**Healthcare Dataset Findings**

* Between the age of 22 and 57 with median age of 40, 50% of the people are less likely to have hypertension and the same is true for heart disease as well. In the box plot, it’s quite evident that 50% of people are at this age group.
* Between the age of 52 and 75 with median age of 62, people are more likely to have hypertension Between the age of 60 and 79 with median age of 70, people are more likely to have heart disease
* There are few outliers in both the scenarios as shown in box plot where people less than 20 years of age could also have hypertension and people less than 33 years old could have heart disease may be due to people being born with heart conditions or other factors.
* Out of a total of 96000 records, about 40000 are male subjects and 56000 are female subjects. So out of this it was found out that higher % of male population is more prone to having these three conditions, with diabetes being at 10.1%, hypertension 8.2%, and heart disease about 6% whereas the numbers for the same for female population are 7.9%, 7.4% and 2.8% respectively. In the graph plotted it shows the count of male and female subjects who have these three conditions.
* To find out correlation between BMI, Blood Glucose Level and diabetes, we have plotted a scatter plot, where it shows a positive correlation between blood glucose level and diabetes, so higher the blood glucose level, more chance that the person is diabetic.
* There is also a weak positive correlation between diabetes and subjects BMI, where if the BMI is greater than 40, more chances of the person being diabetic and in some cases even if the blood glucose level is less, they are still likely to be diabetic as shown in the red dots here.
* When exploring disease likelihood among age groups – here defined as ‘under 30’, ’30-50’, ’50-70’, and ‘over 70’ – the highest concentration of positive results for all 3 diseases (diabetes, hypertension, and heart disease) was in the 50-70 years old range. The next highest range was the over-70 years old group.
* When looking at the correlation between the hbA1c level for patients and the occurrence of the diseases in question, there was a moderate correlation for diabetes (value = 0.406) as opposed to very weak correlation for both hypertension and heart disease (values = 0.081 and 0.068 respectively). From this it can be concluded that hbA1c can potentially be an indicating factor for diabetes but may be disregarded as an indicator for heart concerns.
* There does not appear to be any strong correlation between smoking history and suffering from any of the diseases. Below is a table of correlation coefficients for each group and the range for this set lies between -0.1 and 0.1. This implies that there is no correlation between any of these groups. It’s still interesting to see the ‘Never’ group consistently having the strongest negative correlation while the ‘Former’ group is consistently the highest. This lead us to look deeper into each category.

|  |  |  |  |
| --- | --- | --- | --- |
| **Smoking History** | **Hypertension** | **Heart Disease** | **Diabetes** |
| Never | -0.0229 | -0.0810 | -0.0494 |
| Current | -0.0119 | -0.0062 | -0.0107 |
| Former | 0.0969 | 0.0969 | 0.0789 |
| Ever | 0.0373 | 0.0373 | 0.0056 |
| Not Current | -0.0029 | -0.0029 | -0.0031 |

* After plotting the age distribution for each smoking history category, the charts made it clear that former smokers in made up of a higher age group than the rest others. Below are the averages for each group and former smokers is much higher than the rest.

|  |  |
| --- | --- |
| **Smoking history** | **Mean** |
| Current | 44.103 |
| Ever | 49.147 |
| Never | 43.875 |
| Former | 57.02 |
| Not Current | 47.686 |

* The healthcare data set shows that for patients over 60 years old, 39.3% of them suffer at least 1 of the 3 diseases. We then used pulled population data using the Census API so we could apply our findings to Arizona. With a total of 1,506,346 people over 60 here in AZ, 592,360 suffer at least one of the 3 diseases, assuming the proportion would stay the same.