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


df=pd.read_csv("employees (1).csv")
print(df)
```

```
    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
..         ...     ...         ...             ...     ...      ...
995      Henry     NaN  11/23/2014         6:09 AM  132483   16.655
996    Phillip    Male   1/31/1984         6:30 AM   42392   19.675
997    Russell    Male   5/20/2013        12:39 PM   96914    1.421
998      Larry    Male   4/20/2013         4:45 PM   60500   11.985
999     Albert    Male   5/15/2012         6:24 PM  129949   10.169

    Senior Management                 Team
0                True            Marketing
1                True                  NaN
2               False              Finance
3                True              Finance
4                True       Client Services
..                ...                  ...
995             False          Distribution
996             False              Finance
997             False              Product
998             False  Business Development
999              True                Sales

[1000 rows x 8 columns]
```

```
print(np.size(df))
```

```
8000
```

```
print(np.shape(df))
```

```
(1000, 8)
```

```
print(type(df))
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
print(np.min(df['Salary']))
print(np.max(df['Salary']))
```

```
35013
```

```
print(np.unique(df['Bonus %']))
```

```
 8.461  8.475  8.525  8.567  8.572  8.578  8.611  8.639  8.689  8.696
 8.701  8.723  8.738  8.748  8.756  8.833  8.842  8.86   8.862  8.866
 8.879  8.895  8.932  8.94   8.941  8.945  8.957  8.965  8.992  8.996
 8.999  9.     9.047  9.048  9.061  9.078  9.096  9.127  9.129  9.146
 9.148  9.153  9.16   9.177  9.192  9.212  9.302  9.331  9.34   9.375
 9.402  9.404  9.494  9.54   9.55   9.554  9.555  9.557  9.564  9.601
 9.624  9.631  9.634  9.635  9.64   9.653  9.657  9.66   9.662  9.669
 9.699  9.705  9.712  9.735  9.765  9.769  9.77   9.801  9.828  9.849
 9.851  9.862  9.989  9.999 10.006 10.012 10.048 10.056 10.069 10.081
10.084 10.125 10.146 10.154 10.162 10.166 10.169 10.2   10.247 10.256
10.263 10.28  10.286 10.331 10.366 10.385 10.391 10.411 10.413 10.429
10.458 10.485 10.489 10.494 10.527 10.574 10.615 10.62  10.628 10.664
10.696 10.717 10.736 10.751 10.763 10.833 10.867 10.894 10.895 10.925
10.957 10.966 10.982 10.985 11.048 11.051 11.058 11.105 11.119 11.126
11.159 11.176 11.178 11.196 11.204 11.209 11.211 11.226 11.279 11.295
```

```
19.601 19.612 19.624 19.626 19.642 19.675 19.687 19.691 19.695 19.717
19.731 19.754 19.767 19.78  19.85  19.894 19.897 19.908 19.93  19.934
19.944]
```

```
print(np.mean(df['Salary']))
print(np.median(df['Salary']))
```

```
90662.181
90428.0
```

```
print(np.sum(df['Salary']))
```

```
90662181
```

```
df.head(10)
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
5	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False	Legal
6	Ruby	Female	8/17/1987	4:20 PM	65476	10.012	True	Product

