**The Heart Project**

By Yarden Hodeda

**Table of contacts**

* Opening words
* Questions to discover
* TLDR
* About the process
* Question 1
* Question 2
* Question 3
* Question 4
* *Final words*

**Opening words**

Due to increase in the last decades of heart diseases throughout the population I decided to tackle the situation and to find out what makes the population with heart diseases more prone to getting sick. I wanted to dive deep into the specifics and to find out what characterize people with heart diseases and what can I learn from them as a group.

To analyze this group, I decided to find a dataset related to people with heart diseases and to use my skills with SQL and Power BI to make this report and to give a solid conclusion. I used a dataset from the 2020 annual CDC survey with more than 400k samples of data related to adults health status.

**Questions to Discover**

1. What is the distribution of people with heart diseases based on gander?
2. Does age play are role in heart diseases?
3. What are some common marks of people with heart diseases?
4. Are there any specific diseases that are more common with people that have heart diseases?

**TLDR**

1. We can see according to the study that men are more represented in the heart disease population with 59% compared to women with a representation of 41%.
2. We can see that there is a positive correlation between age and heart diseases, which means that as we get older, we are more prone to have heart diseases.
3. There are several marks that are common with the heart disease group, smoking being the most common of them with 59% of the heart disease group. We can also see that about a third of the group are sleep deprived and 36% don’t do any physical activity.
4. Third of the people with heart diseases in the survey also have diabetes and 18% of them have skin cancer and asthma.

**About The Process**

I downloaded the dataset from Kaggle ([Link](https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease)) and imported it into MySQL Workbench. After making sure all the data was properly imported, I started to run queries within the workbench to explore the data and to find out if there are interesting points that I might find. After the exploration I decided to create the tables I needed to answer my questions and I exported them to a CSV file. The tables were now imported to Power BI, and I started to make a dashboard to visualize the data and to present it neatly.

# **Question 1 - What is the distribution of people with heart diseases based on gander?**

As we can see in the pie chart 58.96% of people with heart diseases are men and 41.04% of people with heart diseases are Women. It is a 3:2 ratio which means that if you are man, you are more likely to have a heart disease compared to a woman.

A screenshot of a cell phone

Description automatically generated with medium confidence

**Question 2 - Does age play are role in heart diseases?**

The age in the survey ranges between 18 to 80 and older. We can see from the result of the survey that as we get older the change of having a heart disease is getting bigger. At the 18-24 age group only 0.5% of the population has a heart disease. we can see the major jump occurs in the space between the 45-49 age group to the 50-54 age group, the number is almost double the previous one. after that there is a steady increase in the percentage. We can see a slight deep in the space between the 70-74 age group and the 75-79 age group but the trend continues strongly and in the 80 or older group we can see the highest number among the groups, almost 20% of the population.

Chart

Description automatically generated

**Question 3 - What are some more common marks of people with heart diseases?**

In the survey we can see a lot of marks such as physical activity, sleep time, smoking, alcohol drinking, physical health, mental health and much more.

From the results of the survey, we can see that there are few marks that are more common within the group of people with heart diseases.

The most common is smoking. A patient is defined as a smoker if he smoked at least 100 cigarettes in their lifetime (note that 100 cigarettes = 5 packs). 59% of the people with heart diseases are smokers.

The second most common mark is not having any physical activity. People who do physical activity are considered as adults who reported doing physical activity or exercise during the past 30 days other than their regular day job. 36% of the people with heart diseases are not doing any physical activity.

The third most common mark is people that are sleep deprived. I considered people as a sleep deprived if they are having less than 7 hours of sleep. This number was taken from the CDC website which states the recommended sleep time for different age groups. Third of the people with heart diseases are also sleep deprived.

Chart

Description automatically generated with low confidence

**Question 4 - Are there any specific diseases that are common with people that have heart diseases?**

The data included several diseases such as asthma, kidney disease, skin cancer, diabetes and stroke. We can see from the result that the most common disease is diabetes. Third of the people with heart diseases are also diabetic. There are also fewer common diseases such as skin cancer and asthma which came in second place with 18%. Followed by them is stroke with 16% and kidney disease with only 13% of the people with heart diseases.

Chart, bar chart

Description automatically generated

**Final words**

Graphical user interface, chart, application

Description automatically generatedIt was really fascinating to see the result from such a large survey. Some of them really took me by surprise, for example the lake of sleep mark that is almost comparable to the lake of physical activity. It was also interesting to see the numbers of the age groups and to realize that we must take care of ourselves and to start as soon as possible to beat the numbers in order to stay healthy.