Task 2

Problem Statement 1&2: (Data Analysis)

Problem Statement 3: (Machine Learning)

Your task is to help John from STA IT services to create a linear regression model in Python or R to predict the CTC/Salary of new hires from the data provided. Here is a snapshot of the case

John

HR Analytics Manager at STA IT services

STA hires hundreds of IT executives and managers for their consulting business. John's team invest a large amount of time and effort in analyze each and every resume to calculate the joining CTC/salary for new hires. John has recently completed a course in Analytics and Machine Learning and he thinks that STA can use Machine Learning to use past data to create data driven process for setting of CTC/salary for new hires

Your task is to

1. Understand the data
2. Transform categorical variables to the numerical variables
3. Apply linear regression to predict the CTC
4. Evaluate the results of the analysis

Excel data files are provided to you to help you with the assignment: Data_file.xlsx

Data_file is a dataset that you will use to train your model.

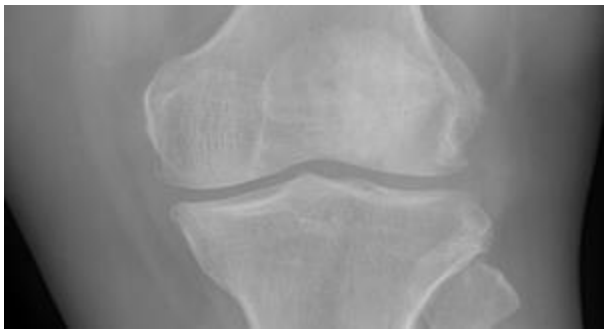
Test Data files:Test_data_file.xlsx

Test_Data_file is a dataset that you will use to test your model and find out the Mean Square Value.

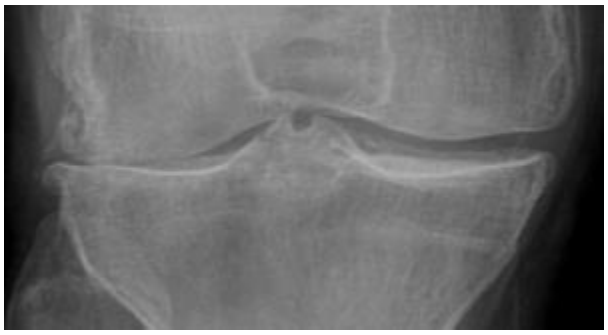
Problem Statement 4: (Machine Learning)

An orthopaedist is a medical doctor specializing in diagnosing and treating disorders related to the skeletal system. Part of their job is to distinguish between a healthy person and a person with Osteoarthritis by looking at their knee X-ray images.

Osteoarthritis: Osteoarthritis, commonly known as wear-and-tear arthritis, is a condition in which the natural cushioning between joints — cartilage — wears away. When this happens, the bones of the joints rub more closely against one another with less of the shock-absorbing benefits of cartilage. The rubbing results in pain, swelling, stiffness, decreased ability to move, and, sometimes, the formation of bone spurs.



Healthy Knee Joint



Osteoarthritis Condition

You have to create a deep learning model that can detect if osteoarthritis is present or not in a given knee X-ray image.

Dataset

The Dataset contains three folders

- Test (845 images)
- Train (2350 images)
- Valid (641 images)

The train is a dataset that you will use to train your model.

The test is a dataset that you will use to test your model and find out the accuracy and confusion matrix.

Each of these folders has two folders

- Test
  - Normal
  - Osteoarthritis
- Train
  - Normal
  - Osteoarthritis
- Valid
  - Normal
  - Osteoarthrit**is**

Task

Your task is to create a deep learning model that can provide the probability of having osteoarthritis for a given knee X-ray image.

Dataset:
https://drive.google.com/drive/folders/1y7AO4RHDLYYvjGMlZX2NBTpSFWho1Jyg?usp=sharing

All the best.