Next steps: Generate code with df | New interactive sheet

```
df.tail(20)
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
980	Kimberly	Female	1/26/2013	12:57 AM	46233	8.862	True	Engineering
981	James	Male	1/15/1993	5:19 PM	148985	19.280	False	Legal
982	Rose	Female	4/6/1982	10:43 AM	91411	8.639	True	Human Resources
983	John	Male	12/23/1982	10:35 PM	146907	11.738	False	Engineering
984	Maria	Female	10/15/2011	4:53 PM	43455	13.040	False	Engineering
985	Stephen	NaN	7/10/1983	8:10 PM	85668	1.909	False	Legal
986	Donna	Female	11/26/1982	7:04 AM	82871	17.999	False	Marketing
987	Gloria	Female	12/8/2014	5:08 AM	136709	10.331	True	Finance
988	Alice	Female	10/5/2004	9:34 AM	47638	11.209	False	Human Resources
989	Justin	NaN	2/10/1991	4:58 PM	38344	3.794	False	Legal
990	Robin	Female	7/24/1987	1:35 PM	100765	10.982	True	Client Services
991	Rose	Male	10/16/1999	8:35 AM	112769	11.625	True	Finance
992	Anthony	Male	5/15/1997	3:53 PM	56450	19.040	True	Engineering
993	Tina	Female	6/21/2013	5:47 PM	98874	4.479	True	Marketing
994	George	Male	6/21/2013	5:47 PM	98874	4.479	True	Marketing

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   First Name        933 non-null    object
 1   Gender            855 non-null    object
 2   Start Date        1000 non-null   object
 3   Last Login Time   1000 non-null   object
 4   Salary            1000 non-null   int64
 5   Bonus %           1000 non-null   float64
 6   Senior Management  933 non-null   object
 7   Team              957 non-null    object
dtypes: float64(1), int64(1), object(6)
memory usage: 62.6+ KB
```

```
df.describe()
```

	Salary	Bonus %
count	1000.000000	1000.000000
mean	90662.181000	10.207555
std	32923.693342	5.528481
min	35013.000000	1.015000
25%	62613.000000	5.401750
50%	90428.000000	9.838500
75%	118740.250000	14.838000

```
df.isnull()
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	True
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
995	False	True	False	False	False	False	False	False
996	False	False	False	False	False	False	False	False
997	False	False	False	False	False	False	False	False
998	False	False	False	False	False	False	False	False
999	False	False	False	False	False	False	False	False

```
df.dropna()
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
994	George	Male	6/21/2013	5:47 PM	98874	4.479	True	Marketing
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product

```
df.dropna(inplace=True)
print(df)
```

```
     First Name  Gender Start Date Last Login Time  Salary  Bonus %  \
0       Douglas    Male   8/6/1993       12:42 PM   97308    6.945
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
5        Dennis    Male  4/18/1987        1:35 AM  115163   10.125
..          ...     ...        ...            ...     ...      ...
994      George    Male  6/21/2013        5:47 PM   98874    4.479
996     Phillip    Male  1/31/1984        6:30 AM   42392   19.675
997     Russell    Male  5/20/2013       12:39 PM   96914    1.421
998       Larry    Male  4/20/2013        4:45 PM   60500   11.985
999      Albert    Male  5/15/2012        6:24 PM  129949   10.169

     Senior Management                Team
0                 True           Marketing
2                False             Finance
3                 True             Finance
4                 True      Client Services
5                False               Legal
..                 ...                 ...
994               True           Marketing
996              False             Finance
997              False             Product
998              False  Business Development
999               True               Sales

[764 rows x 8 columns]
```

```
df['Increment']=df['Salary']*1.01
print(df)
```

```
     First Name  Gender Start Date Last Login Time  Salary  Bonus %  \
0       Douglas    Male   8/6/1993       12:42 PM   97308    6.945
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
5        Dennis    Male  4/18/1987        1:35 AM  115163   10.125
..          ...     ...        ...            ...     ...      ...
994      George    Male  6/21/2013        5:47 PM   98874    4.479
```

```
996    Phillip    Male  1/31/1984         6:30 AM   42392   19.675
997    Russell    Male  5/20/2013        12:39 PM   96914    1.421
998      Larry    Male  4/20/2013         4:45 PM   60500   11.985
999     Albert    Male  5/15/2012         6:24 PM  129949   10.169

    Senior Management                 Team  Increment
0                True            Marketing   98281.08
2               False              Finance  131895.90
3                True              Finance  140092.05
4                True       Client Services  102014.04
5               False                Legal  116314.63
..                ...                  ...        ...
994              True            Marketing   99862.74
996             False              Finance   42815.92
997             False              Product   97883.14
998             False  Business Development   61105.00
999              True                Sales  131248.49

[764 rows x 9 columns]
```

```
df.columns
```

```
Index(['First Name', 'Gender', 'Start Date', 'Last Login Time', 'Salary',
       'Bonus %', 'Senior Management', 'Team', 'Increment'],
      dtype='object')
```

```
print(df['Team'])
```

```
0                 Marketing
2                   Finance
3                   Finance
4           Client Services
5                     Legal
               ...
994               Marketing
996                 Finance
997                 Product
998    Business Development
999                   Sales
Name: Team, Length: 764, dtype: object
```

```
df.fillna(0)
```

```
/tmp/ipython-input-516507702.py:1: FutureWarning: Downcasting object dtype arrays on
  df.fillna(0)
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team	Increme
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing	98281
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance	131895
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance	140092
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services	102014
5	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False	Legal	116314
...	...	...	...	...	...	...	...	...	
994	George	Male	6/21/2013	5:47 PM	98874	4.479	True	Marketing	99862
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product	97883
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development	61105
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales	131248

764 rows × 9 columns

