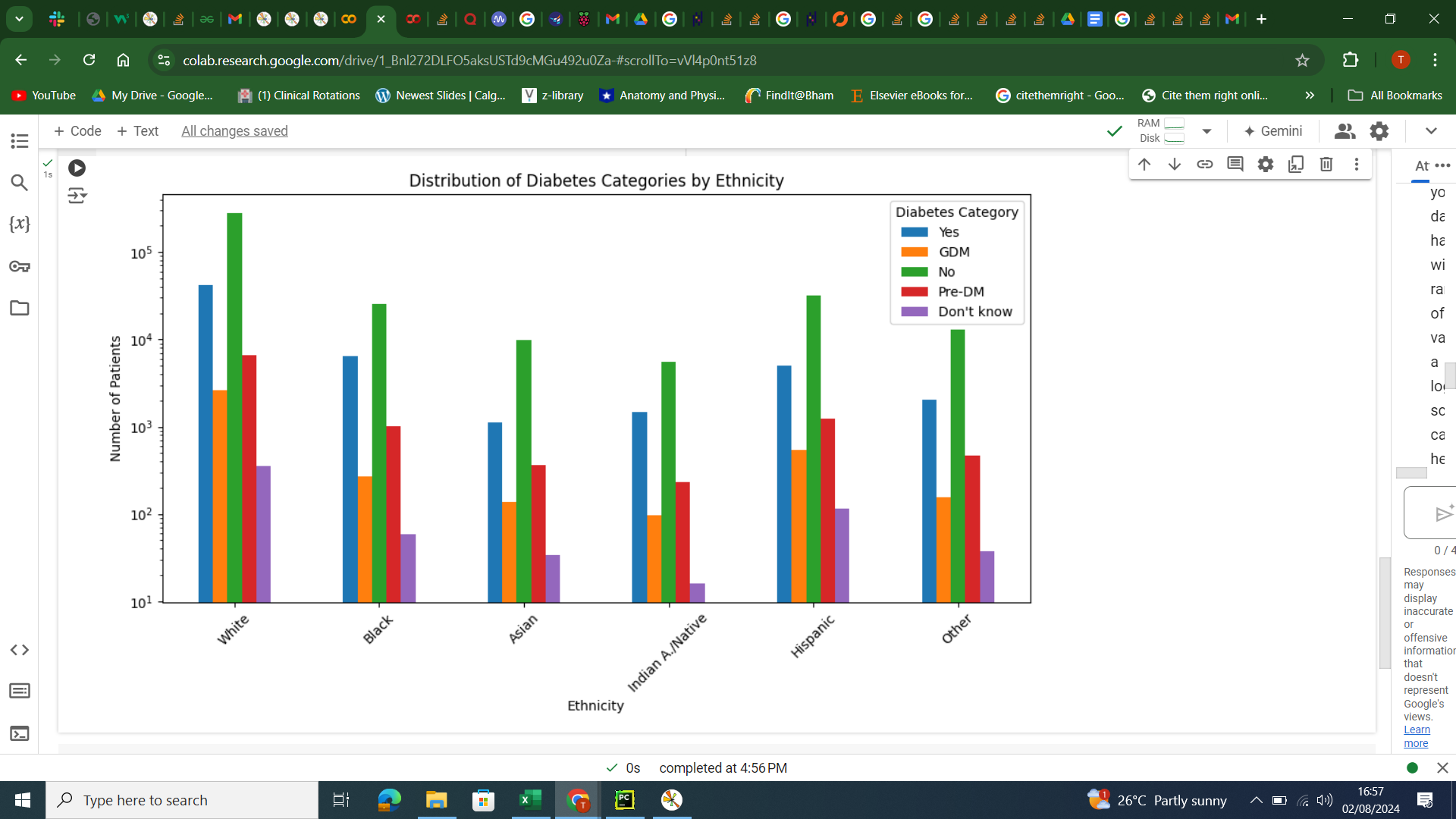
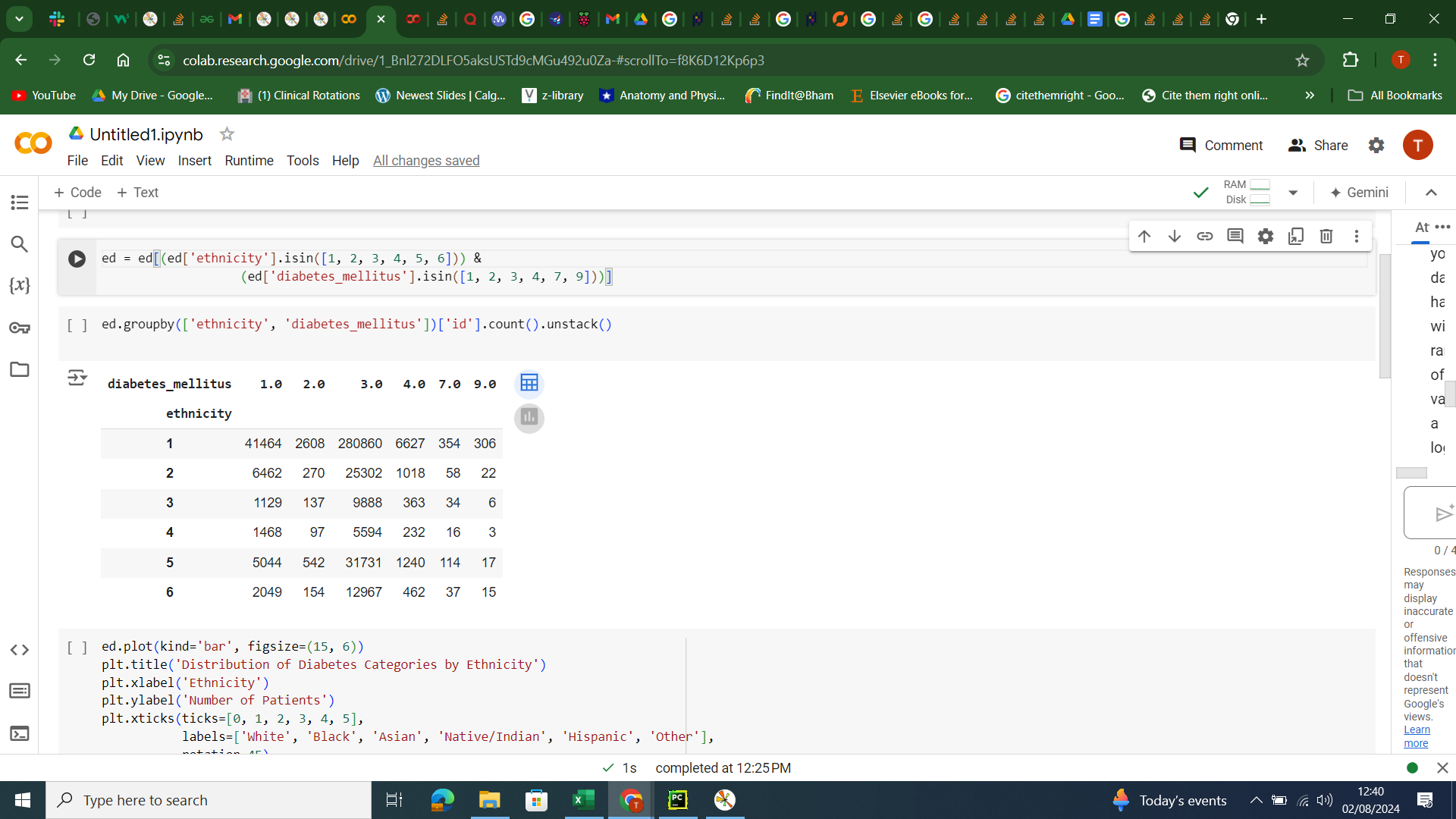
## **DATA VISUALISATION**

