TASK 1

In [5]:
```
import pandas as pd
import numpy as np
```

In [6]:
```
data = pd.read_csv('crx.data' , header = None)
```

In [7]:
```
#task 1
varnames = ['col' + str(s) for s in range(1 , 17)]
```

In [8]:
```
data.columns = varnames
```

In [9]:
```
#task 2
data.tail(10)
```

Out[9]:

	col1	col2	col3	col4	col5	col6	col7	col8	col9	col10	col11	col12	col13	col14	c
680	b	19.50	0.290	u	g	k	v	0.290	f	f	0	f	g	00280	
681	b	27.83	1.000	y	p	d	h	3.000	f	f	0	f	g	00176	
682	b	17.08	3.290	u	g	i	v	0.335	f	f	0	t	g	00140	
683	b	36.42	0.750	y	p	d	v	0.585	f	f	0	f	g	00240	
684	b	40.58	3.290	u	g	m	v	3.500	f	f	0	t	s	00400	
685	b	21.08	10.085	y	p	e	h	1.250	f	f	0	f	g	00260	
686	a	22.67	0.750	u	g	c	v	2.000	f	t	2	t	g	00200	
687	a	25.25	13.500	y	p	ff	ff	2.000	f	t	1	t	g	00200	
688	b	17.92	0.205	u	g	aa	v	0.040	f	f	0	f	g	00280	
689	b	35.00	3.375	u	g	c	h	8.290	f	f	0	t	g	00000	

In [10]:
```
# task 3
# replace ? with np.nan
data = data.replace('?' , np.nan)
```

In [11]:
```
data
```

Out[11]:

	col1	col2	col3	col4	col5	col6	col7	col8	col9	col10	col11	col12	col13	col14	cc
0	b	30.83	0.000	u	g	w	v	1.25	t	t	1	f	g	00202	
1	a	58.67	4.460	u	g	q	h	3.04	t	t	6	f	g	00043	
2	a	24.50	0.500	u	g	q	h	1.50	t	f	0	f	g	00280	
3	b	27.83	1.540	u	g	w	v	3.75	t	t	5	t	g	00100	
4	b	20.17	5.625	u	g	w	v	1.71	t	f	0	f	s	00120	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
685	b	21.08	10.085	y	p	e	h	1.25	f	f	0	f	g	00260	
686	a	22.67	0.750	u	g	c	v	2.00	f	t	2	t	g	00200	
687	a	25.25	13.500	y	p	ff	ff	2.00	f	t	1	t	g	00200	
688	b	17.92	0.205	u	g	aa	v	0.04	f	f	0	f	g	00280	
689	b	35.00	3.375	u	g	c	h	8.29	f	f	0	t	g	00000	

690 rows × 16 columns

In [12]:
```
# task 4
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 690 entries, 0 to 689
Data columns (total 16 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   col1    678 non-null    object 
 1   col2    678 non-null    object 
 2   col3    690 non-null    float64
 3   col4    684 non-null    object 
 4   col5    684 non-null    object 
 5   col6    681 non-null    object 
 6   col7    681 non-null    object 
 7   col8    690 non-null    float64
 8   col9    690 non-null    object 
 9   col10   690 non-null    object 
 10  col11   690 non-null    int64  
 11  col12   690 non-null    object 
 12  col13   690 non-null    object 
 13  col14   677 non-null    object 
 14  col15   690 non-null    int64  
 15  col16   690 non-null    object 
dtypes: float64(2), int64(2), object(12)
memory usage: 86.4+ KB
```

In [13]:
```
# task 5
# + -> P and - -> N
data['col16'] = data['col16'].map({'+' : 'P' , '-' : 'N'})
```

In [14]:
```
data
```

Out[14]:

	col1	col2	col3	col4	col5	col6	col7	col8	col9	col10	col11	col12	col13	col14	cc
0	b	30.83	0.000	u	g	w	v	1.25	t	t	1	f	g	00202	
1	a	58.67	4.460	u	g	q	h	3.04	t	t	6	f	g	00043	
2	a	24.50	0.500	u	g	q	h	1.50	t	f	0	f	g	00280	
3	b	27.83	1.540	u	g	w	v	3.75	t	t	5	t	g	00100	
4	b	20.17	5.625	u	g	w	v	1.71	t	f	0	f	s	00120	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
685	b	21.08	10.085	y	p	e	h	1.25	f	f	0	f	g	00260	
686	a	22.67	0.750	u	g	c	v	2.00	f	t	2	t	g	00200	
687	a	25.25	13.500	y	p	ff	ff	2.00	f	t	1	t	g	00200	
688	b	17.92	0.205	u	g	aa	v	0.04	f	f	0	f	g	00280	
689	b	35.00	3.375	u	g	c	h	8.29	f	f	0	t	g	00000	

690 rows × 16 columns

In [15]:
```
# task 6
cat_columns = [c for c in data.columns if data[c].dtypes == 'O']
data[cat_columns].head()
```

Out[15]:

	col1	col2	col4	col5	col6	col7	col9	col10	col12	col13	col14	col16
0	b	30.83	u	g	w	v	t	t	f	g	00202	P
1	a	58.67	u	g	q	h	t	t	f	g	00043	P
2	a	24.50	u	g	q	h	t	f	f	g	00280	P
3	b	27.83	u	g	w	v	t	t	t	g	00100	P
4	b	20.17	u	g	w	v	t	f	f	s	00120	P