Project on Texas Salary Prediction

TEAM ID : PTID-CDS-FEB-25-2469

PROJECT ID : PRCP-1024-TexasSalaryPrediction

PROJECT NAME : Texas Salary Prediction

Import Basic Modules

```
# import libraryes
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

Reading CSV file

```
df = pd.read_csv('salary.csv')
df.head()
```

```
   AGENCY                                          AGENCY NAME  \
0     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1     212  OFFICE OF COURT ADMINISTRATION                 ...
2     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3     212  OFFICE OF COURT ADMINISTRATION                 ...
4     696  TEXAS DEPARTMENT OF CRIMINAL JUSTICE           ...

                        LAST NAME                       FIRST NAME
MI  \
0  RUCKER                           MORTON                           V

1  RUCKER                           MORTON                           V

2  SPECIA JR                        JOHN                             J

3  SPECIA JR                        JOHN                             J

4  ONTIVEROS                        ESTHER


  CLASS CODE                                        CLASS TITLE  \
0   JD25      JUDGE, RETIRED                                 ...
1   3524      GENERAL COUNSEL IV                             ...
```

```
2   JD25      JUDGE, RETIRED                                  ...
3   3524      GENERAL COUNSEL IV                              ...
4   4504      CORREC  OFFICER IV                              ...

          ETHNICITY           GENDER
STATUS  \
0  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
1  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
2  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
3  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
4  HISPANIC         FEMALE           CRF - CLASSIFIED REGULAR FULL-
TIME

   ... HRLY RATE  HRS PER WK  MONTHLY     ANNUAL  STATE NUMBER
duplicated  \
0  ...  75.96150        29.0  9545.82  114549.84        127717
True
1  ...  81.04454         4.0  1404.77   16857.24        127717
True
2  ...  75.96150        29.0  9545.82  114549.84         59115
True
3  ...  81.04453         4.0  1404.77   16857.24         59115
True
4  ...   0.00000        40.0  3284.27   39411.24        165030
True

  multiple_full_time_jobs  combined_multiple_jobs summed_annual_salary
\
0                     NaN                     NaN            131407.08

1                     NaN                     NaN                  NaN

2                     NaN                     NaN            131407.08

3                     NaN                     NaN                  NaN

4                     1.0                     NaN                  NaN


   hide_from_search
0               NaN
1              True
2               NaN
3              True
4               NaN
```

```
[5 rows x 21 columns]
```

Getting information of dataset

```
df.shape
```

```
(149481, 21)
```

Information of dataset

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149481 entries, 0 to 149480
Data columns (total 21 columns):
 #   Column                   Non-Null Count   Dtype  
---  ------                   --------------   -----  
 0   AGENCY                   149481 non-null  int64  
 1   AGENCY NAME              149481 non-null  object 
 2   LAST NAME                149481 non-null  object 
 3   FIRST NAME               149481 non-null  object 
 4   MI                       149481 non-null  object 
 5   CLASS CODE               149481 non-null  object 
 6   CLASS TITLE              149481 non-null  object 
 7   ETHNICITY                149481 non-null  object 
 8   GENDER                   149481 non-null  object 
 9   STATUS                   149481 non-null  object 
 10  EMPLOY DATE              149481 non-null  object 
 11  HRLY RATE                149481 non-null  float64
 12  HRS PER WK               149481 non-null  float64
 13  MONTHLY                  149481 non-null  float64
 14  ANNUAL                   149481 non-null  float64
 15  STATE NUMBER             149481 non-null  int64  
 16  duplicated               143 non-null     object 
 17  multiple_full_time_jobs  14 non-null      float64
 18  combined_multiple_jobs   97 non-null      object 
 19  summed_annual_salary     16 non-null      float64
 20  hide_from_search         16 non-null      object 
dtypes: float64(6), int64(2), object(13)
memory usage: 23.9+ MB
```

Checking null values

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

```
AGENCY                         0
AGENCY NAME                    0
LAST NAME                      0
```

```
FIRST NAME                      0
MI                              0
CLASS CODE                      0
CLASS TITLE                     0
ETHNICITY                       0
GENDER                          0
STATUS                          0
EMPLOY DATE                     0
HRLY RATE                       0
HRS PER WK                      0
MONTHLY                         0
ANNUAL                          0
STATE NUMBER                    0
duplicated                 149338
multiple_full_time_jobs    149467
combined_multiple_jobs     149384
summed_annual_salary       149465
hide_from_search           149465
dtype: int64
```

Checking for duplicates

```
df.duplicated().sum()
```

```
0
```

There are no duplicates.

```
df_new=df.dropna(axis=1)
df_new.keys()
```

```
Index(['AGENCY', 'AGENCY NAME', 'LAST NAME', 'FIRST NAME', 'MI',
'CLASS CODE',
       'CLASS TITLE', 'ETHNICITY', 'GENDER', 'STATUS', 'EMPLOY DATE',
       'HRLY RATE', 'HRS PER WK', 'MONTHLY', 'ANNUAL', 'STATE
NUMBER'],
      dtype='object')
```

```
print(" \nCount total NaN at each column in a DataFrame : \n\n",
df_new.isnull().sum())
```

```
Count total NaN at each column in a DataFrame :

 AGENCY          0
AGENCY NAME     0
LAST NAME       0
FIRST NAME      0
MI              0
CLASS CODE      0
```

```
CLASS TITLE     0
ETHNICITY       0
GENDER          0
STATUS          0
EMPLOY DATE     0
HRLY RATE       0
HRS PER WK      0
MONTHLY         0
ANNUAL          0
STATE NUMBER    0
dtype: int64
```

Information of dataset after removing Null Values

```
df_new.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149481 entries, 0 to 149480
Data columns (total 16 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   AGENCY         149481 non-null  int64
 1   AGENCY NAME    149481 non-null  object
 2   LAST NAME      149481 non-null  object
 3   FIRST NAME     149481 non-null  object
 4   MI             149481 non-null  object
 5   CLASS CODE     149481 non-null  object
 6   CLASS TITLE    149481 non-null  object
 7   ETHNICITY      149481 non-null  object
 8   GENDER         149481 non-null  object
 9   STATUS         149481 non-null  object
 10  EMPLOY DATE    149481 non-null  object
 11  HRLY RATE      149481 non-null  float64
 12  HRS PER WK     149481 non-null  float64
 13  MONTHLY        149481 non-null  float64
 14  ANNUAL         149481 non-null  float64
 15  STATE NUMBER   149481 non-null  int64
dtypes: float64(4), int64(2), object(10)
memory usage: 18.2+ MB
```

```
df_new.head()
```

```
   AGENCY                                        AGENCY NAME  \
0     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1     212  OFFICE OF COURT ADMINISTRATION                ...
2     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3     212  OFFICE OF COURT ADMINISTRATION                ...
4     696  TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...

                        LAST NAME                        FIRST NAME
```

```
MI  \
0  RUCKER                          MORTON                          V

1  RUCKER                          MORTON                          V

2  SPECIA JR                       JOHN                            J

3  SPECIA JR                       JOHN                            J

4  ONTIVEROS                       ESTHER


  CLASS CODE                                        CLASS TITLE  \
0   JD25      JUDGE, RETIRED                                  ...
1   3524      GENERAL COUNSEL IV                              ...
2   JD25      JUDGE, RETIRED                                  ...
3   3524      GENERAL COUNSEL IV                              ...
4   4504      CORREC  OFFICER IV                              ...

         ETHNICITY           GENDER
STATUS  \
0  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
1  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
2  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
3  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
4  HISPANIC         FEMALE           CRF - CLASSIFIED REGULAR FULL-
TIME

  EMPLOY DATE  HRLY RATE  HRS PER WK  MONTHLY     ANNUAL  STATE NUMBER

0    02/18/88   75.96150        29.0  9545.82  114549.84        127717

1    02/01/15   81.04454         4.0  1404.77   16857.24        127717

2    02/01/20   75.96150        29.0  9545.82  114549.84         59115

3    09/01/18   81.04453         4.0  1404.77   16857.24         59115

4    06/29/20    0.00000        40.0  3284.27   39411.24        165030
```

Rename Columns

```
df_new.columns =
['AGENCY','AGENCYNAME','LASTNAME','FIRSTNAME','MI','CLASSCODE','CLASST
ITLE','ETHNICITY','GENDER','STATUS','EMPLOYDATE','HRLYRATE','HRSPERWK'
,'MONTHLY','ANNUAL','STATENUMBER']
```

```
df_new.head()
```

```
   AGENCY                                          AGENCYNAME  \
0     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1     212  OFFICE OF COURT ADMINISTRATION                 ...
2     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3     212  OFFICE OF COURT ADMINISTRATION                 ...
4     696  TEXAS DEPARTMENT OF CRIMINAL JUSTICE           ...

                          LASTNAME                          FIRSTNAME
MI  \
0  RUCKER                          MORTON                            V

1  RUCKER                          MORTON                            V

2  SPECIA JR                       JOHN                              J

3  SPECIA JR                       JOHN                              J

4  ONTIVEROS                       ESTHER


  CLASSCODE                                          CLASSTITLE  \
0  JD25      JUDGE, RETIRED                                 ...
1  3524      GENERAL COUNSEL IV                             ...
2  JD25      JUDGE, RETIRED                                 ...
3  3524      GENERAL COUNSEL IV                             ...
4  4504      CORREC  OFFICER IV                             ...

         ETHNICITY           GENDER
STATUS  \
0  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
1  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
2  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
3  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
4  HISPANIC         FEMALE           CRF - CLASSIFIED REGULAR FULL-
TIME

  EMPLOYDATE  HRLYRATE  HRSPERWK  MONTHLY     ANNUAL  STATENUMBER
0   02/18/88  75.96150      29.0  9545.82  114549.84       127717
1   02/01/15  81.04454       4.0  1404.77   16857.24       127717
2   02/01/20  75.96150      29.0  9545.82  114549.84        59115
3   09/01/18  81.04453       4.0  1404.77   16857.24        59115
4   06/29/20   0.00000      40.0  3284.27   39411.24       165030
```

Encoding the date

```
import datetime
df_new['EMPLOYDATE']= df_new['EMPLOYDATE'].astype('datetime64[ns]')

from datetime import datetime,date
today=str(date.today())
df_new['Dummy']= datetime.strptime(today, '%Y-%m-%d')

df_new['EMPLOYDATE']= df_new['Dummy']-df_new['EMPLOYDATE']

df_new.drop('Dummy', axis=1, inplace=True)

df_new['EMPLOY_DAY']= df_new['EMPLOYDATE'].dt.day

df_new.head()
```

```
   AGENCY                                          AGENCYNAME  \
0     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1     212  OFFICE OF COURT ADMINISTRATION                 ...
2     241  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3     212  OFFICE OF COURT ADMINISTRATION                 ...
4     696  TEXAS DEPARTMENT OF CRIMINAL JUSTICE           ...

                          LASTNAME                          FIRSTNAME
MI  \
0  RUCKER                          MORTON                            V

1  RUCKER                          MORTON                            V

2  SPECIA JR                       JOHN                              J

3  SPECIA JR                       JOHN                              J

4  ONTIVEROS                       ESTHER


  CLASSCODE                                          CLASSTITLE  \
0  JD25      JUDGE, RETIRED                                 ...
1  3524      GENERAL COUNSEL IV                             ...
2  JD25      JUDGE, RETIRED                                 ...
3  3524      GENERAL COUNSEL IV                             ...
4  4504      CORREC  OFFICER IV                             ...

         ETHNICITY           GENDER
STATUS  \
0  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
1  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
2  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
```

```
3  WHITE             MALE               CTP - CLASSIFIED TEMPORARY PART-
TIME
4  HISPANIC          FEMALE             CRF - CLASSIFIED REGULAR FULL-
TIME

  EMPLOYDATE  HRLYRATE  HRSPERWK  MONTHLY     ANNUAL  STATENUMBER
EMPLOY_DAY
0 1988-02-18  75.96150      29.0  9545.82  114549.84       127717
18
1 2015-02-01  81.04454       4.0  1404.77   16857.24       127717
1
2 2020-02-01  75.96150      29.0  9545.82  114549.84        59115
1
3 2018-09-01  81.04453       4.0  1404.77   16857.24        59115
1
4 2020-06-29   0.00000      40.0  3284.27   39411.24       165030
29
```

Pandas profile importing

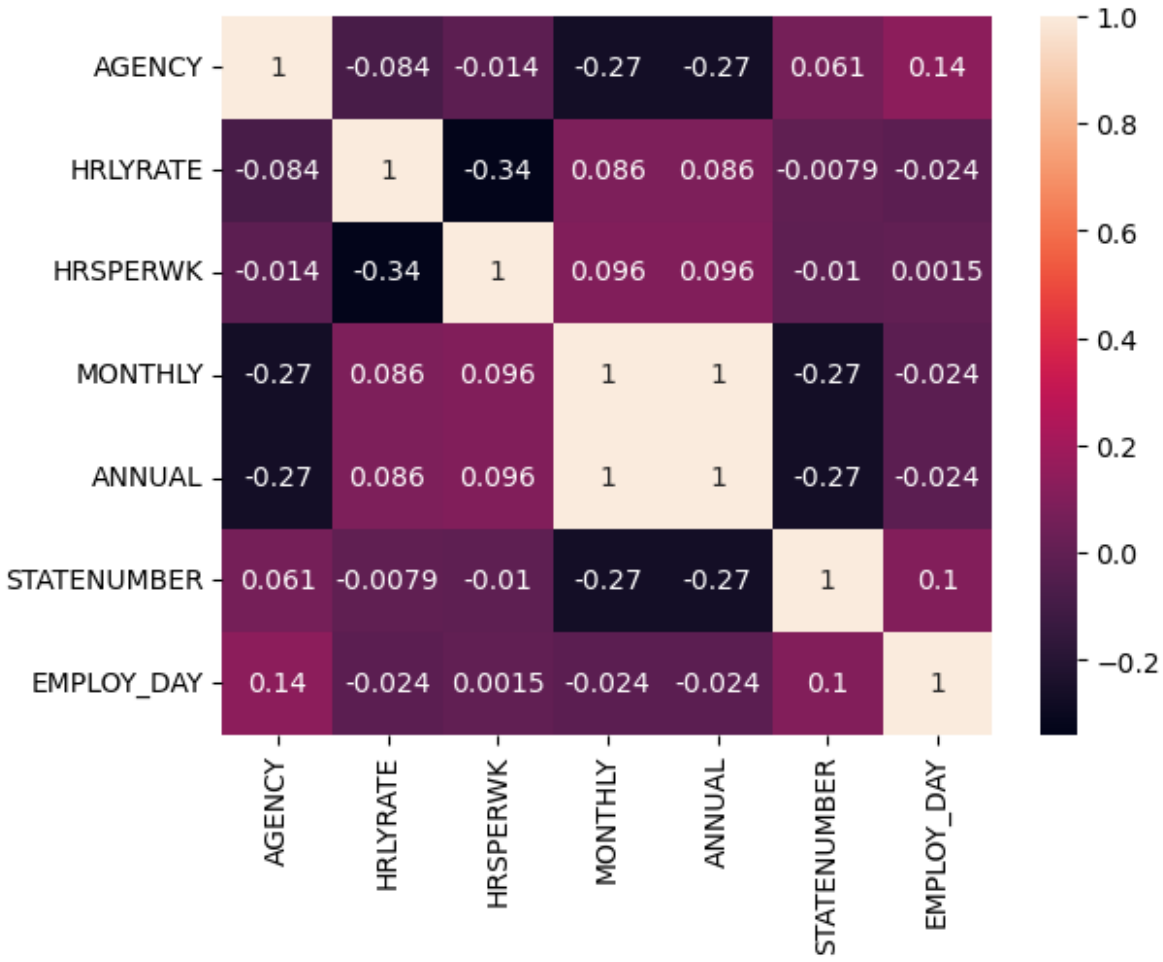
```
#pip install pandas-profiling

numaric_data=df_new.select_dtypes(include=['int64','float64'])
numaric_data.head()
```

```
   AGENCY  HRLYRATE  HRSPERWK  MONTHLY     ANNUAL  STATENUMBER
0     241  75.96150      29.0  9545.82  114549.84       127717
1     212  81.04454       4.0  1404.77   16857.24       127717
2     241  75.96150      29.0  9545.82  114549.84        59115
3     212  81.04453       4.0  1404.77   16857.24        59115
4     696   0.00000      40.0  3284.27   39411.24       165030
```

Bivariate analysis

```
sns.heatmap(df_new.select_dtypes(include='number').corr(),annot=True)
```

```
<Axes: >
```




```
df_new.select_dtypes(include='number').corr().style.background_gradien
t(cmap='coolwarm')
```

```
<pandas.io.formats.style.Styler at 0x1d846187830>
```

```
categorical_data=df_new.select_dtypes(exclude=['int64','float64'])
categorical_data.head()
```

```
                                          AGENCYNAME  \
0  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1  OFFICE OF COURT ADMINISTRATION                ...
2  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3  OFFICE OF COURT ADMINISTRATION                ...
4  TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...

                          LASTNAME                          FIRSTNAME
MI  \
0  RUCKER                          MORTON                            V

1  RUCKER                          MORTON                            V
```

```
2  SPECIA JR                          JOHN                                J

3  SPECIA JR                          JOHN                                J

4  ONTIVEROS                          ESTHER


  CLASSCODE                                              CLASSTITLE  \
0  JD25      JUDGE, RETIRED                                     ...
1  3524      GENERAL COUNSEL IV                                 ...
2  JD25      JUDGE, RETIRED                                     ...
3  3524      GENERAL COUNSEL IV                                 ...
4  4504      CORREC  OFFICER IV                                 ...

          ETHNICITY           GENDER
STATUS  \
0  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
1  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
2  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
3  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
4  HISPANIC         FEMALE           CRF - CLASSIFIED REGULAR FULL-
TIME

  EMPLOYDATE  EMPLOY_DAY
0 1988-02-18         18
1 2015-02-01          1
2 2020-02-01          1
3 2018-09-01          1
4 2020-06-29         29
```

Catagorical encoding

```
categorical_data

                                                  AGENCYNAME  \
0       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1       OFFICE OF COURT ADMINISTRATION                 ...
2       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3       OFFICE OF COURT ADMINISTRATION                 ...
4       TEXAS DEPARTMENT OF CRIMINAL JUSTICE           ...
...                                                    ...
149476  STATE PRESERVATION BOARD                       ...
149477  STATE PRESERVATION BOARD                       ...
149478  STATE PRESERVATION BOARD                       ...
149479  STATE PRESERVATION BOARD                       ...
149480  STATE PRESERVATION BOARD                       ...
```

```
                                LASTNAME                            FIRSTNAME
MI  \
0       RUCKER                              MORTON
V
1       RUCKER                              MORTON
V
2       SPECIA JR                           JOHN
J
3       SPECIA JR                           JOHN
J
4       ONTIVEROS                           ESTHER

...                                   ...                                  ...
..
149476  WESSELS                             JOHN
P
149477  WINDHAM                             EVAN
A
149478  WRIGHT                              DERRICK
C
149479  YOUNG                               DOUGLAS
R
149480  ZUNKER                              GEORGIA
P

       CLASSCODE                                              CLASSTITLE  \
0       JD25      JUDGE, RETIRED                                    ...
1       3524      GENERAL COUNSEL IV                                ...
2       JD25      JUDGE, RETIRED                                    ...
3       3524      GENERAL COUNSEL IV                                ...
4       4504      CORREC  OFFICER IV                                ...
...          ...                                                    ...
149476  6232      SECURITY OFFICER III                              ...
149477  0302      WEB ADMINISTRATOR III                             ...
149478  0130      CUSTOMER SERVICE REP I                            ...
149479  1572      PROGRAM SPECIALIST III                            ...
149480  0130      CUSTOMER SERVICE REP I                            ...

              ETHNICITY           GENDER  \
0       WHITE             MALE
1       WHITE             MALE
2       WHITE             MALE
3       WHITE             MALE
4       HISPANIC          FEMALE
...                  ...                ...
149476  WHITE             MALE
149477  WHITE             FEMALE
149478  WHITE             MALE
149479  WHITE             MALE
```

```
149480  WHITE             FEMALE

                                              STATUS EMPLOYDATE
EMPLOY_DAY
0       URP - UNCLASSIFIED REGULAR PART-TIME     1988-02-18
18
1       CTP - CLASSIFIED TEMPORARY PART-TIME     2015-02-01
1
2       URP - UNCLASSIFIED REGULAR PART-TIME     2020-02-01
1
3       CTP - CLASSIFIED TEMPORARY PART-TIME     2018-09-01
1
4       CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-29
29
...                                          ...        ...          ..
.
149476  CRF - CLASSIFIED REGULAR FULL-TIME       2017-10-30
30
149477  CRF - CLASSIFIED REGULAR FULL-TIME       2015-07-13
13
149478  CRP - CLASSIFIED REGULAR PART-TIME       2012-10-15
15
149479  CRF - CLASSIFIED REGULAR FULL-TIME       1989-09-22
22
149480  CRP - CLASSIFIED REGULAR PART-TIME       2012-02-16
16

[149481 rows x 11 columns]
```

```
for col in df_new.columns:
    print(col,':',len(df_new[col].unique()),'labels')
```

```
AGENCY : 113 labels
AGENCYNAME : 113 labels
LASTNAME : 38675 labels
FIRSTNAME : 23559 labels
MI : 27 labels
CLASSCODE : 1481 labels
CLASSTITLE : 1422 labels
ETHNICITY : 6 labels
GENDER : 2 labels
STATUS : 11 labels
EMPLOYDATE : 6295 labels
HRLYRATE : 206 labels
HRSPERWK : 58 labels
MONTHLY : 40552 labels
ANNUAL : 40554 labels
STATENUMBER : 149465 labels
EMPLOY_DAY : 31 labels
```

```
df_new.drop('AGENCYNAME',axis=1,inplace=True)
```