### 

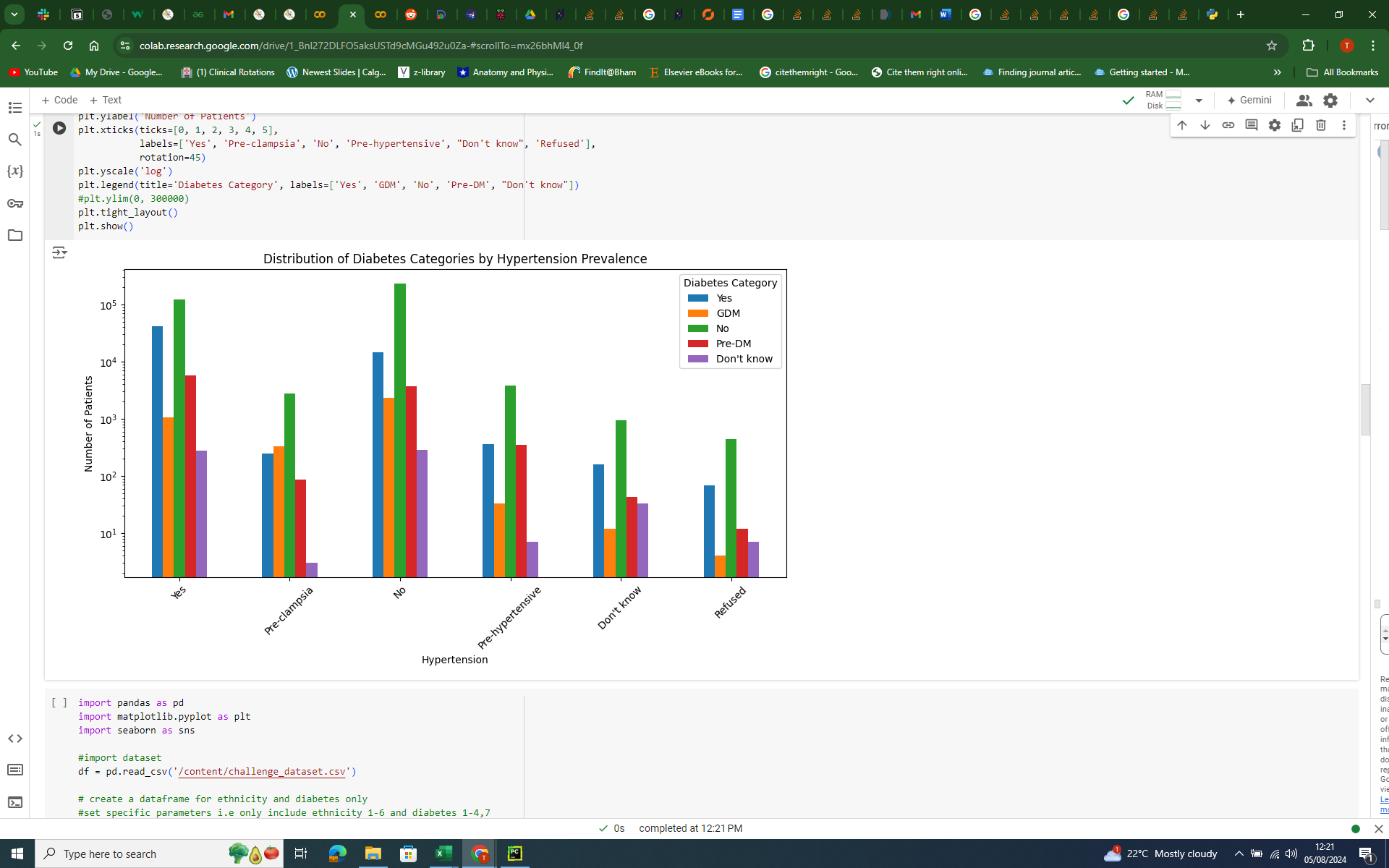
### **Ethnicity & Diabetes graph -**



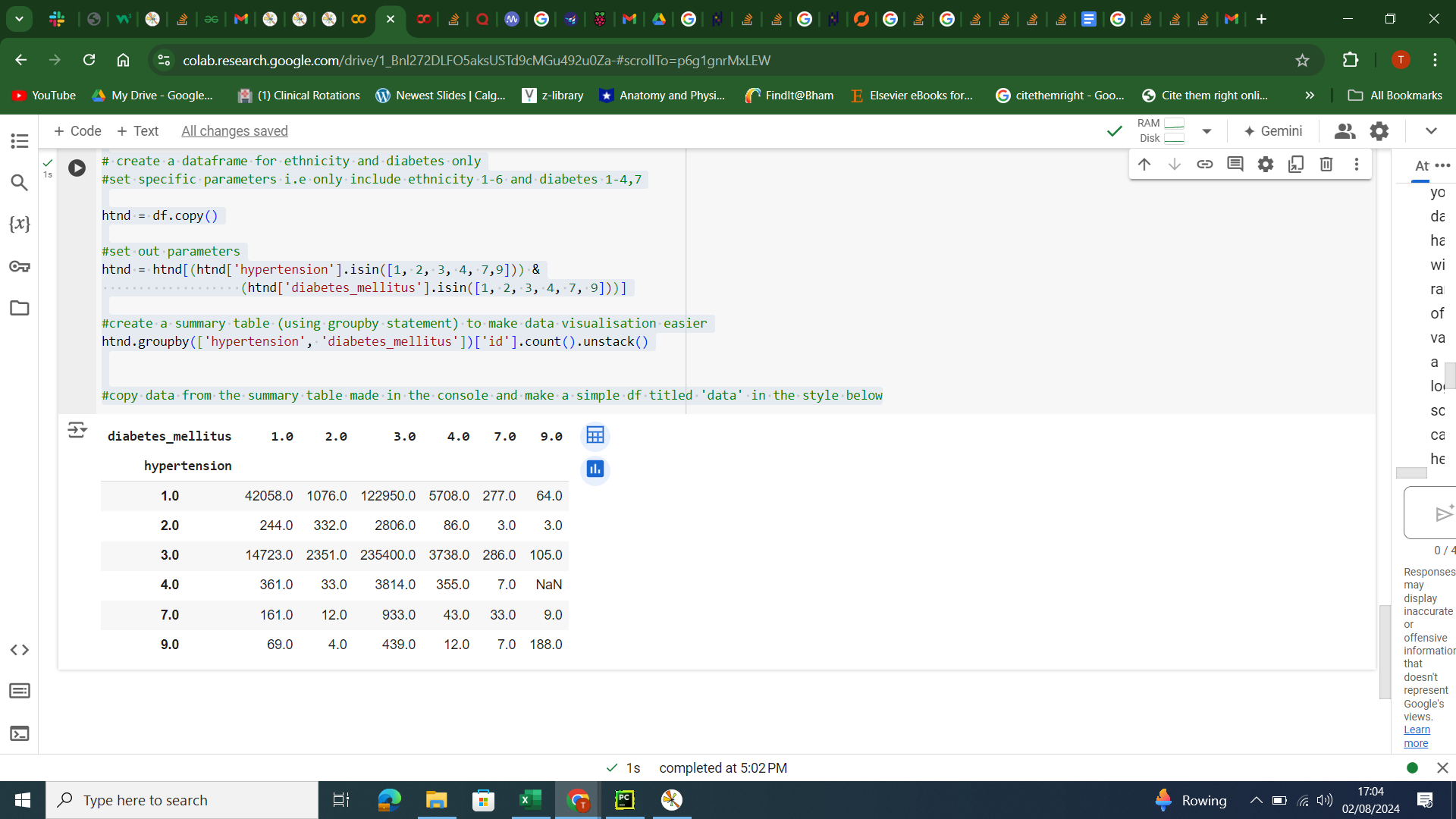
### **Summary table of Ethnicity raw data -**



