

Date : 08/28/2018

Make sure to set header=F since data file don't have column names

We can use `names` function to assign column names First column is the class label. Let's call it `Edibility`

3] Check Mushroom dataframe dimension. Rows and Columns

```
## [1] 8124 23
```

Mushroom dataframe has 8124 rows and 23 columns

```
##
##      e      p
## 4208 3916
```

Out of 8124 samples 4208 samples are edible and 3916 samples are poisonous

5] Subset columns

****Let's take first four attribute along with Edibility class variable**

```
Mushroom.Sub = subset(Mushroom.Data, select=c("Edibility", "CapShape", "CapSurface", "CapColor", "Bruises"))
```

6] Check dimensions of a subset dataframe

```
dim(Mushroom.Sub)
```

```
## [1] 8124    5
```

Subset has 8124 rows and 5 columns. First column is a class variable whihc is Edibility

7] Preview first 10 rows of subset

```
head(Mushroom.Sub,10)
```

```
##      Edibility CapShape CapSurface CapColor Bruises
## 1          p         x           s         n         t
## 2          e         x           s         y         t
## 3          e         b           s         w         t
## 4          p         x           y         w         t
## 5          e         x           s         g         f
## 6          e         x           y         y         t
## 7          e         b           s         w         t
## 8          e         b           y         w         t
## 9          p         x           y         w         t
## 10         e         b           s         y         t
```

8] Preview last 10 rows of subset

```
tail(Mushroom.Sub,10)
```

```
##      Edibility CapShape CapSurface CapColor Bruises
## 8115         p         f           y         c         f
## 8116         e         x           s         n         f
## 8117         p         k           y         n         f
## 8118         p         k           s         e         f
## 8119         p         k           y         n         f
## 8120         e         k           s         n         f
## 8121         e         x           s         n         f
## 8122         e         f           s         n         f
## 8123         p         k           y         n         f
## 8124         e         x           s         n         f
```

9] Replace abbreviations

Replace e with Edible and p with Poisonous in class variable

```
Mushroom.Sub$Edibility[Mushroom.Sub$Edibility=="e"] = "Edible"
Mushroom.Sub$Edibility[Mushroom.Sub$Edibility=="p"] = "Poisonous"
```

10] Check unique values of Class variable Edibility

```
unique(Mushroom.Sub$Edibility)
```

```
## [1] "Poisonous" "Edible"
```

We can see that abbreviations are successfully replaced