Frequency encoding

```
class_encoding=categorical_data['CLASSCODE'].value_counts().to_dict()

class_encoding
```

```
{'4504    ': 9267,
 '4505    ': 7895,
 '4503    ': 4718,
 '5622    ': 3646,
 '5121    ': 3307,
 '9928    ': 2251,
 '5026    ': 1854,
 '5023    ': 1778,
 '0154    ': 1774,
 '4510    ': 1743,
 '0055    ': 1727,
 '0152    ': 1693,
 '5151    ': 1578,
 '0156    ': 1466,
 '5122    ': 1433,
 '5025    ': 1348,
 '2123    ': 1308,
 '2122    ': 1265,
 '0057    ': 1262,
 '1574    ': 1227,
 '0150    ': 1169,
 '5620    ': 1117,
 '0171    ': 1102,
 '9940    ': 1102,
 '1572    ': 1083,
 '1573    ': 1039,
 '0172    ': 1034,
 '4541    ': 1007,
 '5024    ': 976,
 '1603    ': 870,
 '4411    ': 857,
 '1575    ': 857,
 '5505    ': 827,
 '2128    ': 794,
 '1604    ': 771,
 '1621    ': 765,
 '4511    ': 744,
 '5624    ': 741,
 '4412    ': 696,
 '2129    ': 696,
 '2127    ': 690,
```

```
'1601    ': 688,
'5017    ': 682,
'5123    ': 669,
'9022    ': 661,
'0059    ': 627,
'1353    ': 622,
'2119    ': 616,
'1622    ': 589,
'9055    ': 588,
'0256    ': 583,
'4413    ': 569,
'4540    ': 569,
'5702    ': 567,
'2124    ': 556,
'1571    ': 554,
'1602    ': 541,
'1584    ': 537,
'5152    ': 535,
'0173    ': 533,
'4421    ': 529,
'7110    ': 527,
'1600    ': 502,
'0255    ': 494,
'5016    ': 489,
'1570    ': 480,
'0132    ': 476,
'8003    ': 475,
'5541    ': 474,
'D520    ': 472,
'1323    ': 463,
'8103    ': 454,
'0134    ': 447,
'1582    ': 435,
'1576    ': 435,
'3504    ': 430,
'4524    ': 427,
'1913    ': 426,
'1620    ': 425,
'5153    ': 421,
'2130    ': 412,
'8262    ': 408,
'9987    ': 392,
'1014    ': 388,
'5027    ': 386,
'6232    ': 385,
'5543    ': 383,
'2685    ': 373,
'1623    ': 372,
'8109    ': 370,
```

```
'1984    ': 354,
'1354    ': 353,
'5124    ': 351,
'8110    ': 341,
'5542    ': 338,
'3505    ': 336,
'JD25    ': 324,
'9043    ': 322,
'1355    ': 321,
'2683    ': 319,
'5051    ': 314,
'1784    ': 312,
'9308    ': 311,
'5632    ': 310,
'9041    ': 305,
'1586    ': 303,
'0257    ': 302,
'1325    ': 302,
'1324    ': 293,
'4542    ': 293,
'1018    ': 293,
'1075    ': 291,
'1580    ': 282,
'2684    ': 282,
'4522    ': 276,
'1016    ': 275,
'4423    ': 274,
'4512    ': 272,
'0136    ': 269,
'2156    ': 266,
'8005    ': 263,
'1735    ': 256,
'1561    ': 255,
'0158    ': 254,
'2924    ': 253,
'5228    ': 247,
'9935    ': 246,
'2686    ': 245,
'1583    ': 238,
'4422    ': 236,
'0230    ': 236,
'9044    ': 234,
'5052    ': 231,
'1982    ': 230,
'9941    ': 229,
'2131    ': 227,
'5062    ': 222,
'0130    ': 221,
'1020    ': 221,
```

```
'0160    ': 219,
'5005    ': 218,
'5630    ': 217,
'2155    ': 212,
'0244    ': 211,
'5033    ': 206,
'0254    ': 206,
'1625    ': 205,
'0174    ': 202,
'1986    ': 200,
'1284    ': 199,
'1980    ': 199,
'7104    ': 199,
'5154    ': 198,
'5002    ': 197,
'0162    ': 196,
'9045    ': 196,
'3513    ': 195,
'5010    ': 194,
'5018    ': 188,
'5700    ': 187,
'5552    ': 186,
'1560    ': 182,
'4523    ': 181,
'7109    ': 181,
'3020    ': 180,
'1050    ': 179,
'5065    ': 179,
'5053    ': 178,
'1076    ': 173,
'9419    ': 171,
'1158    ': 169,
'8261    ': 169,
'0243    ': 167,
'1785    ': 165,
'3576    ': 164,
'T023    ': 163,
'3514    ': 161,
'6230    ': 161,
'1048    ': 160,
'1022    ': 157,
'1351    ': 157,
'5004    ': 155,
'C300    ': 154,
'7101    ': 154,
'5551    ': 152,
'1624    ': 152,
'1933    ': 152,
'8118    ': 152,
```

```
'0245    ': 152,
'3021    ': 151,
'1352    ': 149,
'9024    ': 148,
'1833    ': 147,
'5526    ': 147,
'2923    ': 147,
'9996    ': 147,
'1012    ': 147,
'1737    ': 144,
'2132    ': 144,
'9956    ': 144,
'9056    ': 143,
'1832    ': 142,
'1912    ': 140,
'5063    ': 140,
'3574    ': 137,
'7102    ': 137,
'T014    ': 134,
'0823    ': 134,
'2154    ': 133,
'5402    ': 132,
'0231    ': 131,
'4076    ': 130,
'3503    ': 128,
'T021    ': 127,
'5132    ': 127,
'1046    ': 127,
'0289    ': 125,
'1073    ': 123,
'1002    ': 123,
'3515    ': 122,
'5050    ': 121,
'5003    ': 121,
'1104    ': 119,
'0138    ': 118,
'5227    ': 118,
'4513    ': 117,
'4366    ': 117,
'5706    ': 117,
'1915    ': 116,
'1061    ': 116,
'1326    ': 116,
'4543    ': 115,
'2157    ': 115,
'5104    ': 115,
'9053    ': 114,
'5133    ': 114,
'1960    ': 113,
```

```
'1868    ': 113,
'5701    ': 111,
'4228    ': 111,
'1552    ': 110,
'5528    ': 110,
'0223    ': 110,
'4231    ': 109,
'5705    ': 109,
'1084    ': 108,
'2914    ': 108,
'8117    ': 107,
'0253    ': 106,
'1157    ': 105,
'4649    ': 105,
'9322    ': 105,
'2689    ': 104,
'9838    ': 104,
'1052    ': 104,
'0164    ': 104,
'1994    ': 102,
'1866    ': 102,
'1356    ': 102,
'8111    ': 101,
'9042    ': 101,
'4550    ': 100,
'6053    ': 100,
'1077    ': 99,
'3567    ': 99,
'3023    ': 98,
'4650    ': 98,
'0608    ': 98,
'0290    ': 98,
'8104    ': 98,
'1280    ': 97,
'5064    ': 96,
'4074    ': 95,
'1733    ': 95,
'1727    ': 95,
'1932    ': 94,
'2153    ': 94,
'4478    ': 93,
'3502    ': 92,
'1063    ': 92,
'3506    ': 90,
'5203    ': 90,
'2843    ': 90,
'1782    ': 90,
'2922    ': 90,
'1159    ': 89,
```

```
'4365    ': 88,
'1550    ': 88,
'4072    ': 88,
'5703    ': 88,
'4230    ': 88,
'1322    ': 88,
'6234    ': 87,
'HOUR    ': 86,
'1630    ': 86,
'1834    ': 86,
'0288    ': 85,
'S010    ': 84,
'1060    ': 84,
'8263    ': 83,
'3578    ': 81,
'2152    ': 81,
'5404    ': 81,
'2731    ': 81,
'2730    ': 80,
'2583    ': 80,
'7403    ': 80,
'3153    ': 80,
'9920    ': 79,
'1588    ': 79,
'1281    ': 79,
'1283    ': 78,
'3154    ': 78,
'1082    ': 78,
'1864    ': 77,
'3511    ': 77,
'1282    ': 77,
'1783    ': 77,
'3512    ': 77,
'6054    ': 76,
'1559    ': 76,
'9309    ': 75,
'5209    ': 75,
'2654    ': 75,
'1108    ': 74,
'0006    ': 73,
'A212    ': 73,
'5229    ': 72,
'8252    ': 72,
'3637    ': 72,
'5730    ': 71,
'9922    ': 71,
'5544    ': 70,
'4363    ': 70,
'5205    ': 70,
```

```
'2653    ': 69,
'3523    ': 69,
'3510    ': 67,
'1553    ': 67,
'9054    ': 67,
'9942    ': 67,
'5105    ': 66,
'1024    ': 65,
'0229    ': 65,
'8007    ': 65,
'2655    ': 64,
'1911    ': 64,
'1830    ': 64,
'1080    ': 64,
'6122    ': 63,
'0287    ': 63,
'9324    ': 62,
'1581    ': 62,
'9972    ': 62,
'6052    ': 62,
'0291    ': 61,
'CC23    ': 61,
'2913    ': 61,
'2640    ': 60,
'0242    ': 60,
'0653    ': 60,
'3660    ': 60,
'4493    ': 60,
'1781    ': 60,
'1914    ': 60,
'4473    ': 59,
'1059    ': 59,
'0606    ': 59,
'1551    ': 59,
'1934    ': 58,
'5704    ': 58,
'5106    ': 58,
'5081    ': 58,
'2842    ': 57,
'0654    ': 57,
'1731    ': 57,
'3531    ': 57,
'1976    ': 57,
'1922    ': 56,
'1286    ': 56,
'1112    ': 56,
'0519    ': 55,
'2652    ': 55,
'1106    ': 55,
```

```
'4674    ': 55,
'4551    ': 54,
'5207    ': 54,
'4084    ': 54,
'1626    ': 54,
'1921    ': 54,
'0331    ': 53,
'1085    ': 53,
'1831    ': 53,
'5040    ': 53,
'3682    ': 52,
'0518    ': 52,
'1000    ': 52,
'9971    ': 52,
'0520    ': 52,
'9323    ': 52,
'3663    ': 51,
'4651    ': 51,
'5529    ': 51,
'3522    ': 51,
'1110    ': 51,
'3530    ': 51,
'1786    ': 50,
'5527    ': 50,
'1974    ': 50,
'6160    ': 49,
'1074    ': 49,
'1935    ': 49,
'4414    ': 49,
'0604    ': 49,
'2086    ': 49,
'0214    ': 48,
'6055    ': 48,
'6162    ': 48,
'0302    ': 48,
'5032    ': 48,
'3171    ': 47,
'8108    ': 47,
'3604    ': 47,
'3524    ': 47,
'2688    ': 46,
'4390    ': 46,
'2064    ': 46,
'3662    ': 46,
'1554    ': 45,
'4416    ': 45,
'9992    ': 45,
'4499    ': 44,
'4018    ': 44,
```

```
'5235    ': 44,
'5082    ': 44,
'5092    ': 43,
'0272    ': 43,
'4438    ': 42,
'2690    ': 42,
'2365    ': 42,
'3683    ': 42,
'1102    ': 42,
'5006    ': 42,
'0313    ': 42,
'4560    ': 41,
'4452    ': 41,
'1739    ': 41,
'5011    ': 41,
'7105    ': 41,
'1860    ': 41,
'0222    ': 41,
'2915    ': 41,
'3572    ': 41,
'0812    ': 41,
'6252    ': 41,
'4083    ': 40,
'0652    ': 40,
'0213    ': 40,
'0224    ': 40,
'9418    ': 40,
'3642    ': 40,
'0236    ': 40,
'1156    ': 40,
'5226    ': 40,
'5111    ': 39,
'3568    ': 39,
'5112    ': 39,
'0228    ': 39,
'9981    ': 39,
'6100    ': 39,
'0237    ': 39,
'4451    ': 39,
'0333    ': 38,
'1729    ': 38,
'2475    ': 38,
'8021    ': 38,
'1992    ': 38,
'0651    ': 38,
'1923    ': 37,
'1133    ': 37,
'1862    ': 37,
'5031    ': 37,
```

```
'0824    ': 37,
'6099    ': 37,
'2084    ': 37,
'1890    ': 36,
'9420    ': 36,
'1100    ': 36,
'1246    ': 36,
'0258    ': 36,
'1285    ': 35,
'5540    ': 35,
'0322    ': 35,
'0284    ': 35,
'1062    ': 35,
'0813    ': 35,
'9945    ': 35,
'3644    ': 35,
'0282    ': 35,
'5400    ': 35,
'5626    ': 35,
'3516    ': 34,
'4526    ': 34,
'1605    ': 34,
'1044    ': 34,
'1558    ': 34,
'1143    ': 34,
'0517    ': 34,
'9906    ': 33,
'8033    ': 33,
'2088    ': 33,
'1142    ': 33,
'0814    ': 33,
'9814    ': 33,
'2366    ': 33,
'1920    ': 33,
'0235    ': 33,
'3610    ': 32,
'1931    ': 32,
'4675    ': 32,
'9816    ': 32,
'0320    ': 32,
'5504    ': 32,
'0314    ': 32,
'7404    ': 31,
'3580    ': 31,
'9960    ': 31,
'3566    ': 31,
'4403    ': 31,
'4648    ': 30,
'4498    ': 30,
'4494    ': 30,
```

```
'1814    ': 30,
'4489    ': 30,
'4544    ': 30,
'3684    ': 30,
'2465    ': 30,
'2844    ': 29,
'4552    ': 29,
'3025    ': 29,
'4464    ': 29,
'0283    ': 29,
'AUD3    ': 29,
'9839    ': 29,
'0334    ': 28,
'5034    ': 28,
'5030    ': 28,
'8032    ': 28,
'2701    ': 28,
'9062    ': 28,
'2703    ': 27,
'4146    ': 27,
'0602    ': 27,
'5107    ': 27,
'1892    ': 27,
'2474    ': 27,
'4216    ': 27,
'3624    ': 27,
'4561    ': 27,
'5091    ': 27,
'0822    ': 26,
'1780    ': 26,
'2264    ': 26,
'4562    ': 26,
'4521    ': 26,
'3672    ': 26,
'6170    ': 26,
'4221    ': 26,
'4226    ': 26,
'4437    ': 26,
'1650    ': 26,
'4492    ': 26,
'0271    ': 26,
'1287    ': 26,
'0650    ': 26,
'1962    ': 26,
'2733    ': 25,
'9950    ': 25,
'1132    ': 25,
'5054    ': 25,
'2161    ': 25,
```

```
'8119    ': 25,
'9704    ': 25,
'1242    ': 24,
'7310    ': 24,
'0170    ': 24,
'1930    ': 24,
'2921    ': 24,
'0303    ': 24,
'6172    ': 24,
'6057    ': 24,
'0241    ': 24,
'9997    ': 24,
'9834    ': 24,
'2266    ': 24,
'2082    ': 23,
'4392    ': 23,
'1144    ': 23,
'0292    ': 23,
'3532    ': 23,
'9806    ': 23,
'9064    ': 23,
'1812    ': 23,
'4223    ': 23,
'P071    ': 23,
'2585    ': 23,
'0335    ': 22,
'0315    ': 22,
'4224    ': 22,
'3681    ': 22,
'9991    ': 22,
'1990    ': 22,
'1631    ': 22,
'3640    ': 22,
'5113    ': 22,
'1872    ': 22,
'1321    ': 21,
'2702    ': 21,
'0285    ': 21,
'0600    ': 21,
'4404    ': 21,
'2732    ': 21,
'2473    ': 21,
'5108    ': 21,
'2584    ': 21,
'5083    ': 21,
'MSA1    ': 21,
'4479    ': 21,
'4017    ': 20,
'7108    ': 20,
```

```
'3525    ': 20,
'2464    ': 20,
'9416    ': 20,
'1825    ': 20,
'6098    ': 20,
'AUD4    ': 19,
'6253    ': 19,
'6010    ': 19,
'4483    ': 19,
'4364    ': 19,
'9812    ': 19,
'9417    ': 19,
'4229    ': 19,
'0273    ': 19,
'4227    ': 19,
'2651    ': 19,
'4127    ': 18,
'6117    ': 18,
'9808    ': 18,
'9961    ': 18,
'AUD5    ': 18,
'6097    ': 18,
'8025    ': 18,
'1141    ': 18,
'9894    ': 18,
'9998    ': 18,
'4222    ': 18,
'5503    ': 18,
'1871    ': 18,
'2476    ': 18,
'2700    ': 18,
'1823    ': 18,
'4344    ': 18,
'2065    ': 17,
'0301    ': 17,
'2682    ': 17,
'0312    ': 17,
'8302    ': 17,
'0212    ': 16,
'2125    ': 16,
'2090    ': 16,
'2742    ': 16,
'6174    ': 16,
'2761    ': 16,
'6095    ': 16,
'AUD2    ': 16,
'8116    ': 16,
'P073    ': 16,
'5090    ': 16,
```

```
'2472    ': 16,
'8254    ': 16,
'9990    ': 16,
'2696    ': 16,
'7354    ': 16,
'5131    ': 16,
'5201    ': 16,
'4459    ': 15,
'5233    ': 15,
'0311    ': 15,
'1894    ': 15,
'9034    ': 15,
'4462    ': 15,
'0655    ': 15,
'6096    ': 15,
'2066    ': 15,
'3635    ': 15,
'9973    ': 15,
'1662    ': 15,
'1134    ': 15,
'4428    ': 15,
'1131    ': 15,
'4082    ': 14,
'0221    ': 14,
'2460    ': 14,
'F023    ': 14,
'0332    ': 14,
'4383    ': 14,
'2802    ': 14,
'0252    ': 14,
'1262    ': 14,
'7106    ': 14,
'0215    ': 14,
'4439    ': 13,
'0516    ': 13,
'4474    ': 13,
'2692    ': 13,
'1824    ': 13,
'AUD6    ': 13,
'2845    ': 13,
'2803    ': 13,
'7409    ': 13,
'8031    ': 13,
'P077    ': 13,
'4142    ': 13,
'2268    ': 13,
'1870    ': 13,
'5134    ': 13,
'4360    ': 13,
```

```
'6502    ': 13,
'7407    ': 12,
'0644    ': 12,
'A348    ': 12,
'6251    ': 12,
'2743    ': 12,
'7103    ': 12,
'5711    ': 12,
'5732    ': 12,
'5232    ': 12,
'9638    ': 12,
'1632    ': 12,
'9832    ': 12,
'4672    ': 12,
'1155    ': 12,
'M001    ': 12,
'0246    ': 12,
'1876    ': 12,
'1244    ': 12,
'9830    ': 12,
'1248    ': 12,
'4342    ': 12,
'2912    ': 12,
'3626    ': 12,
'2694    ': 12,
'5618    ': 12,
'0249    ': 11,
'3569    ': 11,
'4408    ': 11,
'4457    ': 11,
'0324    ': 11,
'3521    ': 11,
'4417    ': 11,
'2135    ': 11,
'4362    ': 11,
'5406    ': 11,
'2805    ': 11,
'0304    ': 11,
'0294    ': 11,
'4453    ': 11,
'1882    ': 11,
'9995    ': 11,
'5109    ': 11,
'2056    ': 11,
'2364    ': 11,
'4218    ': 11,
'5720    ': 11,
'1316    ': 11,
'3630    ': 11,
```

```
'0590    ': 11,
'4078    ': 10,
'4294    ': 10,
'4001    ': 10,
'4405    ': 10,
'2734    ': 10,
'2110    ': 10,
'4465    ': 10,
'9306    ': 10,
'1066    ': 10,
'0248    ': 10,
'1290    ': 10,
'4520    ': 9,
'1272    ': 9,
'9802    ': 9,
'1294    ': 9,
'0238    ': 9,
'2762    ': 9,
'7402    ': 9,
'4394    ': 9,
'0264    ': 9,
'3666    ': 9,
'PT01    ': 9,
'2698    ': 9,
'1642    ': 9,
'0642    ': 9,
'9804    ': 9,
'4002    ': 9,
'4129    ': 9,
'4144    ': 9,
'6056    ': 9,
'1350    ': 9,
'2080    ': 9,
'4225    ': 9,
'5234    ': 9,
'2120    ': 9,
'9907    ': 9,
'0211    ': 8,
'2001    ': 8,
'U102    ': 8,
'3646    ': 8,
'4214    ': 8,
'0317    ': 8,
'9836    ': 8,
'1810    ': 8,
'P072    ': 8,
'5079    ': 8,
'0640    ': 8,
'J010    ': 8,
```

```
'4324    ': 8,
'3151    ': 8,
'7468    ': 8,
'2642    ': 8,
'P076    ': 8,
'2466    ': 8,
'A178    ': 8,
'7470    ': 8,
'3026    ': 8,
'9734    ': 8,
'J015    ': 8,
'2806    ': 8,
'1822    ': 8,
'1068    ': 8,
'1816    ': 8,
'4080    ': 8,
'5119    ': 8,
'AMGR    ': 8,
'0820    ': 7,
'1130    ': 7,
'5142    ': 7,
'5506    ': 7,
'1633    ': 7,
'2804    ': 7,
'4455    ': 7,
'2062    ': 7,
'U107    ': 7,
'0260    ': 7,
'6255    ': 7,
'9993    ': 7,
'4436    ': 7,
'8034    ': 7,
'7012    ': 7,
'4603    ': 7,
'1064    ': 7,
'0821    ': 7,
'0300    ': 7,
'0270    ': 7,
'1660    ': 7,
'6116    ': 7,
'4402    ': 7,
'2182    ': 7,
'4220    ': 7,
'1875    ': 7,
'A341    ': 7,
'2641    ': 7,
'2181    ': 7,
'7352    ': 7,
'7107    ': 7,
```

```
'2704    ': 7,
'4148    ': 7,
'8023    ': 6,
'4531    ': 6,
'2720    ': 6,
'0630    ': 6,
'6243    ': 6,
'0263    ': 6,
'2744    ': 6,
'A342    ': 6,
'2661    ': 6,
'2058    ': 6,
'1661    ': 6,
'B025    ': 6,
'1292    ': 6,
'4426    ': 6,
'9706    ': 6,
'4418    ': 6,
'5628    ': 6,
'4016    ': 6,
'1065    ': 6,
'3665    ': 6,
'S120    ': 6,
'U105    ': 6,
'2705    ': 6,
'A344    ': 6,
'9037    ': 6,
'1261    ': 6,
'4327    ': 6,
'1998    ': 6,
'5616    ': 6,
'9974    ': 6,
'2360    ': 6,
'1145    ': 5,
'9036    ': 5,
'AA01    ': 5,
'5617    ': 5,
'S249    ': 5,
'3150    ': 5,
'4440    ': 5,
'1826    ': 5,
'0265    ': 5,
'P500    ': 5,
'1995    ': 5,
'1842    ': 5,
'4676    ': 5,
'9060    ': 5,
'2663    ': 5,
'1045    ': 5,
```

```
'1877    ': 5,
'3065    ': 5,
'4346    ': 5,
'5315    ': 5,
'5144    ': 5,
'1562    ': 5,
'2740    ': 5,
'2741    ': 5,
'6229    ': 5,
'A177    ': 5,
'4532    ': 5,
'4530    ': 5,
'1263    ': 5,
'2808    ': 5,
'3001    ': 5,
'4533    ': 5,
'1274    ': 5,
'1293    ': 5,
'1034    ': 5,
'5330    ': 5,
'9824    ': 5,
'A343    ': 5,
'1276    ': 5,
'4040    ': 5,
'7464    ': 5,
'2094    ': 5,
'9963    ': 5,
'2662    ': 5,
'1295    ': 5,
'C235    ': 5,
'SRSA    ': 4,
'1140    ': 4,
'7042    ': 4,
'4517    ': 4,
'0295    ': 4,
'2260    ': 4,
'4477    ': 4,
'SRIN    ': 4,
'7405    ': 4,
'0274    ': 4,
'9628    ': 4,
'6120    ': 4,
'6115    ': 4,
'7319    ': 4,
'0217    ': 4,
'MSRA    ': 4,
'PT02    ': 4,
'3622    ': 4,
'3107    ': 4,
```

```
 '5140    ': 4,
 '2305    ': 4,
 '4435    ': 4,
 '3239    ': 4,
 '9035    ': 4,
 '1067    ': 4,
 'P078    ': 4,
 'SCA1    ': 4,
 'P070    ': 4,
 '9999    ': 4,
 '3534    ': 4,
 '4471    ': 4,
 '4385    ': 4,
 '7350    ': 4,
 '7308    ': 4,
 '5212    ': 4,
 '3674    ': 4,
 'QCR2    ': 4,
 '7466    ': 4,
 '1841    ': 4,
 '3692    ': 4,
 '1843    ': 4,
 '6241    ': 3,
 '3004    ': 3,
 '3667    ': 3,
 '4293    ': 3,
 '0646    ': 3,
 ...}
```

```
categorical_data['CLASSCODE']=categorical_data['CLASSCODE'].map(class_
encoding)
categorical_data.head(10)
```

```
                                          AGENCYNAME  \
0  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1  OFFICE OF COURT ADMINISTRATION                ...
2  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3  OFFICE OF COURT ADMINISTRATION                ...
4  TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...
5  OFFICE OF THE ATTORNEY GENERAL                ...
6  TEXAS DEPARTMENT OF TRANSPORTATION            ...
7  TEXAS BEHAVIORAL HEALTH EXECUTIVE COUNCIL     ...
8  BOARD OF EXAMINERS OF PSYCHOLOGISTS           ...
9  DEPARTMENT OF STATE HEALTH SERVICES           ...

                          LASTNAME                       FIRSTNAME
MI  \
0  RUCKER                          MORTON                          V

1  RUCKER                          MORTON                          V
```

```
2  SPECIA JR                          JOHN                                J

3  SPECIA JR                          JOHN                                J

4  ONTIVEROS                          ESTHER

5  ROGERS                             SHAUNA

6  RICHTER                            WILLIAM                             J

7  SPINKS                             DARREL                              D

8  SPINKS                             DARREL                              D

9  ADAMS III                          LEE                                 A


   CLASSCODE                                          CLASSTITLE  \
0        324  JUDGE, RETIRED                                  ...
1         47  GENERAL COUNSEL IV                              ...
2        324  JUDGE, RETIRED                                  ...
3         47  GENERAL COUNSEL IV                              ...
4       9267  CORREC  OFFICER IV                              ...
5        372  DIRECTOR IV                                     ...
6        354  CONTRACT SPEC IV                                ...
7        372  DIRECTOR IV                                     ...
8          1  EXEC DIR, BD OF EXAMS OF PSYCHOLOGISTS          ...
9        463  INSPECTOR III                                   ...

         ETHNICITY           GENDER
STATUS  \
0  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
1  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
2  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
3  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
4  HISPANIC         FEMALE           CRF - CLASSIFIED REGULAR FULL-
TIME
5  HISPANIC         FEMALE           CRF - CLASSIFIED REGULAR FULL-
TIME
6  WHITE            MALE             CRF - CLASSIFIED REGULAR FULL-
TIME
7  WHITE            MALE             CRF - CLASSIFIED REGULAR FULL-
TIME
8  WHITE            MALE             ERP - EXEMPT REGULAR PART-TIME
```

```
9  BLACK            MALE              CRF - CLASSIFIED REGULAR FULL-
TIME

  EMPLOYDATE  EMPLOY_DAY
0 1988-02-18         18
1 2015-02-01          1
2 2020-02-01          1
3 2018-09-01          1
4 2020-06-29         29
5 2020-04-01          1
6 2020-06-22         22
7 2020-03-01          1
8 2020-03-04          4
9 2019-09-01          1
```

```
class_title_encoding=categorical_data['CLASSTITLE'].value_counts().to_
dict()
```

```
class_title_encoding
```

```
{'CORREC  OFFICER IV                              ': 9267,
 'CORREC  OFFICER V                               ': 7895,
 'CORREC OFFCR III                                ': 4718,
 'TEXAS WORKS ADVISOR II                          ': 3646,
 'DIRECT SUPPORT PROFESSIONAL I                   ': 3307,
 'TROOPER                                         ': 2251,
 'CHILD PROTECTIVE SVCS SPEC IV                   ': 1854,
 'CHILD PROTECTIVE SVCS SPEC I                    ': 1773,
 'ADMINISTRATIVE ASST III                         ': 1771,
 'SGT OF CORREC OFFCRS                            ': 1743,
 'CLERK I                                         ': 1727,
 'ADMINISTRATIVE ASST II                          ': 1690,
 'PSYCHIATRIC NURSING ASST I                      ': 1578,
 'ADMINISTRATIVE ASST IV                          ': 1465,
 'DIRECT SUPPORT PROF II                          ': 1433,
 'CHILD PROTECTIVE SVCS SPEC III                  ': 1348,
 'ENGINEERING TECHNICIAN II                       ': 1308,
 'ENGINEERING TECHNICIAN I                        ': 1265,
 'CLERK II                                        ': 1262,
 'PROGRAM SPECIALIST V                            ': 1230,
 'ADMINISTRATIVE ASST I                           ': 1169,
 'TEXAS WORKS ADVISOR I                           ': 1117,
 'LICENSE AND PERMIT SPEC II                      ': 1102,
 'SERGEANT, DPS                                   ': 1102,
 'PROGRAM SPECIALIST III                          ': 1082,
 'PROGRAM SPECIALIST IV                           ': 1039,
 'LICENSE AND PERMIT SPEC III                     ': 1034,
 'PAROLE OFFCR II                                 ': 1007,
 'CHILD PROTECTIVE SVCS SPEC II                   ': 976,
 'MGR IV                                          ': 869,
```

```
'PROGRAM SPECIALIST VI                          ': 856,
'NURSE I                                        ': 854,
'HUMAN SERVICES TECHNICIAN III                  ': 827,
'ENGINEERING SPECIALIST II                      ': 794,
'MGR V                                          ': 769,
'DIRECTOR II                                    ': 765,
'LT OF CORREC OFFCRS                            ': 744,
'TEXAS WORKS ADVISOR III                        ': 741,
'NURSE II                                       ': 696,
'ENGINEERING SPECIALIST III                     ': 696,
'ENGINEERING SPECIALIST I                       ': 690,
'MGR II                                         ': 688,
'FAMILY & PROTECT SCVS SUPR II                  ': 682,
'DIRECT SUPPORT PROF III                        ': 669,
'EQUIPMENT OPERATOR I                           ': 661,
'CLERK III                                      ': 627,
'INVESTIGATOR IV                                ': 622,
'ENGINEERING AIDE                               ': 616,
'DIRECTOR III                                   ': 589,
'MAINTENANCE SUPERVISOR III                     ': 588,
'SYSTEMS ANALYST V                              ': 585,
'NURSE III                                      ': 569,
'PAROLE OFFCR I                                 ': 569,
'H/SRVC SPEC III                                ': 567,
'ENGINEERING TECHNICIAN III                     ': 556,
'PROGRAM SPECIALIST II                          ': 555,
'MGR III                                        ': 541,
'PROGRAM SUPERVISOR V                           ': 537,
'PSYCHIATRIC NURSING ASST II                    ': 535,
'LICENSE AND PERMIT SPEC IV                     ': 533,
'LICENSED VOCATIONAL NURSE I                    ': 528,
'ELECTED OFFICIALS STAFF                        ': 527,
'SYSTEMS ANALYST IV                             ': 504,
'MGR I                                          ': 502,
'FAMILY & PROTECT SCVS SUPR I                   ': 488,
'PROGRAM SPECIALIST I                           ': 480,
'CUSTOMER SERVICE REP II                        ': 476,
'CUSTODIAN I                                    ': 475,
'CHILD SUPPORT OFFCR II                         ': 474,
'DISTRICT JUDGE, ACTIVE                         ': 472,
'INSPECTOR III                                  ': 463,
'FOOD SERVICE WORKER I                          ': 454,
'CUSTOMER SERVICE REP III                       ': 447,
'PROGRAM SPECIALIST VII                         ': 437,
'PROGRAM SUPERVISOR III                         ': 435,
'ATTORNEY III                                   ': 430,
'JUVENILE CORREC OFFCR V                        ': 427,
'INVENTORY & STORE SPEC  III                    ': 426,
'DIRECTOR I                                     ': 425,
```

```
'PSYCHIATRIC NURSING ASST III                    ': 421,
'ENGINEERING SPECIALIST IV                       ': 410,
'LAUNDRY MGR III                                 ': 408,
'ACCOUNTANT II                                   ': 393,
'GAME WARDEN                                     ': 392,
'CHILD PROTECTIVE SVCS SPEC V                    ': 386,
'SECURITY OFFICER III                            ': 385,
'CHILD SUPPORT OFFCR IV                          ': 383,
'NATURAL RESOURCES SPEC IV                       ': 373,
'DIRECTOR IV                                     ': 372,
'FOOD SRVC MGR II                                ': 370,
'CONTRACT SPEC IV                                ': 354,
'INVESTIGATOR V                                  ': 353,
'DIRECT SUPPORT PROF IV                          ': 351,
'FOOD SRVC MGR III                               ': 341,
'CHILD SUPPORT OFFCR III                         ': 338,
'ATTORNEY IV                                     ': 336,
'JUDGE, RETIRED                                  ': 324,
'MAINTENANCE SPECIALIST III                      ': 322,
'INVESTIGATOR VI                                 ': 321,
'NATURAL RESOURCES SPEC II                       ': 319,
'REHAB THERAPY TECHNICIAN II                     ': 314,
'TRAINING SPEC IV                                ': 312,
'TRANS MAINT SPEC IV                             ': 311,
'TEXAS WORKS SUPERVISOR II                       ': 310,
'MAINTENANCE SPECIALIST I                        ': 305,
'PROGRAM SUPERVISOR VI                           ': 303,
'INSPECTOR V                                     ': 302,
'SYSTEMS ANALYST VI                              ': 301,
'ACCOUNTANT IV                                   ': 293,
'PAROLE OFFCR III                                ': 293,
'INSPECTOR IV                                    ': 293,
'PROGRAM SUPERVISOR I                            ': 282,
'NATURAL RESOURCES SPEC III                      ': 282,
'ACCTS EXAM III                                  ': 280,
'JUVENILE CORREC OFFCR III                       ': 276,
'ACCOUNTANT III                                  ': 275,
'CAPT OF CORREC OFFCRS                           ': 272,
'CUSTOMER SERVICE REP IV                         ': 269,
'LICENSED VOCATIONAL NURSE III                   ': 267,
'ENGINEER V                                      ': 266,
'CUSTODIAN II                                    ': 263,
'PROJECT MANAGER IV                              ': 256,
'HR SPEC IV                                      ': 256,
'ADMINISTRATIVE ASST V                           ': 254,
'CLAIMS EXAMINER IV                              ': 253,
'CASE MGR III                                    ': 247,
'CORPORAL                                        ': 246,
'NATURAL RESOURCES SPEC V                        ': 245,
```

```
'PROGRAM SUPERVISOR IV                          ': 238,
'SYSTEMS SUPPORT SPECIALIST III                 ': 238,
'MAINTENANCE SPECIALIST IV                      ': 234,
'REHAB THERAPY TECHNICIAN III                   ': 231,
'CONTRACT SPEC III                              ': 230,
'LIEUTENANT, DPS                                ': 229,
'ENGINEERING SPECIALIST V                       ': 227,
'LICENSED VOCATIONAL NURSE II                   ': 225,
'ACCOUNTANT V                                   ': 223,
'VOC REHAB CNSLR I                              ': 222,
'CUSTOMER SERVICE REP I                         ': 221,
'EXECUTIVE ASST I                               ': 219,
'ADULT PROTECTIVE SVCS SPEC IV                  ': 218,
'TEXAS WORKS SUPERVISOR I                       ': 217,
'ENGINEER IV                                    ': 212,
'PROGRAMMER IV                                  ': 211,
'SYSTEMS ANALYST III                            ': 206,
'DIRECTOR VI                                    ': 205,
'LICENSE AND PERMIT SPEC V                      ': 202,
'PROTECT SRVCS INTAKE SPEC IV                   ': 201,
'CONTRACT SPEC V                                ': 200,
'CONTRACT SPEC II                               ': 199,
'LEGISLATIVE PROFESSIONAL                       ': 199,
'TAX AUDITOR V                                  ': 199,
'PSYCHIATRIC NURSING ASST IV                    ': 198,
'AUDITOR IV                                     ': 197,
'MAINTENANCE SPECIALIST V                       ': 196,
'EXECUTIVE ASST II                              ': 196,
'ASST ATTORNEY GENERAL IV                       ': 195,
'FAMILY SERVICES SPECIALIST I                   ': 194,
'AUDITOR III                                    ': 189,
'ADULT PROTECTIVE SVCS SPEC I                   ': 189,
'FAMILY & PROTECT SVCS SUPR III                 ': 188,
'H/SRVC SPEC I                                  ': 187,
'CHILD SUPPORT TECHN III                        ': 186,
'PROJECT MANAGER III                            ': 182,
'JUVENILE CORREC OFFCR IV                       ': 181,
'ELECTED OFFICIAL                               ': 181,
'VOC REHAB CNSLR IV                             ': 179,
'WORKFORCE DEV SPECIALIST I                     ': 179,
'REHAB THERAPY TECHNICIAN IV                    ': 178,
'ACCTS EXAM IV                                  ': 173,
'MOTOR VEHICLE TECHNICIAN IV                    ': 171,
'LAUNDRY MGR II                                 ': 169,
'BUDGET ANALYST IV                              ': 169,
'PROGRAMMER III                                 ': 167,
'TRAINING SPEC V                                ': 165,
'LEGAL ASSISTANT III                            ': 164,
'PRINCIPALS, TEACHERS, SPRVSR                   ': 163,
```

```
'PROGRAMMER V                                  ': 163,
'ASST ATTORNEY GENERAL V                       ': 161,
'SECURITY OFFICER II                           ': 161,
'ACCOUNTANT VI                                 ': 158,
'INVESTIGATOR II                               ': 157,
'CRIMINAL DISTRICT ATTORNEY                    ': 154,
'ADULT PROTECTIVE SVCS SPEC III                ': 153,
'COOK III                                      ': 152,
'PURCHASER IV                                  ': 152,
'DIRECTOR V                                    ': 152,
'CHILD SUPPORT TECHN II                        ': 152,
'WORKFORCE DEV SPECIALIST II                   ': 150,
'ACCOUNTANT I                                  ': 149,
'INVESTIGATOR III                              ': 149,
'LEG. OFFICIAL/ADMINISTRATOR                   ': 149,
'EQUIPMENT OPERATOR II                         ': 148,
'QUALITY ASSURANCE SPEC I                      ': 147,
"SERGEANT, TEXAS AG'S OFFICE                   ": 147,
'CLAIMS EXAMINER III                           ': 147,
'INFO SPEC IV                                  ': 147,
'AGENT                                         ': 144,
'HR SPEC V                                     ': 144,
'ENGINEERING SPECIALIST VI                     ': 144,
'MAINTENANCE SUPERVISOR IV                     ': 143,
'AUDITOR II                                    ': 142,
'INFO SPEC III                                 ': 142,
'INVENTORY & STORE SPEC  II                    ': 140,
'VOC REHAB CNSLR II                            ': 140,
'LEGAL ASSISTANT II                            ': 137,
'EDUCATION SPECIALIST IV                       ': 134,
'PRINC, TEACHERS, & COACHES                    ': 134,
'ENGINEER III                                  ': 133,
'SYSTEMS SUPPORT SPECIALIST IV                 ': 131,
'SOCIAL WORKER II                              ': 131,
'PUBLIC HLTH & PRVNT SPEC III                  ': 130,
'ATTORNEY II                                   ': 128,
'TEACHER,PRINCIPALS,SUPERVISORS                ': 127,
'QUAL INTEL DISABILITY PROF II                 ': 127,
'NETWORK SPEC III                              ': 125,
'ACCOUNTING TECHNICIAN II                      ': 123,
'ACCTS EXAM I                                  ': 123,
'ASST ATTORNEY GENERAL VI                      ': 122,
'AUDITOR V                                     ': 122,
'REHAB THERAPY TECHNICIAN I                    ': 121,
'ADULT PROTECTIVE SVCS SPEC II                 ': 120,
'CASE MGR II                                   ': 118,
'CUSTOMER SERVICE REP V                        ': 118,
'FINANCIAL EXAM III                            ': 118,
'REGISTERED THERAPIST V                        ': 117,
```

```
'H/SRVC SPEC VII                                ': 117,
'MAJ OF CORREC OFFCRS                           ': 117,
'INVENTORY AND STORE SPEC V                     ': 116,
'INSPECTOR VI                                   ': 116,
'ENGINEER VI                                    ': 115,
'TAXPYR COMPLIANCE OFFCR III                    ': 115,
'PAROLE OFFCR IV                                ': 115,
'VETERANS SERVICES REP I                        ': 115,
'MAINTENANCE SUPERVISOR I                       ': 114,
'QUAL INTEL DISABILITY PROF III                 ': 114,
'CONTRACT ADMIN MANAGER I                       ': 113,
'MANAGEMENT ANALYST V                           ': 113,
'HEALTH SPECIALIST III                          ': 111,
'H/SRVC SPEC II                                 ': 111,
'QUALITY ASSURANCE SPEC III                     ': 110,
'BUSINESS ANALYST III                           ': 110,
'STAFF SRVCS OFFCR III                          ': 110,
'HEALTH ASSISTANT                               ': 109,
'H/SRVC SPEC VI                                 ': 109,
'FINANCIAL ANALYST III                          ': 109,
'RETIRE SYS BENEFITS SPEC III                   ': 108,
'SYSTEMS ANALYST II                             ': 107,
'COOK II                                        ': 107,
'BUDGET ANALYST III                             ': 105,
'INDUSTRIAL SPEC IV                             ': 105,
'VEHICLE DRIVER I                               ': 105,
'FISH AND WILDLIFE TECH II                      ': 104,
'EXECUTIVE ASST III                             ': 104,
'FERRYBOAT DECKHAND I                           ': 104,
'INVESTIGATOR VII                               ': 102,
'MANAGEMENT ANALYST IV                          ': 102,
'PROPERTY MGR III                               ': 102,
'FOOD SRVC MGR IV                               ': 101,
'LEGISLATIVE ADMIN. SUPPORT                     ': 101,
'MAINTENANCE SPECIALIST II                      ': 101,