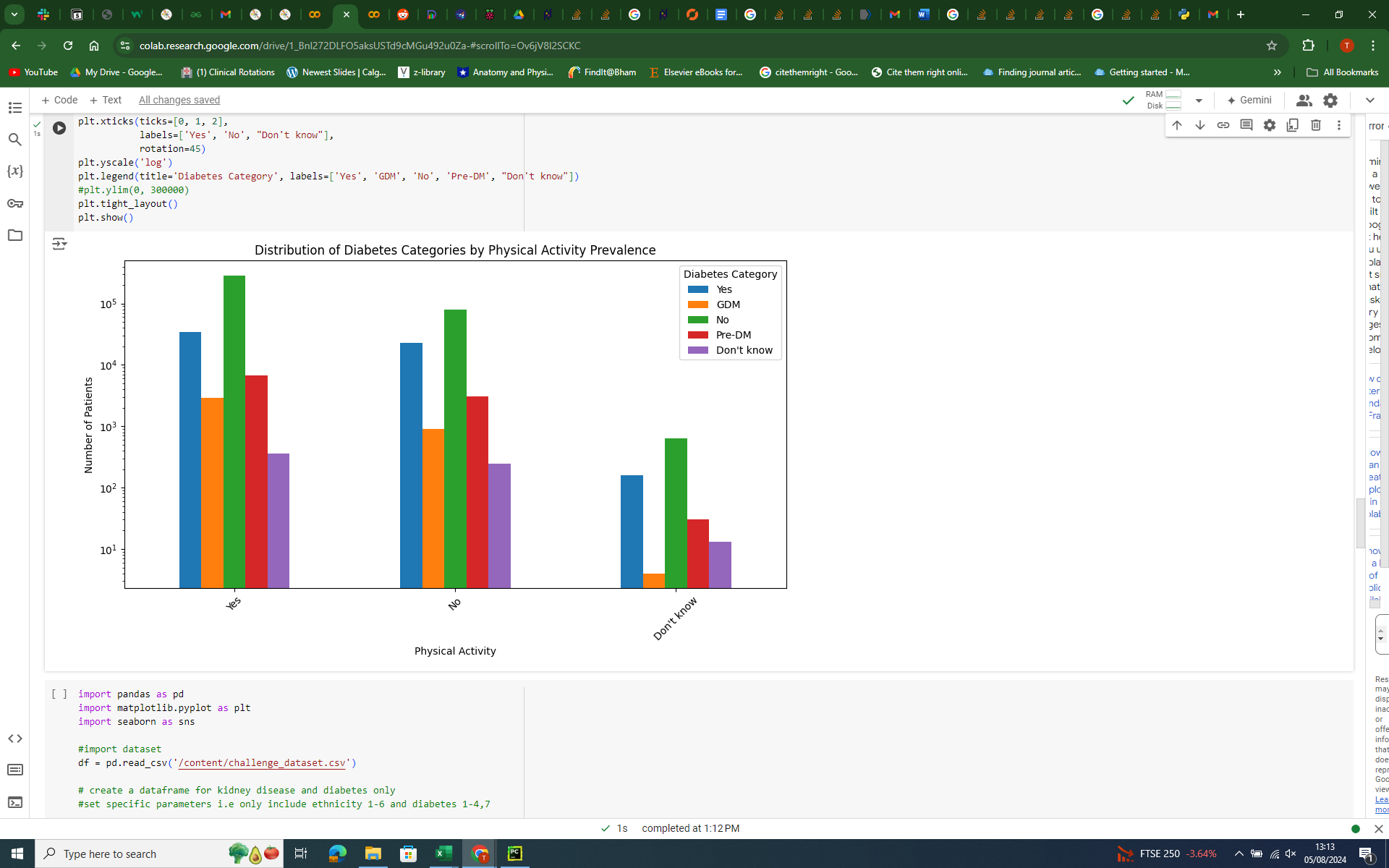
### **HTN & Diabetes graph -**



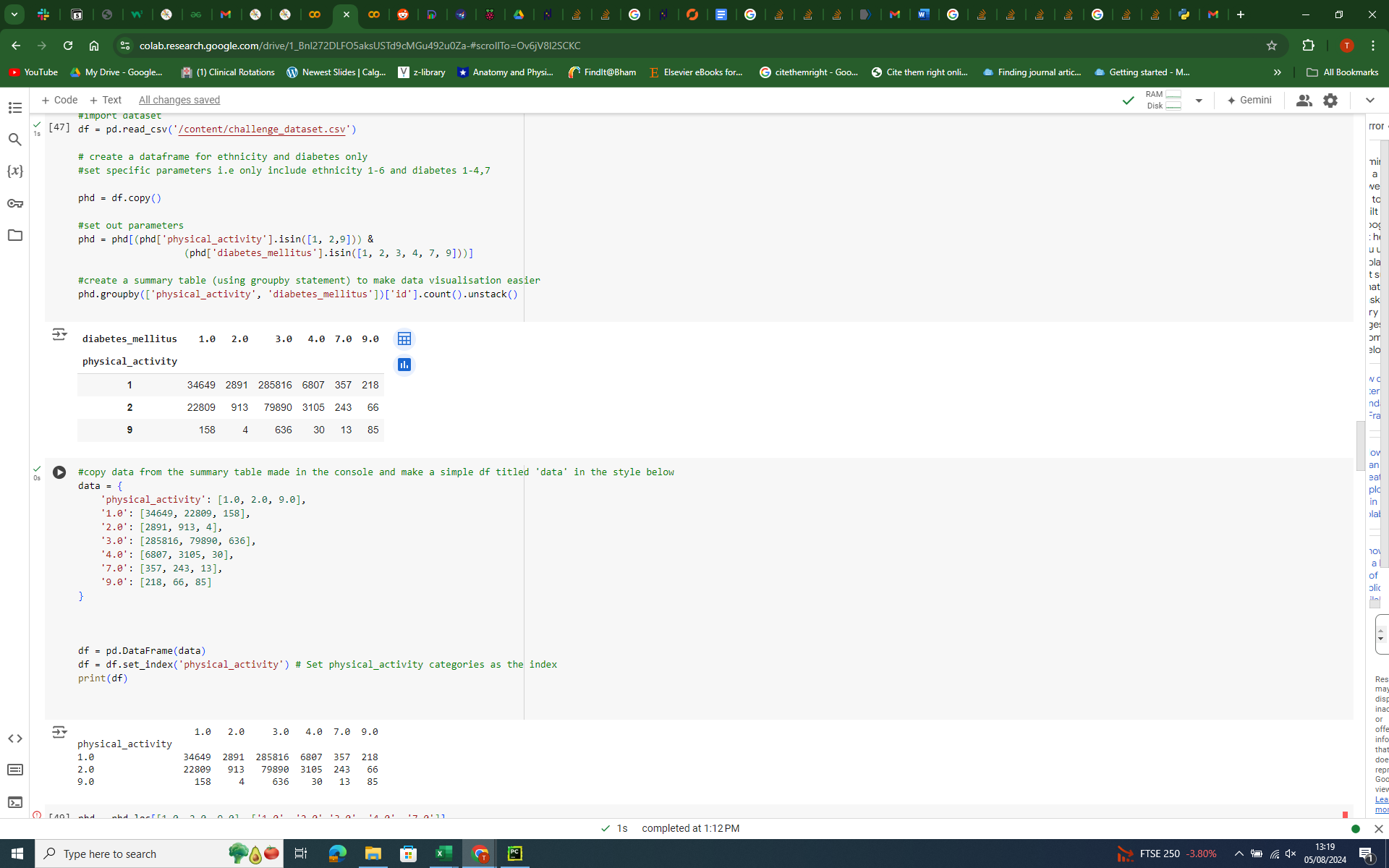
