

lab_work_sol

May 22, 2018

1 Introduction to Data Science

1.1 Lab Work

Date: Apr 19, 2018 10:20AM

Duration: 50 minutes

This work will be graded equally with the other assignments.

You have to solve it yourself individually with no help from others.

You can use internet resources, tutorials, books.

Submit via weonline after you complete.

Make a copy of the file so you would not lose the answers.

1.2 Your Data

Fill out the following information:

First Name:

Last Name:

E-mail:

UID:

1.3 Introduction

In this lab you will investigate Titanic data, a dataset containing demographics and passenger information for 891 of 2224 passengers and crew on board of the Titanic. The data is provided by Kaggle and a description of its structure can be found there.

The question for the analysis suggested in the project details was What factors made people more likely to survive?, it seems indeed one of the most interesting questions one could ask about a disaster, however I'm gonna be a little more specific and try to focus on the following questions: Is there a difference on survival rate between different classes?, Who was more likely to survive, males or females?, Is there a particular age range that tended to survive most? and How did survival rate change across different decks of the ship?.

1.4 Data exploration

1.4.1 Task 1: Import the data

Use the `pandas.read_csv()` function to import the dataset `train.csv`

```
In [2]: # 1.1 imports and setup
import pandas as pd
import scipy as sc
import numpy as np
import matplotlib.pyplot as plt
```

```
In [3]: # 1.2 read datasets
t = pd.read_csv("train.csv")
```

```
In [4]: # 1.3 show the top 10 records
t.head(10)
```

```
Out[4]:
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
5	6	0	3	
6	7	0	1	
7	8	0	3	
8	9	1	3	
9	10	1	2	

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	
5	Moran, Mr. James	male	NaN	0	
6	McCarthy, Mr. Timothy J	male	54.0	0	
7	Palsson, Master. Gosta Leonard	male	2.0	3	
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	
9	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
5	0	330877	8.4583	NaN	Q
6	0	17463	51.8625	E46	S
7	1	349909	21.0750	NaN	S
8	2	347742	11.1333	NaN	S
9	0	237736	30.0708	NaN	C

1.4.2 Task 2: Simple statistics

```
In [5]: #2.1 How many passengers?  
len(t)
```

```
Out[5]: 891
```

```
In [6]: # 2.2 How many male and female passengers?  
t.groupby('Sex').size()
```

```
Out[6]: Sex  
female    314  
male      577  
dtype: int64
```

```
In [7]: #2.3 How many passengers survived and how many died?  
t.groupby('Survived').size()
```

```
Out[7]: Survived  
0      549  
1      342  
dtype: int64
```

```
In [8]: #2.4 What is the survival rate?  
len(t[t.Survived==1])/len(t)
```

```
Out[8]: 0.3838383838383838
```

```
In [9]: # 2.5 What are the survival rates for male and female passengers?  
print('Male survival rate: ', len(t[(t.Survived==1) & (t.Sex=='male')])/len(t[t.Sex=='ma  
print('Female survival rate: ', len(t[(t.Survived==1) & (t.Sex=='female')])/len(t[t.Sex=
```

```
Male survival rate:  0.18890814558058924
```

```
Female survival rate:  0.7420382165605095
```

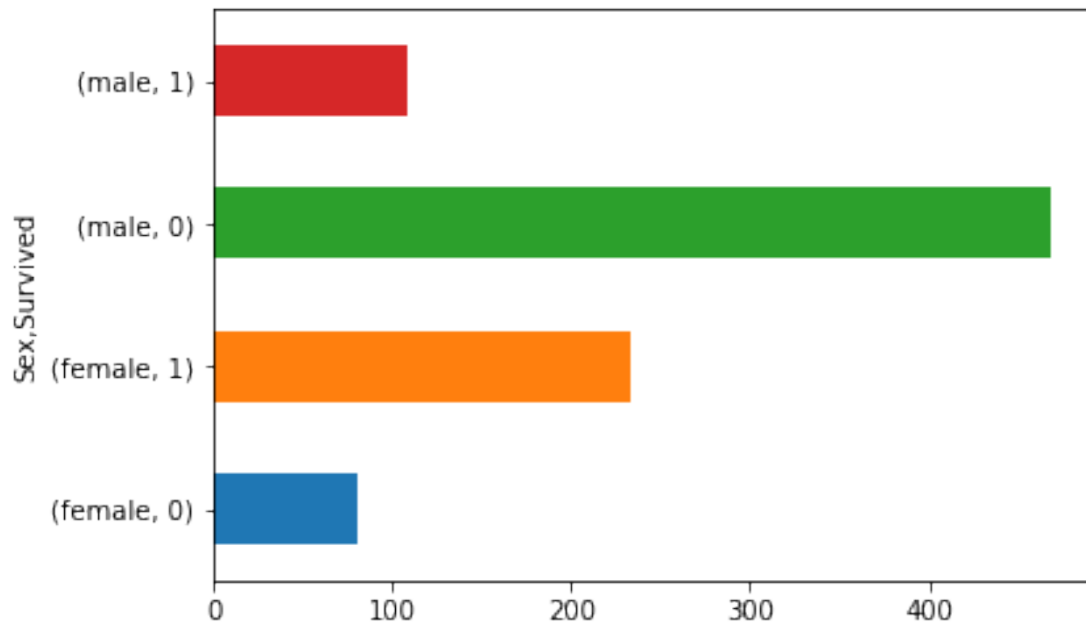
1.4.3 Task 3: Data summarization