'ASST WARDEN                                    ': 100,
'FORENSIC SCIENTIST II                          ': 100,
'ACCTS EXAM V                                   ': 99,
'LEGAL SECRETARY III                            ': 99,
'INDUSTRIAL SPEC V                              ': 98,
'RESEARCH SPEC V                                ': 98,
'WORKFORCE DEV SPECIALIST III                   ': 98,
'FOOD SERVICE WORKER II                         ': 98,
'NETWORK SPEC IV                                ': 98,
'TAX AUDITOR I                                  ': 97,
'VOC REHAB CNSLR III                            ': 96,
'HR SPEC III                                    ': 95,
'PURCHASER III                                  ': 95,
'HR ASST                                        ': 95,
```

```
'PUBLIC HLTH & PRVNT SPEC II                    ': 95,
'ENGINEER II                                    ': 94,
'PSYCHIATRIST III                               ': 93,
'ATTORNEY I                                     ': 92,
'TAXPYR COMPLIANCE OFFCR V                      ': 92,
'INSURANCE SPECIALIST II                        ': 90,
'ATTORNEY V                                     ': 90,
'CLAIMS EXAMINER II                             ': 90,
'RESIDENT SPECIALIST II                         ': 90,
'TRAINING SPEC II                               ': 90,
'BUDGET ANALYST V                               ': 89,
'STAFF SRVCS OFFCR I                            ': 88,
'REGISTERED THERAPIST IV                        ': 88,
'PUBLIC HLTH & PRVNT SPEC I                     ': 88,
'H/SRVC SPEC IV                                 ': 88,
'INSPECTOR II                                   ': 88,
'SECURITY OFFICER IV                            ': 87,
'HOURLY/SUBSTITUTES                             ': 86,
'DEPUTY DIRECTOR I                              ': 86,
'INFO SPEC V                                    ': 86,
'NETWORK SPEC II                                ': 85,
'LAUNDRY MGR IV                                 ': 83,
'SUBSTITUTE WORKERS                             ': 83,
'HEALTH SPECIALIST V                            ': 82,
'SAFETY OFFCR II                                ': 81,
'ENGINEER I                                     ': 81,
'SOCIAL WORKER III                              ': 81,
'LEGAL ASSISTANT IV                             ': 81,
'UNEMPL INSUR CLAIMS EXAM II                    ': 80,
'SAFETY OFFCR I                                 ': 80,
'LIBRARIAN III                                  ': 80,
'SANITARIAN I                                   ': 80,
'TROOPER TRAINEE                                ': 79,
'TAX AUDITOR II                                 ': 79,
'PROGRAM SUPERVISOR VII                         ': 79,
'FINANCIAL ANALYST II                           ': 79,
'RECEPTIONIST                                   ': 79,
'UNEMPL INSUR CLAIMS EXAM III                   ': 78,
'TAX AUDITOR IV                                 ': 78,
'MANAGEMENT ANALYST III                         ': 77,
'TRAINING SPEC III                              ': 77,
'ASST ATTORNEY GENERAL III                      ': 77,
'TAX AUDITOR III                                ': 77,
'ASST ATTORNEY GENERAL II                       ': 77,
'FORENSIC SCIENTIST III                         ': 76,
'PROJECT MANAGER II                             ': 76,
'TRANS MAINT SPEC V                             ': 75,
'ENVIRONMENTAL PROTECT SPEC IV                  ': 75,
'RESIDENT SPECIALIST V                          ': 75,
```

```
'FINANCIAL EXAM V                                  ': 74,
'TAXPYR COMPLIANCE OFFCR II                        ': 73,
'JUSTICE                                           ': 73,
'ASSOCIATE JUDGE                                   ': 73,
'LAUNDRY/SEWING ROOM WORKER I                      ': 72,
'COURT COORDINATOR                                 ': 72,
'CASE MGR IV                                       ': 72,
'GUARDIANSHIP SPECIALIST                           ': 71,
'PROBATIONARY TROOPER                              ': 71,
'CHILD SUPPORT OFFICER V                           ': 70,
'RESIDENT SPECIALIST III                           ': 70,
'REGISTERED THERAPIST II                           ': 70,
'GENERAL COUNSEL III                               ': 69,
'ENVIRONMENTAL PROTECT SPEC III                    ': 69,
'MAINTENANCE SUPERVISOR II                         ': 67,
'CAPTAIN, DPS                                      ': 67,
'ASST ATTORNEY GENERAL I                           ': 67,
'STAFF SRVCS OFFCR IV                              ': 67,
'SYSTEMS SUPPORT SPECIALIST II                     ': 65,
'CUSTODIAN III                                     ': 65,
'ACCOUNTANT VII                                    ': 65,
'INFO SPEC I                                       ': 64,
'ENVIRONMENTAL PROTECT SPEC V                      ': 64,
'INVENTORY & STORE SPEC I                          ': 64,
'VETERANS SERVICES REP II                          ': 63,
'NETWORK SPEC I                                    ': 63,
'CRIME LABORATORY SPEC III                         ': 63,
'DATA ANALYST IV                                   ': 62,
'FORENSIC SCIENTIST I                              ': 62,
'PROGRAM SUPERVISOR II                             ': 62,
'VEHICLE DRIVER III                                ': 62,
'INVESTIGATOR IV - OIG                             ': 62,
'NETWORK SPEC V                                    ': 61,
'RETIRE SYS BENEFITS SPEC II                       ': 61,
'RETIRED COUNTY COURT JUDGE                        ': 61,
'PARK RANGER I                                     ': 60,
'PROGRAMMER II                                     ': 60,
'PHARMACIST II                                     ': 60,
'TRAINING SPEC I                                   ': 60,
'OMBUDSMAN I                                       ': 60,
'INVENTORY & STORE SPEC IV                         ': 60,
'BEHAVIOR ANALYST I                                ': 59,
'STAFF SRVCS OFFCR II                              ': 59,
'RESEARCH SPEC IV                                  ': 59,
'TAXPYR COMPLIANCE OFFCR I                         ': 59,
'H/SRVC SPEC V                                     ': 58,
'DATA ANALYST V                                    ': 58,
'PURCHASER V                                       ': 58,
'VETERANS SERVICES REP III                         ': 58,
```

```
'CHAPLAIN I                                    ': 58,
'CONTRACT SPEC I                               ': 57,
'HEARINGS OFFICER II                           ': 57,
'INSURANCE SPECIALIST I                        ': 57,
'HR SPEC II                                    ': 57,
'FINANCIAL EXAM VII                            ': 56,
'GRANT COORDINATOR III                         ': 56,
'TAX AUDITOR SUPERVISOR                        ': 56,
'FINANCIAL EXAM IV                             ': 55,
'ENVIRONMENTAL PROTECT SPEC II                 ': 55,
'PLANNER IV                                    ': 55,
'AGRICULTURE SPEC IV                           ': 55,
'EPIDEMIOLOGIST III                            ': 54,
'RESIDENT SPECIALIST IV                        ': 54,
'GRANT COORDINATOR II                          ': 54,
'WARDEN I                                      ': 54,
'DOCUMENT SERVICES TECH I                      ': 53,
'FINANCIAL ANALYST IV                          ': 53,
'ADULT CHILD CARE LIC SPEC I                   ': 53,
'INFO SPEC II                                  ': 53,
'COMPLIANCE ANALYST II                         ': 52,
'PLANNER V                                     ': 52,
'INVESTIGATOR III - OIG                        ': 52,
'PLANNER III                                   ': 52,
'ACCOUNTING TECHNICIAN I                       ': 52,
'VEHICLE DRIVER II                             ': 52,
'DIRECTOR VII                                  ': 52,
'HEARINGS OFFICER I                            ': 51,
'OMBUDSMAN III                                 ': 51,
'GENERAL COUNSEL II                            ': 51,
'INDUSTRIAL SPEC VI                            ': 51,
'QUALITY ASSURANCE SPEC IV                     ': 51,
'CONTRACT TECHN III                            ': 50,
'TRAINING SPEC VI                              ': 50,
'FINANCIAL EXAM VI                             ': 50,
'QUALITY ASSURANCE SPEC II                     ': 50,
'PURCHASER VI                                  ': 49,
'ACCTS EXAM II                                 ': 49,
'CRIME ANALYST I                               ': 49,
'NURSE IV                                      ': 49,
'RESEARCH SPEC III                             ': 49,
'RIGHT OF WAY AGENT IV                         ': 49,
'PROTECT SRVCS INTAKE SPEC III                 ': 48,
'WEB ADMINISTRATOR III                         ': 48,
'CRIME ANALYST II                              ': 48,
'FORENSIC SCIENTIST IV                         ': 48,
'LAW CLERK                                     ': 47,
'UNEMPL INSUR SPEC I                           ': 47,
'FOOD SRVC MGR I                               ': 47,
```

```
'GENERAL COUNSEL IV                                ': 47,
'APPRAISER II                                      ': 46,
'FINANCIAL ANALYST I                               ': 46,
'DATA BASE ADMINISTRATOR IV                        ': 46,
'HEALTH PHYSICIST I                                ': 46,
'FISH AND WILDLIFE TECH I                          ': 46,
'OMBUDSMAN II                                      ': 46,
'PUBLIC HEALTH NURSE I                             ': 45,
'CAPT, GAME WARDEN                                 ': 45,
'STAFF SRVCS OFFCR V                               ': 45,
'CHAPLAIN II                                       ': 44,
'PHARMACY TECHNICIAN II                            ': 44,
'VOLUNTEER SERVICES COORD IV                       ': 44,
'DIETETIC & NUTRITION SPEC III                     ': 43,
'EXECUTIVE DIRECTOR                                ': 43,
'GEOGRAPHIC INFO SPEC III                          ': 43,
'REHABILITATION TEACHER III                        ': 43,
'FISH AND WILDLIFE TECH III                        ': 42,
'PHYSICIAN III                                     ': 42,
'ADULT PROTECTIVE SVCS SPEC V                      ': 42,
'COMPLIANCE ANALYST III                            ': 42,
'GEOSCIENTIST IV                                   ': 42,
'SYSTEMS ADMINISTRATOR IV                          ': 42,
'FAMILY SERVICES SPECIALIST II                     ': 41,
'LEGAL ASSISTANT I                                 ': 41,
'FINANCIAL EXAM II                                 ': 41,
'RETIRE SYS BENEFITS SPEC IV                       ': 41,
'LEGISLATIVE PARAPROFESSIONAL                      ': 41,
'MANAGEMENT ANALYST I                              ': 41,
'STATE PARK POLICE OFFICER                         ': 41,
'ADV PRACTICE RGD NURSE II                         ': 41,
'BUSINESS ANALYST II                               ': 41,
'TEACHER AIDE I                                    ': 41,
'HR SPEC VI                                        ': 41,
'COUNSEL SUBSTITUTE I                              ': 41,
'DATA BASE ADMINISTRATOR III                       ': 40,
'MOTOR VEHICLE TECHNICIAN III                      ': 40,
'ADMINISTRATIVE LAW JUDGE II                       ': 40,
'CASE MGR I                                        ': 40,
'INFO TECH SECURITY ANALYST II                     ': 40,
'DATA ANALYST III                                  ': 40,
'BUSINESS ANALYST IV                               ': 40,
'BUDGET ANALYST II                                 ': 40,
'EPIDEMIOLOGIST II                                 ': 40,
'SUBSTANCE ABUSE COUNSELOR I                       ': 39,
'LEGAL SECRETARY IV                                ': 39,
'SYSTEMS SUPPORT SPECIALIST I                      ': 39,
'ADV PRACTICE RGD NURSE I                          ': 39,
'PROBATIONARY GAME WARDEN                          ': 39,
```

```
'POLICE COMMUNICATIONS OPER VI                    ': 39,
'INFO TECH SECURITY ANALYST III                   ': 39,
'SUBSTANCE ABUSE COUNSELOR II                     ': 39,
'PROPERTY MGR II                                  ': 38,
'HR SPEC I                                        ': 38,
'CHEMIST IV                                       ': 38,
'DATA ANALYST II                                  ': 38,
'CUSTODIAL MGR I                                  ': 38,
'DOCUMENT SERVICES TECH III                       ': 38,
'MANAGEMENT ANALYST II                            ': 37,
'GRANT COORDINATOR IV                             ': 37,
'INVESTMENT ANALYST IV                            ': 37,
'RIGHT OF WAY AGENT III                           ': 37,
'POLICE COMMUNICATIONS OPER V                     ': 37,
'EDUCATION SPECIALIST V                           ': 37,
'REIMBURSEMENT OFFICER III                        ': 36,
'LEGISLATIVE ADMIN SUPPORT                        ': 36,
'SYSTEMS ANALYST VII                              ': 36,
'MOTOR VEHICLE TECHNICIAN V                       ': 36,
'GOVT RELATIONS SPECIALIST I                      ': 36,
'CYBERSECURITY ANALYST II                         ': 35,
'CHILD SUPPORT OFFCR I                            ': 35,
'TAX AUDITOR VI                                   ': 35,
'MAJOR, DPS                                       ': 35,
'TELECOMMUNICATIONS SPEC I                        ': 35,
'TAXPYR COMPLIANCE OFFCR IV                       ': 35,
'TEXAS WORKS ADVISOR IV                           ': 35,
'TELECOMMUNICATIONS SPEC III                      ': 35,
'TEACHER AIDE II                                  ': 35,
'ADMINISTRATIVE LAW JUDGE III                     ': 35,
'SOCIAL WORKER I                                  ': 35,
'PROJECT MANAGER I                                ': 34,
'PLANNER II                                       ': 34,
'ASST ATTORNEY GENERAL VII                        ': 34,
'DORM SUPERVISOR I                                ': 34,
'MANAGER VI                                       ': 34,
'INFO TECH SECURITY ANALYST I                     ': 33,
'GRANT COORDINATOR I                              ': 33,
'GEOSCIENTIST V                                   ': 33,
'HVAC MECHANIC II                                 ': 33,
'TEACHER AIDE III                                 ': 33,
'PORTFOLIO MANAGER III                            ': 33,
'PILOT INVESTIGATOR II                            ': 33,
'PORTFOLIO MANAGER IV                             ': 33,
'AUDITOR I                                        ': 33,
'RIGHT OF WAY AGENT V                             ': 33,
'CYBERSECURITY ANALYST I                          ': 32,
'GROUNDSKEEPER III                                ': 32,
'PURCHASER II                                     ': 32,
'HUMAN SERVICES TECHNICIAN II                     ': 32,
```

```
'AGRICULTURE SPEC V                          ': 32,
'SYSTEMS ADMINISTRATOR V                     ': 32,
'HVAC MECHANIC III                           ': 32,
'COURT LAW CLERK I                           ': 32,
'SGT, TABC                                   ': 31,
'LIBRARIAN IV                                ': 31,
'LEGAL SECRETARY II                          ': 31,
'LEGAL ASSISTANT V                           ': 31,
'PAROLE OFFCR V                              ': 30,
'HYDROLOGIST IV                              ': 30,
'FINANCIAL EXAM I                            ': 30,
'CREATIVE MEDIA DESIGNER III                 ': 30,
'INDUSTRIAL SPEC III                         ': 30,
'PHARMACY TECHNICIAN I                       ': 30,
'DENTAL HYGIENIST I                          ': 30,
'PHARMACIST III                              ': 30,
'PROTECT SRVCS INTAKE SPEC II                ': 30,
'WORKFORCE DEV SPECIALIST IV                 ': 29,
'INSURANCE SPECIALIST III                    ': 29,
'WARDEN II                                   ': 29,
'PSYCHOLOGIST II                             ': 29,
'TELECOMMUNICATIONS SPEC II                  ': 29,
'FERRYBOAT DECKHAND II                       ': 29,
'COMPLIANCE ANALYST IV                       ': 29,
'DOCUMENT SERVICES TECH IV                   ': 28,
'GROUNDSKEEPER II                            ': 28,
'ELECTRONICS TECHNICIAN II                   ': 28,
'PARK SUPERINTENDENT II                      ': 28,
'PROTECT SRVCS INTAKE SPEC I                 ': 28,
'MEDICAL TECHNOLOGIST III                    ': 28,
'PROTECT SRVCS INTAKE SPEC V                 ': 28,
'RESEARCH SPEC II                            ': 27,
'LABORATORY TECHNICIAN III                   ': 27,
'VETERANS SERVICES REP IV                    ': 27,
'MOLECULAR BIOLOGIST III                     ': 27,
'PARK SUPERINTENDENT IV                      ': 27,
'COUNSEL SUBSTITUTE II                       ': 27,
'REHABILITATION TEACHER II                   ': 27,
'DPTY CLERK III                              ': 27,
'GOVT RELATIONS SPECIALIST II                ': 27,
'CHEMIST III                                 ': 27,
'TAX AUDITOR MANAGER                         ': 26,
'DATA ANALYST I                              ': 26,
'MICROBIOLOGIST I                            ': 26,
'GEOGRAPHIC INFO SPEC II                     ': 26,
'EDUCATION SPECIALIST III                    ': 26,
'CONTRACT ADMIN MANAGER II                   ': 26,
'PORTFOLIO PROJECT MANAGER I                 ': 26,
'ARCHITECT II                                ': 26,
```

```
'HEALTH SPECIALIST I                                ': 26,
'PHARMACIST I                                       ': 26,
'COUNSEL SUBSTITUTE III                             ': 26,
'BENEFIT REVIEW OFFICER II                          ': 26,
'JUVENILE CORREC OFFCR II                           ': 26,
'TRAINING ASST                                      ': 26,
'PHYSICIAN II                                       ': 26,
'CRIMINAL INTEL ANALYST I                           ': 26,
'SAFETY OFFCR IV                                    ': 25,
'COOK IV                                            ': 25,
'RADIO COMMUNICATIONS TECH II                       ': 25,
'REHAB THERAPY TECHNICIAN V                         ': 25,
'PROBATIONARY AGENT                                 ': 25,
'DISTRICT ENGINEER                                  ': 25,
'INVESTMENT ANALYST III                             ': 25,
'CLAIMS EXAMINER I                                  ': 24,
'FERRYBOAT SPECIALIST III                           ': 24,
'LICENSE AND PERMIT SPEC I                          ': 24,
'FORENSIC SCIENTIST VI                              ': 24,
"LIEUTENANT, TEXAS AG'S OFFICE                      ": 24,
'WEB ADMINISTRATOR IV                               ': 24,
'REIMBURSEMENT OFFICER I                            ': 24,
'PROGRAMMER I                                       ': 24,
'CRIMINAL INTEL ANALYST II                          ': 24,
'PURCHASER I                                        ': 24,
'ARCHEOLOGIST III                                   ': 24,
'ARCHITECT III                                      ': 24,
'ANALYST I                                          ': 23,
'MICROBIOLOGIST III                                 ': 23,
'RIGHT OF WAY AGENT II                              ': 23,
'SANITARIAN III                                     ': 23,
'ELECTRONICS TECHNICIAN III                         ': 23,
'CREATIVE MEDIA DESIGNER II                         ': 23,
'NETWORK SPECIALIST VI                              ': 23,
'HEARINGS OFFICER III                               ': 23,
'HEALTH PHYSICIST II                                ': 23,
'ELECTRICIAN III                                    ': 23,
'PORTFOLIO MANAGER V                                ': 23,
'LT, GAME WARDEN                                    ': 22,
'PROPERTY MGR I                                     ': 22,
'MICROBIOLOGIST IV                                  ': 22,
'ADMINISTRATIVE LAW JUDGE I                         ': 22,
'TECHNICAL WRITER III                               ': 22,
'DEPUTY DIRECTOR II                                 ': 22,
'COMPLIANCE ANALYST I                               ': 22,
'SYSTEMS ADMINISTRATOR VI                           ': 22,
'DOCUMENT SERVICES TECH V                           ': 22,
'SUBSTANCE ABUSE COUNSELOR III                      ': 22,
'MANAGING SENIOR AUDITOR                            ': 21,
```

```
'INSPECTOR I                                    ': 21,
'SANITARIAN II                                  ': 21,
'CHEMIST II                                     ': 21,
'SAFETY OFFCR III                               ': 21,
'CHAPLAIN III                                   ': 21,
'RESEARCH SPEC I                                ': 21,
'PARK SUPERINTENDENT III                        ': 21,
'TELECOMMUNICATIONS SPEC IV                     ': 21,
'PSYCHIATRIST IV                                ': 21,
'VETERANS SERVICES REP V                        ': 21,
'POLICE COMMUNICATIONS OPER IV                  ': 20,
'HYDROLOGIST III                                ': 20,
'LEGISLATIVE TECHNICIAN                         ': 20,
'MARKETING SPECIALIST IV                        ': 20,
'DIRECTOR                                       ': 20,
'GENERAL COUNSEL V                              ': 20,
'MOTOR VEHICLE TECHNICIAN I                     ': 20,
'MEDICAL TECHNOLOGIST IV                        ': 20,
'DIETETIC & NUTRITION SPEC II                   ': 20,
'HVAC MECHANIC I                                ': 19,
'MOTOR VEHICLE TECHNICIAN II                    ': 19,
'DENTAL ASSISTANT II                            ': 19,
'ENVIRONMENTAL PROTECT SPEC I                   ': 19,
'SERGEANT ST PARK POLICE OFF                    ': 19,
'PROGRAM DIRECTOR                               ': 19,
'HEALTH SPECIALIST II                           ': 19,
'GEOGRAPHIC INFO SPEC IV                        ': 19,
'REGISTERED THERAPIST III                       ': 19,
'HEALTH SPECIALIST IV                           ': 19,
'POLICE COMMUNICATIONS OPER III                 ': 18,
'SENIOR DIRECTOR                                ': 18,
'FINGERPRINT ANALYST III                        ': 18,
'CUSTODIAL MGR III                              ': 18,
'TECH WRITER II                                 ': 18,
'VETERINARIAN II                                ': 18,
'CHEMIST V                                      ': 18,
'PORTFOLIO MANAGER II                           ': 18,
"CAPTAIN, TEXAS AG'S OFFICE                     ": 18,
'ELECTRICIAN IV                                 ': 18,
'HUMAN SERVICES TECHNICIAN I                    ': 18,
'PARK SUPERINTENDENT I                          ': 18,
'SERGEANT, TDI                                  ': 18,
'MARKETING SPECIALIST II                        ': 18,
'ORTHOPEDIC EQUIP TECHN II                      ': 18,
'LT, TABC                                       ': 18,
'MICROBIOLOGIST II                              ': 18,
'APPRAISER III                                  ': 17,
'WEB ADMINISTRATOR II                           ': 17,
'NATURAL RESOURCES SPEC I                       ': 17,
```

```
'MANAGER                                          ': 17,
'SYSTEMS ADMINISTRATOR III                        ': 17,
'BARBER/COSMETOLOGIST                             ': 17,
'LIBRARY ASST III                                 ': 16,
'LAUNDRY/SEWING ROOM WORKER III                   ': 16,
'CHEMIST I                                        ': 16,
'SGT, GAME WARDEN                                 ': 16,
'QUAL INTEL DISABILITY PROF I                     ': 16,
'RIGHT OF WAY AGENT VI                            ': 16,
'RISK MGMT SPEC III                               ': 16,
'REHABILITATION TEACHER I                         ': 16,
'ASSIST PARK SUPERINTENDENT III                   ': 16,
'RESCUE SPEC I                                    ': 16,
'POLICE COMMUNICATIONS OPER I                     ': 16,
'CRIMINAL INTEL ANALYST III                       ': 16,
'DATA BASE ADMINISTRATOR II                       ': 16,
'COOK I                                           ': 16,
'APPRAISER IV                                     ': 15,
'REGIONAL SUPERVISOR- OIG                         ': 15,
'VOLUNTEER SERVICES COORD II                      ': 15,
'CHIEF JUSTICE                                    ': 15,
'PSYCHOLOGIST I                                   ': 15,
'CLERK OF THE COURT                               ': 15,
'SYSTEMS ANALYST I                                ': 15,
'PROJECT MGT SPECIALIST III                       ': 15,
'DENTIST III                                      ': 15,
'RESPIRATORY CARE PRACTITIONER                    ': 15,
'INVESTMENT ANALYST II                            ': 15,
'RESIDENT SPECIALIST I                            ': 15,
'INVESTMENT ANALYST V                             ': 15,
'POLICE COMMUNICATIONS OPER II                    ': 15,
'SYSTEMS ADMINISTRATOR II                         ': 15,
'GOVT RELATIONS SPEC III                          ': 15,
'DATA ANALYST VI                                  ': 15,
'AC & BOILER OPER I                               ': 15,
'ACTUARY I                                        ': 14,
'DOCUMENT SERVICES TECH II                        ': 14,
'DISTRICT JUDGE, FORMER                           ': 14,
'LOAN SPECIALIST III                              ': 14,
'BUSINESS ANALYST I                               ': 14,
'LEGISLATIVE PROTECTIVE SERVICE                   ': 14,
'MEDICAL TECHNICIAN I                             ': 14,
'ASSISTANT DIRECTOR                               ': 14,
'EPIDEMIOLOGIST I                                 ': 14,
'HYDROLOGIST II                                   ': 14,
'DATA BASE ADMINISTRATOR V                        ': 13,
'ANALYST IV                                       ': 13,
'QUAL INTEL DISABILITY PROF IV                    ': 13,
'MARKETING SPECIALIST III                         ': 13,
```

```
'AUDITOR VI                                    ': 13,
'PHYSICIAN IV                                  ': 13,
'ARCHIVIST III                                 ': 13,
'ARCHITECT IV                                  ': 13,
'PLANNER I                                     ': 13,
'TECH WRITER I                                 ': 13,
'LABORATORY TECHNICIAN I                       ': 13,
'INSURANCE SPECIALIST IV                       ': 13,
'ACTUARY II                                    ': 13,
'ASSIST PARK SUPERINTENDENT I                  ': 13,
'BEHAVIOR ANALYST II                           ': 13,
'MILITARY SPECIALIST III                       ': 13,
'PROGRAMMER VI                                 ': 13,
'AIRCRAFT MECHANIC II                          ': 12,
'FERRYBOAT SPECIALIST II                       ': 12,
'AGRICULTURE SPEC II                           ': 12,
'EDITOR II                                     ': 12,
'REIMBURSEMENT OFFICER II                      ': 12,
'ECONOMIST III                                 ': 12,
'FERRYBOAT SPECIALIST I                        ': 12,
'GUARDIANSHIP SUPERVISOR                       ': 12,
'VOLUNTEER SERVICES COORD I                    ': 12,
'LEG. SERVICE/MAINTENANCE                      ': 12,
'REIMBURSEMENT OFFICER IV                      ': 12,
'ARCHIVIST II                                  ': 12,
'RISK MGMT SPEC IV                             ': 12,
'GROUNDSKEEPER I                               ': 12,
'PROB ST PARK POLICE OFFICER                   ': 12,
'ORTHOPEDIC EQUIP TECHN I                      ': 12,
'INTERPRETER III                               ': 12,
'ASSIST PARK SUPERINTENDENT II                 ': 12,
'DPTY CLERK IV                                 ': 12,
'DEPUTY DIRECTOR III                           ': 12,
'TRANSITION COORDINATOR II                     ': 12,
'BUDGET ANALYST I                              ': 12,
'RETIRE SYS BENEFITS SPEC I                    ': 12,
'CYBERSECURITY ANALYST III                     ': 11,
'HUMAN RIGHTS OFFICER I                        ': 11,
'BUSINESS CONTINUITY COORD I                   ': 11,
'INFO TECH AUDITOR III                         ': 11,
'SOCIAL WORKER IV                              ': 11,
'BOILER INSPECTOR II                           ': 11,
'REGISTERED THERAPIST ASST                     ': 11,
'ACTUARY IV                                    ': 11,
'GEOSCIENTIST III                              ': 11,
'PROG AN III                                   ': 11,
'MAJOR, GAME WARDEN                            ': 11,
'DENTIST II                                    ': 11,
'MOLECULAR BIOLOGIST IV                        ': 11,
```

```
'LEGAL SECRETARY V                                ': 11,
'VETERANS SERVICES REP VI                         ': 11,
'LEG COUN I                                       ': 11,
'CHIEF DEPUTY CLERK                               ': 11,
'LAND SURVEYOR III                                ': 11,
'MEDICAL RESEARCH SPECIALIST                      ': 11,
'ENGINEERING TECHNICIAN IV                        ': 11,
"GOVERNOR'S ADVISOR III                           ": 11,
'GENERAL COUNSEL I                                ': 11,
'RESEARCH AND STATS TECH I                        ': 11,
'WEB ADMINISTRATOR V                              ': 11,
'PSYCHOLOGIST III                                 ': 10,
'SAFETY OFFICER V                                 ': 10,
'DIETETIC TECHN I                                 ': 10,
'PUBLIC HLTH & PRVNT SPEC IV                      ': 10,
'RADIOLOGICAL TECHNOLOGIST III                    ': 10,
'SENIOR SYSTEMS ANALYST                           ': 10,
'INFORMATION SECURITY OFFICER                     ': 10,
'MEDICAL TECHNOLOGIST V                           ': 10,
'REGISTERED THERAPIST I                           ': 10,
'INFO TECH AUDITOR II                             ': 10,
'SR LEG COUNSEL                                   ': 10,
'TRANS MAINT SPEC II                              ': 10,
'PAYROLL ASSISTANT                                ': 10,
'TAX ANALYST II                                   ': 10,
'LEG COUN IV                                      ': 9,
'INVESTIGATOR I                                   ': 9,
'HEALTH PHYSICIST III                             ': 9,
'ECONOMIST II                                     ': 9,
'APPN CONTROL OFFICER II                          ': 9,
'LABORATORY TECHNICIAN II                         ': 9,
'ASSIST PARK SUPERINTENDENT IV                    ': 9,
'RESCUE SPEC II                                   ': 9,
'PAYROLL SPECIALIST IV                            ': 9,
'LIBRARIAN II                                     ': 9,
'POLICY ANALYST I                                 ': 9,
'COMMISSIONER                                     ': 9,
'OMBUDSMAN V                                      ': 9,
'DIETETIC TECHN II                                ': 9,
'COMPUTER OPERATIONS SPEC V                       ': 9,
'LICENSED VOCA NURSE TRAINEE                      ': 9,
'VOLUNTEER SERVICES COORD III                     ': 9,
'ELECTRICIAN II                                   ': 9,
'RIGHT OF WAY AGENT I                             ': 9,
'ELECTRICIAN I                                    ': 9,
'VETERINARIAN III                                 ': 9,
'FORENSIC SCIENTIST V                             ': 9,
'DEPUTY ATTORNEY GENERAL                          ': 9,
'PILOT INVESTIGATOR III                           ': 9,
```

```
'WORKFORCE DEV SPECIALIST V                    ': 8,
'MASTER ADMIN LAW JUDGE I                      ': 8,
'CREATIVE MEDIA DESIGNER IV                    ': 8,
'TAX ANALYST IV                                ': 8,
'ANALYST II                                    ': 8,
'UNEMPL INSUR CLAIMS EXAM I                    ': 8,
'DOC DEL/FILING ASST                           ': 8,
'AUDIT MANAGER                                 ': 8,
'PUBLIC HLTH & PRVNT SPEC V                    ': 8,
'CHAPLAINCY SERVICES ASSISTANT                 ': 8,
'FERRYBOAT SPECIALIST IV                       ': 8,
'INVESTMENT ATTORNEY                           ': 8,
'HYDROLOGIST V                                 ': 8,
'CURATOR III                                   ': 8,
'CURATOR II                                    ': 8,
'CREATIVE MEDIA DESIGNER I                     ': 8,
'DATA ARCHITECT I                              ': 8,
'PARK RANGER III                               ': 8,
'ECONOMIST I                                   ': 8,
'ADMINISTRATIVE ASSISTANT II                   ': 8,
'ACCOUNT REPRESENTATIVE                        ': 8,
'MARKETING SPECIALIST I                        ': 8,
'JUDGE                                         ': 8,
'LEG OFFICE CONSULTANT II                      ': 8,
'DATA BASE ADMINISTRATOR I                     ': 8,
'MOLECULAR BIOLOGIST II                        ': 8,
'ACTUARY V                                     ': 8,
'EQUIPMENT MAINT TECH II                       ': 8,
'TAXPAYER COMPLIANCE OFFICER VI                ': 7,
'ASST COMM, GAME WARDEN                        ': 7,
'COMPUTER OPERATIONS SPEC I                    ': 7,
'CAPTAIN ST PARK POLICE OFF                    ': 7,
'RECREATION PROG SPEC II                       ': 7,
'HUMAN SERVICES TECHNICIAN IV                  ': 7,
'LEGISLATIVE SKILLED CRAFT                     ': 7,
'LABORATORY TECHNICIAN IV                      ': 7,
'DENTIST I                                     ': 7,
'GROUNDSKEEPER IV                              ': 7,
'14 CUSTOMER SERVICES REP                      ': 7,
'ANALYST III                                   ': 7,
'MEDICAL TECHNOLOGIST II                       ': 7,
'PARK RANGER II                                ': 7,
'PHYSICIAN I                                   ': 7,
'PARK SUPERINTENDENT V                         ': 7,
'20 PROGRAM SPECIALIST III                     ': 7,
'GEOGRAPHIC INFO SPEC I                        ': 7,
'QUALITY ASSURANCE ANALYST II                  ': 7,
'INVESTMENT ANALYST I                          ': 7,
'APPRAISER I                                   ': 7,
```

```
'ACTUARY III                                    ': 7,
'PROJECT LEADER                                 ': 7,
'DEPUTY DIRECTOR IV                             ': 7,
'DRAFTING TECHNICIAN I                          ': 7,
'EDUCATION SPECIALIST I                         ': 7,
'EDITOR I                                       ': 7,
'EDUCATION SPECIALIST II                        ': 7,
'WEB ADMINISTRATOR I                            ': 7,
'DRAFTING TECHNICIAN II                         ': 7,
'PROJECT MGT SPECIALIST I                       ': 7,
'PROG IV                                        ': 7,
'FINGERPRINT ANALYST II                         ': 7,
'DIETETIC & NUTRITION SPEC I                    ': 6,
'PAYROLL SPECIALIST II                          ': 6,
'LOAN SPECIALIST II                             ': 6,
