data_preprocessing_vis

July 28, 2022

```
Traceback (most recent call last):
  File "/usr/bin/jupyter-nbconvert", line 11, in <module>
    load_entry_point('nbconvert==5.6.1', 'console_scripts', 'jupyter-
nbconvert')()
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 490, in
load_entry_point
    return get_distribution(dist).load_entry_point(group, name)
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 2854, in
load_entry_point
    return ep.load()
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 2445, in
load
    return self.resolve()
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 2451, in
resolve
    module = __import__(self.module_name, fromlist=['__name__'], level=0)
  File "/usr/lib/python3/dist-packages/nbconvert/__init__.py", line 4, in
<module>
    from .exporters import *
  File "/usr/lib/python3/dist-packages/nbconvert/exporters/__init__.py", line 3,
in <module>
    from .html import HTMLExporter
  File "/usr/lib/python3/dist-packages/nbconvert/exporters/html.py", line 12, in
<module>
    from jinja2 import contextfilter
  File "/usr/lib/python3/dist-packages/jinja2/__init__.py", line 33, in <module>
    from jinja2.environment import Environment, Template
  File "/usr/lib/python3/dist-packages/jinja2/environment.py", line 15, in
<module>
    from jinja2 import nodes
  File "/usr/lib/python3/dist-packages/jinja2/nodes.py", line 23, in <module>
    from jinja2.utils import Markup
  File "/usr/lib/python3/dist-packages/jinja2/utils.py", line 656, in <module>
    from markupsafe import Markup, escape, soft_unicode
ImportError: cannot import name 'soft_unicode' from 'markupsafe'
(/projects/d5a3aa23-997d-4ffc-977a-1dc20c583e62/.local/lib/python3.8/site-
packages/markupsafe/__init__.py)
```

```
Traceback (most recent call last):
```

```
  File "/usr/bin/jupyter-nbconvert", line 11, in <module>
    load_entry_point('nbconvert==5.6.1', 'console_scripts', 'jupyter-
nbconvert')()
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 490, in
load_entry_point
    return get_distribution(dist).load_entry_point(group, name)
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 2854, in
load_entry_point
    return ep.load()
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 2445, in
load
    return self.resolve()
  File "/usr/lib/python3/dist-packages/pkg_resources/__init__.py", line 2451, in
resolve
    module = __import__(self.module_name, fromlist=['__name__'], level=0)
  File "/usr/lib/python3/dist-packages/nbconvert/__init__.py", line 4, in
<module>
    from .exporters import *
  File "/usr/lib/python3/dist-packages/nbconvert/exporters/__init__.py", line 3,
in <module>
    from .html import HTMLExporter
  File "/usr/lib/python3/dist-packages/nbconvert/exporters/html.py", line 12, in
<module>
    from jinja2 import contextfilter
  File "/usr/lib/python3/dist-packages/jinja2/__init__.py", line 33, in <module>
    from jinja2.environment import Environment, Template
  File "/usr/lib/python3/dist-packages/jinja2/environment.py", line 15, in
<module>
    from jinja2 import nodes
  File "/usr/lib/python3/dist-packages/jinja2/nodes.py", line 23, in <module>
    from jinja2.utils import Markup
  File "/usr/lib/python3/dist-packages/jinja2/utils.py", line 656, in <module>
    from markupsafe import Markup, escape, soft_unicode
ImportError: cannot import name 'soft_unicode' from 'markupsafe'
(/projects/d5a3aa23-997d-4ffc-977a-1dc20c583e62/.local/lib/python3.8/site-
packages/markupsafe/__init__.py)
```

[4]:

```
   Age Sex ChestPainType  RestingBP  Cholesterol  FastingBS RestingECG  MaxHR  \
0   40   M           ATA        140          289          0     Normal    172
1   49   F           NAP        160          180          0     Normal    156
2   37   M           ATA        130          283          0         ST     98
3   48   F           ASY        138          214          0     Normal    108
4   54   M           NAP        150          195          0     Normal    122

  ExerciseAngina  Oldpeak ST_Slope  HeartDisease
0              N      0.0       Up             0
1              N      1.0     Flat             1
```

```
2              N       0.0       Up              0
3              Y       1.5     Flat              1
4              N       0.0       Up              0
```

