```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df=pd.read_csv(r"C:\Users\ASUS\Documents\pythonStack\DS_PR\employees.csv - employees.csv.csv");

df.head()
```

```
  First Name  Gender Start Date Last Login Time  Salary  Bonus %  \
0    Douglas    Male   8/6/1993        12:42 PM   97308    6.945
1     Thomas    Male  3/31/1996         6:53 AM   61933    4.170
2      Maria  Female  4/23/1993        11:17 AM  130590   11.858
3      Jerry    Male   3/4/2005         1:00 PM  138705    9.340
4      Larry    Male  1/24/1998         4:47 PM  101004    1.389

  Senior Management            Team
0              True       Marketing
1              True             NaN
2             False         Finance
3              True         Finance
4              True  Client Services
```

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

```
First Name            67
Gender               145
Start Date             0
Last Login Time        0
Salary                 0
Bonus %                0
Senior Management     67
Team                  43
dtype: int64
```

```
df.describe(include='all')
```

```
       First Name  Gender  Start Date Last Login Time         Salary  \
count         933     855        1000            1000    1000.000000
unique        200       2         972             720            NaN
top       Marilyn  Female  10/30/1994         1:35 PM            NaN
freq           11     431           2               5            NaN
mean          NaN     NaN         NaN             NaN   90662.181000
std           NaN     NaN         NaN             NaN   32923.693342
min           NaN     NaN         NaN             NaN   35013.000000
25%           NaN     NaN         NaN             NaN   62613.000000
50%           NaN     NaN         NaN             NaN   90428.000000
75%           NaN     NaN         NaN             NaN  118740.250000
max           NaN     NaN         NaN             NaN  149908.000000

            Bonus % Senior Management            Team
```

```
count   1000.000000               933              957
unique          NaN                 2               10
top             NaN              True  Client Services
freq            NaN               468              106
mean      10.207555               NaN              NaN
std        5.528481               NaN              NaN
min        1.015000               NaN              NaN
25%        5.401750               NaN              NaN
50%        9.838500               NaN              NaN
75%       14.838000               NaN              NaN
max       19.944000               NaN              NaN
```

```
df.shape
```

```
(1000, 8)
```

```
print(df.dtypes)
```

```
First Name            object
Gender                object
Start Date            object
Last Login Time       object
Salary                 int64
Bonus %              float64
Senior Management     object
Team                  object
dtype: object
```

```
df['Start Date']
```

```
0        8/6/1993
1       3/31/1996
2       4/23/1993
3        3/4/2005
4       1/24/1998
          ...
995    11/23/2014
996     1/31/1984
997     5/20/2013
998     4/20/2013
999     5/15/2012
Name: Start Date, Length: 1000, dtype: object
```

```
df['Start Date']=pd.to_datetime(df['Start Date'])
```

```
df['Start Date']
```

```
0     1993-08-06
1     1996-03-31
2     1993-04-23
3     2005-03-04
4     1998-01-24
```

```
        ...
995   2014-11-23
996   1984-01-31
997   2013-05-20
998   2013-04-20
999   2012-05-15
Name: Start Date, Length: 1000, dtype: datetime64[ns]
```

```
df['Last Login Time']
```

```
0      12:42 PM
1       6:53 AM
2      11:17 AM
3       1:00 PM
4       4:47 PM
         ...
995     6:09 AM
996     6:30 AM
997    12:39 PM
998     4:45 PM
999     6:24 PM
Name: Last Login Time, Length: 1000, dtype: object
```

```
df[df['Start Date'] > '2000-01-01']
```

```
    First Name  Gender Start Date Last Login Time  Salary  Bonus %  \
3        Jerry    Male 2005-03-04         1:00 PM  138705    9.340
7          NaN  Female 2015-07-20        10:43 AM   45906   11.598
8       Angela  Female 2005-11-22         6:29 AM   95570   18.523
9      Frances  Female 2002-08-08         6:51 AM  139852    7.524
13       Gary    Male 2008-01-27        11:40 PM  109831    5.831
..        ...     ...        ...             ...     ...      ...
994     George    Male 2013-06-21         5:47 PM   98874    4.479
995      Henry     NaN 2014-11-23         6:09 AM  132483   16.655
997    Russell    Male 2013-05-20        12:39 PM   96914    1.421
998      Larry    Male 2013-04-20         4:45 PM   60500   11.985
999     Albert    Male 2012-05-15         6:24 PM  129949   10.169

    Senior Management                 Team
3                True              Finance
7                 NaN              Finance
8                True          Engineering
9                True  Business Development
13              False                Sales
..                ...                  ...
994              True            Marketing
995             False         Distribution
997             False              Product
998             False  Business Development
999              True                Sales
```

```
[479 rows x 8 columns]
```

```
df['Last Login Time']
```

```
0      12:42 PM
1       6:53 AM
2      11:17 AM
3       1:00 PM
4       4:47 PM
         ...
995     6:09 AM
996     6:30 AM
997    12:39 PM
998     4:45 PM
999     6:24 PM
Name: Last Login Time, Length: 1000, dtype: object
```

```
df['Last Login Time'] = pd.to_datetime(df['Last Login Time'])
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_15336\482447781.py:1: UserWarning:
Could not infer format, so each element will be parsed individually, falling
back to `dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.
  df['Last Login Time'] = pd.to_datetime(df['Last Login Time'])
```

```
df['Last Login Time']
```

```
0     2025-05-01 12:42:00
1     2025-05-01 06:53:00
2     2025-05-01 11:17:00
3     2025-05-01 13:00:00
4     2025-05-01 16:47:00
              ...
995   2025-05-01 06:09:00
996   2025-05-01 06:30:00
997   2025-05-01 12:39:00
998   2025-05-01 16:45:00
999   2025-05-01 18:24:00
Name: Last Login Time, Length: 1000, dtype: datetime64[ns]
```

```
df['Senior Management']
```

```
0       True
1       True
2      False
3       True
4       True
       ...
995    False
996    False
997    False
```

```
998    False
999     True
Name: Senior Management, Length: 1000, dtype: object
```

```
df['Gender']=df['Gender'].astype('category')
```

```
df['Gender']
```

```
0        Male
1        Male
2      Female
3        Male
4        Male
        ...
995       NaN
996      Male
997      Male
998      Male
999      Male
Name: Gender, Length: 1000, dtype: category
Categories (2, object): ['Female', 'Male']
```

```
df['Team']
```

```
0                 Marketing
1                       NaN
2                   Finance
3                   Finance
4           Client Services
               ...
995            Distribution
996                 Finance
997                 Product
998    Business Development
999                   Sales
Name: Team, Length: 1000, dtype: object
```

```
df['Team']=df['Team'].astype('category')
```

```
(df['Team']=="Finance").sum()
```

```
102
```

```
male=df['Gender']=='Male'
```

```
male.sum()
```

```
424
```

```
df['Senior Management'].dropna()
```

```
0       True
1       True
```

