Out[9]: dtype('O')
```

## 2 发现缺失值

### 2.1 df.isnull()

```
In [10]: heros_null=pd.read_csv("user_inforNaN.csv", index_col="Unnamed: 0")
heros_null
```

Out[10]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

判断是否为空值：dataframe.isnull()或者dataframe.notnull()

```
In [11]: heros_null.isnull()
```

Out[11]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	True	True
2	False	False	False	False	False	False	False	True
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	True	True
5	False	False	False	False	False	False	True	True
6	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False

当然，一般我们是对单列变量的缺失值进行观察,并利用返回的布尔型Series对DataFrame进行筛选：

```
In [12]: heros_null.Height.isnull()

heros_null[heros_null.Height.isnull()]
```

```
Out[12]: 0    False
1     True
2    False
3    False
4     True
5     True
6    False
7    False
Name: Height, dtype: bool
```

```
Out[12]:
```

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN

## 2.2 df.notnull()

```
In [13]: heros_null.Weight.notnull()
heros_null[heros_null.Weight.notnull()]
```

```
Out[13]: 0     True
1    False
2    False
3     True
4    False
5    False
6     True
7     True
Name: Weight, dtype: bool
```

```
Out[13]:
```

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

## 3 剔除缺失值df.dropna()

```
In [14]: heros_null
```

Out[14]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

直接对DataFrame对象使用.dropna()方法会删除掉所有带有缺失值的行（返回新表，并不是改变原表）。

```
In [15]: heros_null.dropna()
```

Out[15]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

如果里面增加axis参数，指定第二坐标轴，就会默认删掉出现缺失值的列。

```
In [19]: heros_null.dropna(axis=1)
```

Out[19]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械

df.dropna( )中的参数how和thresh是设置删除行列标准的参数：

### 3.1 参数how

In [16]: heros\_null.dropna(axis=1,how="all")

#how="all"只有当该列（或行）全都为缺失值时，才会将该列删除  
#how="any"只有当该列（或行）有一个缺失值，就会将该列删除

Out[16]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

### 3.2 参数thresh

thresh参数设置的是：你想至少留下多少非缺失值！

In [17]: heros\_null.dropna(axis=0,thresh=7)

#一共8列，至少留下7列，就是说，如果有一行中有两个缺失值，就删掉

Out[17]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

### 3.3 参数subset

subset设定一个子集，子集中的列作为剔除缺失值的参考列：

In [18]:

heros\_null.dropna(axis=0,how="any")

Out[18]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

In [19]:

heros\_null.dropna(axis=0,how="any",subset=["Birthplace","Weapon"])

Out[19]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

### 3.4 练习

如何删除heros\_null中缺失值超过超过45%的字段。

In [23]:

heros\_null.shape  
heros\_null.shape[0]

Out[23]: (8, 8)

Out[23]: 8

In [22]:

Out[22]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0

## 4 填充缺失值df.fillna()

In [24]: heros\_null

Out[24]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

### 4.1 固定值填充

In [26]: heros\_null.fillna(0) *#用0来填充缺失值*

Out[26]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	0.0	0.0
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	0.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	0.0	0.0
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	0.0	0.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

填充均值容易受到异常值的影响，所以可以使用：

- 中位数：Series.median()
- 众数：Series.mode() 但并不是所有数据集都有众数，如[1,2,3,4,5]



In [32]: heros\_null

Out[32]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

In [43]: median=heros\_null.Height.median()  
mode=heros\_null.Weight.median()  
median  
mode  
  
heros\_null.fillna(value={"Height":median,"Weight":mode})

Out[43]: 168.0

Out[43]: 67.0

Out[43]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	168.0	67.0
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	67.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	168.0	67.0
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	168.0	67.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

## 4.2 从前往后填充

In [44]: heros\_null.fillna(method="ffill") #forward-fill

Out[44]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	168.0	66.0
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	66.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	172.0	72.0
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	172.0	72.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

### 4.3 从后往前填充

In [45]: heros\_null.fillna(method="bfill") #back-fill

Out[45]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	178.0	72.0
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.0	72.0
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	168.0	68.0
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	168.0	68.0
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

除了通过 fillna 方法来填充缺失值外，还可以通过 interpolate 方法来填充。默认情况下使用线性差值，可以设置 method 参数来改变方式。

In [46]: heros\_null.interpolate()

Out[46]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	纽约	蜘蛛感应	168.000000	66.000000
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	173.000000	68.000000
2	本尼迪克特·康伯巴奇	奇异博士	36	男	费城	魔法	178.000000	70.000000
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.000000	72.000000
4	本·阿弗莱克	蝙蝠侠	40	男	哥谭	有钱	170.666667	70.666667
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	169.333333	69.333333
6	盖尔·加朵	神奇女侠	2000	女	天堂岛	弑神者	168.000000	68.000000
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.000000	65.000000

## 4.4 df.replace()

None、np.nan、NaT（时间数据类型的缺失值） 这些都是缺失值。

但是这些在 Pandas 的眼中是缺失值，有时候在我们人类的眼中，某些异常值我们也会当做缺失值来处理。例如，在我们的存储的用户信息中，假定我们限定用户都是青年，年龄字段出现了80，我们就可以认为这是一个异常值。

再比如，我们都知道性别分为男性（male）和女性（female），在记录用户性别的时候，对于未知的用户性别都记为了“unknown”,很明显，我们也可以认为“unknown”是缺失值。

此外，有的时候会出现空白字符串，这些也可以认为是缺失值。

对于上面的这种情况，我们可以使用 replace 方法来替换缺失值。

In [47]: heros\_null.iloc[:,4]=“unknown”  
heros\_null

Out[47]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	unknown	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	unknown	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	unknown	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	unknown	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

比如，如果我们想要将上面DataFrame中所有的"unknown"都转为pandas能识别的缺失值形式np.nan，可以使用df.replace()的方法：

In [48]:

heros\_null.replace("unknown", np. nan)

Out[48]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	NaN	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	NaN	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	NaN	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	NaN	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0

当然，指定到"Birthplace"这一列来进行替换缺失值，也是可以的：

In [49]:

heros\_null["Birthplace"]=heros\_null["Birthplace"].replace("unknown", np. nan)  
heros\_null

Out[49]:

	Actor or Actress	Hero Name	Age	Sex	Birthplace	Weapon	Height	Weight
0	荷兰弟	蜘蛛侠	19	男	NaN	蜘蛛感应	168.0	66.0
1	乔什·布洛林	灭霸	3000	男	泰坦星球	暴君屠刀	NaN	NaN
2	本尼迪克特·康伯巴奇	奇异博士	36	男	NaN	魔法	178.0	NaN
3	小罗伯特·唐尼	钢铁侠	42	男	纽约	纳米战甲	172.0	72.0
4	本·阿弗莱克	蝙蝠侠	40	男	NaN	有钱	NaN	NaN
5	克里斯·海姆斯沃斯	索尔	1505	男	阿斯加德	暴风战斧	NaN	NaN
6	盖尔·加朵	神奇女侠	2000	女	NaN	弑神者	168.0	68.0
7	斯嘉丽	黑寡妇	35	女	斯大林格勒	枪械	165.0	65.0