'GENERAL COUNSEL                                ': 6,
'PARK SUPERINTENDENT VI                         ': 6,
'PROG III                                       ': 6,
'PAROLE BOARD MEMBER                            ': 6,
'SENIOR MANAGING DIRECTOR                       ': 6,
'PROJECT MGT SPECIALIST II                      ': 6,
'REGIONAL MANAGER- OIG                          ': 6,
'JUVENILE CORREC OFFCR I                        ': 6,
'HALFWAY HOUSE SUPERINTENDENT                   ': 6,
'RADIO COMMUNICATIONS TECH III                  ': 6,
'LAND SURVEYOR IV                               ': 6,
'OMBUDSMAN IV                                   ': 6,
'CUSTODIAL MGR II                               ': 6,
'TAX ANALYST I                                  ': 6,
'PUBLIC HEALTH NURSE III                        ': 6,
'LEG OFFICE CONSULTANT I                        ': 6,
'INFRASTRUCTURE SPECIALIST I                    ': 6,
'RISK MGMT SPEC V                               ': 6,
'STATISTICIAN IV                                ': 6,
'TEXAS WORKS ADVISOR V                          ': 6,
'LIFEGUARD                                      ': 6,
'AC & BOILER OPER IV                            ': 6,
'TOXICOLOGIST I                                 ': 6,
'INTERPRETER I                                  ': 6,
'EMERG MGT PROGRAM COORD IV                     ': 6,
'GEOSCIENTIST II                                ': 6,
'FLEET MANAGER III                              ': 6,
'COMPUTER OPERATIONS SPEC IV                    ': 6,
'EXHIBIT TECHNICIAN                             ': 5,
'YOUTH FACILITY ASST SUPT                       ': 5,
'HALFWAY HOUSE ASST SUPT                        ': 5,
'YOUTH FACILITY SUPERINTENDENT                  ': 5,
'PROPERTY MANAGER IV                            ': 5,
'MARKETING SPECIALIST V                         ': 5,
```

```
'LEG COUN III                                    ': 5,
'MULTIMEDIA TECHNICIAN III                       ': 5,
'RES SPEC II                                     ': 5,
'AGRICULTURE SPEC VI                             ': 5,
'INTERPRETER II                                  ': 5,
'ADMINISTRATIVE ASSISTANT                        ': 5,
'EDITOR III                                      ': 5,
'PORTFOLIO MANAGER VI                            ': 5,
'PAYROLL SPECIALIST V                            ': 5,
'LIBRARY ASST II                                 ': 5,
'COMMISSION MEMBERS                              ': 5,
'COLLECTIONS SPECIALIST                          ': 5,
'ORTHOPEDIC EQUIP TECHN III                      ': 5,
'RECREATION PROG SPEC III                        ': 5,
'ELECTRONICS TECHNICIAN I                        ': 5,
'PROOFREADER I                                   ': 5,
'TOXICOLOGIST III                                ': 5,
'TOXICOLOGIST II                                 ': 5,
'CRIMINAL DISTRICT ATTORNEYCNP                   ': 5,
'SECURITY OFFICER I                              ': 5,
'APPN CONTROL OFFICER IV                         ': 5,
'COMPUTER OPERATIONS SPEC VI                     ': 5,
'MICROBIOLOGIST V                                ': 5,
'APPN CONTROL OFFICER III                        ': 5,
'INDEPENDENT AUDIT REVIEWER III                  ': 5,
'UTILITY SPECIALIST II                           ': 5,
'PAYROLL SPECIALIST III                          ': 5,
'ASSISTANT COMMISSIONER                          ': 5,
'MAJOR, TABC                                     ': 5,
'AC & BOILER OPER III                            ': 5,
'SENIOR ANALYST                                  ': 5,
'CHIEF ACTUARY                                   ': 5,
'SECTION MANAGER                                 ': 5,
'REVIEWER I                                      ': 5,
'SENIOR CUSTOMER SERVICE REPRES                  ': 5,
'LEG OFFICIAL/ADMINISTRATOR                      ': 5,
'COMPLIANCE SPECIALIST                           ': 5,
'PHYSICIAN ASSISTANT                             ': 5,
'RISK MGMT SPEC II                               ': 5,
'PROJECT MANAGER V                               ': 5,
'PLUMBER IIII                                    ': 5,
'RISK MGMT SPEC I                                ': 5,
'LOAN SPECIALIST IV                              ': 5,
'RESIDENT PHYSICIAN                              ': 4,
'HISTORIAN III                                   ': 4,
'MEDICAL TECHNICIAN III                          ': 4,
'AC & BOILER OPER II                             ': 4,
'PUBLIC HEALTH NURSE II                          ': 4,
'MOLECULAR BIOLOGIST V                           ': 4,
```

```
 'TAX ANALYST III                                    ': 4,
 'DATA TRANSCRIPTION TECH III                        ': 4,
 'MULTIMEDIA TECHNICIAN II                           ': 4,
 'LIBRARY ASST I                                     ': 4,
 'MULTIMEDIA TECHNICIAN IV                           ': 4,
 "MAJOR, TEXAS AG'S OFFICE                          ": 4,
 'DPTY CLERK II                                      ': 4,
 'RESEARCH EDITOR                                    ': 4,
 'ARCHEOLOGIST II                                    ': 4,
 'ARCHIVIST I                                        ': 4,
 'CURATOR I                                          ': 4,
 'ADP EQUIPMENT OPERATOR IV                          ': 4,
 'BUSINESS CONTINUITY COORD II                       ': 4,
 'ANALYST                                            ': 4,
 'FINGERPRINT ANALYST I                              ': 4,
 'PSYCHIATRIST II                                    ': 4,
 'ARCHITECT I                                        ': 4,
 'MED TECHNOLOGIST TRAINEE                           ': 4,
 'CRIME LABORATORY SPEC I                            ': 4,
 'RECREATION PROG SPEC I                             ': 4,
 'GEOGRAPHIC INFO SPEC V                             ': 4,
 'DATA OFFICER                                       ': 4,
 'SENIOR INVESTIGATOR                                ': 4,
 'SENIOR PROJECT MANAGER                             ': 4,
 'MED FEE DISPUTE OFFICER II                         ': 4,
 'LEG ED IV                                          ': 4,
 'AIRCRAFT PILOT II                                  ': 4,
 'SR. CLASSIFICATION ANALYST                         ': 4,
 'PORTFOLIO MANAGER I                                ': 4,
 'HEARINGS OFFICER V                                 ': 4,
 'PSYCHOLOGICAL ASSOCIATE IV                         ': 4,
 'MANAGING SR RISK ASSESSMEN AUD                     ': 4,
 'BENEFIT REVIEW OFFICER III                         ': 4,
 'INFRASTRUCTURE SPECIALIST II                       ': 4,
 'SR QUALITY CONTROL REVIEWER                        ': 4,
 'SENIOR ACCOUNTANT                                  ': 4,
 '21 BUSINESS ANALYST                                ': 3,
 'PROG AN II                                         ': 3,
 ...}
```

```
categorical_data['CLASSTITLE']=categorical_data['CLASSTITLE'].map(clas
s_title_encoding)
categorical_data.head(10)
```

```
                                        AGENCYNAME  \
0  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1  OFFICE OF COURT ADMINISTRATION                ...
2  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3  OFFICE OF COURT ADMINISTRATION                ...
4  TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...
```

```
5  OFFICE OF THE ATTORNEY GENERAL                ...
6  TEXAS DEPARTMENT OF TRANSPORTATION            ...
7  TEXAS BEHAVIORAL HEALTH EXECUTIVE COUNCIL     ...
8  BOARD OF EXAMINERS OF PSYCHOLOGISTS           ...
9  DEPARTMENT OF STATE HEALTH SERVICES           ...

                          LASTNAME                          FIRSTNAME
MI  \
0  RUCKER                              MORTON                              V

1  RUCKER                              MORTON                              V

2  SPECIA JR                           JOHN                                J

3  SPECIA JR                           JOHN                                J

4  ONTIVEROS                           ESTHER

5  ROGERS                              SHAUNA

6  RICHTER                             WILLIAM                             J

7  SPINKS                              DARREL                              D

8  SPINKS                              DARREL                              D

9  ADAMS III                           LEE                                 A


   CLASSCODE  CLASSTITLE        ETHNICITY            GENDER  \
0        324         324  WHITE             MALE
1         47          47  WHITE             MALE
2        324         324  WHITE             MALE
3         47          47  WHITE             MALE
4       9267        9267  HISPANIC          FEMALE
5        372         372  HISPANIC          FEMALE
6        354         354  WHITE             MALE
7        372         372  WHITE             MALE
8          1           1  WHITE             MALE
9        463         463  BLACK             MALE

                                       STATUS EMPLOYDATE  EMPLOY_DAY
0  URP - UNCLASSIFIED REGULAR PART-TIME     1988-02-18          18
1  CTP - CLASSIFIED TEMPORARY PART-TIME     2015-02-01           1
2  URP - UNCLASSIFIED REGULAR PART-TIME     2020-02-01           1
3  CTP - CLASSIFIED TEMPORARY PART-TIME     2018-09-01           1
4  CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-29          29
5  CRF - CLASSIFIED REGULAR FULL-TIME       2020-04-01           1
6  CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-22          22
7  CRF - CLASSIFIED REGULAR FULL-TIME       2020-03-01           1
```

```
8  ERP - EXEMPT REGULAR PART-TIME           2020-03-04           4
9  CRF - CLASSIFIED REGULAR FULL-TIME       2019-09-01           1
```

```
df_new.replace(' ','other',inplace=True)
```

```
ordinal_labels=df_new.groupby(['MI'])
['ANNUAL'].mean().sort_values().index
ordinal_labels
```

```
Index(['O', 'U', 'N', 'I', 'Y', 'Q', 'other', 'Z', 'L', 'M', 'D', 'V',
'S',
       'A', 'R', 'G', 'K', 'J', 'E', 'F', 'C', 'T', 'P', 'B', 'X',
'W', 'H'],
      dtype='object', name='MI')
```

```
ordinal_labels2={k:i for i,k in enumerate(ordinal_labels,0)}
ordinal_labels2
```

```
{'O': 0,
 'U': 1,
 'N': 2,
 'I': 3,
 'Y': 4,
 'Q': 5,
 'other': 6,
 'Z': 7,
 'L': 8,
 'M': 9,
 'D': 10,
 'V': 11,
 'S': 12,
 'A': 13,
 'R': 14,
 'G': 15,
 'K': 16,
 'J': 17,
 'E': 18,
 'F': 19,
 'C': 20,
 'T': 21,
 'P': 22,
 'B': 23,
 'X': 24,
 'W': 25,
 'H': 26}
```

```
#MI_encoding=df2_cat['MI'].value_counts().to_dict()

#df2_cat['MI']=df2_cat['MI'].map(MI_encoding)
#df2_cat.head(5)
```

```
df_new['MI']=df_new['MI'].map(ordinal_labels2)
df_new.head(10)
```

```
   AGENCY                          LASTNAME
FIRSTNAME  MI  \
0     241  RUCKER                                MORTON
11
1     212  RUCKER                                MORTON
11
2     241  SPECIA JR                             JOHN
17
3     212  SPECIA JR                             JOHN
17
4     696  ONTIVEROS                             ESTHER
6
5     302  ROGERS                                SHAUNA
6
6     601  RICHTER                               WILLIAM
17
7     510  SPINKS                                DARREL
10
8     520  SPINKS                                DARREL
10
9     537  ADAMS III                             LEE
13

  CLASSCODE                                              CLASSTITLE  \
0  JD25      JUDGE, RETIRED                                      ...
1  3524      GENERAL COUNSEL IV                                  ...
2  JD25      JUDGE, RETIRED                                      ...
3  3524      GENERAL COUNSEL IV                                  ...
4  4504      CORREC  OFFICER IV                                  ...
5  1623      DIRECTOR IV                                         ...
6  1984      CONTRACT SPEC IV                                    ...
7  1623      DIRECTOR IV                                         ...
8  E178      EXEC DIR, BD OF EXAMS OF PSYCHOLOGISTS              ...
9  1323      INSPECTOR III                                       ...

          ETHNICITY           GENDER
STATUS  \
0  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
1  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
2  WHITE            MALE             URP - UNCLASSIFIED REGULAR PART-
TIME
3  WHITE            MALE             CTP - CLASSIFIED TEMPORARY PART-
TIME
4  HISPANIC         FEMALE           CRF - CLASSIFIED REGULAR FULL-
TIME
```

```
5  HISPANIC          FEMALE            CRF - CLASSIFIED REGULAR FULL-
TIME
6  WHITE             MALE              CRF - CLASSIFIED REGULAR FULL-
TIME
7  WHITE             MALE              CRF - CLASSIFIED REGULAR FULL-
TIME
8  WHITE             MALE              ERP - EXEMPT REGULAR PART-TIME

9  BLACK             MALE              CRF - CLASSIFIED REGULAR FULL-
TIME

  EMPLOYDATE  HRLYRATE  HRSPERWK   MONTHLY     ANNUAL  STATENUMBER
EMPLOY_DAY
0 1988-02-18  75.96150      29.0   9545.82  114549.84       127717
18
1 2015-02-01  81.04454       4.0   1404.77   16857.24       127717
1
2 2020-02-01  75.96150      29.0   9545.82  114549.84        59115
1
3 2018-09-01  81.04453       4.0   1404.77   16857.24        59115
1
4 2020-06-29   0.00000      40.0   3284.27   39411.24       165030
29
5 2020-04-01   0.00000      40.0  12899.00  154788.00      1177785
1
6 2020-06-22   0.00000      40.0   5835.50   70026.00      1085586
22
7 2020-03-01   0.00000      40.0  10000.00  120000.00       147334
1
8 2020-03-04  49.40717      20.0   4281.95   51383.40       147334
4
9 2019-09-01   0.00000      40.0   3447.25   41367.00       129635
1
```

```
categorical_data['ETHNICITY']=df_new['ETHNICITY']
categorical_data['STATUS']=df_new['STATUS']

categorical_data.drop('FIRSTNAME',axis=1,inplace=True)

categorical_data.head(10)
```

```
                                          AGENCYNAME  \
0  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
1  OFFICE OF COURT ADMINISTRATION                ...
2  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...
3  OFFICE OF COURT ADMINISTRATION                ...
4  TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...
5  OFFICE OF THE ATTORNEY GENERAL                ...
6  TEXAS DEPARTMENT OF TRANSPORTATION            ...
7  TEXAS BEHAVIORAL HEALTH EXECUTIVE COUNCIL     ...
```

```
8  BOARD OF EXAMINERS OF PSYCHOLOGISTS           ...
9  DEPARTMENT OF STATE HEALTH SERVICES           ...

                          LASTNAME MI  CLASSCODE  CLASSTITLE
ETHNICITY  \
0  RUCKER                           V        324         324  WHITE

1  RUCKER                           V         47          47  WHITE

2  SPECIA JR                        J        324         324  WHITE

3  SPECIA JR                        J         47          47  WHITE

4  ONTIVEROS                                9267        9267  HISPANIC

5  ROGERS                                    372         372  HISPANIC

6  RICHTER                          J        354         354  WHITE

7  SPINKS                           D        372         372  WHITE

8  SPINKS                           D          1           1  WHITE

9  ADAMS III                        A        463         463  BLACK


            GENDER                                        STATUS
EMPLOYDATE  \
0  MALE             URP - UNCLASSIFIED REGULAR PART-TIME     1988-02-
18
1  MALE             CTP - CLASSIFIED TEMPORARY PART-TIME     2015-02-
01
2  MALE             URP - UNCLASSIFIED REGULAR PART-TIME     2020-02-
01
3  MALE             CTP - CLASSIFIED TEMPORARY PART-TIME     2018-09-
01
4  FEMALE           CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-
29
5  FEMALE           CRF - CLASSIFIED REGULAR FULL-TIME       2020-04-
01
6  MALE             CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-
22
7  MALE             CRF - CLASSIFIED REGULAR FULL-TIME       2020-03-
01
8  MALE             ERP - EXEMPT REGULAR PART-TIME           2020-03-
04
9  MALE             CRF - CLASSIFIED REGULAR FULL-TIME       2019-09-
01

   EMPLOY_DAY
```

```
0          18
1           1
2           1
3           1
4          29
5           1
6          22
7           1
8           4
9           1
```

```
categorical_data.drop('LASTNAME',axis=1,inplace=True)
```

```
categorical_data.head(10)
```

```
                                          AGENCYNAME MI  CLASSCODE  \
0  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  V        324
1  OFFICE OF COURT ADMINISTRATION                ...  V         47
2  COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  J        324
3  OFFICE OF COURT ADMINISTRATION                ...  J         47
4  TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...          9267
5  OFFICE OF THE ATTORNEY GENERAL                ...           372
6  TEXAS DEPARTMENT OF TRANSPORTATION            ...  J        354
7  TEXAS BEHAVIORAL HEALTH EXECUTIVE COUNCIL     ...  D        372
8  BOARD OF EXAMINERS OF PSYCHOLOGISTS           ...  D          1
9  DEPARTMENT OF STATE HEALTH SERVICES           ...  A        463

   CLASSTITLE        ETHNICITY           GENDER  \
0         324  WHITE            MALE
1          47  WHITE            MALE
2         324  WHITE            MALE
3          47  WHITE            MALE
4        9267  HISPANIC         FEMALE
5         372  HISPANIC         FEMALE
6         354  WHITE            MALE
7         372  WHITE            MALE
8           1  WHITE            MALE
9         463  BLACK            MALE

                                    STATUS EMPLOYDATE  EMPLOY_DAY
0  URP - UNCLASSIFIED REGULAR PART-TIME     1988-02-18          18
1  CTP - CLASSIFIED TEMPORARY PART-TIME     2015-02-01           1
2  URP - UNCLASSIFIED REGULAR PART-TIME     2020-02-01           1
3  CTP - CLASSIFIED TEMPORARY PART-TIME     2018-09-01           1
4  CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-29          29
5  CRF - CLASSIFIED REGULAR FULL-TIME       2020-04-01           1
6  CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-22          22
7  CRF - CLASSIFIED REGULAR FULL-TIME       2020-03-01           1
8  ERP - EXEMPT REGULAR PART-TIME           2020-03-04           4
9  CRF - CLASSIFIED REGULAR FULL-TIME       2019-09-01           1
```

```
categorical_data['GENDER']=df_new['GENDER']
from sklearn.preprocessing import LabelEncoder
encoder=LabelEncoder()
categorical_data['GENDER']=encoder.fit_transform(categorical_data['GEN
DER'])

categorical_data

                                                AGENCYNAME MI
CLASSCODE  \
0       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  V
324
1       OFFICE OF COURT ADMINISTRATION                ...  V
47
2       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  J
324
3       OFFICE OF COURT ADMINISTRATION                ...  J
47
4       TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...
9267
...                                                   ... ..         ..
.
149476  STATE PRESERVATION BOARD                      ...  P
385
149477  STATE PRESERVATION BOARD                      ...  A
48
149478  STATE PRESERVATION BOARD                      ...  C
221
149479  STATE PRESERVATION BOARD                      ...  R
1083
149480  STATE PRESERVATION BOARD                      ...  P
221

        CLASSTITLE        ETHNICITY  GENDER  \
0              324  WHITE                 1
1               47  WHITE                 1
2              324  WHITE                 1
3               47  WHITE                 1
4             9267  HISPANIC              0
...            ...                  ...     ...
149476         385  WHITE                 1
149477          48  WHITE                 0
149478         221  WHITE                 1
149479        1082  WHITE                 1
149480         221  WHITE                 0

                                          STATUS EMPLOYDATE
EMPLOY_DAY
0       URP - UNCLASSIFIED REGULAR PART-TIME     1988-02-18
18
```

```
1       CTP - CLASSIFIED TEMPORARY PART-TIME     2015-02-01
1
2       URP - UNCLASSIFIED REGULAR PART-TIME     2020-02-01
1
3       CTP - CLASSIFIED TEMPORARY PART-TIME     2018-09-01
1
4       CRF - CLASSIFIED REGULAR FULL-TIME       2020-06-29
29
...                                          ...        ...          ..
.
149476  CRF - CLASSIFIED REGULAR FULL-TIME       2017-10-30
30
149477  CRF - CLASSIFIED REGULAR FULL-TIME       2015-07-13
13
149478  CRP - CLASSIFIED REGULAR PART-TIME       2012-10-15
15
149479  CRF - CLASSIFIED REGULAR FULL-TIME       1989-09-22
22
149480  CRP - CLASSIFIED REGULAR PART-TIME       2012-02-16
16

[149481 rows x 9 columns]
```

```
encoded_categorcal_data=pd.get_dummies(categorical_data,columns=['ETHN
ICITY','STATUS'],drop_first=True,)

encoded_categorcal_data
```

```
                                              AGENCYNAME MI
CLASSCODE  \
0       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  V
324
1       OFFICE OF COURT ADMINISTRATION                ...  V
47
2       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  J
324
3       OFFICE OF COURT ADMINISTRATION                ...  J
47
4       TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...
9267
...                                                   ... ..         ..
.
149476  STATE PRESERVATION BOARD                      ...  P
385
149477  STATE PRESERVATION BOARD                      ...  A
48
149478  STATE PRESERVATION BOARD                      ...  C
221
149479  STATE PRESERVATION BOARD                      ...  R
1083
```

```
149480  STATE PRESERVATION BOARD                        ...  P
221

        CLASSTITLE  GENDER EMPLOYDATE  EMPLOY_DAY  ETHNICITY_ASIAN
\
0              324       1 1988-02-18          18
False
1               47       1 2015-02-01           1
False
2              324       1 2020-02-01           1
False
3               47       1 2018-09-01           1
False
4             9267       0 2020-06-29          29
False
...            ...     ...        ...         ...
...
149476         385       1 2017-10-30          30
False
149477          48       0 2015-07-13          13
False
149478         221       1 2012-10-15          15
False
149479        1082       1 1989-09-22          22
False
149480         221       0 2012-02-16          16
False

        ETHNICITY_BLACK            ETHNICITY_HISPANIC         ...  \
0                         False                       False  ...
1                         False                       False  ...
2                         False                       False  ...
3                         False                       False  ...
4                         False                        True  ...
...                         ...                         ...  ...
149476                    False                       False  ...
149477                    False                       False  ...
149478                    False                       False  ...
149479                    False                       False  ...
149480                    False                       False  ...

        STATUS_CRP - CLASSIFIED REGULAR PART-TIME        \
0                                                   False
1                                                   False
2                                                   False
3                                                   False
4                                                   False
...                                                   ...
149476                                              False
149477                                              False
```

```
149478                                                  True
149479                                                 False
149480                                                  True

        STATUS_CTF - CLASSIFIED TEMPORARY FULL-TIME      \
0                                                      False
1                                                      False
2                                                      False
3                                                      False
4                                                      False
...                                                      ...
149476                                                 False
149477                                                 False
149478                                                 False
149479                                                 False
149480                                                 False

        STATUS_CTP - CLASSIFIED TEMPORARY FULL-TIME      \
0                                                      False
1                                                      False
2                                                      False
3                                                      False
4                                                      False
...                                                      ...
149476                                                 False
149477                                                 False
149478                                                 False
149479                                                 False
149480                                                 False

        STATUS_CTP - CLASSIFIED TEMPORARY PART-TIME      \
0                                                      False
1                                                       True
2                                                      False
3                                                       True
4                                                      False
...                                                      ...
149476                                                 False
149477                                                 False
149478                                                 False
149479                                                 False
149480                                                 False

        STATUS_ERF - EXEMPT REGULAR FULL-TIME            \
0                                                      False
1                                                      False
2                                                      False
3                                                      False
4                                                      False
...                                                      ...
```

```
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_ERP - EXEMPT REGULAR PART-TIME             \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_URF - UNCLASSIFIED REGULAR FULL-TIME       \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_URP - UNCLASSIFIED REGULAR PART-TIME       \
0                                                        True
1                                                       False
2                                                        True
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_UTF - UNCLASSIFIED TEMPORARY FULL-TIME     \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
```

```
4                                                     False
...                                                     ...
149476                                                False
149477                                                False
149478                                                False
149479                                                False
149480                                                False

        STATUS_UTP - UNCLASSIFIED TEMPORARY PART-TIME
0                                                     False
1                                                     False
2                                                     False
3                                                     False
4                                                     False
...                                                     ...
149476                                                False
149477                                                False
149478                                                False
149479                                                False
149480                                                False

[149481 rows x 22 columns]
```

```
encoded_categorcal_data['MI']=df_new['MI']
```

```
encoded_categorcal_data
```

```
                                                AGENCYNAME  MI
CLASSCODE  \
0       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  11
324
1       OFFICE OF COURT ADMINISTRATION                ...  11
47
2       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  17
324
3       OFFICE OF COURT ADMINISTRATION                ...  17
47
4       TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...   6
9267
...                                                   ...  ..         .
..
149476  STATE PRESERVATION BOARD                      ...  22
385
149477  STATE PRESERVATION BOARD                      ...  13
48
149478  STATE PRESERVATION BOARD                      ...  20
221
149479  STATE PRESERVATION BOARD                      ...  14
1083
149480  STATE PRESERVATION BOARD                      ...  22
```

```
221

        CLASSTITLE  GENDER EMPLOYDATE  EMPLOY_DAY  ETHNICITY_ASIAN
\
0              324       1 1988-02-18          18
False
1               47       1 2015-02-01           1
False
2              324       1 2020-02-01           1
False
3               47       1 2018-09-01           1
False
4             9267       0 2020-06-29          29
False
...            ...     ...        ...         ...
...
149476         385       1 2017-10-30          30
False
149477          48       0 2015-07-13          13
False
149478         221       1 2012-10-15          15
False
149479        1082       1 1989-09-22          22
False
149480         221       0 2012-02-16          16
False

        ETHNICITY_BLACK            ETHNICITY_HISPANIC          ...  \
0                         False                       False  ...
1                         False                       False  ...
2                         False                       False  ...
3                         False                       False  ...
4                         False                        True  ...
...                         ...                         ...  ...
149476                    False                       False  ...
149477                    False                       False  ...
149478                    False                       False  ...
149479                    False                       False  ...
149480                    False                       False  ...

        STATUS_CRP - CLASSIFIED REGULAR PART-TIME        \
0                                                  False
1                                                  False
2                                                  False
3                                                  False
4                                                  False
...                                                  ...
149476                                             False
149477                                             False
149478                                              True
```

```
149479                                                  False
149480                                                   True

        STATUS_CTF - CLASSIFIED TEMPORARY FULL-TIME      \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_CTP - CLASSIFIED TEMPORARY FULL-TIME      \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_CTP - CLASSIFIED TEMPORARY PART-TIME      \
0                                                       False
1                                                        True
2                                                       False
3                                                        True
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_ERF - EXEMPT REGULAR FULL-TIME            \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
```

```
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_ERP - EXEMPT REGULAR PART-TIME             \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_URF - UNCLASSIFIED REGULAR FULL-TIME       \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_URP - UNCLASSIFIED REGULAR PART-TIME       \
0                                                        True
1                                                       False
2                                                        True
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_UTF - UNCLASSIFIED TEMPORARY FULL-TIME     \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
```

```
...                                                    ...
149476                                                False
149477                                                False
149478                                                False
149479                                                False
149480                                                False

        STATUS_UTP - UNCLASSIFIED TEMPORARY PART-TIME
0                                                     False
1                                                     False
2                                                     False
3                                                     False
4                                                     False
...                                                     ...
149476                                                False
149477                                                False
149478                                                False
149479                                                False
149480                                                False

[149481 rows x 22 columns]
```

```
after_EDA_data=pd.concat([encoded_categorcal_data,numaric_data],axis=1
)
```

```
after_EDA_data
```

```
                                                AGENCYNAME  MI
CLASSCODE  \
0       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  11
324
1       OFFICE OF COURT ADMINISTRATION                ...  11
47
2       COMPTROLLER OF PUBLIC ACCOUNTS, JUDICIARY SECT...  17
324
3       OFFICE OF COURT ADMINISTRATION                ...  17
47
4       TEXAS DEPARTMENT OF CRIMINAL JUSTICE          ...   6
9267
...                                                   ...  ..         .
..
149476  STATE PRESERVATION BOARD                      ...  22
385
149477  STATE PRESERVATION BOARD                      ...  13
48
149478  STATE PRESERVATION BOARD                      ...  20
221
149479  STATE PRESERVATION BOARD                      ...  14
1083
149480  STATE PRESERVATION BOARD                      ...  22
```