[5]:
```
Index(['Age', 'Sex', 'ChestPainType', 'RestingBP', 'Cholesterol', 'FastingBS',
       'RestingECG', 'MaxHR', 'ExerciseAngina', 'Oldpeak', 'ST_Slope',
       'HeartDisease'],
      dtype='object')
```

```
(918, 12)
(918, 12)
```

#Age is age

#Sex
is gender [F = 0, M = 1]

#ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic] (Angina is a type of chest pain caused by reduced blood flow to heart)

#Resting BP is resting blood pressure (mm hg) (Normal blood pressure is less than 120/80 mm hg)

#Cholesterol is serum cholesterol (mm /dl) (Serum cholestrol is total cholestrol in body)

#Fasting BS is fasting blood sugar (1 if its >120 mg/dl; otherwise its zero) <=== discrete values (Blood sugar is the amt of sugar in the blood, it is the main source of energy)

#Resting ECG is a test that measures the electrical activity of the heart. Used to detect irregular heartbeats.

#Max Heart Rate is Max Heart Rate Id guess (between 60 and 202)

#Exercise Angina is pain in the chest that comes on with exercise, stress, or other things that make the heart work harder (yes or no question)

#Old peak is no idea (ST [Numeric value measured in depression] ??????) a finding on an electrocardiogram, where the trace in the ST segment is abnormally low below the baseline.

#A normal ST_Slope has a slight upward concavity. Flat, downsloping, or depressed ST segments may indicate coronary ischemia.

#Heart Disease is either 1 or 0 (yes or no) probably for the machine to know if its right or wrong () or to indicate that heart disease was/ was not a factor in heart failure?

#http://www.imperialendo.co.uk/Newskills/ecg/ECG1.html

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Age             918 non-null    int64
 1   Sex             918 non-null    object
```

```
 2   ChestPainType   918 non-null    object
 3   RestingBP       918 non-null    int64
 4   Cholesterol     918 non-null    int64
 5   FastingBS       918 non-null    int64
 6   RestingECG      918 non-null    object
 7   MaxHR           918 non-null    int64
 8   ExerciseAngina  918 non-null    object
 9   Oldpeak         918 non-null    float64
 10  ST_Slope        918 non-null    object
 11  HeartDisease    918 non-null    int64
dtypes: float64(1), int64(6), object(5)
memory usage: 86.2+ KB
['Normal' 'ST' 'LVH']
['ATA' 'NAP' 'ASY' 'TA']
['Up' 'Flat' 'Down']
['N' 'Y']
```

[8]:
```
   Age  Sex  ChestPainType  RestingBP  Cholesterol  FastingBS  RestingECG  \
0   40    1              0        140          289          0           0
1   49    0              1        160          180          0           0
2   37    1              0        130          283          0           1
3   48    0              2        138          214          0           0
4   54    1              1        150          195          0           0

