

---

# HEART DISEASE: CAUSES AND PREVENTION ANALYSIS

DATA ANALYSIS SUMMARY AND RECOMMENDED ACTIONS

## TABLE OF CONTENTS

*Our team 3*

---

*Background information on client organization 4*

---

*Data and variables used in the analysis 5*

---

*Goal of the analysis 6*

---

*Validity, reliability and collinearity checks 7*

---

*Results of the analysis 8*

---

*Recommended actions based on results 9*

---

*Appendix A: Team members and contributions 10*

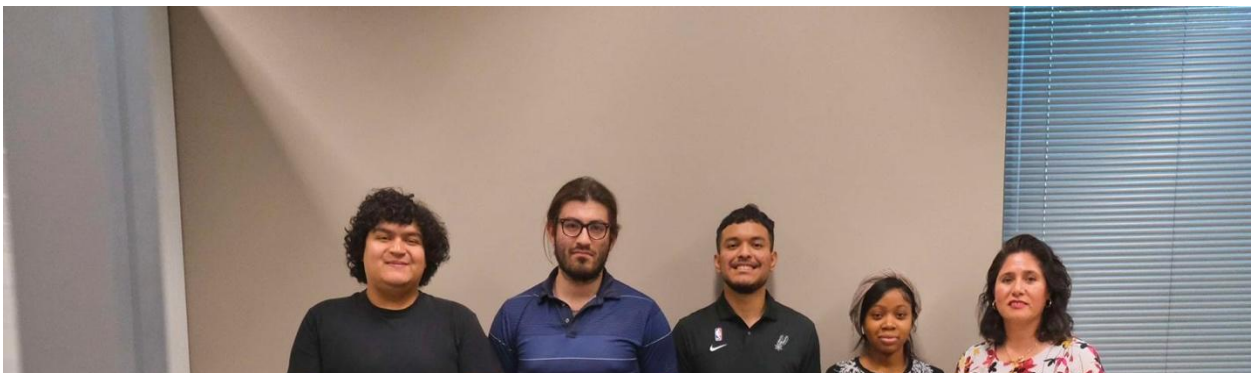
---

*Appendix B: Letter from client organization 11*

---

## OUR TEAM

---



**Team introduction follows from the initial individual on the left to the right.**

**Javier Alvarado:** I am the first of my family to attend University in the United States, after living in Mexico for eleven years I am now an undergraduate student in my last year in this university, pursuing a degree in Management Information Systems and Data Analytics, hoping to one day become an IT consultant and help as many people as possible with any technical issues they might have. I am also very passionate about graphic design, video game development and audiovisual editing.

**Diego Cenicerros Dominguez:** I am a senior studying Management Information Systems and Data Analytics. While I am first and foremost an athlete at heart, having competed in various

cross country and powerlifting events. I've also developed a jack of all trades skillset from working numerous jobs (currently a front desk supervisor at a Hilton hotel), and plan on pursuing a software developer role that specializes in the MERN stack upon graduation.

**Felipe Contreras:** I'm a senior here at the University of Texas A&M International, majoring in management information systems. This is currently my last semester before I graduate. After graduation, I plan to continue to work on my social media agency and perhaps add my new skill sets learned from class as a data analyst to my entrepreneurial career.

**Simileoluwa Odunuga:** I am a senior soon to graduate, majoring in Management Information Systems and Data Analytics, concerning my educational background, I plan on furthering myself in my academic pursuit. I recently became accepted into the Baccalaureate Accelerated Program (4 + 1), it has become the best and most suitable path to achieving my goals.

**Miriam E Loera:** I am a senior pursuing a Management Information Systems and Data Analytics major. I only have worked at Walmart and a Jewelry store; I have always been a housewife. I decided to finish my degree, and I will graduate by the end of the fall of this year, 2023. I don't have any plans, but we'll see what the future repairs are for me.

---

## BACKGROUND INFORMATION ON CLIENT ORGANIZATION

---

The University of Texas Health Science Center at Houston (UTHealth) is a public academic health center located in Houston, Texas, United States. Established in 1972 by the Texas Legislature as the University of Texas Health Science Center at Houston (UTHSC-H). After this, in the year 2013, the institution was renamed to UTHealth to better reflect its diverse programs and academic activities.

