```
# Load the libraries

import pandas as pd


df=pd.read_csv("/content/nba.csv")


# Analyse the Data : Metadata
# Data Attributes

df.dtypes
```

```
Name         object
Team         object
Number      float64
Position     object
Age         float64
Height       object
Weight      float64
College      object
Salary      float64
dtype: object
```

```
# Dataset Size [R X C]

df.shape
```

```
(458, 9)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 458 entries, 0 to 457
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Name      457 non-null    object 
 1   Team      457 non-null    object 
 2   Number    457 non-null    float64
 3   Position  457 non-null    object 
 4   Age       457 non-null    float64
 5   Height    457 non-null    object 
 6   Weight    457 non-null    float64
 7   College   373 non-null    object 
 8   Salary    446 non-null    float64
dtypes: float64(4), object(5)
memory usage: 32.3+ KB
```

```
df.isnull().sum()
```

```
Name         1
Team         1
Number       1
Position     1
Age          1
Height       1
Weight       1
College     85
Salary      12
dtype: int64
```

```
df["Number"].describe()
```

```
count    457.000000
mean      17.678337
std       15.966090
min        0.000000
25%        5.000000
50%       13.000000
75%       25.000000
max       99.000000
Name: Number, dtype: float64
```

```
df["Team"].describe()
```

```
count                     457
unique                     30
top       New Orleans Pelicans
freq                       19
Name: Team, dtype: object
```

```
df["Position"].unique()
```

```
array(['PG', 'SF', 'SG', 'PF', 'C', nan], dtype=object)
```

```
df["College"].unique()
```

```
array(['Texas', 'Marquette', 'Boston University', 'Georgia State', nan,
       'LSU', 'Gonzaga', 'Louisville', 'Oklahoma State', 'Ohio State',
       'Washington', 'Kentucky', 'North Carolina', 'Arizona',
       'Georgia Tech', 'Cincinnati', 'Miami (FL)', 'Stanford', 'Syracuse',
       'Saint Louis', 'Kansas', 'Georgetown', 'Texas A&M', 'UCLA', 'UNLV',
       'Wichita State', "Saint Joseph's", 'Notre Dame', 'Norfolk State',
       'Duke', 'Murray State', 'Tennessee State', 'Bowling Green',
       'Purdue', 'Wake Forest', 'Michigan', 'Missouri', 'USC',
       'Villanova', 'Rider', 'Utah', 'Belmont', 'Davidson', 'Vanderbilt',
       'Michigan State', 'Florida', 'Washington State', 'Arizona State',
       'Oklahoma', 'Wyoming', "St. John's", 'Maryland', 'Wisconsin',
       'Utah Valley', 'North Carolina State', 'UC Santa Barbara',
       'Baylor', 'Connecticut', 'Oregon State', 'New Mexico', 'Oregon',
       'Creighton', 'Arkansas', 'Memphis', "Saint Mary's", 'Tennessee',
       'Alabama', 'Georgia', 'Colorado', 'Boston College', 'Temple',
       'Fresno State', 'IUPUI', 'Eastern Washington', 'Western Michigan',
       'Virginia', 'Northeastern', 'Western Kentucky', 'Nevada',
       'Illinois', 'Kansas State', 'Charleston', 'Clemson',
       'Blinn College', 'Providence', 'Detroit', 'Rhode Island',
       'California', 'Cleveland State', 'Iowa State', 'Florida State',
       'Long Beach State', 'Penn State', 'Indiana', 'San Diego State',
       'Western Carolina', 'Houston', 'Xavier', 'Old Dominion',
       'Minnesota', 'Louisiana Tech', 'Bucknell', 'Pittsburgh',
       'Virginia Commonwealth', 'Harvard', 'Marshall', 'Iowa',
       'St. Bonaventure', 'Louisiana-Lafayette', 'Colorado State',
       'Virginia Tech', 'DePaul', 'Morehead State', 'Central Michigan',
       'Weber State', 'Lehigh', 'Westchester CC', 'Dayton', 'Butler'],
      dtype=object)
```

```
# Data itself
# Display Top 5 Records
```

```
df.head()
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99.0	SF	25.0	6-6	235.0	Marquette	6796117.0
2	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28.0	SG	22.0	6-5	185.0	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0