   MaxHR  ExerciseAngina  Oldpeak  ST_Slope  HeartDisease
0    172               0      0.0         0             0
1    156               0      1.0         1             1
2     98               0      0.0         0             0
3    108               1      1.5         1             1
4    122               0      0.0         0             0
```

[10]:
```
0    547
1    371
Name: ExerciseAngina, dtype: int64
```

[11]:
```
    Age  Sex  ChestPainType  RestingBP  Cholesterol  FastingBS  RestingECG  \
0    40    1              0        140          289          0           0
1    49    0              1        160          180          0           0
2    37    1              0        130          283          0           1
3    48    0              2        138          214          0           0
4    54    1              1        150          195          0           0
5    39    1              1        120          339          0           0
6    45    0              0        130          237          0           0
7    54    1              0        110          208          0           0
8    37    1              2        140          207          0           0
9    48    0              0        120          284          0           0
10   37    0              1        130          211          0           0
```

11	58	1	0	136	164	0	1
12	39	1	0	120	204	0	0
13	49	1	2	140	234	0	0
14	42	0	1	115	211	0	1
15	54	0	0	120	273	0	0
16	38	1	2	110	196	0	0
17	43	0	0	120	201	0	0
18	60	1	2	100	248	0	0
19	36	1	0	120	267	0	0

	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	172	0	0.0	0	0
1	156	0	1.0	1	1
2	98	0	0.0	0	0
3	108	1	1.5	1	1
4	122	0	0.0	0	0
5	170	0	0.0	0	0
6	170	0	0.0	0	0
7	142	0	0.0	0	0
8	130	1	1.5	1	1
9	120	0	0.0	0	0
10	142	0	0.0	0	0
11	99	1	2.0	1	1
12	145	0	0.0	0	0
13	140	1	1.0	1	1
14	137	0	0.0	0	0
15	150	0	1.5	1	0
16	166	0	0.0	1	1
17	165	0	0.0	0	0
18	125	0	1.0	1	1
19	160	0	3.0	1	1

This visualizations shows a boxplot of cholesterol between the two groups of people with and without heart disease. The maximum cholesterol levels for people with heart disease is 603, while the maximum cholesterol levels for people without heart disease is 564, indicating that cholesterol levels are extreme in people with heart disease.

For the people with heart disease boxplot, 50% of the data is between 212 and 283 cholesterol because that's the bounds for the box. Then most of the rest of the data is between 110 and 388.

For the people without heart disease, the cholesterol is between 203 and 269. Then most of the rest of the data is between 126 and 365.

There seems to be not a big difference between these two groups. That is because the ranges for the boxplots are essentially the same and the outliers for both groups are very similar as well. The people with heart disease are slightly more spread out than the people without heart disease but that could be because of our sample size. We would think that cholesterol would not be a very important feature for distinguishing between people with and without heart disease.

- Specific things
  - Specific values to reference
- Broad things
  - Compar visualizations
    * What's similar?
    * What's different
  - Trends that you see

The heartrate boxplot indicates that heartrate is overall slower in people with heart disease, while people without heart disease have higher heart rates. For example, the minimum and maximum for people without heart disease had a significantly higher heart rate in comparison to those with heart disease. This may show that Heart Disease is often associated with lower heart rate, because it cannot function as well.

This visualization shows the Maximum heart rate for two groups, one with heart disease, and one without. In the heart disease boxplot, 50% of the data is between 112 and 144.5, and the rest of the data is between 67 and 182

In the group without heart disease, the boxplot shows that 50% of the data is between 134 and 165, and the rest of the data is between 90 and 202

The group without heart disease seems to have a much higher heart rate compared to the group with heart disease. From this data, we can conclude that heart rate might be an important factor in determining if someone has heart disease.

```
We can see that there is not much correlation between heart disease and
RestingBP, because both box plots are very similar. So we can conclude that
resting blood pressure does not differ much in people with heart disease vs
people without. The first 75% of the data does not seem to be much different from
each other. The last 25% of the data seems to be slightly different, but not too
much to make a conclusion.
```

```
This boxplot shows the restingBP for two groups, one with heart disease and one
without. In the heart disease boxplot, 50% of the data is between 120 and 145,
and the rest of the data is between 92 and 180
```

```
In the group without heart disease, the boxplot shows that 50% of the data is
between 120 and 140, and the rest of the data is between 94 and 170
```

