# **Chronic kidney disease**

**Hussein adel Kamal 42010102**

**Mina timothaws 42010626**

**Mina yousef hakim 42010716**

**Salma amr. 42010025**

**Ahmed sobhy. 41810141**

**Mahmoud Saad. 41920014**

Chronic kidney disease   
This dataset is talking about chronic kidney disease and it contains 25 columns and 400 rows   
Problems   
There are many problems in this dataset   
-missing values   
-change from categorical to numerical   
-checking for the outliers   
-change the datatypes to the datatypes that we need it   
And we need to visualize it by asking some questions about age and about the attributes of the dataset and what is the relation of the attributes with the ckd   
And then we make the machine learning and we used the logistic regression method (classification) because we have to predict 0 or 1 (0 for non ckd and 1 for ckd )

The attributes   
• age - age   
  
• bp - blood pressure   
  
• sg - specific gravity   
  
• al - albumin   
  
• su - sugar   
  
• rbc - red blood cells   
  
• pc - pus cell   
  
• pcc - pus cell clumps   
  
• ba - bacteria   
  
• bgr - blood glucose random   
  
• bu - blood urea   
  
• sc - serum creatinine   
  
• sod - sodium   
  
• pot - potassium   
  
• hemo - hemoglobin   
  
• pcv - packed cell volume   
  
• wc - white blood cell count   
  
• rc - red blood cell count   
  
• htn - hypertension   
  
• dm - diabetes mellitus   
  
• cad - coronary artery disease   
  
• appet - appetite   
  
• pe - pedal edema   
  
• ane - anemia   
  
• class - class  
  
0 for normal

1 for abnormal

0 for no

1 for yes

In class 0 for non ckd and 1 for ckd

Every attribute of this attributes we will discuss it and its relation with ckd and the distribution of it with and without ckd .

And finally we made a dashboard by powerbi