```
# Display Bottom 5 Records
```

```
df.tail()
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
453	Shelvin Mack	Utah Jazz	8.0	PG	26.0	6-3	203.0	Butler	2433333.0
454	Raul Neto	Utah Jazz	25.0	PG	24.0	6-1	179.0	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21.0	C	26.0	7-3	256.0	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24.0	C	26.0	7-0	231.0	Kansas	947276.0
457	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
df.tail(1)
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
457	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
# 1) Data Cleaning : Missing Values
# Use of Global Constant : "Unknown"

df["Name"].fillna("Unknown")
```

```
0      Avery Bradley
1        Jae Crowder
2       John Holland
3        R.J. Hunter
4      Jonas Jerebko
             ...
453     Shelvin Mack
454        Raul Neto
455     Tibor Pleiss
456      Jeff Withey
457          Unknown
Name: Name, Length: 458, dtype: object
```

```
df["College"].fillna("Not Mentioned")
```

```
0                 Texas
1             Marquette
2      Boston University
3          Georgia State
4          Not Mentioned
              ...
453               Butler
454        Not Mentioned
455        Not Mentioned
456               Kansas
457        Not Mentioned
Name: College, Length: 458, dtype: object
```

```
df["College"].fillna("Not Mentioned",inplace=True)
```

```
df.fillna("Not Available",inplace=True)
```

```
df.isnull().sum()
```

```
Name        0
Team        0
Number      0
Position    0
Age         0
Height      0
Weight      0
College     0
Salary      0
dtype: int64
```

```
# 2) Data Reduction
# Reduce Attributes ( Delete Columns ) [Axis=1]
# Reduce Observations ( Delete Rows )  [Axis=0]

# Drop Attribute

df.drop("Salary",axis=1,inplace=True)
```

```
df.head()
```

	Name	Team	Number	Position	Age	Height	Weight	College
0	Avery Bradley	Boston Celtics	0	PG	25	6-2	180	Texas
1	Jae Crowder	Boston Celtics	99	SF	25	6-6	235	Marquette
2	John Holland	Boston Celtics	30	SG	27	6-5	205	Boston University
3	R.J. Hunter	Boston Celtics	28	SG	22	6-5	185	Georgia State
4	Jonas Jerebko	Boston Celtics	8	PF	29	6-10	231	Not Mentioned

```
# Drop Record with index 3

df.drop(3, axis=0)
```

	Name	Team	Number	Position	Age	Height	Weight	College
0	Avery Bradley	Boston Celtics	0	PG	25	6-2	180	Texas
1	Jae Crowder	Boston Celtics	99	SF	25	6-6	235	Marquette
2	John Holland	Boston Celtics	30	SG	27	6-5	205	Boston University
4	Jonas Jerebko	Boston Celtics	8	PF	29	6-10	231	Not Mentioned
5	Amir Johnson	Boston Celtics	90	PF	29	6-9	240	Not Mentioned
...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	6-3	203	Butler
454	Raul Neto	Utah Jazz	25	PG	24	6-1	179	Not Mentioned
455	Tibor Pleiss	Utah Jazz	21	C	26	7-3	256	Not Mentioned
456	Jeff Withey	Utah Jazz	24	C	26	7-0	231	Kansas
457	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Mentioned

457 rows × 8 columns

```
# 3) Data Transformation

from sklearn import preprocessing
e=preprocessing.LabelEncoder()

e.fit_transform(df["Position"].unique())
# 'PG : 3', 'SF : 4', 'SG : 5', 'PF : 2', 'C : 0','Not Available : 1'
```

```
array([3, 4, 5, 2, 0, 1])
```