

Data Source: The dataset used in question #1 of homework #1 is acquired from Kaggle.

<https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>

Data is sliced since the file is too big to upload to moodle.

Data Description: The dataset used in question #1 is a medical dataset, which contains 18 features. Features are as follows:

- **HeartDisease (categorical data – “yes” or “no”):** This column indicates if the person has a heart disease or not.
- **BMI (qualitative data):** This feature gives the body mass index of the person.
- **Smoking (categorical data – “yes” or “no”):** This feature indicates if the person is smoking or not.
- **Stroke (categorical data – “yes” or “no”):** This feature indicates if the person has a stroke history or not.
- **PhysicalHealth (qualitative data):** This feature gives the number of days over the past 30 days at which the person had a physical illness or injury.
- **MentalHealth (qualitative data):** This feature gives the number of days over the past 30 days at which the person’s mental health was not good.
- **DiffWalking (categorical data – “yes” or “no”):** This feature indicates if the person has serious difficulty walking or climbing stairs.
- **Sex (categorical data – “male” or “female”):** This feature indicates if the person is male or female.
- **AgeCategory (categorical data):** This feature gives the age category of the person.
- **Diabetic (categorical data – “yes” or “no”):** This feature indicates if the person has diabetes or not.
- **PhysicalActivity (categorical data – “yes” or “no”):** This feature indicates if the person did a physical activity or exercise other than regular job in the past 30 days.
- **GenHealth (categorical data):** This feature gives the general health of the person.
- **SleepTime (qualitative data):** This feature gives the average number of hours the person sleeps in a 24-hour period.
- **Asthma (categorical data – “yes” or “no”):** This feature indicates if the person has asthma or not.
- **KidneyDisease (categorical data – “yes” or “no”):** This feature indicates if the person ever had a kidney disease excluding kidney stones, bladder infection and incontinence or not.
- **SkinCancer (categorical data – “yes” or “no”):** This feature indicates if the person ever had skin cancer or not.

The dataset contains 319795 rows.

Libraries used: Pandas, numpy, seaborn and matplotlib libraries are used.