```
print(df['Last Login Time'],df['Start Date'])
```

```
0      12:42 PM
2      11:17 AM
3       1:00 PM
4       4:47 PM
5       1:35 AM
         ...
994     5:47 PM
996     6:30 AM
997    12:39 PM
998     4:45 PM
999     6:24 PM
Name: Last Login Time, Length: 764, dtype: object 0       8/6/1993
2      4/23/1993
3       3/4/2005
4      1/24/1998
5      4/18/1987
         ...
994    6/21/2013
996    1/31/1984
997    5/20/2013
998    4/20/2013
999    5/15/2012
Name: Start Date, Length: 764, dtype: object
```

```
print(df.loc[0,'Salary'])
```

```
97308
```

```
print(df.iloc[2,5])
```

```
9.34
```

```
df.values
```

```
array([['Douglas', 'Male', '8/6/1993', ..., True, 'Marketing', 98281.08],
       ['Maria', 'Female', '4/23/1993', ..., False, 'Finance', 131895.9],
       ['Jerry', 'Male', '3/4/2005', ..., True, 'Finance', 140092.05],
       ...,
       ['Russell', 'Male', '5/20/2013', ..., False, 'Product', 97883.14],
       ['Larry', 'Male', '4/20/2013', ..., False, 'Business Development',
        61105.0],
       ['Albert', 'Male', '5/15/2012', ..., True, 'Sales', 131248.49]],
      dtype=object)
```

```
print(df[df['Salary']>50000])
```

```
    First Name  Gender Start Date Last Login Time  Salary  Bonus %  \
0      Douglas    Male   8/6/1993        12:42 PM   97308    6.945
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
5       Dennis    Male  4/18/1987         1:35 AM  115163   10.125
..         ...     ...        ...             ...     ...      ...
993       Tina  Female  5/15/1997         3:53 PM   56450   19.040
994     George    Male  6/21/2013         5:47 PM   98874    4.479
997    Russell    Male  5/20/2013        12:39 PM   96914    1.421
998      Larry    Male  4/20/2013         4:45 PM   60500   11.985
999     Albert    Male  5/15/2012         6:24 PM  129949   10.169

    Senior Management                 Team  Increment
0                True            Marketing   98281.08
2               False              Finance  131895.90
3                True              Finance  140092.05
4                True       Client Services  102014.04
5               False                Legal  116314.63
..                ...                  ...        ...
993              True          Engineering   57014.50
994              True            Marketing   99862.74
997             False              Product   97883.14
998             False  Business Development   61105.00
999              True                Sales  131248.49