```
In [10]: # 3.1 Show the sex/survived counts as a table  
t[['Sex', 'Survived']].groupby(['Sex', 'Survived']).size()
```

```
Out[10]: Sex      Survived  
female  0           81  
         1          233  
male    0          468  
         1          109  
dtype: int64
```

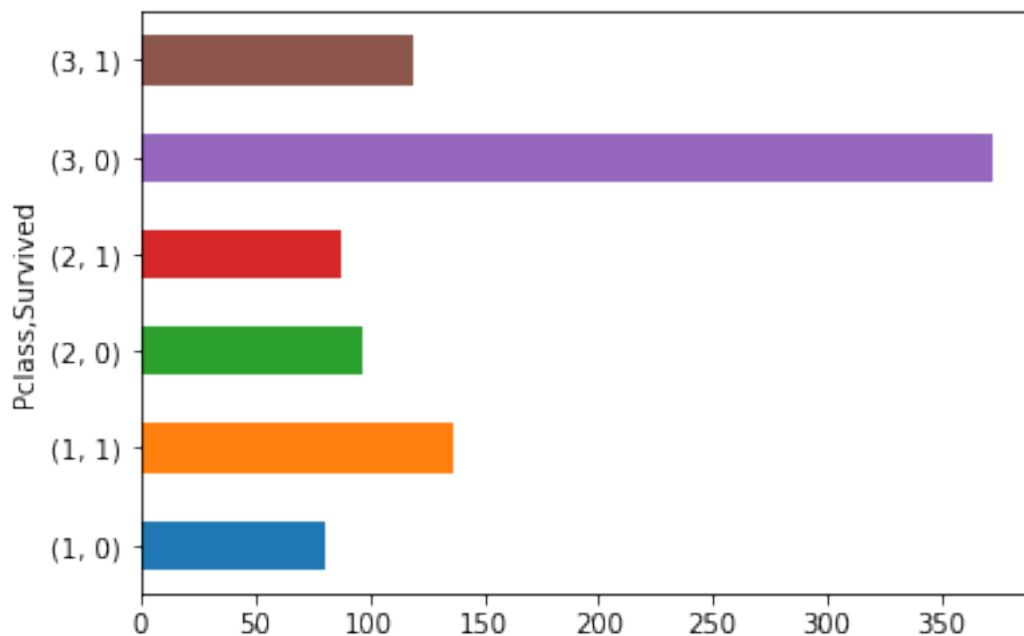
```
In [11]: # 3.2 Plot the sex/survived counts as a bar graph  
t[['Sex', 'Survived']].groupby(['Sex', 'Survived']).size().plot.barh()
```

Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x1156b3550>



In [12]: *# 3.3 Plot the pclass/survived counts as a bar graph*
t[['Pclass', 'Survived']].groupby(['Pclass', 'Survived']).size().plot.barh()

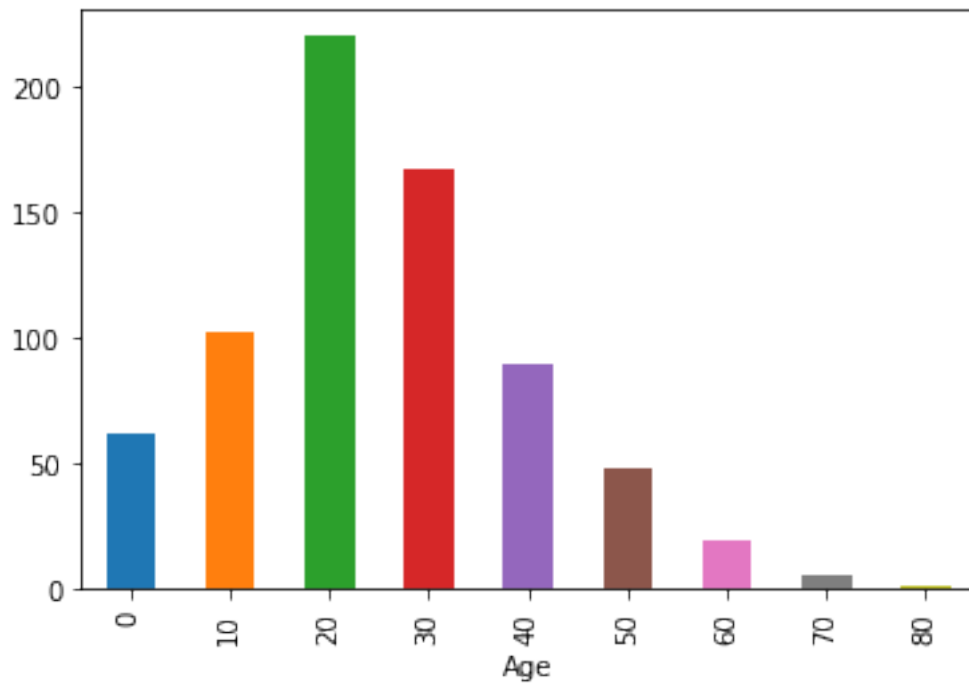
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x110e60860>