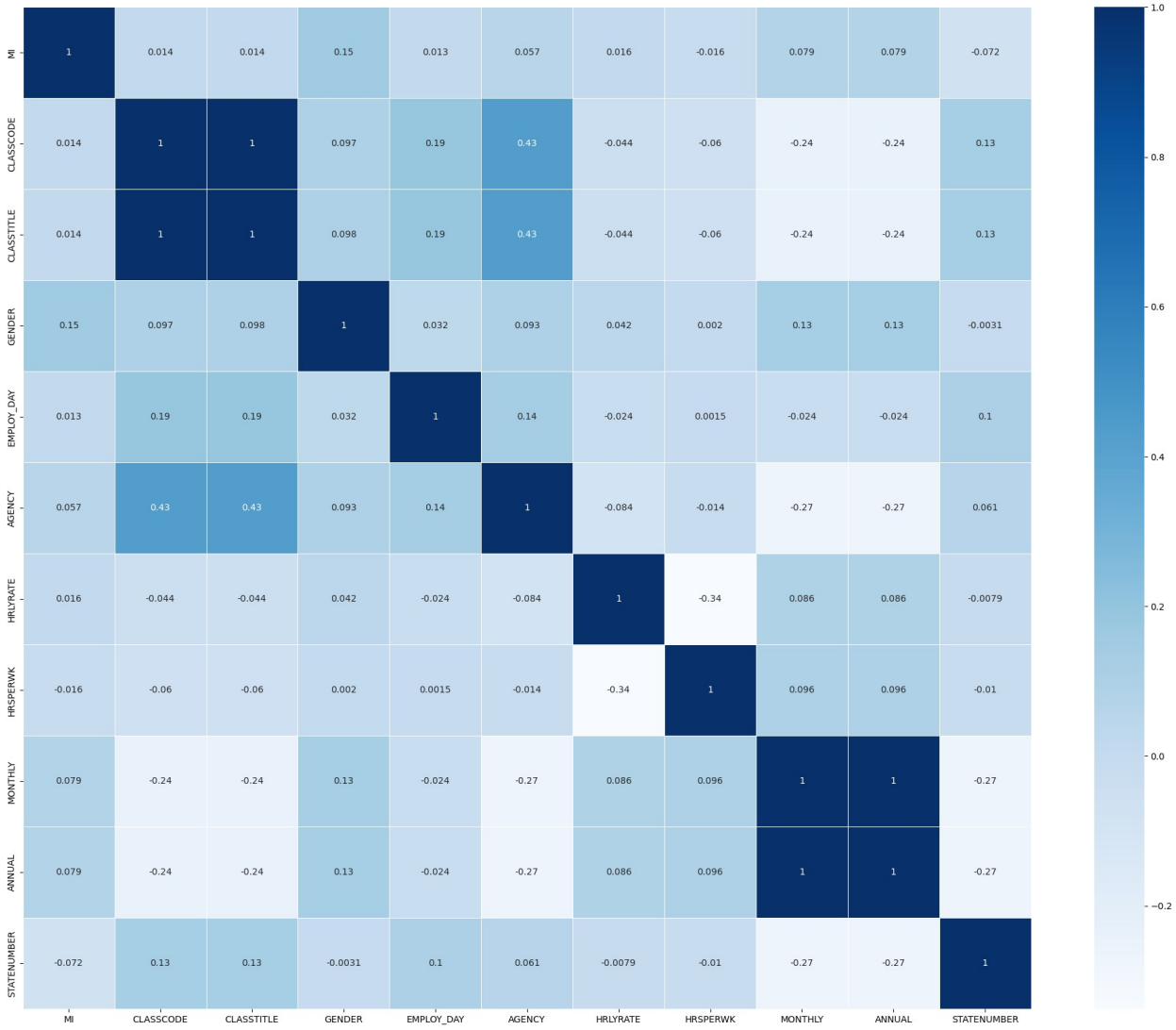
```
221

        CLASSTITLE  GENDER EMPLOYDATE  EMPLOY_DAY  ETHNICITY_ASIAN
\
0              324       1 1988-02-18          18
False
1               47       1 2015-02-01           1
False
2              324       1 2020-02-01           1
False
3               47       1 2018-09-01           1
False
4             9267       0 2020-06-29          29
False
...            ...     ...        ...         ...
...
149476         385       1 2017-10-30          30
False
149477          48       0 2015-07-13          13
False
149478         221       1 2012-10-15          15
False
149479        1082       1 1989-09-22          22
False
149480         221       0 2012-02-16          16
False

        ETHNICITY_BLACK            ETHNICITY_HISPANIC          ...  \
0                         False                       False  ...
1                         False                       False  ...
2                         False                       False  ...
3                         False                       False  ...
4                         False                        True  ...
...                         ...                         ...  ...
149476                    False                       False  ...
149477                    False                       False  ...
149478                    False                       False  ...
149479                    False                       False  ...
149480                    False                       False  ...

        STATUS_URF - UNCLASSIFIED REGULAR FULL-TIME      \
0                                                  False
1                                                  False
2                                                  False
3                                                  False
4                                                  False
...                                                  ...
149476                                             False
149477                                             False
149478                                             False
```

```
149479                                                  False
149480                                                  False

        STATUS_URP - UNCLASSIFIED REGULAR PART-TIME      \
0                                                        True
1                                                       False
2                                                        True
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_UTF - UNCLASSIFIED TEMPORARY FULL-TIME    \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_UTP - UNCLASSIFIED TEMPORARY PART-TIME    AGENCY
HRLYRATE  \
0                                                       False     241
75.96150
1                                                       False     212
81.04454
2                                                       False     241
75.96150
3                                                       False     212
81.04453
4                                                       False     696
0.00000
...                                                       ...     ...
...
149476                                                  False     809
0.00000
149477                                                  False     809
0.00000
149478                                                  False     809
12.93000
149479                                                  False     809
```

```
0.00000
149480                                                False     809
11.74000

        HRSPERWK  MONTHLY     ANNUAL  STATENUMBER
0           29.0  9545.82  114549.84       127717
1            4.0  1404.77   16857.24       127717
2           29.0  9545.82  114549.84        59115
3            4.0  1404.77   16857.24        59115
4           40.0  3284.27   39411.24       165030
...          ...      ...        ...          ...
149476      40.0  2899.00   34788.00       770781
149477      40.0  5500.00   66000.00       847431
149478      20.0  1120.60   13447.20        34266
149479      40.0  5744.16   68929.92       123490
149480      20.0  1017.46   12209.52       103583

[149481 rows x 28 columns]
```

```
import seaborn as sns
corr1 = after_EDA_data.select_dtypes(include='number').corr()
fig, ax = plt.subplots(figsize=(25,20))
sns.heatmap(corr1,linewidths=.5,cmap= 'Blues', ax=ax,annot=True)
```

```
<Axes: >
```



