

**Portfolio Project # 02**

**COURSE: DATA ANALYTICS**

**Cardiovascular Diseases Risk Prediction Using Personal Lifestyle Factors**

SUBMITTED BY

NAZISH ABDULLAH

**COHORT 12**

SUBMISSION DATE: 16-08-2024

**Introduction**

Heart disease remains a leading cause of morbidity and mortality worldwide, emphasizing the critical need for effective preventive strategies. Identifying and understanding the risk factors associated with heart disease can provide valuable insights into how lifestyle and behavioral modifications can mitigate risk and promote cardiovascular health.

**Dataset**

The "Cardiovascular Diseases Risk Prediction Dataset," available on Kaggle, is a comprehensive resource designed to facilitate research and development in the field of cardiovascular risk assessment. This dataset encompasses a range of features relevant to cardiovascular health, including lifestyle, and medical history variables. The dataset has over 300,000 respondents' input and 19 variables.

|  |  |
| --- | --- |
| **Fields** | **Detail** |
| General Health | Would you say that in general, your health is |
| Checkup | About how long has it been since you last visited a doctor for a routine checkup? |
| Exercise | During the past month, did you participate in any physical activities? |
| Heart Disease | Respondents who reported having coronary heart disease |
| Skin Cancer | Respondents who reported having skin cancer |
| Other Cancer | Respondents who reported having any other types of cancer |
| Depression | Respondents who reported having a depressive disorder |
| Diabetes | Respondents reported having diabetes. |
| Arthritis | Respondents that reported having an Arthritis |
| Gender | Male/Female |
| Age Category | Age Range of respondents |
| Height\_(cm) | Height of respondents |
| Weight\_(kg) | Weight of respondents |
| BMI | Body Mass Index |
| Smoking History | Does the respondent smoke? |
| Alcohol Consumption | What is the frequency of alcohol intake every month? |
| Fruit Consumption | Fruit Consumption in last month |
| Green Vegetable Consumption | Number of intakes of Green vegetables during the last month |
| Fried Potato Consumption | Times fried potato was consumed by the respondent |

**Data Wrangling**

Dataset is uploaded on Google Collab to perform analysis using Python. The dataset csv file was cleaned from null and redundant values.

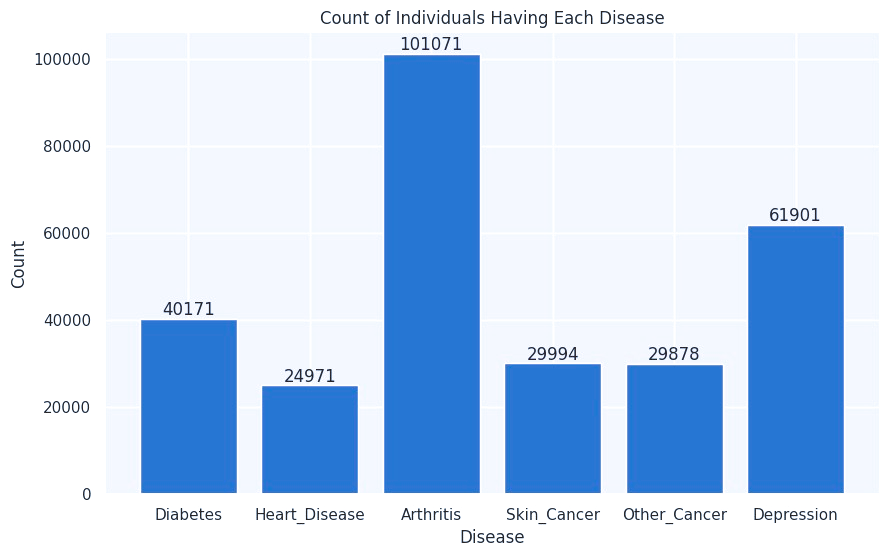
In the diabetes column, the occurrence of a temporary rise in sugar level was replaced by No.

The major transformation had to be made in numerical data types, by grouping and naming them into categories for analysis purposes. In fields of fruit and vegetable consumption, the number represented the frequency of its consumption in a month. Therefore to reach meaningful insight the numbers were grouped in four ranges **Never, Seldom, Frequent, and Regular.**

Once the data was checked by running df.head(), df.info(), df.tail(), df.describe commands, it was saved as a cleaned.modified.csv file.