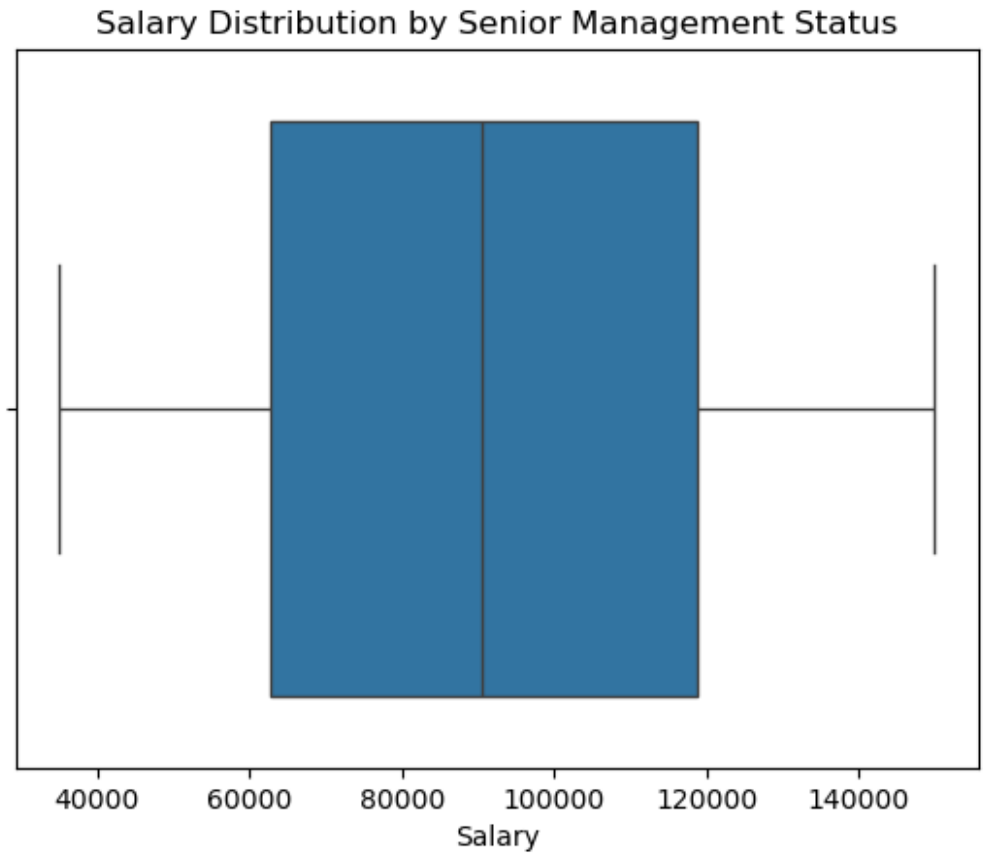
```
2      False
3       True
4       True
       ...
995    False
996    False
997    False
998    False
999     True
Name: Senior Management, Length: 933, dtype: object
```

```
df['Senior Management']
```

```
0       True
1       True
2      False
3       True
4       True
       ...
995    False
996    False
997    False
998    False
999     True
Name: Senior Management, Length: 1000, dtype: object
```

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
sns.boxplot(x=df['Salary'])
plt.title('Salary Distribution by Senior Management Status')
plt.show()
```


Salary Distribution by Senior Management Status
40000
60000
80000
100000
120000
140000
Salary