Data preprocessing

```
from scipy.stats import skew
print(numaric_data['HRLYRATE'].skew())
numaric_data['HRSPERWK'].skew()
```

```
15.91809864054434
```

```
-8.379277707265828
```

```
skwed_data_check=numaric_data
```

```
skwed_data_check
```

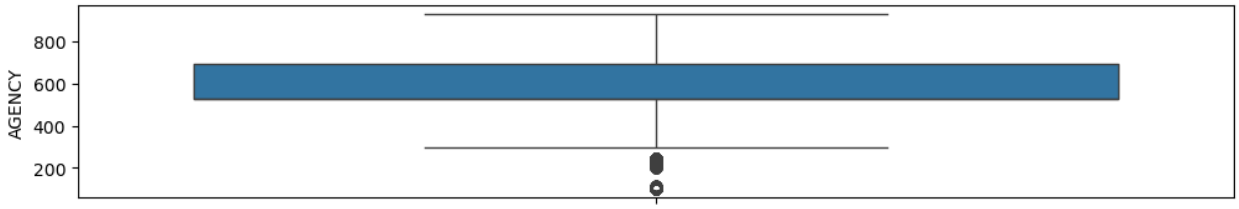
	AGENCY	HRLYRATE	HRSPERWK	MONTHLY	ANNUAL	STATENUMBER
0	241	75.96150	29.0	9545.82	114549.84	127717
1	212	81.04454	4.0	1404.77	16857.24	127717
2	241	75.96150	29.0	9545.82	114549.84	59115

```
3          212  81.04453       4.0  1404.77   16857.24        59115
4          696   0.00000      40.0  3284.27   39411.24       165030
...        ...       ...       ...      ...        ...          ...
149476     809   0.00000      40.0  2899.00   34788.00       770781
149477     809   0.00000      40.0  5500.00   66000.00       847431
149478     809  12.93000      20.0  1120.60   13447.20        34266
149479     809   0.00000      40.0  5744.16   68929.92       123490
149480     809  11.74000      20.0  1017.46   12209.52       103583

[149481 rows x 6 columns]
```