[654 rows x 9 columns]
```

```
df[(df['Bonus %']>0.5)&(df['Salary']<50000)]
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team	Incr
14	Kimberly	Female	1/14/1999	7:13 AM	41426	14.543	True	Finance	418
26	Craig	Male	2/27/2000	7:45 AM	37598	7.757	True	Marketing	379
38	Stephanie	Female	9/13/1986	1:52 AM	36844	5.574	True	Business Development	372
52	Todd	Male	2/18/1990	2:41 AM	49339	1.695	True	Human Resources	498
63	Matthew	Male	1/2/2013	10:33 PM	35203	18.040	False	Human Resources	355
...	...	...	...	...	...	...	...	...	
969	Linda	Female	2/4/2010	8:49 PM	44486	17.308	True	Engineering	449
980	Kimberly	Female	1/26/2013	12:57 AM	46233	8.862	True	Engineering	466
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance	428

410 rows × 9 columns

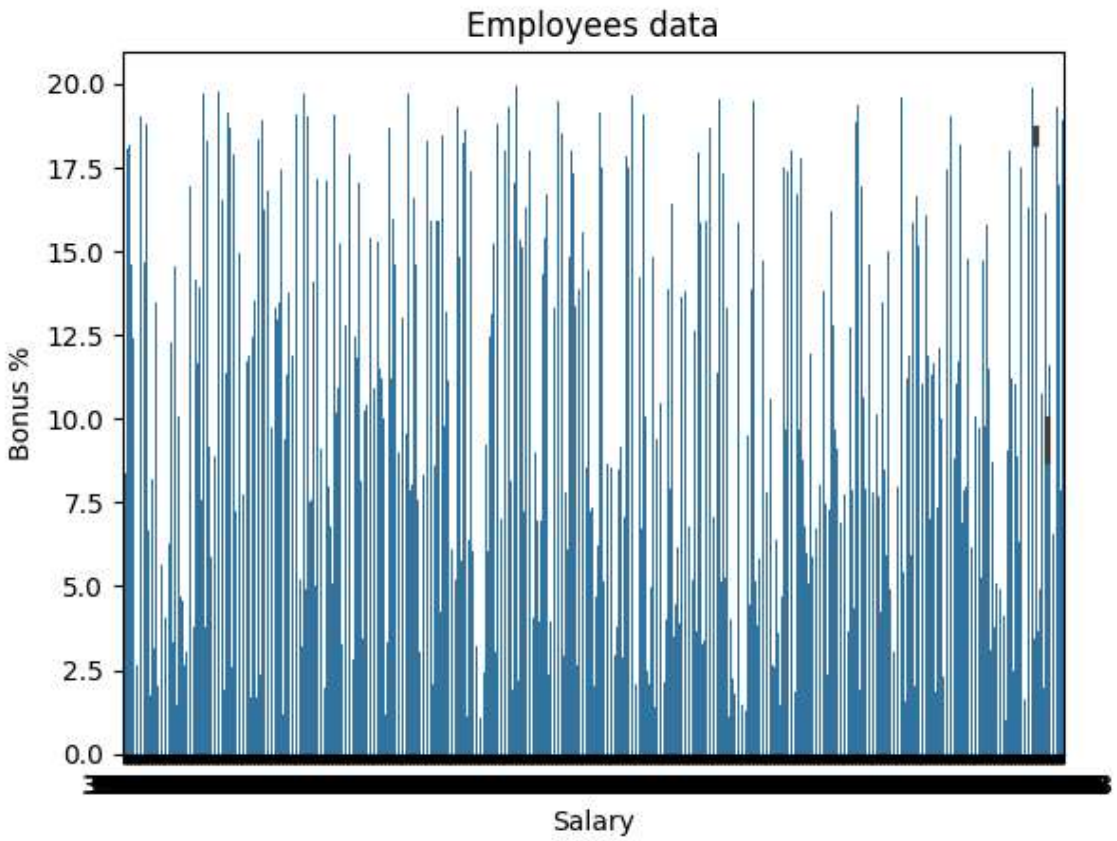
```
df.drop(columns=['Increment'],inplace=True)
print(df)
```

```
    First Name  Gender Start Date Last Login Time  Salary  Bonus %  \
0      Douglas    Male   8/6/1993         12:42 PM   97308    6.945
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
5       Dennis    Male  4/18/1987          1:35 AM  115163   10.125
..         ...     ...        ...              ...     ...      ...
994     George    Male  6/21/2013          5:47 PM   98874    4.479
996    Phillip    Male  1/31/1984          6:30 AM   42392   19.675
997    Russell    Male  5/20/2013         12:39 PM   96914    1.421
998      Larry    Male  4/20/2013          4:45 PM   60500   11.985
999     Albert    Male  5/15/2012          6:24 PM  129949   10.169

    Senior Management                Team
0                True             Marketing
2               False               Finance
3                True               Finance
4                True       Client Services
5               False                 Legal
..                ...                   ...
994              True             Marketing
996             False               Finance
997             False               Product
998             False  Business Development
999              True                 Sales

[764 rows x 8 columns]
```

```
df.sort_values(by=['Salary'],ascending=False,inplace=True)
print(df)
```