UTHealth is an advocate and is strongly dedicated to improving human health through education, research, and patient care. It offers a wide range of programs in health sciences, including medicine, nursing, public health, dentistry, and biomedical sciences. The institution's primary mission is to educate the next generation of healthcare professionals and scientists, conduct innovative research, and provide exceptional patient care.

UTHealth has six schools which include McGovern Medical School, School of Dentistry, School of Nursing, School of Public Health, School of Biomedical Informatics, and School of Health Professions. Each school offers a variety of degree programs and areas of study, ranging from undergraduate degrees to doctoral programs.

In addition to its educational programs, UTHealth is a leading research institution with more than 1,000 active research projects across its schools. The institution conducts research in a wide range of areas, including cancer, cardiovascular disease, neurology, psychiatry, infectious diseases, and genetics. UTHealth is also home to several research centers and institutes, including the Center for Clinical and Translational Sciences, the Center for Health Promotion and Prevention Research, and the Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases.

UTHealth has a strong commitment to patient care, with several affiliated hospitals and clinics in the Houston area. These include UT Physicians, which provides comprehensive primary and speciality care to patients, and the UTHealth School of Dentistry Clinic, which offers dental care services to the community. Overall, UTHealth is a renowned academic health center dedicated to improving human health through education, research, and patient care. With its diverse programs and commitment to innovation, UTHealth is a leading institution in the field of health sciences, both in Texas and around the world.

According to its 2020 financial report, the University of Texas Health Science Center at Houston (UTHealth) had a total revenue of \$1.08 billion and a total expense of \$1.06 billion, resulting in a surplus of \$16.4 million. It's important to note that this financial information is subject to change and may not accurately reflect their current revenue or profits. In the year 2022, UTHealth's financial report the university had over 11,000 employees. As of August 31st 2022 with current assets of over \$100 million and total operating revenue of \$1.06 million. It's important to note that this number may have changed since then due to factors such as hiring, retirement, and turnover.

---

## DATA AND VARIABLES USED IN THE ANALYSIS

---

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey conducted by the Centers for Disease Control and Prevention (CDC). The study collects data from over 400,000 Americans each year on health-related risk behaviors, chronic health issues, usage of preventative treatments, and other topics. The core dataset contains 330 traits and answers from 441,455 people. These aspects include direct questions posed to participants as well as computed variables depending on individual participant replies. This dataset comprises 253,680 cleaned BRFSS 2015 survey responses that will be used largely for the binary categorization of heart disease.

This dataset does not have a significant class imbalance. 229,787 respondents do not have/have never had heart disease, whereas 23,893 have. Heart disease is one of the most prevalent long-term illnesses in the United States, impacting millions of people each year. Heart disease is the leading cause of mortality in the United States, claiming over 600,000 lives each year. The CDC has identified three major risk factors for heart disease: high blood pressure, high cholesterol, and smoking. Approximately half of all Americans have one or more of these three risk factors. The age coding was a formula designed to make the data more intelligible by categorizing distinct age categories, ages 18 to 24 are labelled as 1, 25 to 29 as 2, and so on.

Warpls was used to create the path model, which changed one of the variables into a logistic regression variable. We made use of Warppls was used to create the path model, which changed one of the variables into a logistic regression variable. When the variable is dichotomous, we utilized logistics regression to estimate probability. It is a classification strategy that works well

for predicting categorical response variables, particularly binary classification. It also aids in the detection of data abnormalities such as Simpson's paradox. Simpson paradox occurs when the correlations and path coefficients have opposite signs. This is an incident we are attempting to prevent since it establishes erroneous associations that can lead to wrong conclusions.

To understand the relationship between the variables, they must be properly defined.

Heart Disease: In this case, heart disease was our main variable, which led to it being the focus of our analysis. When we worked with our main variable, we were exploring its relationship to the other variables in the analysis. This is why “heart disease” had a very important role in our research as we went forward, we discussed how it relates to the other variables such as education, Income, Physical Activity, BMI, if the individual is a Smoker, and if the individual consumed fruits daily.

The relationship between education and heart disease is linked to how individuals with lower levels of education may be more likely to develop heart disease. This can be associated with many factors, such as the differences in lifestyles and habits. This can include how the individual manages their diet and exercise routines. This can also be related to how individuals with lower levels of education are more likely to be exposed to situations where they are under stress and therefore that leads to poor mental health, both contributing to the development of heart disease.