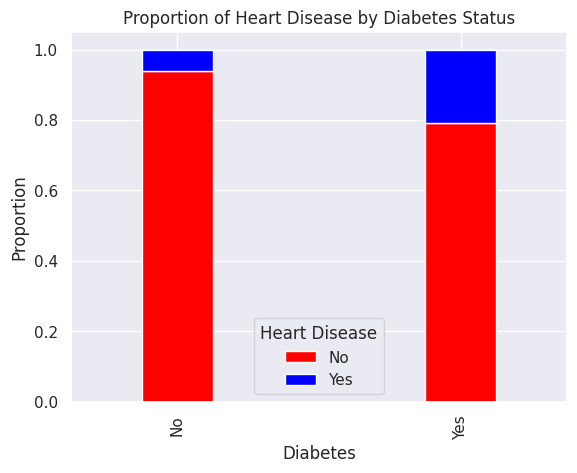
**Data Analysis**

1. Statistics of different disease reported by respondents**.**

****

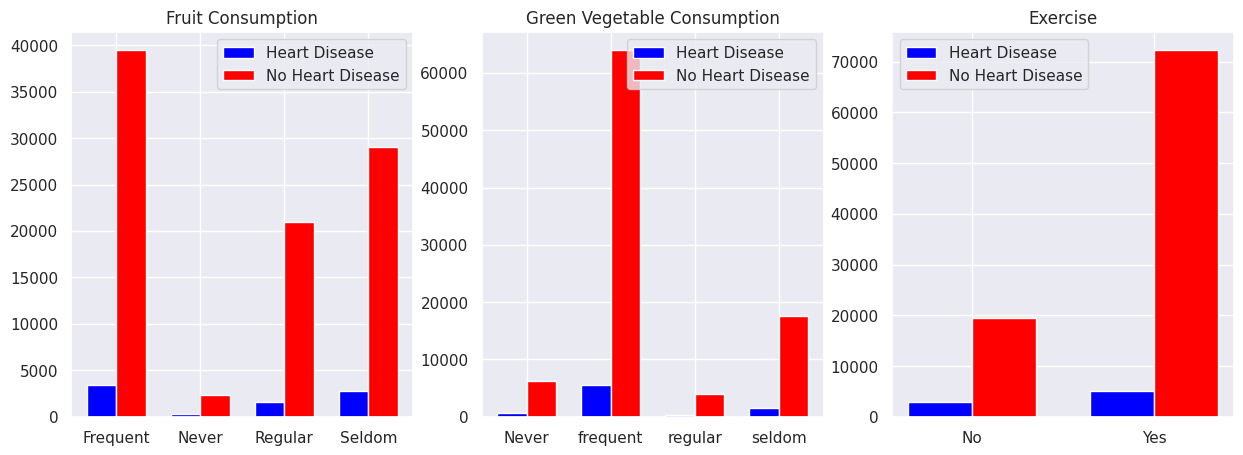
**Insight:** As illustrated in graph, over 100,000 respondents are suffering from Arthritis, followed by depression and diabetes. There are 24,971 respondents who have CVD.

1. Relationship between Diabetes and CVD.



**Insight:** Among 24,971 respondents having CVDs, the majority also suffer from Diabetes.

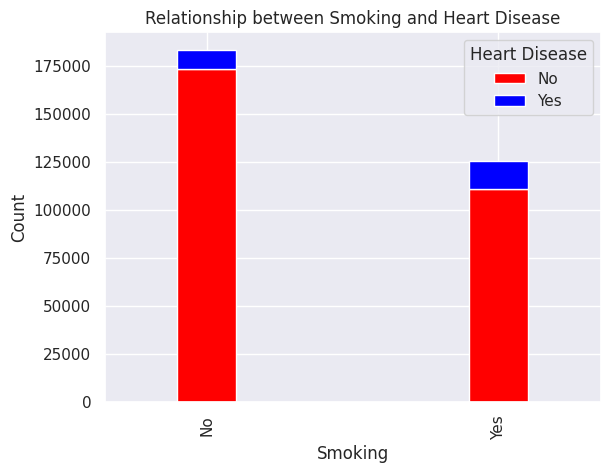
1. Which are controllable factors to ensure prevention of heart diseases.

:

**Insights:** In the following illustrations we can see that among individuals who eat vegetables and fruits on regular basis very less percentage suffer from CVDs. Regular exercise and physical activities are mandatory to main overall as well as health of the heart.

1. How does smoking habit impact heart?

|  |  |  |
| --- | --- | --- |
| **Smoking** | **Heart Disease** | |
| **No** | No | 94% |
|  | Yes | 6% |
| **Yes** | No | 88% |
|  | Yes | 12% |



**Insight:** Out of the total respondents around 184,000 do not have history of smoking.

**Recommendations**

* **Eat a Heart-Healthy Diet**
* Emphasize fruits, vegetables, whole grains, and healthy fats
* **Maintain a Healthy Weight**
* Balance caloric intake with physical activity
* **Engage in Regular Exercise**
* Aim for 150 minutes of moderate activity per week
* Include strength training twice a week
* **Avoid Smoking & Limit Alcohol**
* Quit smoking and reduce alcohol consumption
* **Control Cholesterol & Blood Sugar**
* Follow dietary guidelines and monitor levels
* **Get Regular Health Checkups**
* Regular exams and follow medical advice
* **Stay Socially & Mentally Active**
* Maintain social connections and manage stress

**Link for dataset**

[Cardiovascular Diseases Risk Prediction Dataset (kaggle.com)](https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset)

**Conclusion**

In analyzing the cardiovascular disease risk prediction dataset, we have uncovered several key insights into the factors contributing to cardiovascular risk. The data wrangling process highlighted the importance of handling missing values, ensuring consistent data types, and addressing categorical variables effectively. These insights can guide targeted interventions and inform preventive strategies by focusing on high-impact risk factors.

.