Prediction of Heart Failure

Heart disease is one of the most concerning health issues in modern times. Heart disease is used to refer to an umbrella of illnesses that include coronary artery disease, high blood pressure, and stroke. Another of those conditions being congestive heart failure. It is a condition in which the heart fails to pump blood to its full ability. An illness that plagues millions around the world, it is necessary to examine some of the factors and circumstances that can potentially bring about heart failure. The dataset of interest uses 12 factors including age, sex, cholesterol levels, and blood pressure to predict the likelihood one experiences heart failure. It contains subjects with a fair mix of numerical and categorical measures, each of which either contributes to an increased or decreased likelihood of the event's occurrence. The dataset is certainly not comprehensive in its inclusion of factors that can predict heart failure but is representative with the elements that are included.

With the examination and visualization of this dataset, the goal was to answer the primary question of which specific elements lead to the prediction of heart failure. Over the course of creating the visualization, the question became even more specified to focus on heart related factors. Even though age and sex are predictors of the event, looking deeper into measures such as chest pain, heart rate, and the prevalence of heart disease can provide more insight. The data reflects the details of over 900 males and females, primarily male, with a key emphasis on measures related to heart health that point towards heart failure.The narrative will take the viewer through a series of projections that show how individual and combined measures play a role in the likelihood of a heart disease prediction.

The story is meant for those of all ages and walks of life but is directed even more towards people and their family members with a history of heart failure. This demographic of people will carry the most interest in the presentation and be more invested in learning from the visualization narrative. The narrative is meant to bring further awareness of the topic of heart failure and inform those at risk of the potential factors that lead to the event. It is also important to note that all viewers may not agree with or be intrigued by the narrative. There is a strong possibility that any of the viewers may become discouraged by the data presented. Ultimately, the chief goal is to display factual details and measures that concern one's likelihood of experiencing heart failure.

In order to properly visualize the data of interest, it was necessary to transform the data into unique displays that present the content of the data accurately. Various aspects of the data were alteered into graphs, such as the boxplot and the bargraph to display the distribution of a single variable. The scatterplot showed the interaction between two quantitive variables in the dataset and was the most effective transformation for this subset. The versatile graph was able to represent where a subject placed in the distribution in terms of two quantities. It was very useful in not only the visualization of the data, but also connecting information together regarding the subjects.

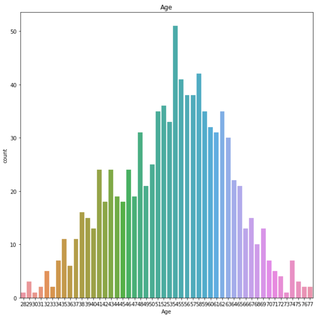
In [55]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [26]:
```
df = pd.read_excel ('heart.xlsx')
```

In [61]:
```
plt.figure(figsize = (22,22))
plt.subplot(2,2,1)
sns.countplot('Age', data = df)
plt.title('Age')
```

```
/Users/ikennaobi/opt/anaconda3/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[61]: Text(0.5, 1.0, 'Age')

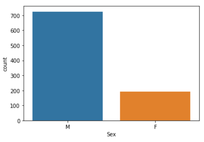


This plot was created to show the age distribution of the 917 subjects recorded for a prediction of experiencing heart failure. The histogram was chosen because it is able to quantify several different ages and show the outliers and maximum, minunums.

In [53]:
```
sns.countplot('Sex', data = df)
```

```
/Users/ikennaobi/opt/anaconda3/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[53]: <AxesSubplot:xlabel='Sex', ylabel='count'>

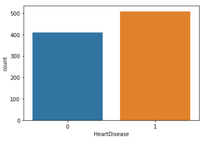


This graph was created to display the gender split of the subjects recorded. The bar graph is used to display categorical data. In this case, a split between Male and Female subjects. In which males outnumber female subjects 4:1. The distribution between gender is an important detail to note considering the likelihood of the projection.