### **Summary table of HTN raw data -**



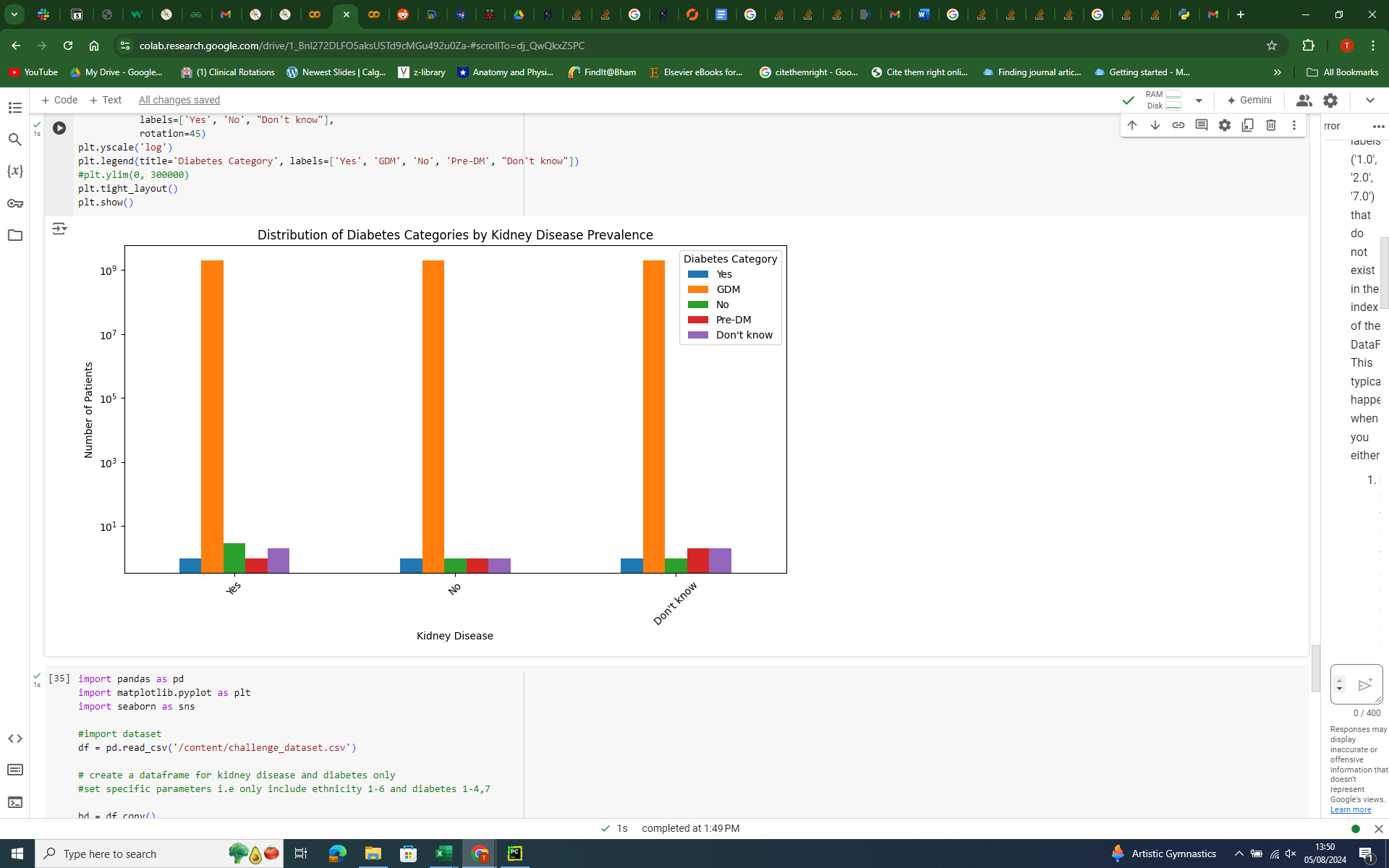
### **Physical Activity & Diabetes