```
There does not seem to be much of a significant difference between these two
groups in terms of RestingBP. From this, we would not think RestingBP has much
of an indication on whether or not someone has heart disease or not.
```

The correlation I found in this box plot was that people who have heart disease have a higher age. For people with heart disease, the data shows that they have a larger in the age boxplot compared to those who don't. In the heart disease box plot, each quartile is significantly larger than their corresponding quartile in the non-heart disease boxplot. This shows a clear correlation between age, and whether or not someone has heart disease. These boxplots indicate that heart disease is somewhat correlated with Age, Maximum heart rate, and Cholesterol.

This visualization is a box plot that shows the correlation between age and heart disease, one group with heart disease and the other without. In the heart disease boxplot, 50% of the data is between 51and 62, and the rest of the data is between 35 and 77

In the group without heart disease, the boxplot shows that 50% of the data is between 43 and 57, and the rest of the data is between 28 and 76.

This shows a clear correlation between age, and whether or not someone has heart disease. We would think from this data, that the higher the age, the more likely it is to get heart disease.

[17]:
```
2    496
1    203
0    173
3     46
Name: ChestPainType, dtype: int64
```

[18]:
```
1    508
0    410
Name: HeartDisease, dtype: int64
```

[19]:
```
1      1
3      2
8      2
11     0
13     2
      ..
912    2
913    3
914    2
915    2
916    0
Name: ChestPainType, Length: 508, dtype: int64
```

[20]:
```
0      0
2      0
4      1
5      1
6      0
      ..
903    0
904    0
906    0
910    0
917    1
Name: ChestPainType, Length: 410, dtype: int64
```

[21]:
```
0      0
1      1
2      0
4      0
5      0
      ..
```

```
911    1
913    1
914    1
916    1
917    0
Name: HeartDisease, Length: 547, dtype: int64
```

[22]:
```
3      1
8      1
11     1
13     1
23     1
      ..
907    1
908    1
909    1
912    1
915    1
Name: HeartDisease, Length: 371, dtype: int64
```

Atypical Angina (ATA) = 0

Non-Anginal Pain (NAP) = 1

Asymptomatic (ASY) = 2

Typical Angina (TA) = 3

Overall, 54% of patients were asymptomatic and experienced no chest pain. The second most prevalent pain was non-anginal pain, followed by atypical angina. A small minority experienced typical angina. In patients with heart disease, an overwhelming majority were asymptomatic, with only 22.86% experiencing any pain. A minority of patients with heart disease had typical angina and atypical angina. However, in patients without heart disease, a third suffered non-anginal pain and another third suffered atypical angina. Asymptomatic chest pain made up a quarter of patients without heart disease, and a minority suffered typical angina. Among both groups, typical angina was the least common type of chest pain. However, a much larger percentage of patients with heart disease were asymptomatic compared to patients without heart disease.

No = 0

Yes = 1

A majority of patients did not experience exercise angina. However, in patients with heart disease, an overwhelming majority (85.2%) experienced exercise angina. In great contrast, about two thirds of patients without heart disease did not experience exercise angina and around a third did. Heart disease could be a factor that contributes to exercise angina, but exercise angina is not necessarily an indicator of heart disease considering a considerable amount of patients without heart disease experienced it as well.

This correlation matrix is comparing all columns in the table to see how correlated they are in other. This meaning, that a strong correlation means that there is a significant trend found in relation to another, and then scaling down until you get to neglibible in which they barely affect each other to no correlation where they have no affect at all on each other. Overall for most of the values there was either practically no correlation, a neglegible correlation, or a weak correlation. The highest instances would be described as a moderate correlation(if you take the definition of a moderate correlation to be between 0.5 and 0.7).These instances came in the relationship between ST_Slope and old peak with a correlation of 0.502, and a moderate correlation between heart disease and ST_Slope of 0.599. ST_Slope is measured during a heart test and is supposed to go up when one does exercise. Therefore, an abnormal slope would be either staying flat or trending downward which could be a sign of heart disease. There are then a number of high low correlations coming to either 0.4 up to 0.494. The overall trend was that the bulk of the columns were correlated to each other between the correlation number of the absolute value 0.09 to the absolute value 0.35. There was a wide spread of values being both positive, meaning when one increases, the other can increase, and negative, which means when one value increases the other decreases. There were also some factors with an especially low correlation towards everything, namely RestingECG which had a highest value resulting in a weak correlation of 0.213 but often below 0.1.