Outliers

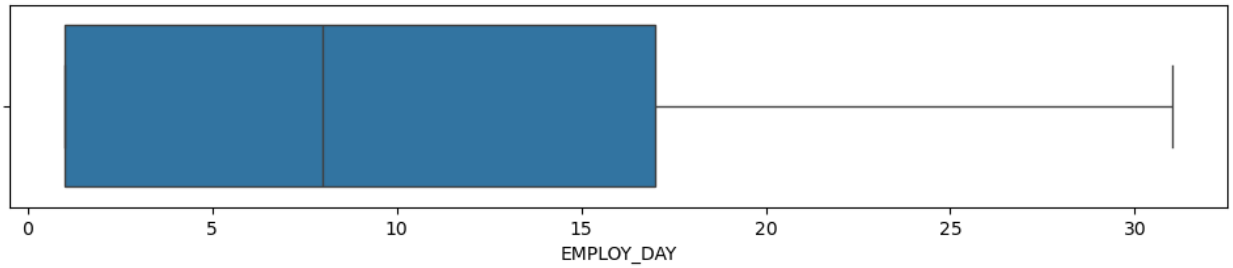
```
plt.figure(figsize=(12,2))
sns.boxplot(numaric_data['AGENCY'])
```

```
<Axes: ylabel='AGENCY'>
```



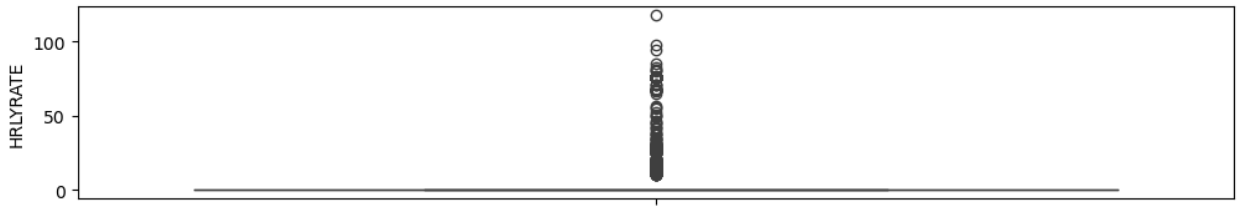

```
plt.figure(figsize=(12, 2))
sns.boxplot(x=after_EDA_data['EMPLOY_DAY'])
```

```
<Axes: xlabel='EMPLOY_DAY'>
```



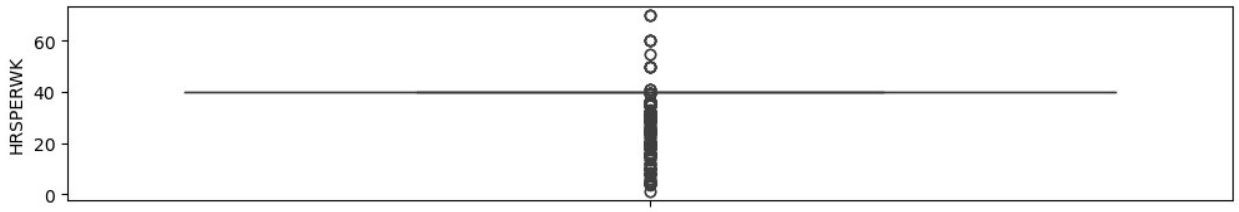

```
plt.figure(figsize=(12,2))
sns.boxplot(numaric_data['HRLYRATE'])
```

```
<Axes: ylabel='HRLYRATE'>
```