graph -**



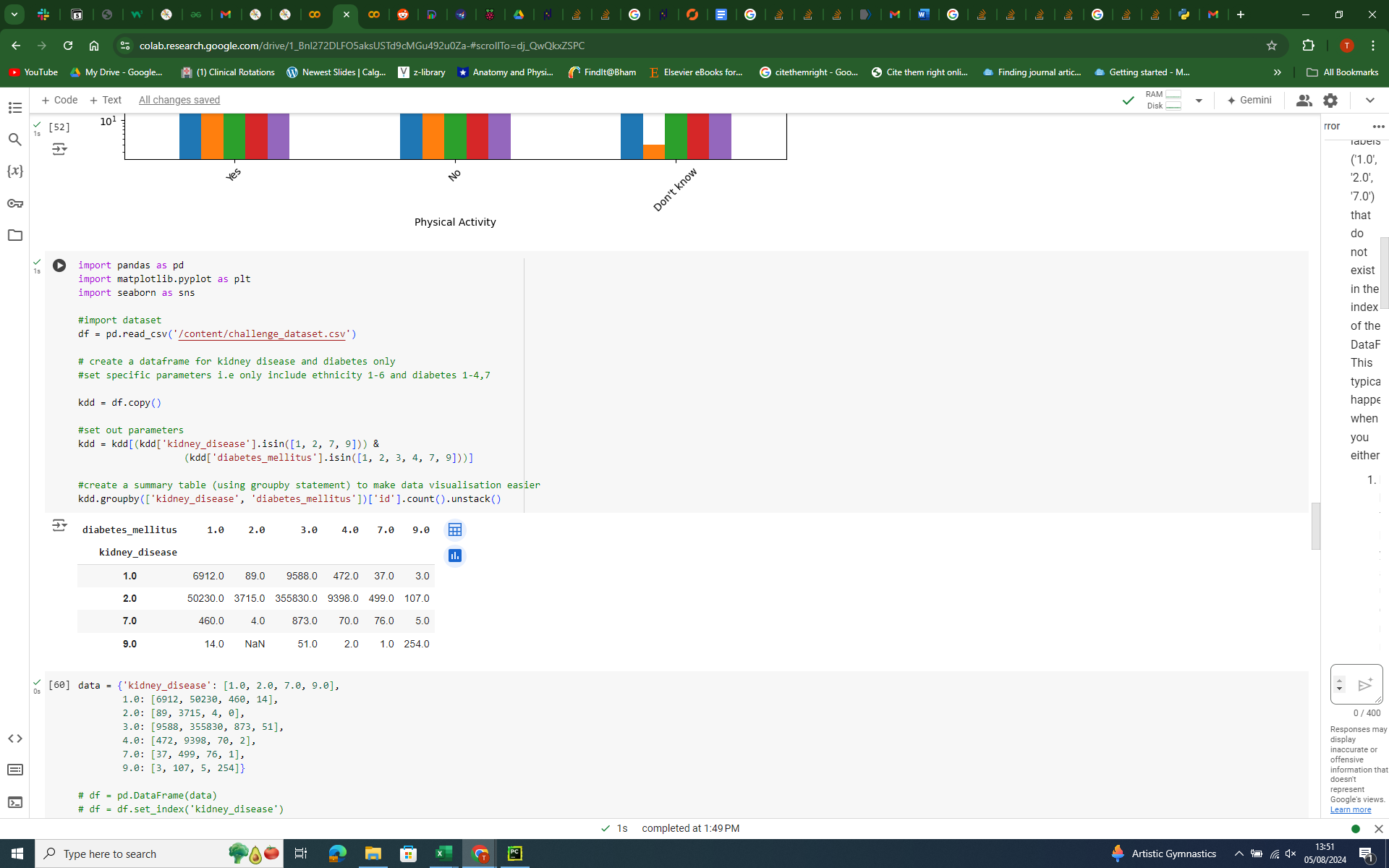
### **Summary table of physical activity raw data -**