```
/tmp/ipykernel_472/2339356345.py:4: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
Analysis: This is a scatter matrix which shows a scatter plot comparing all the
continuous inputs in the data frame. I think this matrix shows that although
heart disease is complicated and nothing is exactly one to one, there are general
trends with certain traits that increase risks. For example, in the Age and MaxHR
scatter plot, we see that people with heart disease generally trend towards being
older and having a lower MaxHR. We can also see that people with a high Oldpeak
tend to have heart disease. Some variables also have limited relationships with
others. For example, we can see that broadly speaking, the higher the age the
higher the resting RestingBP
```

This histogram above displays the total count of the sex of participants within the data set. There were 193 females and 725 males, making a total of 918 data entries. This data should be taken into consideration when considering the accuracy of the results that can be predicted with this data. The predicted results for the males will likely be more accurate because there is more data to base the predictions off of.

The histogram above shows the ages of the participants of the data set. The youngest participants were twenty-eight to twenty-nine years old, while the oldest participants were seventy-six to seventy-seven years old. We can see that the greatest amount of participants was in the age group of fifty-four to fifty-five year olds, with a total of 92 participants.

The turning point within the dataset is ages forty-six to forty-seven. When you compare the data of the ages of those younger than forty-six to forty-seven, more people do NOT have heart failure. Whereas after ages forty-six to forty-seven, more people had heart failure than people who did not.

The peak of the count of people who had heart failure was centered at around ages fifty-four to sixty-three with age ranges of fifty-six to fifty-seven and ages fifty-eight to fifty-nine both having the greatest count of 51 participants having heart failure. The distribution of the graph of people without heart failure is distributed differently. The highest count of people who did NOT have heart failure was 45 participants ages fifty-four to fifty-five. The graph of people without heart failure is less smooth, with a greater change between age ranges.

The graph of the cholesterol level of people with and without heart disease are very similar. The data for cholesterol levels was measured in millimoles per deciliters. Some other data to consider is that there was outliers in this graph. An example of this was some participants had a cholesterol level of 0. I am not certain that this measurement was an error so I decided to not include that data point in the range of this graph. For people with and without heart failure, both of the graphs peaked at around the same level with people who had heart disease at a cholesterol level of 200-219 mm/dL., and those without peaking at 220-239 mm/dL..

The graph of the resting blood pressure of people with and without heart disease was also very similar. Between each range, for both people with and without heart disease, the count of the participants that had a certain level at which their resting blood pressure increased and decreased between ranges. Every five, the count would increase, and then the next five, it would decrease.

Based on these two graphs, I would conclude that Cholesterol Level and Resting Blood Pressure do not play a great factor in predicting whether a person has heart failure, as the graphs of people with and without heart failure for both categories are too similar.

Maximum Heart Rate could be a factor that could greatly influence whether someone has heart disease or not. The graph of people without heart failure is skewed to having a higher maximum heart rate than the graph of people with heart failure which is skewed to the lower end. Both graphs are placed on the same scale so that the counts of the graphs can be compared. While the counts of people who did Not have heart failure are lower than the counts of people who did HAVE heart failure, the graphs give a clear indication people who do not have heart failure have a higher maximum heart rate.

Again, this first graph illustrates the Sex of participants with and without heart failure. Like the histogram of just the distribution of the sex without relation to heart failure, there are more males than females in the study. However, what the previous histogram didn't reveal was that in this data set, more males had heart disease than did not which compared to females had more females without heart disease than who did.

The final histogram reveals that a normal Old Peak should be 0. 253 participants without heart failure had an Old Peak of 0- 0.1 while 129 participants with heart failure had the same Old Peak. This data reveals that while Old Peak should be considered when determining heart failure, it is also due to other factors. If Old Peak were a factor that determined heart failure, less people with heart failure would have an Old Peak measurement of 0-0.1.

To do:

- First check for null values
- Check if you need to drop any columns
- Convert columns that are strings to values
- Find which columns are continuous and which are categorical

Visualizations/Charts to make:

- Pie charts (Emma)
- Scatterplots (Scatter plot matrix?) (Thomas)
  - Continuous values only please
- Boxplots OR Vase plots (Jeffrey and Kenju)
- Histograms of continous values and categorical (try to separate it by sex) (Halli)
- CORRELATION MATRIX :open_mouth: (Try to look up what that is first) (Rhone)
- Sunburst(Thomas)
- Contour Maps as well

```
(746, 12)
```

```
/tmp/ipykernel_472/157527307.py:2: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

/tmp/ipykernel_472/157527307.py:3: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

For Max Heart Rate, it seems like there may be a correlation between higher heart rate in individuals with heart disease and individuals with no heart disease. However, looking at the female group especially, it appears that the difference may be insignificant due to the large overlap at the peaks of the graphs.

