

MSIS 670 FINAL PROJECT REPORT (Group 5)

FACTORS AFFECTING HEART DISEASES

For the final project of Business Intelligence, we have decided to base our work on a dataset collected from **Kaggle** that provides an overview on the different attributes (both physical and mental) that may cause heart diseases. The dataset comes from the CDC and is a significant piece of the Behavioral Risk Factor Surveillance System (BRFSS), which conducts yearly phone studies to assemble information on the wellbeing status of U.S. inhabitants.

Introduction to the dataset

According to the CDC, heart disease is one of the leading causes of death for people of most races in the US (African Americans, American Indians and Alaska Natives, and white people). The main area of focus is on the causes of heart diseases. Our data source considers several key indicators like:

- Health & Fitness Indicators: level of physical activities, obesity in terms of BMI, drinking / smoking habits etc.
- Mental health indicators: Good / bad mental health days in the last 30 days, sleep time, level of physical activities etc.
- Underlying additional ailments: Diabetes, asthma, kidney disease and skin cancer.
- Additional miscellaneous attributes: gender, age.

The dataset retrieved from **Kaggle** was the clean version and therefore did not require any additional cleaning to be performed.

Story telling by Data Visualization (using Tableau)

The idea of the project revolves around covering the different factors that affect the occurrence of heart diseases. To begin with, in order to get the basic idea of the project, our first visual representation includes a comparison between physical health, physical activity and general health.

From the visualization obtained it can be observed that people having low physical activity levels and a poor general health are seen to be having a greater number of illness days, whereas people with excellent levels of physical activity and general health have very low number of average illness days.

This visualization gives us an idea about what the data set covers and the main areas it revolves around that affect the occurrence of heart diseases.

As our main question for the project focuses on the usage of insights from the visualizations on insurance companies / market in particular, this preliminary analysis can be used to obtain the basic idea for the sale of policies.

The next attributes being taken into consideration are the difficulty in walking, age category, and, how the combination of these two attributes affects the occurrence of heart diseases.

It can be seen from the visualization that as the range of age category increases across the dataset, the probability of individuals prone to heart diseases increases across the dataset. For e.g Starting from the age category of 18 – 29, the probability of heart diseases is very low when compared with the higher age category of 70 – 79 that has a probability of **14.18%**.

Another insight that can be generated from the visualization would be the increase in the probability of heart diseases when the person also suffers from difficulty in walking. If the age category of 70 - 79 is taken into consideration, it can be observed that the probability of getting heart disease almost doubles for the same age category (**27.02%**) when the person also suffers from difficulty in walking.

From an insurance company's point of view, this observation can be used to design policies that cater to a higher age category like 70 – 79 etc, so that the premium rates can be increased for this particular age category and higher returns can be obtained by the company.

Moving on to the next visualization, a comparison is drawn between the occurrence of heart disease based on the genders. For this visualization, we have taken two pie charts into consideration. It can be observed from the charts that males are more prone to heart diseases (**10.62%**) when compared with females (**6.69%**).

The previous strategy for the insurance company can be applied to this case as well by designing policies that cover / cater effectively and on a larger scale to the male population in order to generate higher returns.

Question 1: What are the main factors responsible for heart diseases?

For the dashboard preparation, we have made use of different filters that can be used to navigate through the dataset and the visualization. When the filter settings are set to female, heart diseases (all), age category (above 50), the insights generated from the dashboard indicate that males who suffer from difficulty in walking and have a higher age category are more prone to heart diseases. For e.g., taking the 50 or older age category into consideration, a drastic increase in the percentage of heart diseases can be observed when the sex changes from female to male.

Moving on to the next part of storytelling, alcohol consumption in individuals is now taken into consideration. The pie chart thus generated indicates that the probability of heart diseases is less in people who consume alcohol.

Question 2: How do lifestyle choices affect the occurrence of heart disease in a person?

This insight does not align with the basic idea that is attached to alcohol consumption and heart diseases. The popular medical belief is that alcohol consumption increases the chances of heart diseases in individuals. However, our visualization indicates no such data. One of the reasons responsible for this could be the unavailability of adequate data in terms of alcohol consumption levels i.e., if a person is a heavy drinker or consumes just one drink per day. The availability of this data could help generate better useful insights.

Smoking is a major cause of heart disease and causes approximately one of every four deaths from CVD, according to the 2014 Surgeon General's Report on smoking and health. Smoking increases the formation of plaque in blood vessels. Coronary Heart Disease Occurs when arteries that carry blood to the heart muscle are narrowed by plaque or blocked by clots. Chemicals in cigarette smoke cause the blood to thicken and form clots inside veins and arteries.