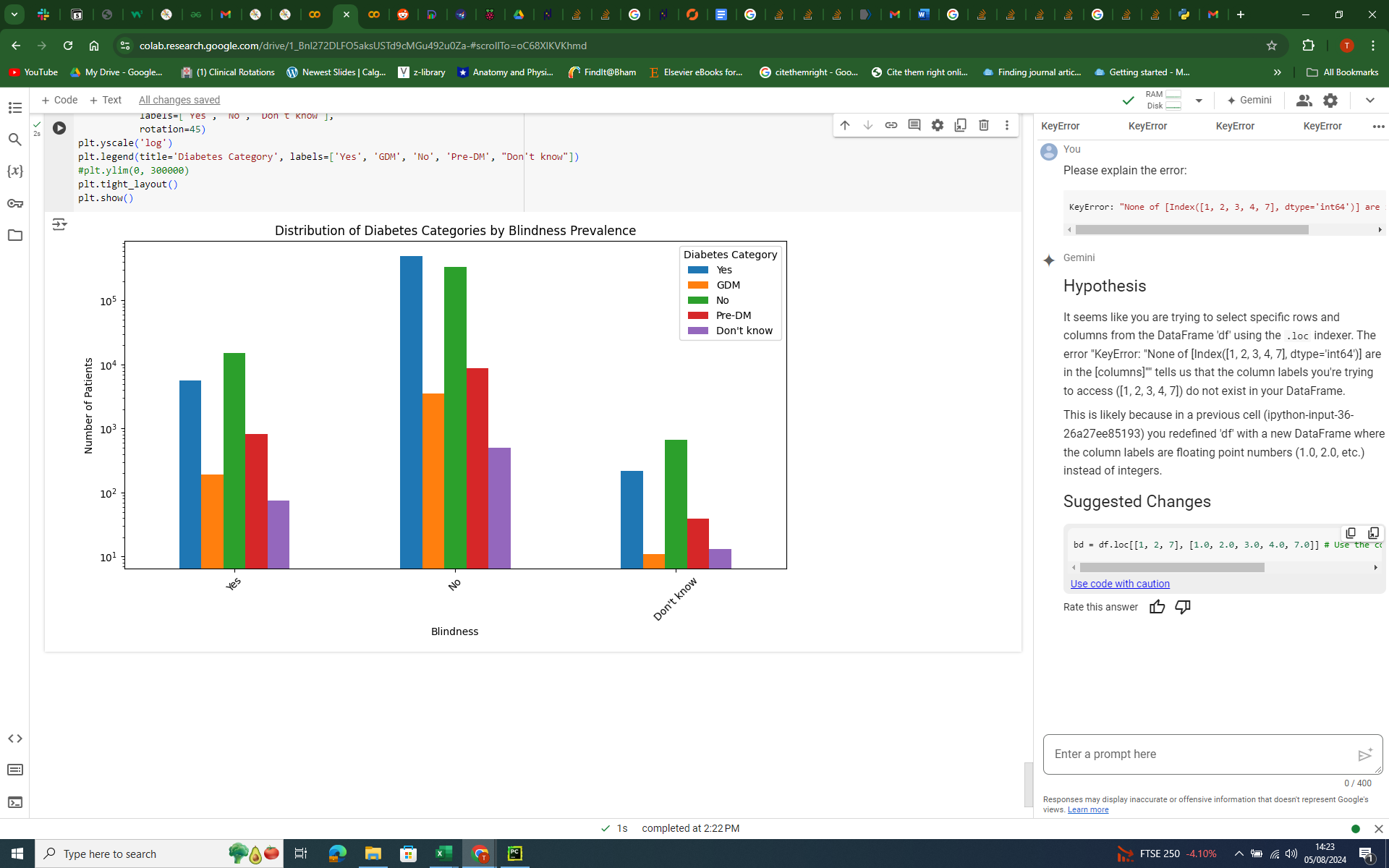
### **Kidney Disease & Diabetes graph -**