There are various ways that income and heart disease are related in this analysis. In the case of this analysis, income can be seen as a risk factor, an individual with lower income would have poor access to health care. This lower income will also lead to this individual having an unhealthy diet, and possibly higher levels of stress. Additionally, in our analysis, education has a strong relationship with income. An individual who pursues a higher level of education is more

likely to earn more income, this can lead to many lifestyle changes and factors that can lower the risk of developing heart disease.

Physical activity and heart disease have a very well-established relationship. An individual who involves themselves in physical activity regularly will most likely not encounter any problems against heart disease, compared to an individual who follows a sedentary lifestyle, which could lead to a risk of developing heart disease. In our analysis, we established that physical activity could assist in the reduction of the risk of heart disease in many ways. For example, following a lifestyle of regular physical activity can help lower the blood pressure of an individual, as well as reducing the risk of obesity and diabetes. It is also important to note that physical activity can improve cardiovascular functions, which is very important, because it can increase the ability to pump blood efficiently. Contrary to all of this, an individual who follows a lifestyle that lacks physical activity will most likely develop heart disease. This can be caused by high blood pressure, high cholesterol levels, obesity and, as previously mentioned, diabetes.

In the relationship between age and heart disease, BMI plays a moderating role. Age has no connection to education because Education, being a time-consuming process, has no relation to age since it takes time and is continuous, hence a connection being nonexistent.

Moving on to the relationship between BMI and heart disease, this is also considered to be a very well-established relationship. Because BMI is known to be a measure of body fat related to an individual's height and weight, it is known that a high BMI can lead to an increase in the risk of developing various health conditions, in this case including heart disease. Having excessive body fat around the waist can increase the risk of developing heart disease in an individual. If an individual possesses the physical properties that have been previously mentioned, this can also



lead to high blood pressure, high cholesterol, and diabetes. All these factors are also related to inflammation and damage to the blood vessels, which can further increase the risk of heart disease.

Furthermore, smoking is known to be a risk factor for heart disease. It is commonly known that smoking can increase the risk of developing heart disease by damaging the blood vessels of the individual who commonly smokes, additionally increasing their blood pressure, and reducing the amount of oxygen that reaches the heart of that individual. In our analysis, this meant that an individual who regularly smokes is more likely to develop a heart disease opposed to those who do not smoke.

Lastly, we took a look at the relationship between fruits and heart disease. Fruits are considered to be a part of a healthy diet and are often associated with reducing the risk of developing heart disease in the future. This is all because fruits are rich in nutrients like fiber, vitamins, and minerals, that help to improve the heart health of individuals. In our analysis, we discussed how the increased consumption of fruits is associated with a lower risk of heart disease.

## GOAL OF THE ANALYSIS

---

The main objective of our analysis "Heart Disease: causes and prevention analysis," which heart disease is one of the leading causes of death in the United States, is to establish that although we have seven variables that have an impact on the cause of heart disease, even so, the two main most vital variables since they influence the prevention of heart disease, are the education that each one of us has about this disease and the income that each one of us receives.

Although doctors analyze factors such as age to diagnose heart disease, and it is a factor that we cannot change, education is something that we can help with; it is a significant variable that would reduce the risk of heart disease since preventive measures, can solve many problems and thus try to avoid symptoms (chest pain, neck pain, shortness of breath, numbness, weakness or cold in the legs or arms). It is also good to know the causes and risk factors for heart disease; smoking, alcohol consumption, obesity (high BMI), high blood pressure, diabetes, and not doing enough physical activity are some of the critical risk factors for these diseases. Heart diseases significantly impact older adults, and they hardly have any information about this disease. It is significant because people with higher education have a lower risk of developing heart disease; as they studied, they have a lower body mass index, more rigorous physical activities, and better perceived physical health.

In our analysis, we also found that income is another of the main variables for heart disease since the risk of cardiovascular disease (CVD) is more significant due to a poor diet (high consumption of salt, sugar, and fat), whether unhealthy, without consumption of fruits and vegetables, the main reason of this is because of the income. The decrease in income can also

cause the loss of health insurance, avoiding doctor appointments, and increasing stress or depression. With low income, you can also prevent regular medical exams, such as checking blood pressure and cholesterol levels and diabetes screening exams. However, with that income, you still need to make sure you buy your medicines as prescribed by the healthcare provider and follow a plan for a healthy lifestyle.