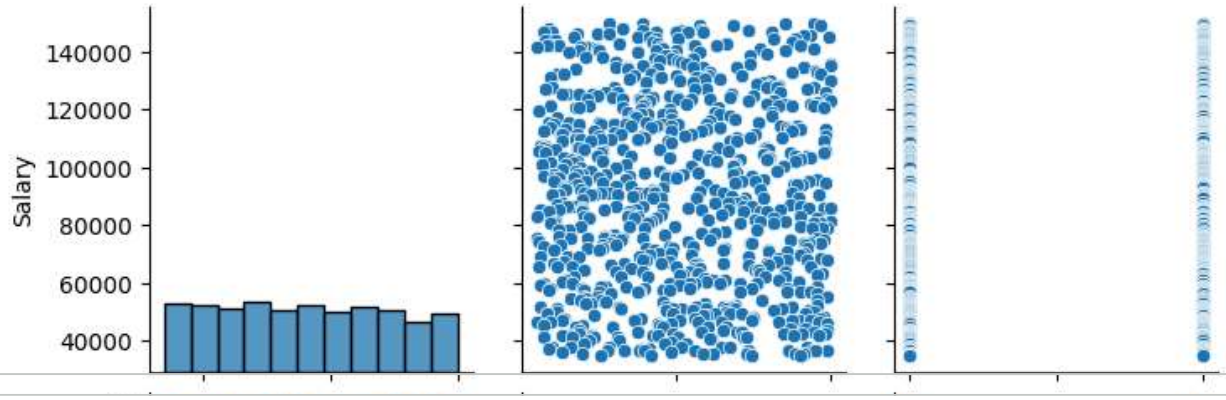
```
    First Name  Gender Start Date Last Login Time  Salary  Bonus %  \
644  Katherine  Female  8/13/1996        12:21 AM  149908   18.912
429       Rose  Female  5/28/2015         8:40 AM  149903    5.630
828    Cynthia  Female  7/12/2006         8:55 AM  149684    7.864
160      Kathy  Female  3/18/2000         7:26 PM  149563   16.991
793     Andrea  Female  7/22/1999         9:25 AM  149105   13.707
..         ...     ...        ...             ...     ...      ...
650    Cynthia  Female   7/5/1986         1:24 AM   35381   11.749
63     Matthew    Male   1/2/2013        10:33 PM   35203   18.040
82      Steven    Male  3/30/1980         9:20 PM   35095    8.379
238      Kevin    Male  3/25/1982         7:31 AM   35061    5.128
576    Michael    Male  7/30/1993         5:35 PM   35013   14.879

    Senior Management            Team
644             False         Finance
429             False  Human Resources
828             False         Product
160              True         Finance
793              True    Distribution
..                ...             ...
650             False         Finance
63              False  Human Resources
82               True  Client Services
238             False           Legal
576             False         Product

[764 rows x 8 columns]
```

```
sns.barplot(data=df,x='Salary',y='Bonus %')
plt.title("Employees data")
plt.show()
```




```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7842090627e0>
```



Start coding or generate with AI.