# Topic: K-Nearest Neighbor

**Instructions**

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name: Batch Id:**  **Topic: K-Nearest Neighbor**

1. **Business Problem**
   1. **Objective**
   2. **Constraints (if any)**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its Data type and its relevance to the model building, if not relevant provide reasons and provide description of the feature.**

**Using R and Python codes perform:**

1. **Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Univariate analysis**
   3. **Bivariate analysis**
2. **Model Building**
   1. **Build the model on the scaled data (try multiple options)**
   2. **Perform KNN, and use cross validation techniques to get N-neighbors**
   3. **Train and Test the data and perform cross validation techniques, compare accuracies, precision and recall and explain about them.**
   4. **Briefly explain the model output in the documentation.**

1. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**

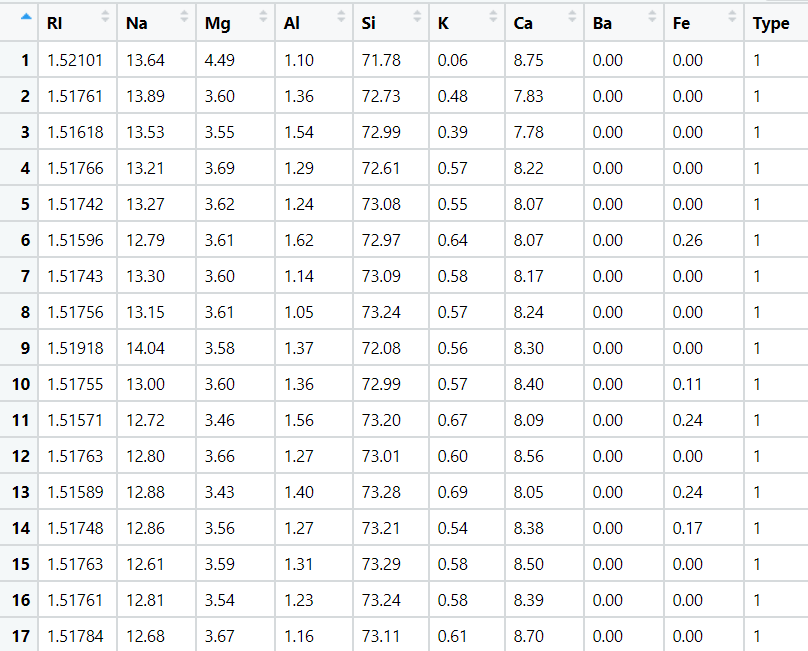
# Note:

The assignment should be submitted in the following format:

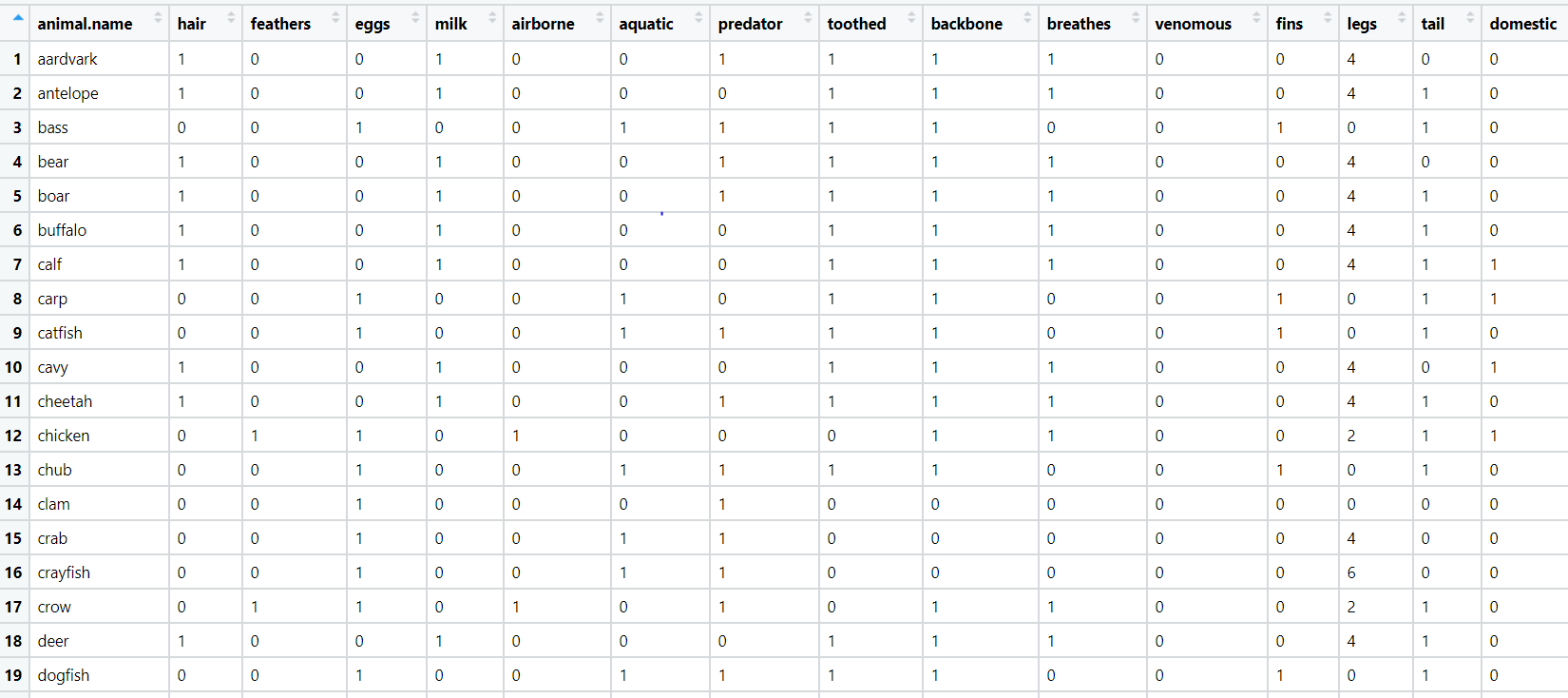
* R code
* Python code
* Code Modularization should be maintained
* Documentation of the model building (elaborating on steps mentioned above)

**Problem Statement: -**

A glass manufacturing plant, uses different Earth elements to design a new glass based on customer requirements for that they would like to automate the process of classification as it’s a tedious job to manually classify it, help the company reach its objective by correctly classifying the Earth elements, by using KNN Algorithm



Problem Statement: -

A National Park, in India is dealing with a problem of segregation of its species based on the different attributes it has so that they can have cluster of species together rather than manually classify them, they have taken painstakingly collected the data and would like you to help them out with a classification model for their business objective to be achieved, by using KNN Algorithm classify the different species and draft your inferences in the documentation.