Data-607 Week-1 Assignment

Student Name: Sachid Deshmukh

Date: 08/28/2018

- GitHub Location for rmd file
- GitHub Location for pdf file
- RPub location of published file

1] Read Mushroom data from csv file

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

Mushroom.Data = read.csv("https://raw.githubusercontent.com/mlforsachid/MSDSQ1/master/Data607/Week1/aga

2] Assign column names

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

```
names(Mushroom.Data) = c("Edibility", "CapShape", "CapSurface", "CapColor", "Bruises", "Odor", "GrillAt
```

3] Check Mushroom dataframe dimension. Rows and Columns

```
dim(Mushroom.Data)
## [1] 8124 23
```

Mushroom dataframe has 8124 rows and 23 columns

4] Check Edibility proportion

```
table(Mushroom.Data$Edibility)
```

```
## 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

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

6 Check dimensions of a subset dataframe

```
dim(Mushroon.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(Mushroon.Sub,10)

## Edibility CapShape CapSurface CapColor Bruises
```

ππ		Laibility	capsnape	Capbullace	Capcoror	DIGIDES
##	1	р	X	s	n	t
##	2	е	X	s	У	t
##	3	е	b	S	W	t
##	4	р	X	У	W	t
##	5	е	X	S	g	f
##	6	е	X	У	У	t
##	7	е	b	S	W	t
##	8	е	b	У	W	t
##	9	р	X	У	W	t
##	10	е	b	s	У	t

8] Preview last 10 rows of subset

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

```
Edibility CapShape CapSurface CapColor Bruises
##
## 8115
                p
                                      у
## 8116
                                                        f
## 8117
                          k
                р
                                      у
## 8118
                р
## 8119
                                                        f
                          k
                p
                                      У
                                               n
## 8120
                          k
                                                        f
## 8121
                е
                                      s
                                                        f
                          X
                                               n
## 8122
                                                        f
                          f
## 8123
                          k
                                                        f
                                               n
                p
                                      У
## 8124
```

9] Replace abbreviations

Replace e with Edible and p with Poisonous in class variable

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

10] Check unique values of Class variable Edibility

unique(Mushroon.Sub\$Edibility)

[1] "Poisonous" "Edible"

We cas see that abbreaviatoins are successfully replaced