1. Download the `faulty_steel_plate.csv` file from the class website. This file contains 34 columns with the first 27 columns being the input features measuring various fault characteristics. The last seven columns are fault types. The first six of these seven appear to be one-hot encoded whereas the last column (column 34) is the "Other Faults" type.

   We wish to use the last column as the response variable given the first 27 columns as the input features. Since there are two classes for this variable, we can treat this as a binary classification problem.

   For this mini-project, perform the following tasks using Python:

   - Analyze the data. Plot the number of each fault type using bar charts.
   - Remove the columns that may not be relevant for predictive purposes.
   - Develop a knn model for predicting the response variable `Other Faults`. Consider a range of values for k and determine the optimal value.
   - Consider both Manhattan and Euclidean distance measures. (Do you know why Manhattan Distance is called Manhattan Distance?)
   - Provide plots of accuracy versus k and confusion matrices for each case.
   - Conduct a feature importance study and extract the most important features (if any). This topic will be discussed in the class, this Tuesday.

   For some guidance on how to analyze the data, please see `https://www.kaggle.com/krishnamsheth31/faulty-steel-plate-classification/notebook`. Sriganesh will provide guidance on how to submit your results through the class discussion forum.