In our analysis, a comparison is drawn between smoking and heart disease. The probability for people who smoke and have heart disease is higher than the people who don't smoke and have a heart disease. The visualization shows that the probability levels almost increase by **20%** from non-smokers to smokers. This indicates via bar graphs that smokers have higher chances of heart diseases when compared to non-smokers.

From the insurance company / market point of view, this insight can help companies to design policies that cater and take into consideration a wider range of health concerns and not just focus on alcohol consumption by the individuals.

Question 3: How does mental health in combination with sleep time affect the occurrence of heart disease in a person?

Moving on to the next visualization, sleep provides time for the body to restore and recharge, playing a key role in nearly all aspects of physical health. According to the CDC, most adults need at least 7 hours of sleep each night. However, more than 1 in 3 American adults say they don't get the recommended amount of sleep. While this may be fine for a day or two, not getting enough sleep over time can lead to serious health problems—and make certain health problems worse.

Based on our visualization that is based on number of hours of sleep and heart diseases it can be observed that individuals who have a sleep time of 6 - 10 hours per day, are less prone to heart

diseases when compared with individuals having poor sleep cycles of 0 - 3 hours or higher sleep cycle of 12+ hours.

Heart disease and mental health are closely linked. Depression can be as big a risk factor for coronary heart disease as smoking, high cholesterol levels and high blood pressure. Depression can also affect the recovery of people with coronary heart disease and increase their risk of further heart problems.

Now taking mental health into consideration, our visualization indicates that individuals who have 12+ bad mental health days in a month are more at risk for heart diseases. This insight can be used by insurance companies to provide add ons to the existing policies to individuals that suffer from mental health issues.

Question 4: How does the combination of underlying ailments affect the chances of developing heart diseases?

If an individual is diabetic, the risk of developing cardiovascular disease is more than double that of the general population, according to the American Heart Association Trusted Source. For people with type 2 diabetes, heart disease is the most common cause of death. Understanding the connection between diabetes and heart disease from our dashboard, we can see that people with diabetes have a 16% more chance, whereas people who are borderline diabetic have 50% less chance of being prone to heart diseases.

Multimorbidity, defined as the coexistence of 2 or more chronic diseases, is a common phenomenon especially in older people. For e.g., when the kidneys don't work well, more stress is put on the heart. When someone has kidney diseases, their heart needs to pump harder to get blood to the kidneys. This can lead to heart disease.

From the visualization obtained it can be observed that the individuals that do not have any underlying health ailments are less prone to heart diseases. As we move gradually across the table and the underlying health ailments increase in individuals, the trend can be observed that individuals with a number of underlying health ailments / multimorbidity have significantly higher chances of heart diseases.

There is also a comparison drawn between the occurrence of heart disease and the history of previous stroke. The bar graphs represent higher probability of heart disease in individuals who have a history of previous stroke.

Taking this trend and insights into consideration from an insurance company / market point of view, policies can be designed in order to generate higher returns in case of individuals suffering from multiple health ailments / multimorbidity.

Question 5: How can the insights derived from the previous questions be used to prevent heart disease in the age category that is most prone to it?

The insights obtained from the data shows people above the age of 50 are more prone to heart disease than the people below the age of 50. Reducing risk means maintaining a healthy weight, getting plenty of exercise, good quality sleep and not smoking.

The visualizations obtained show there's a strong connection between mental health, sleep and overall physical health, specifically cardiovascular health.

Lifestyle factors such as diet and exercise may reduce the risk of heart disease, both directly and indirectly, by fostering better sleep. This, in turn, may lead to a better outlook and more energy – the very best kind of feedback loop.

All the above insights can be used by health insurance providers determining if an individual is a high risk or not.

For example, if the insurance company actuaries review a certain individuals for one year and determine that they are in a low risk category and only charges very minimal premiums that year, but then by the end of the year they see a rise in high losses, or claims payouts, it will cause them to review their results and change the premium they charge for individuals in the new year. In this case, our analysis could come in handy to determine which individuals' premium costs need to be reviewed again. Those individuals will then see rate increases as a result.

In Massachusetts alone over 70% of all employers offer health insurance as a benefit to their employees. Most of these employers pay part of the premium and also offer a choice of several health plans. So the above analysis could help choose the health plan that is best for an individual. Insurance premiums usually have a base calculation. Then, based on personal information and location, insurance companies could offer discounts that are added to the base premium that reduces cost.

TABLE OF CONTRIBUTIONS

Description	Person Responsible
Project Proposal	Honey Johnson, Sushobhit Lutade, Aaditi Shyam
Data Set & Problem Statement	Adithya Sajith Kumar, Ramandeep Kaur
Project Strategy	Sushobhit Lutade, Honey Johnson, Adithya Sajith Kumar
Analysis/Insights	Aaditi Shyam, Ramandeep Kaur