Interestingly, it seems like cholesterol levels are very similar in all groups. While the values in the male group with heart disease have a larger range than its sibling, the female group's graphs are the opposite. Additionally, the distributions overlap almost entirely, with the heart disease group only being slightly higher in cholesterol levels than the no heart disease group.

I believe the graphs indicate nothing about the no heart disease group; however, it more indicates that people are more likely to get heart disease as an adult than as a child, since the data peaks at around 55 years of age for males and 60 for females.

[43]:
```
Index(['Age', 'Sex', 'ChestPainType', 'RestingBP', 'Cholesterol', 'FastingBS',
       'RestingECG', 'MaxHR', 'ExerciseAngina', 'Oldpeak', 'ST_Slope',
       'HeartDisease'],
      dtype='object')
```

[45]:
```
   ChestPainType        Age       Sex   RestingBP  Cholesterol  FastingBS  \
0              0  49.242775  0.653179  130.624277   233.046243   0.109827
1              1  53.310345  0.738916  130.960591   197.438424   0.201970
```

```
2              2  54.959677  0.858871  133.229839   186.645161   0.284274
3              3  54.826087  0.782609  136.413043   207.065217   0.282609

   RestingECG       MaxHR  ExerciseAngina   Oldpeak  ST_Slope  HeartDisease
0    0.421965  150.208092        0.098266  0.307514  0.219653      0.138728
1    0.625616  143.236453        0.251232  0.674877  0.532020      0.354680
2    0.633065  128.477823        0.598790  1.162702  0.826613      0.790323
3    0.869565  147.891304        0.130435  1.036957  0.652174      0.434783
```

[46]:

```
   ChestPainType        Age       Sex  RestingBP  Cholesterol  FastingBS  \
0              0   9.259754  0.477340  16.861711    69.266406   0.313581
1              1   9.608023  0.440311  19.412878   103.869348   0.402463
2              2   8.763468  0.348506  18.580961   122.058634   0.451523
3              3  11.449026  0.417029  19.059644    83.383292   0.455243

   RestingECG      MaxHR  ExerciseAngina   Oldpeak  ST_Slope  HeartDisease
0    0.716080  22.281042        0.298538  0.611138  0.455287      0.346666
1    0.837094  25.608501        0.434794  0.940512  0.590946      0.479599
2    0.803047  23.483317        0.490638  1.135079  0.573997      0.407489
3    0.909425  23.126923        0.340503  1.120587  0.640048      0.501206
```

```
/projects/d5a3aa23-997d-4ffc-977a-1dc20c583e62/miniconda3/envs/ds_env/lib/python
3.8/site-packages/plotly/express/_core.py:1637: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a
future version. Use pandas.concat instead.

/projects/d5a3aa23-997d-4ffc-977a-1dc20c583e62/miniconda3/envs/ds_env/lib/python
3.8/site-packages/plotly/express/_core.py:1637: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a
future version. Use pandas.concat instead.

/projects/d5a3aa23-997d-4ffc-977a-1dc20c583e62/miniconda3/envs/ds_env/lib/python
3.8/site-packages/plotly/express/_core.py:1637: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a
future version. Use pandas.concat instead.
```

This sunburst chart shows the percentage of certain groups that are a part of other groups. For example, the majority of people who have heart disease has ASY chest pain type, and people who have ASY chest pain type also tend to be older. This is contrasted by people who don't have heart disease, and it turns out that people who have ASY chest pain type are a minority in that group.

- Is there anything that you find interesting?
-