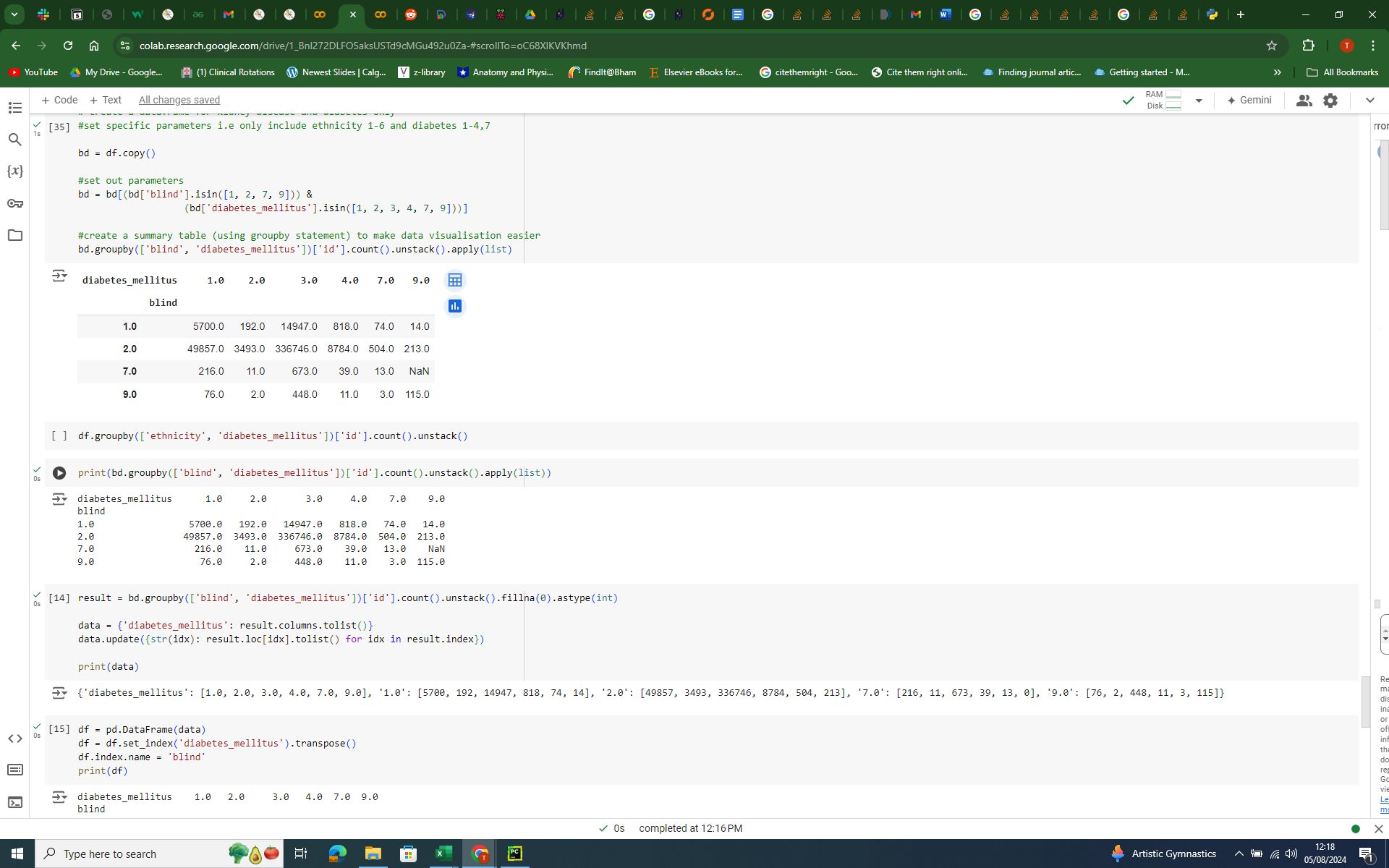
### **Summary table of Kidney disease raw data -**



