**PRODIGY\_DS\_02**

**TASK - 02**

OBJECTIVE : To perform data cleaning and exploratory data analysis(EDA) on a dataset.

To explore relationships among variables and identify pattern and trends in the data.

DATASET : A dataset containing patient’s medical history has been taken having 28 columns and 5191 values .

DATA CLEANING and EDA : Data cleaning and EDA has been done using python jupyter notebook and same has been uploaded named “PRODIGY\_DS\_02.ipynb”.

RESULTS AND KEY INSIGHTS :

1. The mean height, weight and BMI are 62.85 kg, 1.61 m and 24.16 kg/m2.
2. More females have been diagnosed than males.
3. People in the age group of 50-59,40-49 are more vulnerable to get a disease.
4. 78% of the diagnosed people are Chinese.
5. Around 54% of diagnosed people have primary and GCE level educational background.
6. Females who exercise less than 30 mins are more prone to get a disease.
7. Diabetes, high blood pressure and high cholesterol is more prominent in the age group 60-69 years.
8. Health Risk is critical for 2% people.
9. 2.7% and 1.7% people have a history of heart attack and stroke.
10. Around 53% of females are overweight.
11. Around 97% of people consume fruits.
12. 10.8 % people smoke.
13. Around 75% people are unhealthy.
14. Weight and BMI are highly correlated.
15. Females of the age group 50-59 years are diagnosed more.