**Your Task - 6:** Identifying unneeded columns in a dataset and how to drop them Identifying unneeded columns in a dataset requires a combination of data analysis, domain knowledge, and the specific goals of your analysis.

To

Rishabh Goyal,

From

Pampana Jai Kiran.

Sir, I have successfully completed the task 6.

Code:

import pandas as pd

data = pd.read\_csv('/content/Practice dataset 5.csv')

df = data

df.info()

df.describe()

df

columns\_to\_drop = ['hypertension', 'smoking\_history', 'bmi', 'HbA1c\_level', 'blood\_glucose\_level']

df = df.drop(columns=columns\_to\_drop)

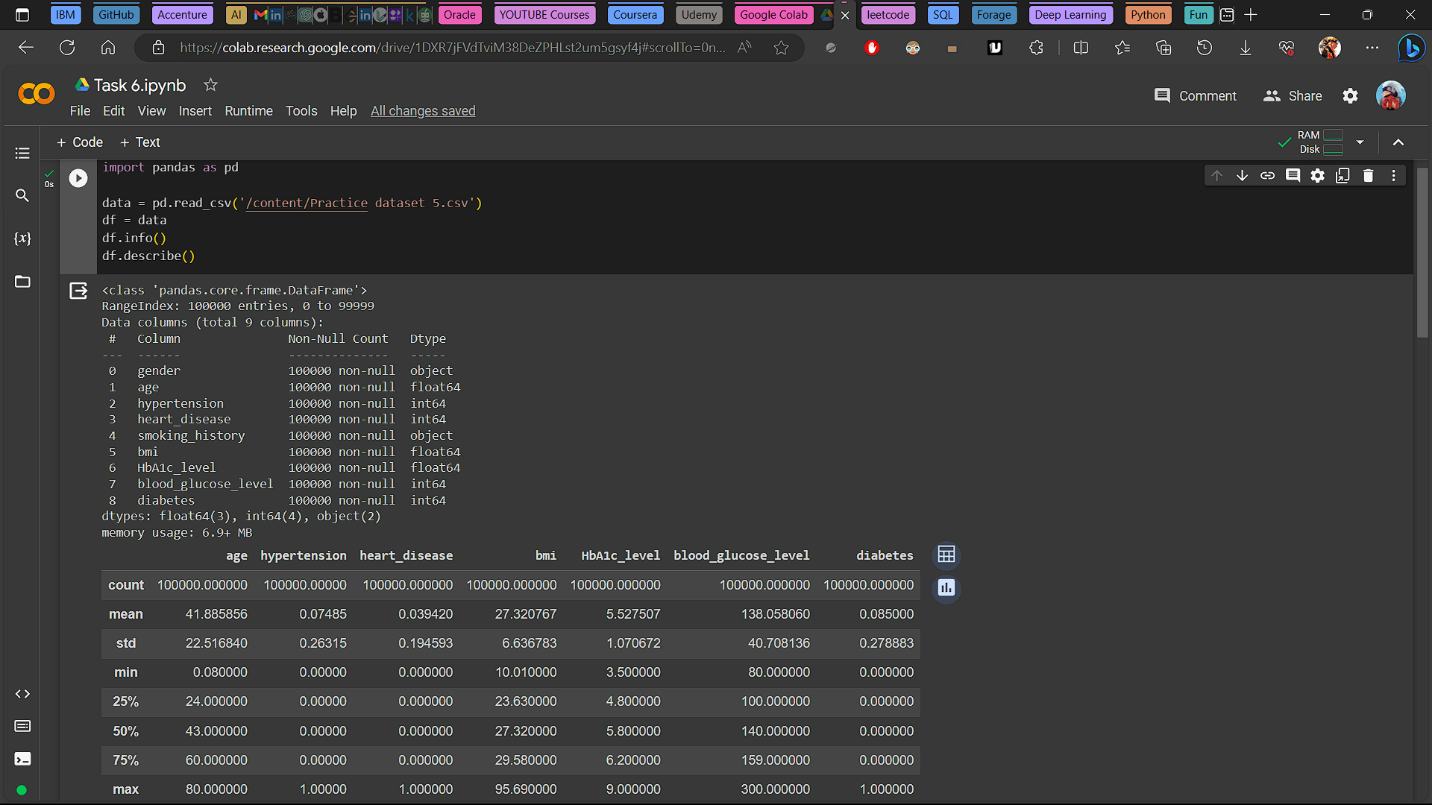
df

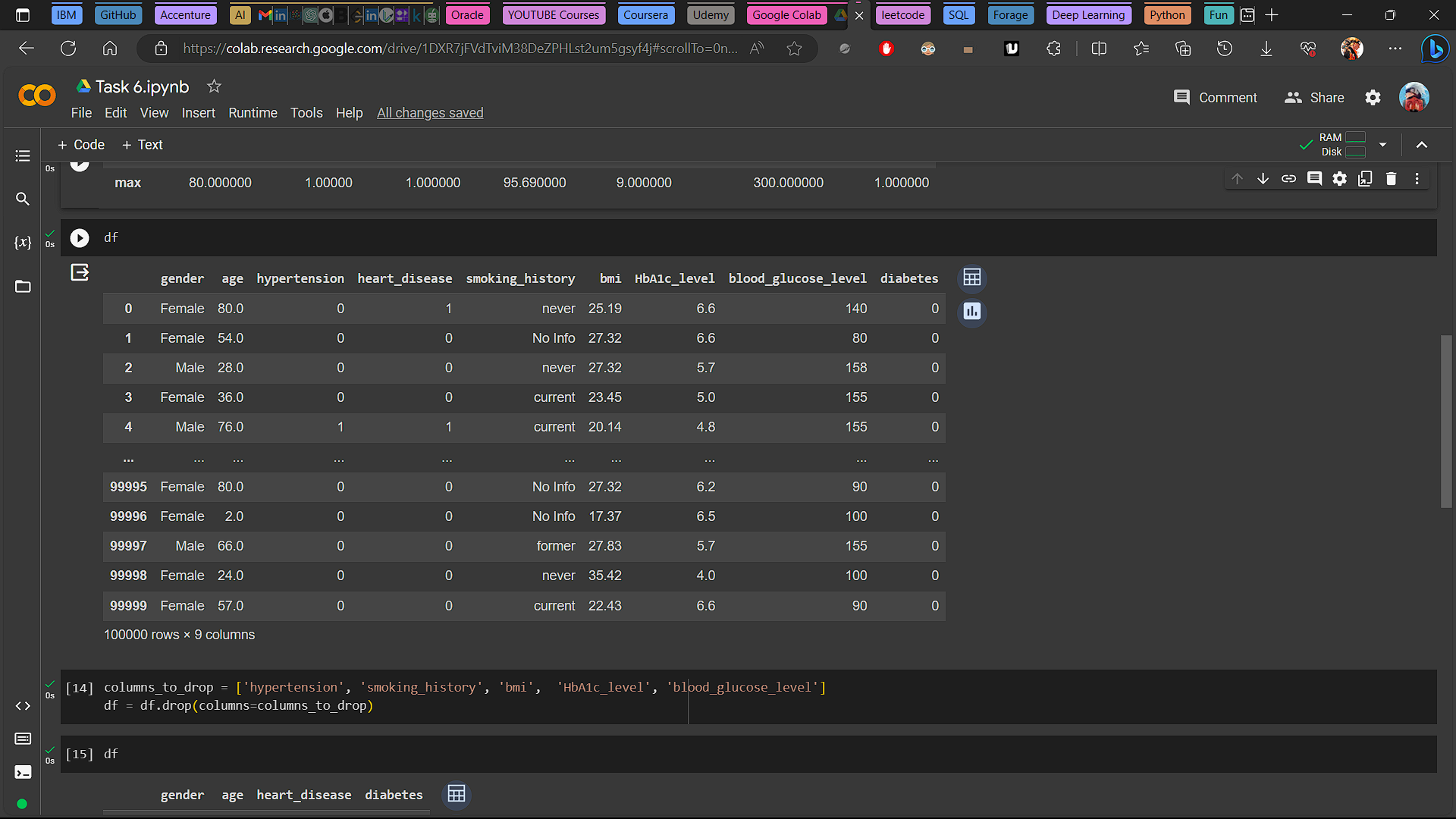
value\_counts = data['gender'].value\_counts()