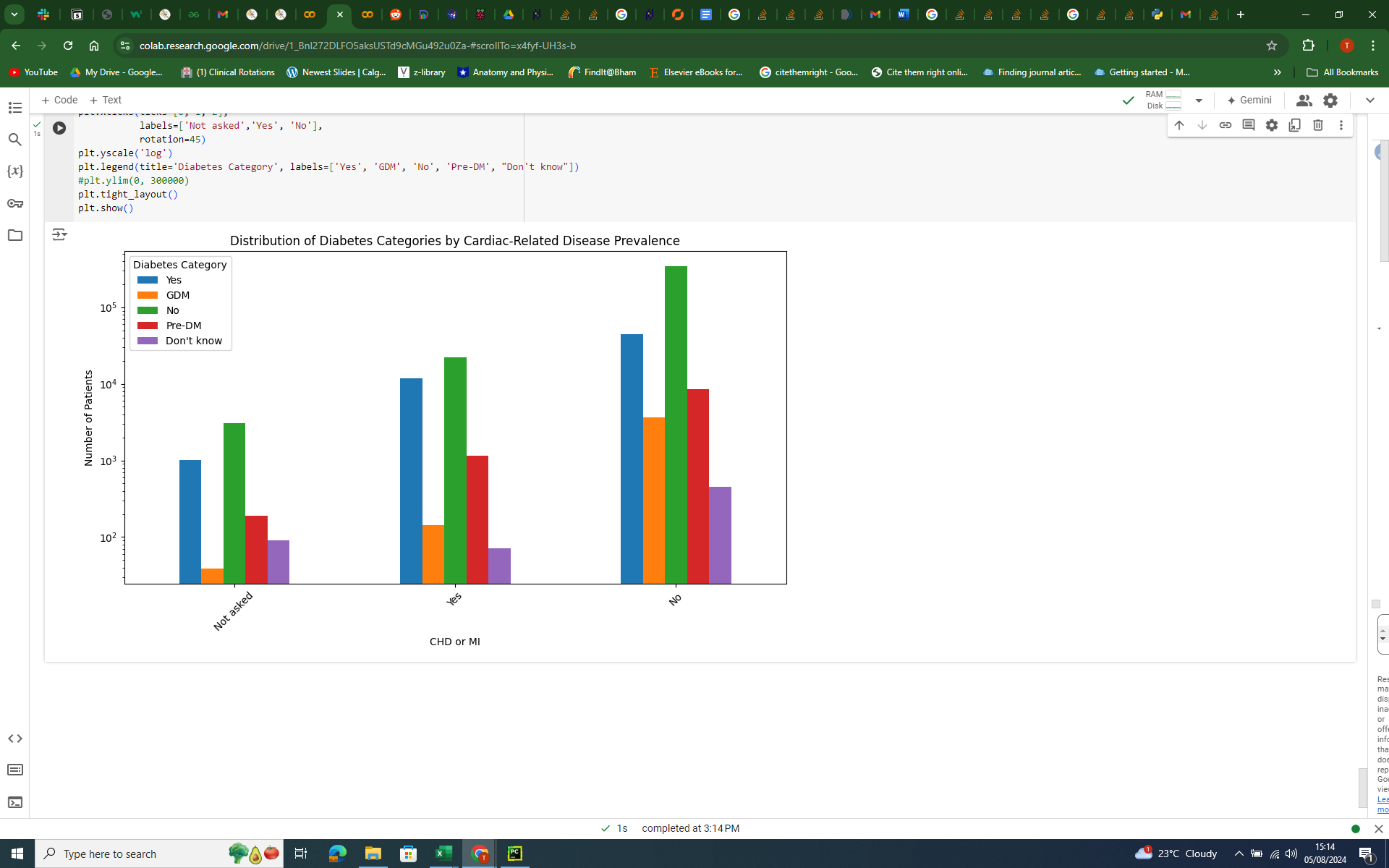
### **Blindness & Diabetes graph -**



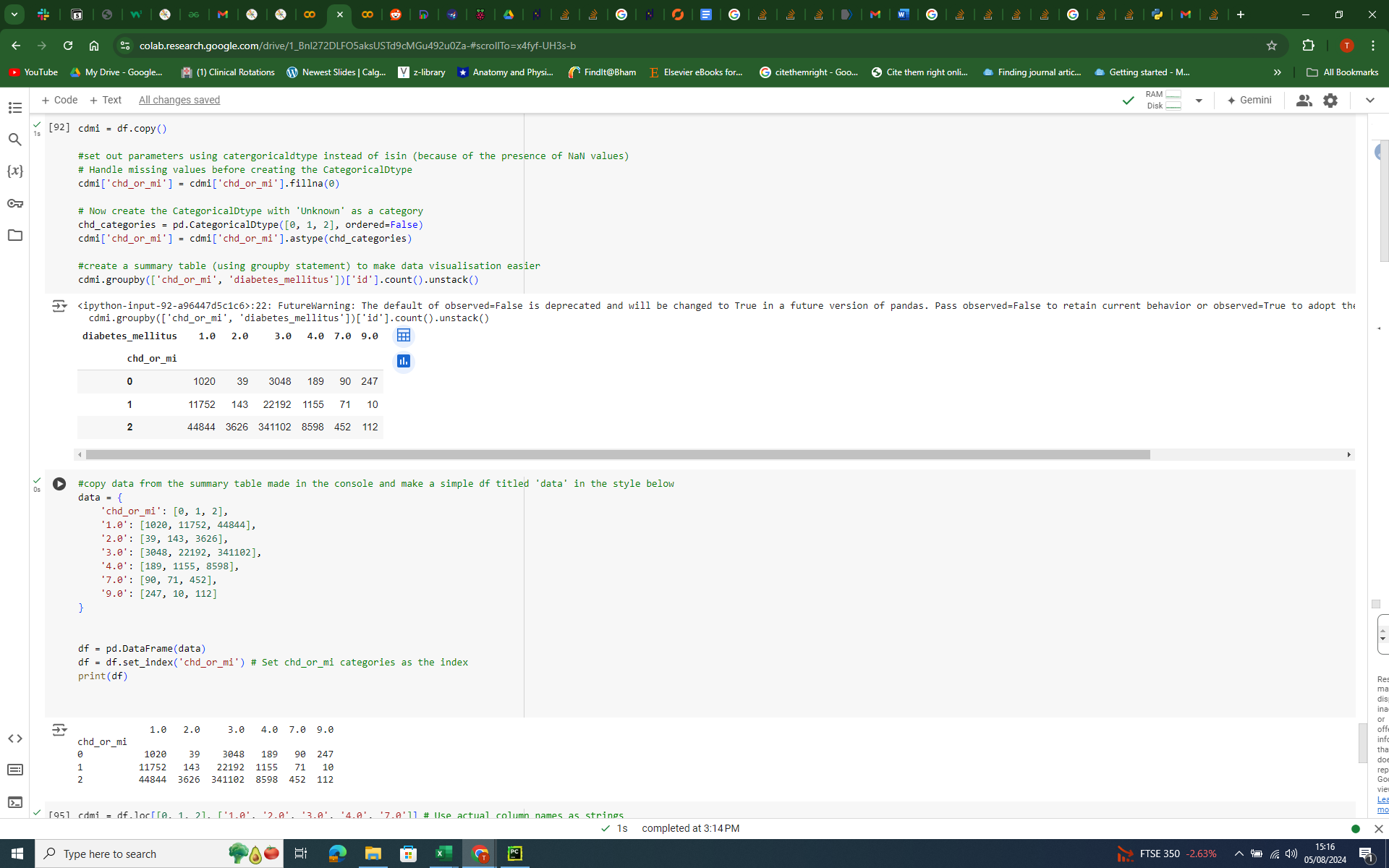
### **Summary table of blindness raw data -**



### **CHD/MI & Diabetes graph -**



### **Summary table of CHD/MI raw data -**



Further considerations

* May have to make scatter plots to describe relationships

## **Linear Regression**

* Purpose - to calculate risk?? of developing a certain comorbidity based on prevalence of the comorbidity amongst diabetic patients
* Prediction question -
  + how accurately can we predict the likelihood of developing ‘x’ comorbidity given you have ‘y’ (diabetes)
* Inferential question -
  + how accurately can we estimate the effect of diabetes ‘x’ on the different comorbidities
* Assess accuracy

## **Logistic Regression**