# Preprocesamiento - Discretización

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
myData = pd.read_csv('myData_p.csv', header=0, low_memory=False)
```

In [5]:
```
edad = pd.cut(myData['AGE'],[0,10,20,60,98])
print(edad)
print(pd.value_counts(edad))
```

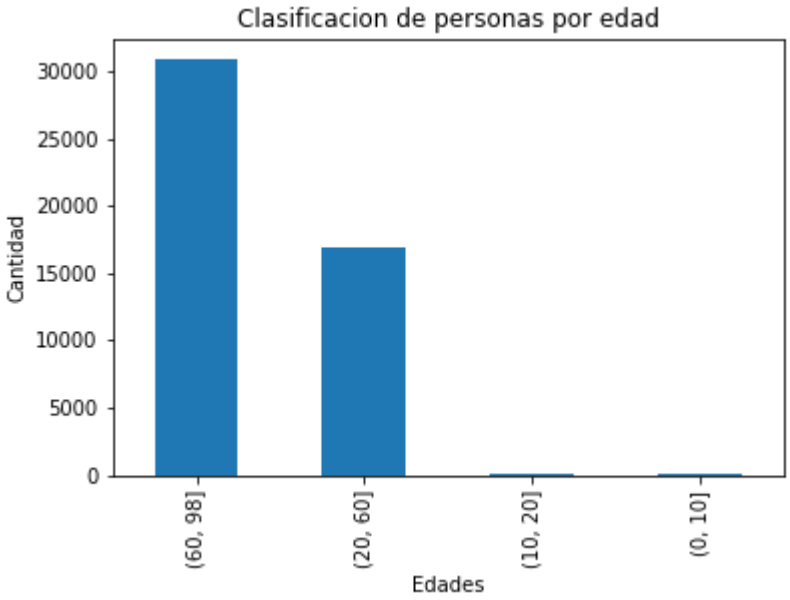
```
0        (20, 60]
1        (60, 98]
2        (60, 98]
3        (60, 98]
4        (60, 98]
           ...
47687    (60, 98]
47688    (60, 98]
47689    (60, 98]
47690    (20, 60]
47691    (60, 98]
Name: AGE, Length: 47692, dtype: category
Categories (4, interval[int64, right]): [(0, 10] < (10, 20] < (20, 60] < (60, 98]]
(60, 98]    30810
(20, 60]    16843
(10, 20]       26
(0, 10]        13
Name: AGE, dtype: int64
```

In [6]:
```
print(myData['AGE'].mean())
```

```
61.73142809392496
```

In [9]:
```
plot = pd.value_counts(edad).plot(kind='bar',title='Clasificacion de personas por edad
plot.set_ylabel('Cantidad')
plot.set_xlabel('Edades')
```

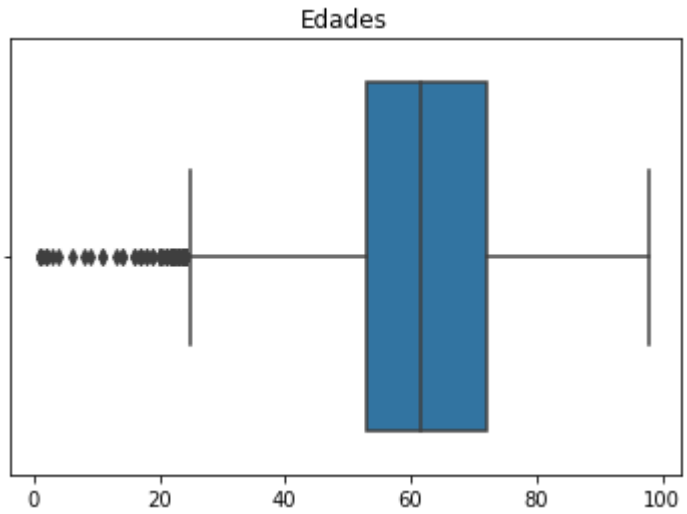
Out[9]:
```
Text(0.5, 0, 'Edades')
```




## Visualizando con gráfico de cajas

In [11]:
```
import seaborn as sns
sns.boxplot(list(myData['AGE']))
plt.title('Edades')
plt.show()
```

```
C:\Users\hdavi\anaconda3\envs\jupyterlab-3.3.2\lib\site-packages\seaborn\_decorators.
py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```




## Creando un subconjunto personas con mayor a 40 años

In [14]:
```
datasetMayores = myData[myData.AGE >= 40]
print(len(myData))
print(len(datasetMayores))
```

```
47692
43990
```

In [15]:
```
datasetMayores.to_csv("datasetMayores.csv", index=False)
```

In [ ]:
```

```