In [64]:
```
sns.countplot('HeartDisease', data = df)
```

```
/Users/ikennaobi/opt/anaconda3/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

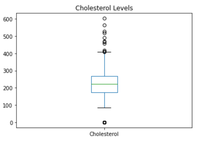
Out[64]: <AxesSubplot:xlabel='HeartDisease', ylabel='count'>



Visualizing the split between subjects who had heart disease vs those who did not, was a key factor in the study. Heart failure falls under the umbrella of illnesses under heart disease. The bar graph was a strong fit for this subject of categorical data. The data is showing more subjects in the study than not were experiencing heart disease.

In [85]:
```
df['Cholesterol'].plot.box()
plt.title('Cholesterol Levels')
```

Out[85]: Text(0.5, 1.0, 'Cholesterol Levels')



The cholesterol was another important measure to record in the study because of its relation to heart failure and other heart diseases. The box plot was chosen as a simple way to display the cholesterol levels of the subjects. It contains the minimum, maximum, median, and outliers in a unique structure.

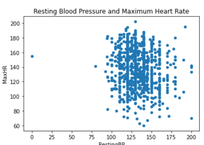
In [62]:
```
df.plot.scatter(x='Age', y='Cholesterol')
plt.title('Age vs Cholesterol Levels')
```

Out[62]: Text(0.5, 1.0, 'Age vs Cholesterol Levels')

This visualization of the interaction between age and cholesterol levels is a complement to the "Cholesterol Levels" distribution graph. We get more detail about what cholesterol levels look like at different ages. Cholesterol being an improtant factor in the development of heart disease. The scatterplot was chosen because of its ability to show the interaction between two quantitative measures in one subject.

In [47]:
```
df.plot.scatter(x='RestingBP', y='MaxHR')
plt.title('Resting Blood Pressure and Maximum Heart Rate')
```

Out[47]: Text(0.5, 1.0, 'Resting Blood Pressure and Maximum Heart Rate')

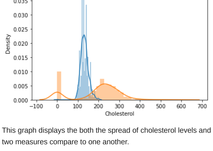


This visualization shows the interaction between an individual subject's blood pressure at rest and their resting heart rate. This interaction was a measure created through the combination of two quantitative measures in connection to the physiology of the heart and systems related to it. The scatterplot is able to show great detail about the where the subjects are as a whole and where individuals subjects lie in respect to the fellow subjects.

In [92]:
```
sns.distplot(df['RestingBP'])
sns.distplot(df['Cholesterol']);
plt.title('Resting Blood Pressure and Cholesterol')
```

```
/Users/ikennaobi/opt/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/Users/ikennaobi/opt/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[92]: Text(0.5, 1.0, 'Resting Blood Pressure and Cholesterol')



This graph displays the both the spread of cholesterol levels and resting blood pressure using the histogram. These two measures are closely related in their effects. This visualization is able to show how the two measures compare to one another.

The seven visualizations provided an alternative method of communicating data to the viewer besides the usual format of tables. The graphs were meant to convey the data on different scale, showing information that spreads the distribution of subjects in terms of categorical,quantitative, and interactive measures and where they stand relative to other subjects. The various methods chosen to visualize the data were picked for their particular abilities to communicate each measure of data. The bar graphs compared categorical variables. The box plots and histograms were used to show the distribution of a single one measure of a subject. Lastly, the scatterplots were chosen because they connect two variables of a subject and provide the position of the combination compared to other subjects.

The measures that were selected for display are among the most essential when considering the likelihood that the prediction of heart failure is correct. While the data did primarily record males, the data is still able to show where females stand in the distribution. The majority of the difficulties came from designing a visualization that was able to show the most essential measures in the most best display at hand. Transforming data that had limitations on the possible formats was an important consideraiton when creating the story. In addition to the conversion of the data records, ordering the particular displays in a way that introduced the graphs that were the most straightforward and then follow up using graphs that had multiple pieces of information.

The dataset chosen was suitable for the narratation of a story using visualization. The viewer is able to get a more broader concept of the data, including trends that provides a general sense to how the data should be communicated. The dataset allowed for the combination of metrics that had a strong correlation to each other and a strong contribution to the heart failure prediction likelihood. An addition to this visualization narrative might include further interactions with potentially more than 2 variables and a more insightful look at supporting measures.

In [ ]:
```
```