The Governing Bodies of the Pan American Health Organization (PAHO) adopt technical guidelines and recommendations on the formulation of health-related legislation, consolidating it in a strategy on the subject. The objectives of compiling the recommendations related to the health of the Governing Bodies of PAHO in terms of legislation, reviewing the concepts and links primary relationships between legislation, public health, and human rights, determining national trends and challenges regarding legislation, and finally, point out the principles, values, vision, the objectives, and lines of action of a strategy on the bill to be approved.

With this analysis, we prove the causes and preventions of heart diseases, showing that this disease is not inevitable since, with a good education and a good income, one can obtain the best practices to prevent, treat and manage heart diseases to positively impact the entire burden attributed to these diseases and thus maintain a healthy lifestyle.

---

## VALIDITY, RELIABILITY AND COLLINEARITY CHECKS

In order to measure and ensure our model and analysis are valid, two things must be established: collinearity checks and the P values.

Initially, we have to take into account the Variance Inflation Factors (VIF) which measure the degree of collinearity amongst the variables. It takes two main forms in vertical and lateral collinearity. With that said, we ensured validation of the coefficients with a technique called full linearity which combines vertical and lateral types of collinearity, effectively measuring both simultaneously. In addition, the coefficients below the 3.3 threshold suggests the existence of no multicollinearity in the model, with 5 being the more relaxed and lenient threshold.

Path coefficients									
	HrtDis	BMI	Smoker	PhysActv	Fruits	Age	Educ	Income	BMI*Age
HrtDis		0.178	0.179	-0.228	-0.114	0.168	-0.310	-0.314	0.203
BMI							-0.135		
Smoker							-0.128		
PhysActv							0.195		
Fruits							0.096		
Age									
Educ									
Income							0.472		
BMI*Age									

As one can see, the coefficient values are all below the 3.3 threshold both vertically and laterally with the only exception being that of income and education. However, it still falls under the more “relaxed” threshold of 5. Thus, ensuring the important precondition of collinearity in our analysis that allows the interpretation of the latent variables to measure different things.

Furthermore, to check the validity of our data, we must first consider the convergent validity. Convergent validity is a measure of the quality of the measurement instrument. Instruments typically fall under the questionnaire or survey range. Moreover, an instrument is said to have good convergent validity if the questioned statements are understood by the participants in the same way the designers of the questions intended them to be understood. With that said, the best way to ensure convergent validity is if the P values associated with the loadings are equal to or below the 0.5 threshold.

P values									
	HrtDis	BMI	Smoker	PhysActv	Fruits	Age	Educ	Income	BMI*Age
HrtDis		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
BMI							<0.001		
Smoker							<0.001		
PhysActv							<0.001		
Fruits							<0.001		
Age									
Educ									
Income							<0.001		
BMI*Age									

When the P values are analyzed one can see that all the values share the same score of 0.1. This ensures our convergence validity and solidifies the idea that the respondents of the survey clearly understood the question. Thus, solidifying the validity and reliability of our analysis and model.

## RESULTS OF THE ANALYSIS

---

Before the creation of our model, we first ensured no instance of Simpson's paradox existed in our data, and we used Logistic regression to convert the dichotomous variable into a probability. As a result, our Warp PLS model then shows that all P values significantly influence heart disease development because all our P values are below the established threshold of 0.5. The beta coefficients of each variable demonstrate their impact on heart disease likelihood. Excessive physical activity may increase coronary artery disease risk, despite its overall health benefits. A higher body mass index for age increases heart disease risk. Smoking and fruit consumption have a lesser impact but should still be considered in prevention. These findings emphasize the importance of a healthy lifestyle in preventing heart disease.

Education and heart disease also have a direct link that is negatively correlated. The order of importance for education indirectly influencing heart disease moderating variables is based on their beta coefficient. The order goes as follows:

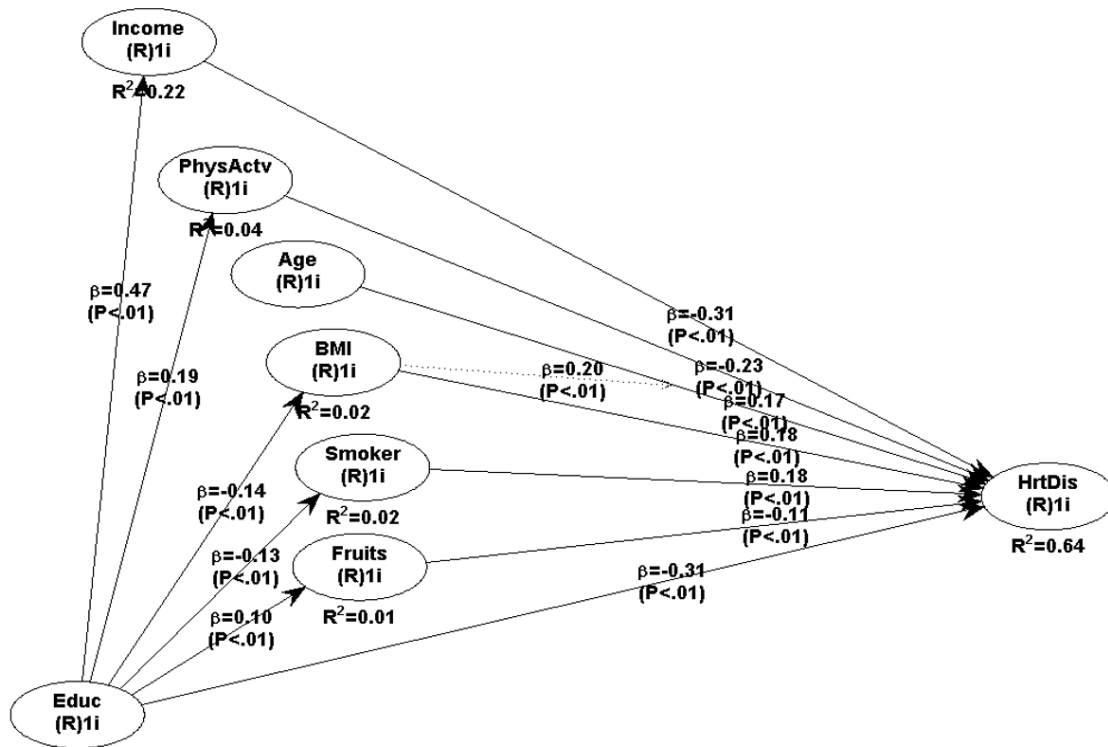
1. Income
2. Physical activity
3. BMI in relation to the age of a person
4. Being a smoker
5. Consuming fruits

The beta coefficient must be significant on both sides of moderating variables to ensure the validity of the correlation, which they currently all are. Higher education is positive to income,

which is negative for heart disease. Physical activity is also positive with higher education and negative to heart disease. Education has a negative influence with respect to BMI and smoking because they are less likely to partake in self-destructive habits. Fruits reduce the chance of heart disease due to their antioxidants.

Although age could've been a leading variable towards getting heart disease, it wasn't connected through our model because age does not change based on our current predictor. For example, regardless of a person's age, there are many instances where young adults and children have suffered from cardiovascular disease due to their heart disease being passed down to their parent's genes. This phenomenon is called an inherited cardiac condition (ICC), where genetic cardiac conditions can pass through heart conditions and other conditions such as high blood pressure and high blood cholesterol. For this particular reason, we couldn't formally calculate if age can be one of the main predictors of getting heart disease since the results may vary differently from person to person.

Our results have shown that Education has the most positive beta coefficient linked towards income, with a mark of 0.47. This result concludes that Education is a strong predictor of income. Education as a whole also has a solid connection to having a better BMI, food consumption, and higher chances of being a smoker. This shows that Education is the main independent variable of our model due to all the strong connections it has with other variables. Based on Education's positive relationship with not getting heart disease, we can further investigate what preferred education levels, whether secondary (high school) or university education, show more signs of getting a heart disease.





---

## RECOMMENDED ACTIONS BASED ON RESULTS

---

For better improvement towards getting better results in our linear regression model and having an improved chance of not getting heart disease, we strongly recommend that the average person increase their physical activity. Our analysis showed that not having enough physical activity was a strong link to getting heart disease. Many of us will more than likely not be able to have total control over the income, BMI, and choice of foods we want to have in our everyday life, so due to this reason, physical activity is easily the most accessible variable to control of out of all five. We recommend at least 30 mins physical activity a day, three times a week, to reduce the chances of getting heart disease. Not only will the average person reduce their chances of getting cardiovascular disease, but they will also reduce their chance of getting other diseases like obesity, diabetes, and osteoporosis. Simple exercises like walking in the park or playing basketball for a few minutes have significantly improved people's overall health and fitness compared to not doing anything at all. Overall, recommending that more people consistently do physical activities throughout the week will reduce the chances of getting heart disease and improve our data analysis.