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

df=pd.read_csv(r"C:\Users\ASUS\Documents\pythonStack\DS_PR\3_employees.csv");

df.head()
```

```
   EMPLOYEE_ID FIRST_NAME  LAST_NAME     EMAIL  PHONE_NUMBER  HIRE_DATE  \
0          198     Donald   OConnell  DOCONNEL  650.507.9833  21-JUN-07
1          199    Douglas      Grant    DGRANT  650.507.9844  13-JAN-08
2          200   Jennifer     Whalen   JWHALEN  515.123.4444  17-SEP-03
3          201    Michael  Hartstein  MHARTSTE  515.123.5555  17-FEB-04
4          202        Pat        Fay      PFAY  603.123.6666  17-AUG-05

     JOB_ID  SALARY COMMISSION_PCT MANAGER_ID  DEPARTMENT_ID
0  SH_CLERK    2600              -         124             50
1  SH_CLERK    2600              -         124             50
2   AD_ASST    4400              -         101             10
3    MK_MAN   13000              -         100             20
4    MK_REP    6000              -         201             20
```

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

```
EMPLOYEE_ID       0
FIRST_NAME        0
LAST_NAME         0
EMAIL             0
PHONE_NUMBER      0
HIRE_DATE         0
JOB_ID            0
SALARY            0
COMMISSION_PCT    0
MANAGER_ID        0
DEPARTMENT_ID     0
dtype: int64
```

```
df.dtypes
```

```
EMPLOYEE_ID        int64
FIRST_NAME        object
LAST_NAME         object
EMAIL             object
PHONE_NUMBER      object
HIRE_DATE         object
JOB_ID            object
SALARY             int64
COMMISSION_PCT    object
MANAGER_ID        object
DEPARTMENT_ID      int64
dtype: object
```

```
df.shape
```

```
(50, 11)
```

```
df['HIRE_DATE'] = pd.to_datetime(df['HIRE_DATE'])
```

```
df['HIRE_DATE']
```

```
0    2007-06-21
1    2008-01-13
2    2003-09-17
3    2004-02-17
4    2005-08-17
5    2002-06-07
6    2002-06-07
7    2002-06-07
8    2002-06-07
9    2003-06-17
10   2005-09-21
11   2001-01-13
12   2006-01-03
13   2007-05-21
14   2005-06-25
15   2006-02-05
16   2007-02-07
17   2002-08-17
18   2002-08-16
19   2005-09-28
20   2005-09-30
21   2006-03-07
22   2007-12-07
23   2002-12-07
24   2003-05-18
25   2005-12-24
26   2005-07-24
27   2006-11-15
28   2007-08-10
29   2004-07-18
30   2005-04-10
31   2003-05-01
32   2005-10-10
33   2007-11-16
34   2005-07-16
35   2006-09-28
36   2007-01-14
37   2008-03-08
38   2005-08-20
39   2005-10-30
40   2005-02-16
41   2007-04-10
42   2004-06-14
```

```
43   2006-08-26
44   2007-12-12
45   2008-02-06
46   2003-07-14
47   2005-10-26
48   2006-02-12
49   2006-04-06
Name: HIRE_DATE, dtype: datetime64[ns]
```

```
df['COMMISSION_PCT'] = pd.to_numeric(df['COMMISSION_PCT'], errors='coerce')
```

```
from scipy.stats import zscore
import seaborn as sns
import matplotlib.pyplot as plt
```

```
SALARY_Zs = zscore(df['SALARY'])
```

```
(df['SALARY']==5000).sum()
```

```
0
```

```
sns.boxplot(data=df[['SALARY', 'SALARY_Zs']])
plt.title("Original vs Z-Score Normalized Salary")
plt.show()
```