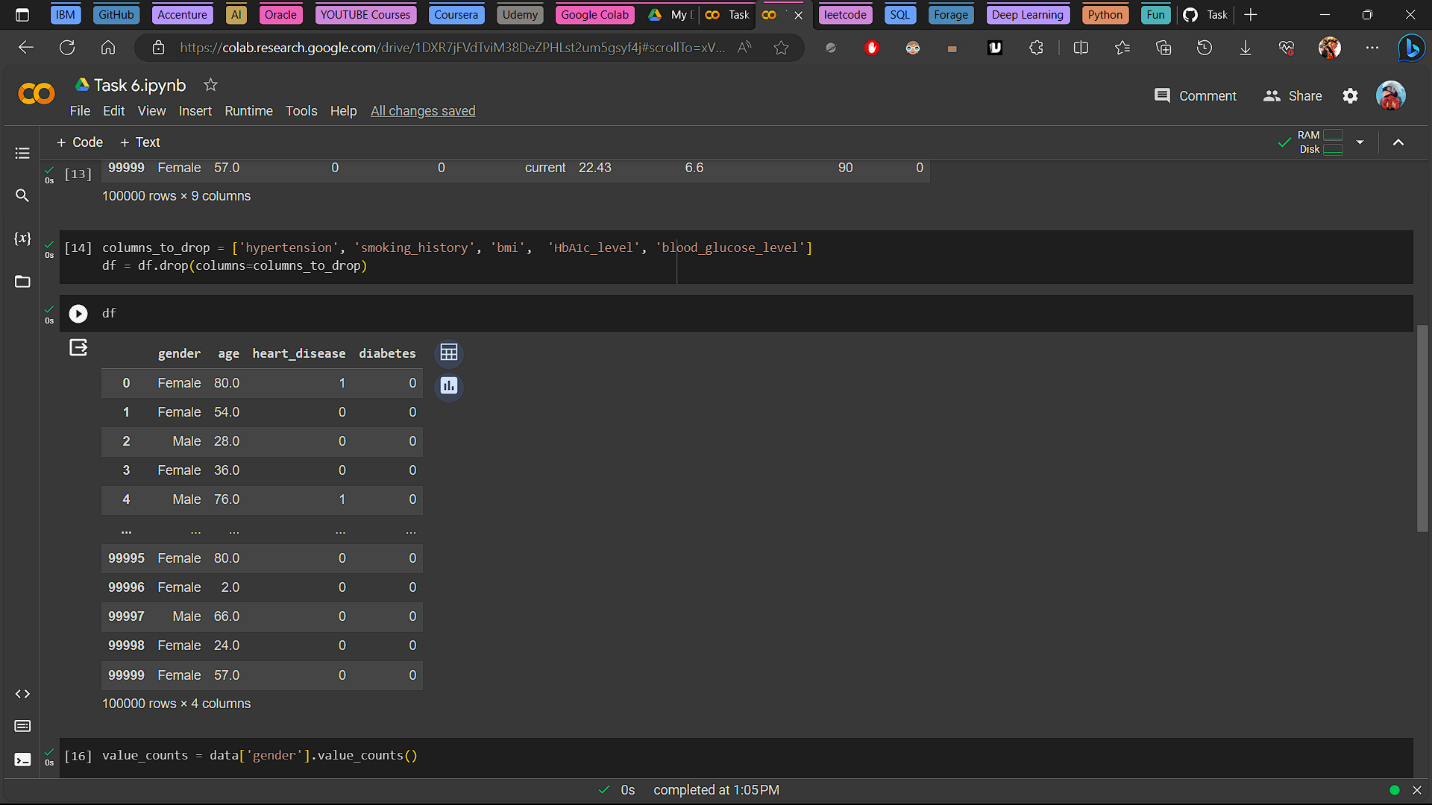
print(value\_counts)

