DATASET LINK:

https://www.kaggle.com/uciml/mushroom-classification

A. Describe the dataset in your own words.

**Objective**

Classify whether if a mushroom is safe to eat or poisonous?

**Overview**

23 species of gilled mushrooms from the Agaricus and Lepiota family are observed in the dataset. The dataset contains 23 features of mushroom and the first feature is whether the mushroom is edible or poisonous.

**Attributes and observations**

The dataset contains 23 features and 8124 observations, all are categorical (factor) data.

**Missing values**

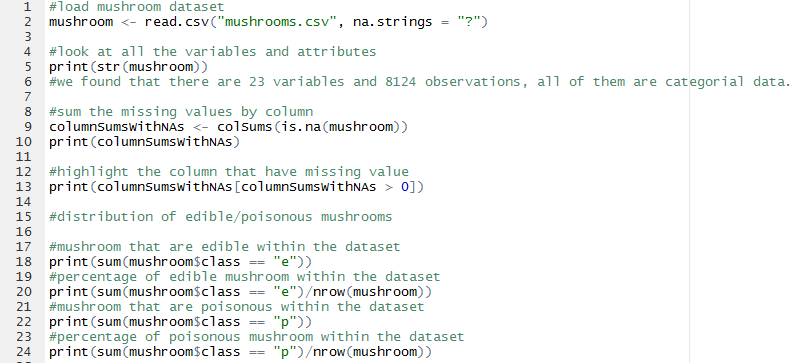
There are 2480 missing values in the dataset, where the missing values is denoted as “?” and are found in “stalk-root”, the 11th column.

**Distribution of edible/poisonous mushroom**

-- edible : 4208 observations (51.8%)

-- poisonous : 3916 observations (48.2%)

-- total : 8124 observations (100.0%)



B. What possible insights can be obtained from mining the chosen dataset?

To obtain common features that are found in edible and poisonous mushrooms for building a prediction model.