```
In [13]: # 3.4 Find the passenger counts for age groups (0-9, 10-19, 20-29, ...), display as tab
         t1 = t[t.Age.notnull()]
         t2 = (t1[['Age']]/10).astype('int64')*10
         t3 = t2.groupby('Age').size()
         print(t3)
         t3.plot.bar()
```

```
Age
0      62
10     102
20     220
30     167
40      89
50      48
60      19
70       6
80       1
dtype: int64
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1158485c0>
```



```
In [14]: # 3.5 Names of 1st class female passengers who survived
t[(t.Pclass==1) & (t.Survived==1)]['Name']
```

```
Out[14]: 1      Cumings, Mrs. John Bradley (Florence Briggs Th...
3          Futrelle, Mrs. Jacques Heath (Lily May Peel)
11          Bonnell, Miss. Elizabeth
23          Sloper, Mr. William Thompson
31      Spencer, Mrs. William Augustus (Marie Eugenie)
52          Harper, Mrs. Henry Sleeper (Myna Haxtun)
55          Woolner, Mr. Hugh
61          Icard, Miss. Amelie
88          Fortune, Miss. Mabel Helen
97          Greenfield, Mr. William Bertram
136         Newsom, Miss. Helen Monypeny
151         Pears, Mrs. Thomas (Edith Wearne)
166         Chibnall, Mrs. (Edith Martha Bowerman)
187      Romaine, Mr. Charles Hallace ("Mr C Rolmane")
194         Brown, Mrs. James Joseph (Margaret Tobin)
195         Lurette, Miss. Elise
209         Blank, Mr. Henry
215         Newell, Miss. Madeleine
218         Bazzani, Miss. Albina
224         Hoyt, Mr. Frederick Maxfield
230      Harris, Mrs. Henry Birkhardt (Irene Wallach)
248         Beckwith, Mr. Richard Leonard
256         Thorne, Mrs. Gertrude Maybelle
257         Cherry, Miss. Gladys
258         Ward, Miss. Anna
268      Graham, Mrs. William Thompson (Edith Jenkins)
269         Bisette, Miss. Amelia
275         Andrews, Miss. Kornelia Theodosia
290         Barber, Miss. Ellen "Nellie"
291      Bishop, Mrs. Dickinson H (Helen Walton)
...
710      Mayne, Mlle. Berthe Antonine ("Mrs de Villiers")
712         Taylor, Mr. Elmer Zebley
716         Endres, Miss. Caroline Louise
724         Chambers, Mr. Norman Campbell
730         Allen, Miss. Elisabeth Walton
737         Lesurer, Mr. Gustave J
740         Hawksford, Mr. Walter James
742         Ryerson, Miss. Susan Parker "Suzette"
759      Rothes, the Countess. of (Lucy Noel Martha Dye...
763         Carter, Mrs. William Ernest (Lucile Polk)
765         Hogeboom, Mrs. John C (Anna Andrews)
779      Robert, Mrs. Edward Scott (Elisabeth Walton Mc...
781         Dick, Mrs. Albert Adrian (Vera Gillespie)
796         Leader, Dr. Alice (Farnham)
```