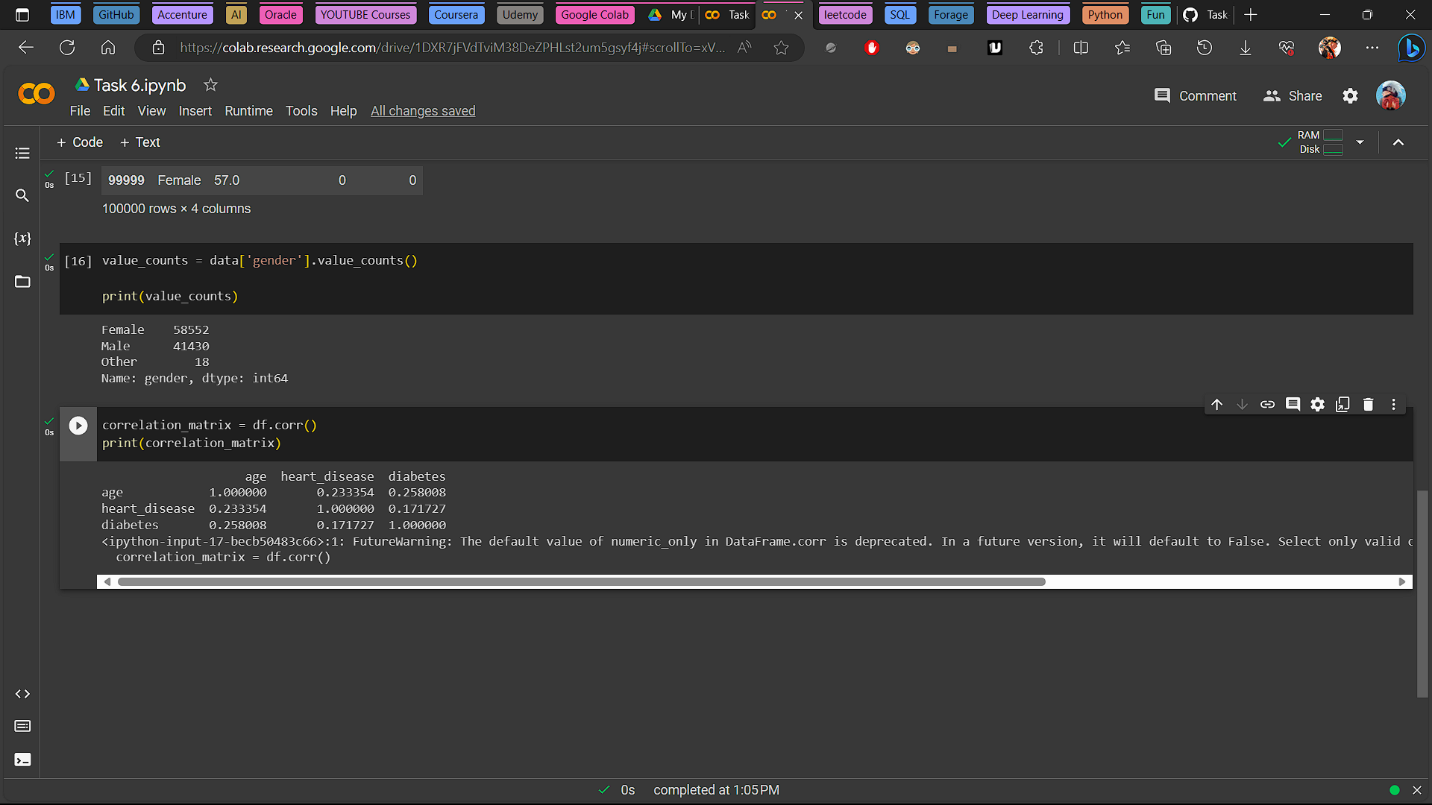
correlation\_matrix = df.corr()

print(correlation\_matrix)









**Data Analysis Report**

In this Analysis we have gone through the data and understand the nature of the data. And made a GOAL to show the analysis on the increase risk of Diabetes and Heart Disease in the data.

To Achieve that I dropped some columns from the data which are 'hypertension', 'smoking\_history', 'bmi',  'HbA1c\_level', 'blood\_glucose\_level' from the dataset to concentrate more on the ‘Age’, ‘Heart disease’ and ’Diabetes’ in the people.

We have used data.Value\_counts() function to get the following data analysis.

**Gender Distribution:**

Female: 58,552

Male: 41,430

Other: 18

**Observations**

The majority of individuals in the dataset identify as female, with a count of 58,552.

The dataset also includes a significant number of male individuals, with a count of 41,430.

A very small number of individuals are classified as 'Other,' with a count of 18.

**Correlation Analysis Report**

Correlation Among Age, Heart Disease, and Diabetes

**Summary**

In this analysis, we calculated the Pearson correlation matrix for the dataset among the 'age,' 'heart\_disease,' and 'diabetes' variables from the dataset.

**Correlation Matrix:**

**Age Heart\_disease Diabetes**

**Age**  1.000000 0.233354 0.258008

**Heart\_disease** 0.233354 1.000000 0.171727

**Diabetes**  0.258008 0.171727 1.000000

**Observations**

**Age and Heart Disease:** There is a positive correlation of approximately 0.23 between age and the presence of heart disease. This suggests that as individuals age, there is a moderate increase in the likelihood of having heart disease.

**Age and Diabetes:** Age is positively correlated with diabetes (correlation coefficient ≈ 0.26). This indicates that as individuals age, there is a moderate increase in the likelihood of having diabetes.

**Heart Disease and Diabetes:** Heart disease and diabetes also exhibit a positive correlation (correlation coefficient ≈ 0.17), indicating a moderate relationship between these two health conditions.