Analytics of Business Intelligence Fall 2018, Practice Set # 3

Use the table below named *titanic*, to answer the questions from #1-5.

Survived	Pclass	Name	Sex	Age
0	3	Mr. Owen Harris Braund	male	22
1	1	Mrs. John Bradley (Florence Briggs Thayer) Cumings	female	38
1	3	Miss. Laina Heikkinen	female	26
1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35
0	3	Mr. William Henry Allen	male	35
0	3	Mr. James Moran	male	27
0	1	Mr. Timothy J McCarthy	male	54
0	3	Master. Gosta Leonard Palsson	male	2
1	3	Mrs. Oscar W (Elisabeth Vilhelmina Berg) Johnson	female	27
1	2	Mrs. Nicholas (Adele Achem) Nasser	female	14
1	3	Miss. Marguerite Rut Sandstrom	female	4
1	1	Miss. Elizabeth Bonnell	female	58

1. **(Code) Convert the Survived, Pclass and Sex columns into categorical columns.**

2. **(Code) Create a bar graph, where the x-axis is Sex and each individual bar is divided by those that suvived and didn't survive.**

3. **(Code) Create a bar graph, where the x-axis is Pclass and each individual bar is divided by those that suvived and didn't survive. Create two subgraphs side by side, one for males and the other for females.**

4. **(Code) Change the histogram below into bars of 2.**

```
ggplot(titanic, aes(x=Age)) + geom_histogram(binwidth = 5) + theme_bw() +
labs(y= "Passenger Count",x="Age (5)", title = "Titanice Survival Rates")
```

5. **You have training data with 10 rows. You want to build a decision tree model that classifies what kind of fruit you have based on the attributes. The target column has 2 pineapples, 4 oranges, 3 plums, and 1 cherry. What is the initial entropy score of your data set?**

6. **Given the code below, what is a?**

```
z = 3:1
a=0

for(i in z)
{
  a=a-i
}
```

7. **Given the code below, what is b?**

```
x = 1:4
b=c(2,4,6,8)

for(i in x)
{
  if((i/2)>3)
  {
    b[NROW(b)]=i/2
  }
  else
  {
    b[(i/2)+2*i]=i/2
  }

}
```

Use the table *mdf* below to answer questions from #8-13.

Ticker	Date	Total Assets	Total Debt	Price
FB	201703	150	30	4
FB	201706	100	25	1
FB	201709	500	100	3
FB	201712	100	20	5
AAPL	201703	200	100	2
AAPL	201706	150	30	4
AAPL	201709	300	15	6
AAPL	201712	100	10	8
TSLA	201703	250	50	4
TSLA	201706	200	40	10
TSLA	201709	350	70	6
TSLA	201712	150	50	5

8. Given the code below, what is tt?

```
z = 1:4
tt=head(mdf$Price,5)
r=seq(2,10,2)

for(i in z)
{
  for(x in r[c(1,3,5)])
  {
    tt[NROW(r)]=i+x
  }
}
```

9. Given the code below, what is b?

```
v = head(mdf$Price,9)
z = tail(v,5)

b=c()
for(i in z)
{
  if((i-1)>=3 && (i-4)<=3)
  {
    b[NROW(b)+1]=i
  }
  else
  {
    b=c(b,i)
  }
}
```

10. Given the code below, what is v?

```
v = c(2,4,10,5,2,8,12)
z=seq(2,12,mdf[mdf$Ticker=='AAPL' & mdf$Date==201703,'Price'])
v[c(2:3,6)]=z[c(1,4:5)]
```

11. Given the code below, what is v?

```
stock=head(mdf,4)
stockc = as.vector(stock[,2])
ret=c()

for(i in 1:(NROW(stockc)-1))
{
  ret[i]=(stockc[i+1]/stockc[i]) - 1
}
v=ret[1]
```

12. Given the code below, what is v?

```
stock=tail(mdf,NROW(mdf)/4)
stock2=head(stock,NROW(stock)-1)
stockc = as.vector(stock2[,4])
ret=0

for(i in stockc)
{
  ret=ret+i
}
v=ret
```

13. Given the code below, what is v?

```
stock = as.vector(mdf$Price)
max_v = max(head(stock,mdf[mdf$Ticker=='TSLA' & mdf$Date==201703,'Price']))
m_loc = which.max(tail(stock,-3))
min_v = min(stock[1:m_loc])
v = (max_v/min_v)
```