```

802             Carter, Master. William Thornton II
809     Chambers, Mrs. Norman Campbell (Bertha Griggs)
820     Hays, Mrs. Charles Melville (Clara Jennings Gr...
829             Stone, Mrs. George Nelson (Martha Evelyn)
835             Compton, Miss. Sara Rebecca
839             Marechal, Mr. Pierre
842             Serepeca, Miss. Augusta
849     Goldenberg, Mrs. Samuel L (Edwiga Grabowska)
853             Lines, Miss. Mary Conover
856     Wick, Mrs. George Dennick (Mary Hitchcock)
857             Daly, Mr. Peter Denis
862     Swift, Mrs. Frederick Joel (Margaret Welles Ba...
871     Beckwith, Mrs. Richard Leonard (Sallie Monypeny)
879     Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)
887             Graham, Miss. Margaret Edith
889             Behr, Mr. Karl Howell
Name: Name, Length: 136, dtype: object

```

In [15]: t1

```

Out[15]:
   PassengerId  Survived  Pclass
0             1         0       3
1             2         1       1
2             3         1       3
3             4         1       1
4             5         0       3
6             7         0       1
7             8         0       3
8             9         1       3
9            10         1       2
10            11         1       3
11            12         1       1
12            13         0       3
13            14         0       3
14            15         0       3
15            16         1       2
16            17         0       3
18            19         0       3
20            21         0       2
21            22         1       2
22            23         1       3
23            24         1       1
24            25         0       3
25            26         1       3
27            28         0       1
30            31         0       1
33            34         0       2
34            35         0       1

```

35	36	0	1
37	38	0	3
38	39	0	3
..	...	...	...
856	857	1	1
857	858	1	1
858	859	1	3
860	861	0	3
861	862	0	2
862	863	1	1
864	865	0	2
865	866	1	2
866	867	1	2
867	868	0	1
869	870	1	3
870	871	0	3
871	872	1	1
872	873	0	1
873	874	0	3
874	875	1	2
875	876	1	3
876	877	0	3
877	878	0	3
879	880	1	1
880	881	1	2
881	882	0	3
882	883	0	3
883	884	0	2
884	885	0	3
885	886	0	3
886	887	0	2
887	888	1	1
889	890	1	1
890	891	0	3

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	
6	McCarthy, Mr. Timothy J	male	54.0	0	
7	Palsson, Master. Gosta Leonard	male	2.0	3	
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	
9	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	
10	Sandstrom, Miss. Marguerite Rut	female	4.0	1	
11	Bonnell, Miss. Elizabeth	female	58.0	0	
12	Saunderscock, Mr. William Henry	male	20.0	0	

13	Andersson, Mr. Anders Johan	male	39.0	1
14	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0
15	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0
16	Rice, Master. Eugene	male	2.0	4
18	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1
20	Fynney, Mr. Joseph J	male	35.0	0
21	Beesley, Mr. Lawrence	male	34.0	0
22	McGowan, Miss. Anna "Annie"	female	15.0	0
23	Sloper, Mr. William Thompson	male	28.0	0
24	Palsson, Miss. Torborg Danira	female	8.0	3
25	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38.0	1
27	Fortune, Mr. Charles Alexander	male	19.0	3
30	Uruchurtu, Don. Manuel E	male	40.0	0
33	Wheadon, Mr. Edward H	male	66.0	0
34	Meyer, Mr. Edgar Joseph	male	28.0	1
35	Holverson, Mr. Alexander Oskar	male	42.0	1
37	Cann, Mr. Ernest Charles	male	21.0	0
38	Vander Planke, Miss. Augusta Maria	female	18.0	2
..	...	...	...	...
856	Wick, Mrs. George Dennick (Mary Hitchcock)	female	45.0	1
857	Daly, Mr. Peter Denis	male	51.0	0
858	Baclini, Mrs. Solomon (Latifa Qurban)	female	24.0	0
860	Hansen, Mr. Claus Peter	male	41.0	2
861	Giles, Mr. Frederick Edward	male	21.0	1
862	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48.0	0
864	Gill, Mr. John William	male	24.0	0
865	Bystrom, Mrs. (Karolina)	female	42.0	0
866	Duran y More, Miss. Asuncion	female	27.0	1
867	Roebling, Mr. Washington Augustus II	male	31.0	0
869	Johnson, Master. Harold Theodor	male	4.0	1
870	Balkic, Mr. Cerin	male	26.0	0
871	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1
872	Carlsson, Mr. Frans Olof	male	33.0	0
873	Vander Cruyssen, Mr. Victor	male	47.0	0
874	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0	1
875	Najib, Miss. Adele Kiamie "Jane"	female	15.0	0
876	Gustafsson, Mr. Alfred Ossian	male	20.0	0
877	Petroff, Mr. Nedelio	male	19.0	0
879	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0
880	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0
881	Markun, Mr. Johann	male	33.0	0
882	Dahlberg, Miss. Gerda Ulrika	female	22.0	0
883	Banfield, Mr. Frederick James	male	28.0	0
884	Sutehall, Mr. Henry Jr	male	25.0	0
885	Rice, Mrs. William (Margaret Norton)	female	39.0	0
886	Montvila, Rev. Juozas	male	27.0	0
887	Graham, Miss. Margaret Edith	female	19.0	0
889	Behr, Mr. Karl Howell	male	26.0	0

890 Dooley, Mr. Patrick male 32.0 0

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/02. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
6	0	17463	51.8625	E46	S
7	1	349909	21.0750	NaN	S
8	2	347742	11.1333	NaN	S
9	0	237736	30.0708	NaN	C
10	1	PP 9549	16.7000	G6	S
11	0	113783	26.5500	C103	S
12	0	A/5. 2151	8.0500	NaN	S
13	5	347082	31.2750	NaN	S
14	0	350406	7.8542	NaN	S
15	0	248706	16.0000	NaN	S
16	1	382652	29.1250	NaN	Q
18	0	345763	18.0000	NaN	S
20	0	239865	26.0000	NaN	S
21	0	248698	13.0000	D56	S
22	0	330923	8.0292	NaN	Q
23	0	113788	35.5000	A6	S
24	1	349909	21.0750	NaN	S
25	5	347077	31.3875	NaN	S
27	2	19950	263.0000	C23 C25 C27	S
30	0	PC 17601	27.7208	NaN	C
33	0	C.A. 24579	10.5000	NaN	S
34	0	PC 17604	82.1708	NaN	C
35	0	113789	52.0000	NaN	S
37	0	A./5. 2152	8.0500	NaN	S
38	0	345764	18.0000	NaN	S
...	...	...	...	...	...
856	1	36928	164.8667	NaN	S
857	0	113055	26.5500	E17	S
858	3	2666	19.2583	NaN	C
860	0	350026	14.1083	NaN	S
861	0	28134	11.5000	NaN	S
862	0	17466	25.9292	D17	S
864	0	233866	13.0000	NaN	S
865	0	236852	13.0000	NaN	S
866	0	SC/PARIS 2149	13.8583	NaN	C
867	0	PC 17590	50.4958	A24	S
869	1	347742	11.1333	NaN	S
870	0	349248	7.8958	NaN	S
871	1	11751	52.5542	D35	S
872	0	695	5.0000	B51 B53 B55	S

873	0	345765	9.0000	NaN	S
874	0	P/PP 3381	24.0000	NaN	C
875	0	2667	7.2250	NaN	C
876	0	7534	9.8458	NaN	S
877	0	349212	7.8958	NaN	S
879	1	11767	83.1583	C50	C
880	1	230433	26.0000	NaN	S
881	0	349257	7.8958	NaN	S
882	0	7552	10.5167	NaN	S
883	0	C.A./SOTON 34068	10.5000	NaN	S
884	0	SOTON/OQ 392076	7.0500	NaN	S
885	5	382652	29.1250	NaN	Q
886	0	211536	13.0000	NaN	S
887	0	112053	30.0000	B42	S
889	0	111369	30.0000	C148	C
890	0	370376	7.7500	NaN	Q

[714 rows x 12 columns]

In [27]: (t1.Age // 10 * 10)

Out[27]:

0	20.0
1	30.0
2	20.0
3	30.0
4	30.0
6	50.0
7	0.0
8	20.0
9	10.0
10	0.0
11	50.0
12	20.0
13	30.0
14	10.0
15	50.0
16	0.0
18	30.0
20	30.0
21	30.0
22	10.0
23	20.0
24	0.0
25	30.0
27	10.0
30	40.0
33	60.0
34	20.0

35	40.0
37	20.0
38	10.0
	...
856	40.0
857	50.0
858	20.0
860	40.0
861	20.0
862	40.0
864	20.0
865	40.0
866	20.0
867	30.0
869	0.0
870	20.0
871	40.0
872	30.0
873	40.0
874	20.0
875	10.0
876	20.0
877	10.0
879	50.0
880	20.0
881	30.0
882	20.0
883	20.0
884	20.0
885	30.0
886	20.0
887	10.0
889	20.0
890	30.0

Name: Age, Length: 714, dtype: float64