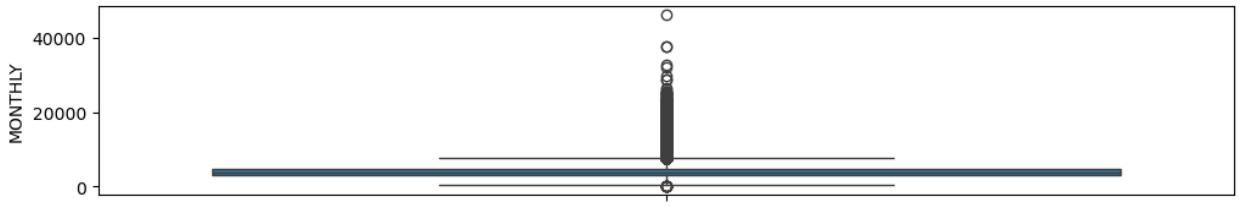

```
plt.figure(figsize=(12,2))
sns.boxplot(numaric_data['HRSPERWK'])
```

```
<Axes: ylabel='HRSPERWK'>
```



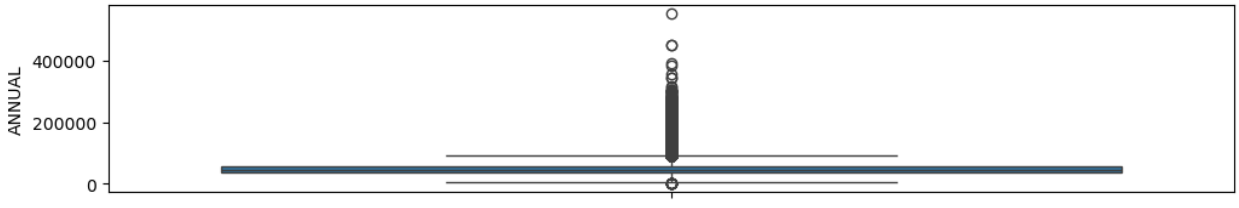

```
plt.figure(figsize=(12,2))
sns.boxplot(numaric_data['MONTHLY'])
```

```
<Axes: ylabel='MONTHLY'>
```



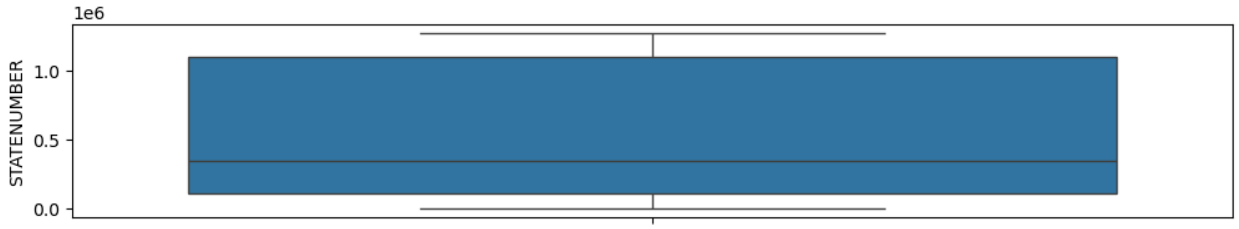

```
plt.figure(figsize=(12,2))
sns.boxplot(numaric_data['ANNUAL'])
```

```
<Axes: ylabel='ANNUAL'>
```




```
plt.figure(figsize=(12,2))
sns.boxplot(numaric_data['STATENUMBER'])
```

```
<Axes: ylabel='STATENUMBER'>
```




```
print(after_EDA_data['EMPLOY_DAY'].median())
print(numaric_data['HRLYRATE'].median())
print(numaric_data["HRSPERWK"].median())
print(numaric_data["MONTHLY"].median())
print(numaric_data["ANNUAL"].median())
```

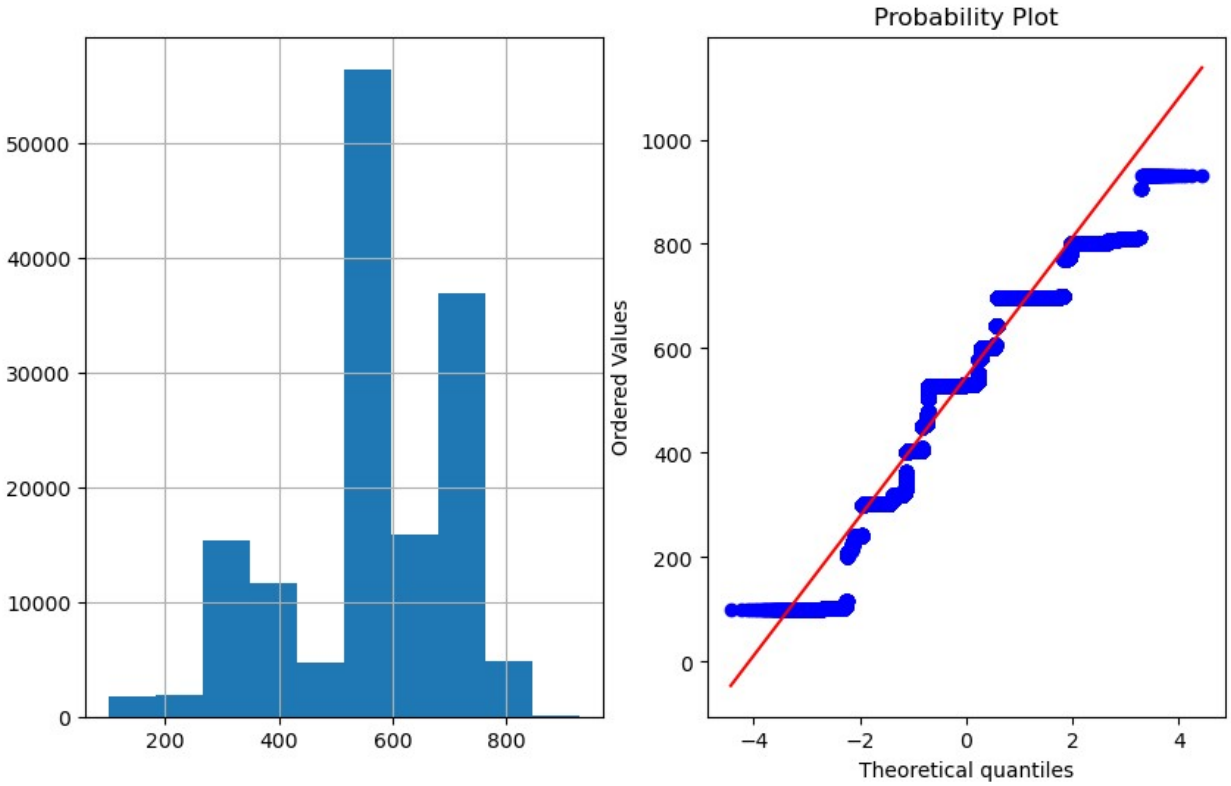
```
8.0
0.0
40.0
3720.17
44642.04
```

```
numaric_data.drop('MONTHLY',axis=1)
```

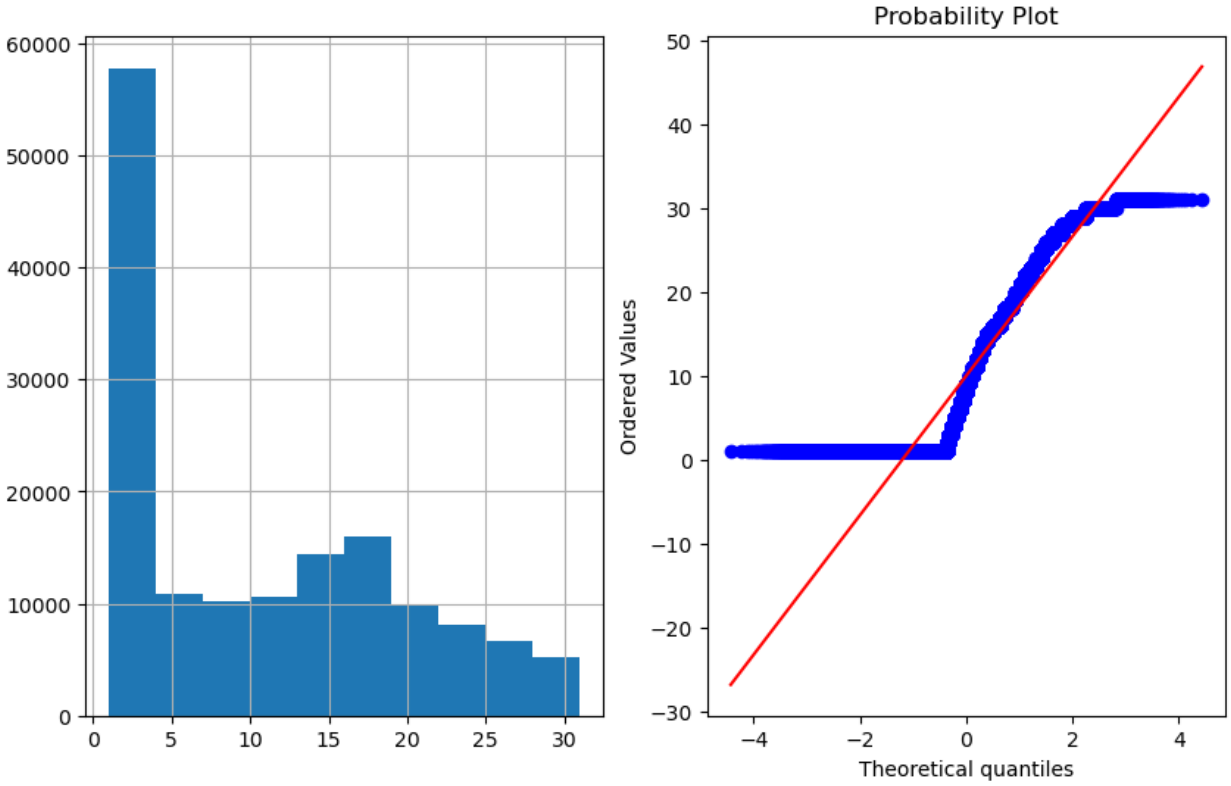
	AGENCY	HRLYRATE	HRSPERWK	ANNUAL	STATENUMBER
0	241	75.96150	29.0	114549.84	127717
1	212	81.04454	4.0	16857.24	127717
2	241	75.96150	29.0	114549.84	59115
3	212	81.04453	4.0	16857.24	59115
4	696	0.00000	40.0	39411.24	165030
...	...	...	...	...	...
149476	809	0.00000	40.0	34788.00	770781
149477	809	0.00000	40.0	66000.00	847431
149478	809	12.93000	20.0	13447.20	34266
149479	809	0.00000	40.0	68929.92	123490
149480	809	11.74000	20.0	12209.52	103583

```
[149481 rows x 5 columns]
```

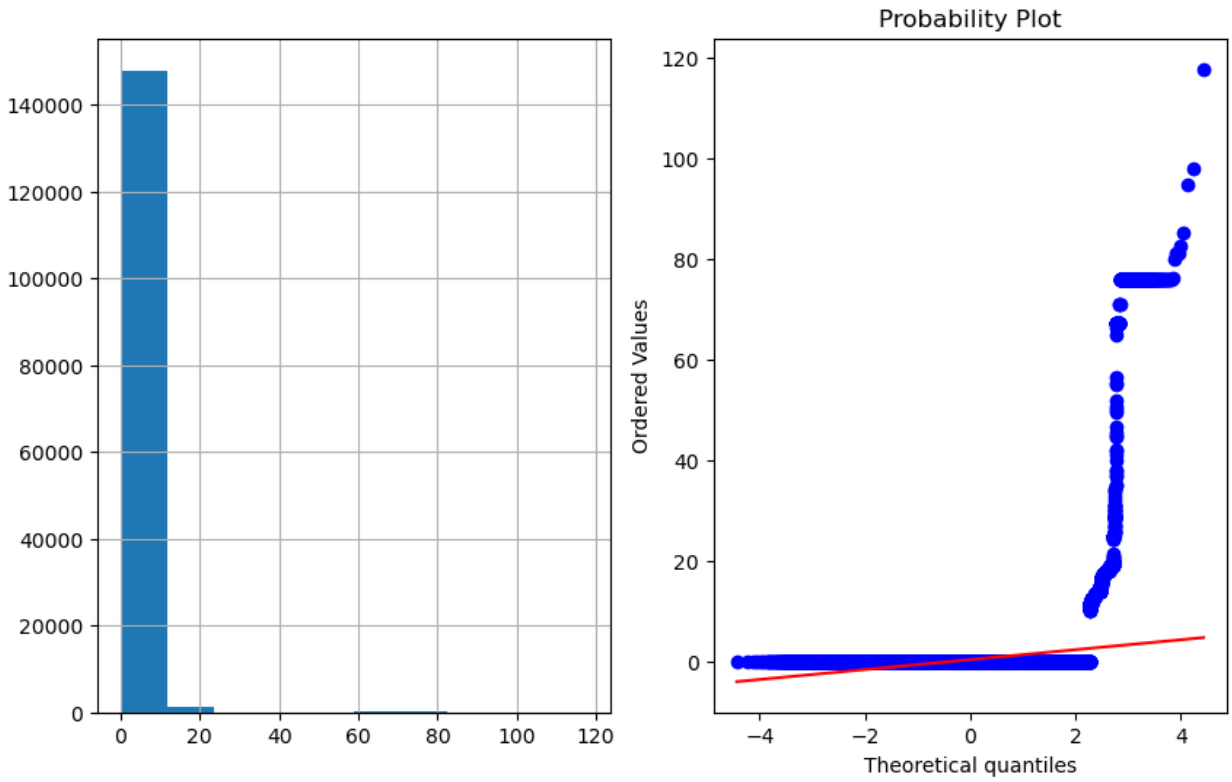
```
import scipy.stats as stat
import pylab
#### Q-Q plot
def plot_data(numaric_data,feature):
    plt.figure(figsize=(10,6))
    plt.subplot(1,2,1)
    numaric_data[feature].hist()
    plt.subplot(1,2,2)
    stat.probplot(numaric_data[feature],dist='norm',plot=pylab)
    plt.show()

plot_data(numaric_data,'AGENCY')
```



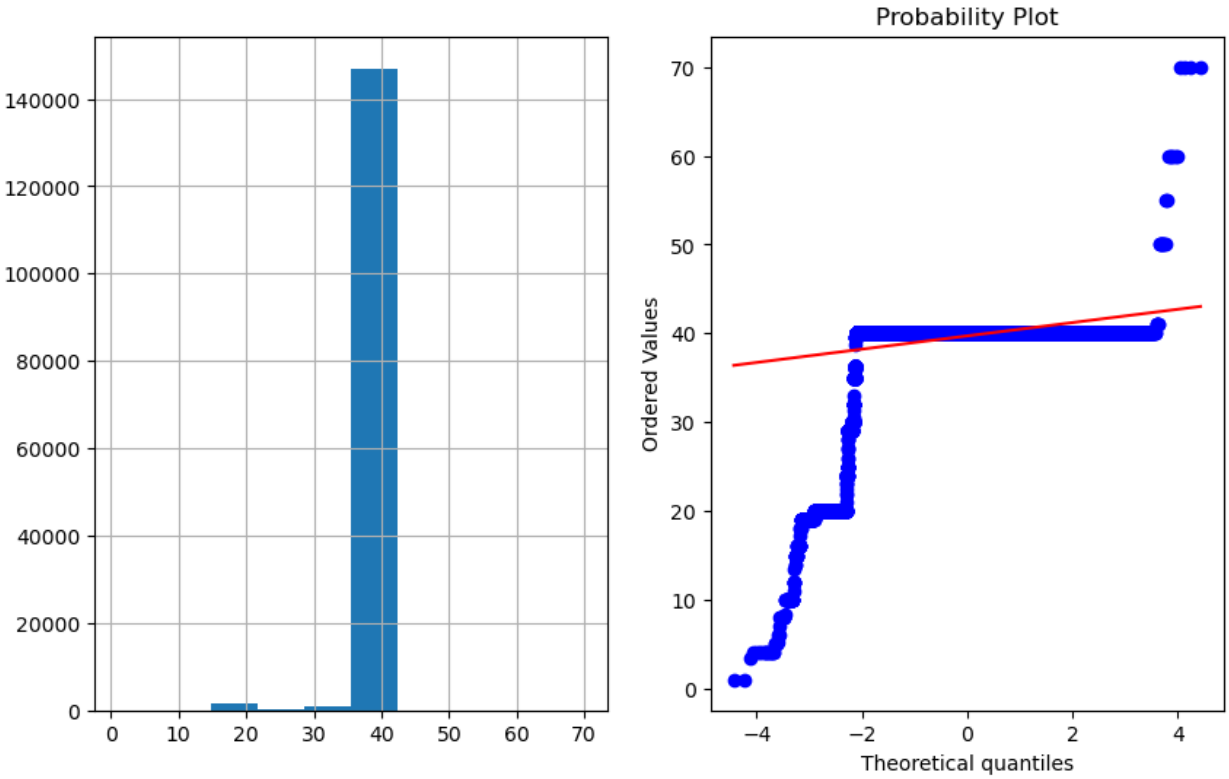

```
plot_data(after_EDA_data,'EMPLOY_DAY')
```



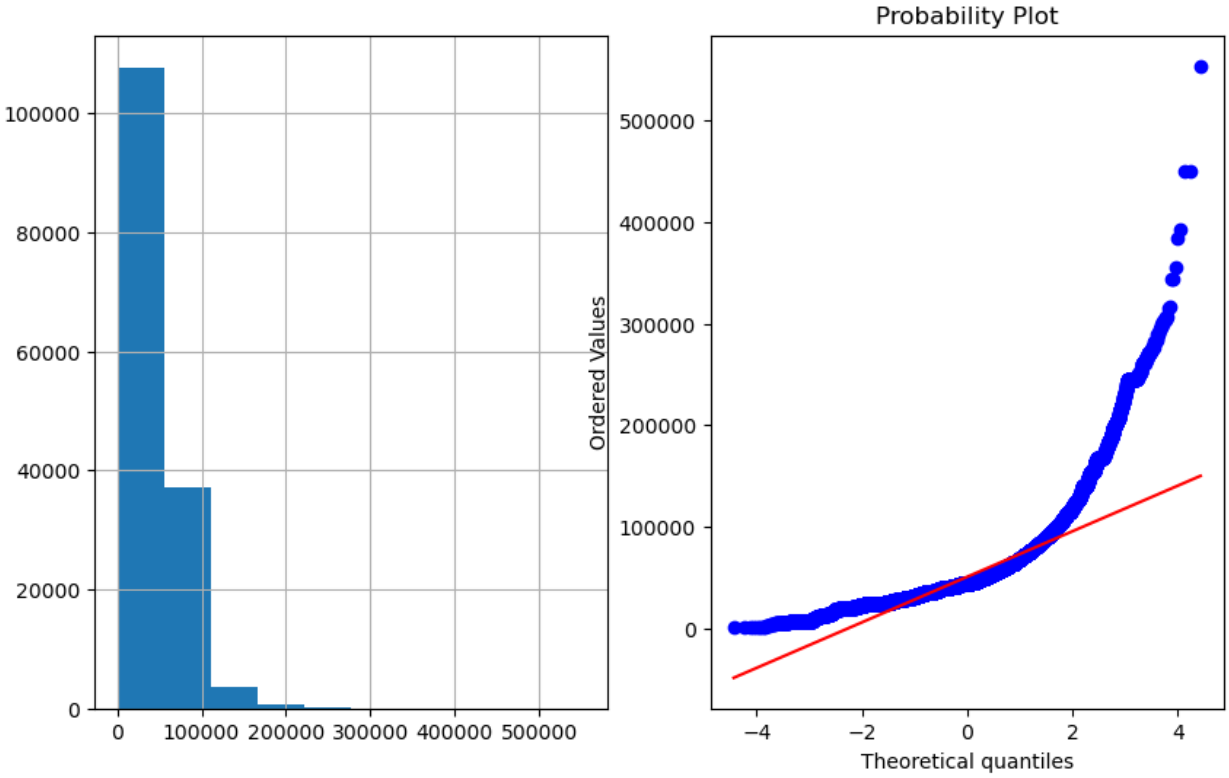
```
plot_data(numaric_data,'HRLYRATE')
```



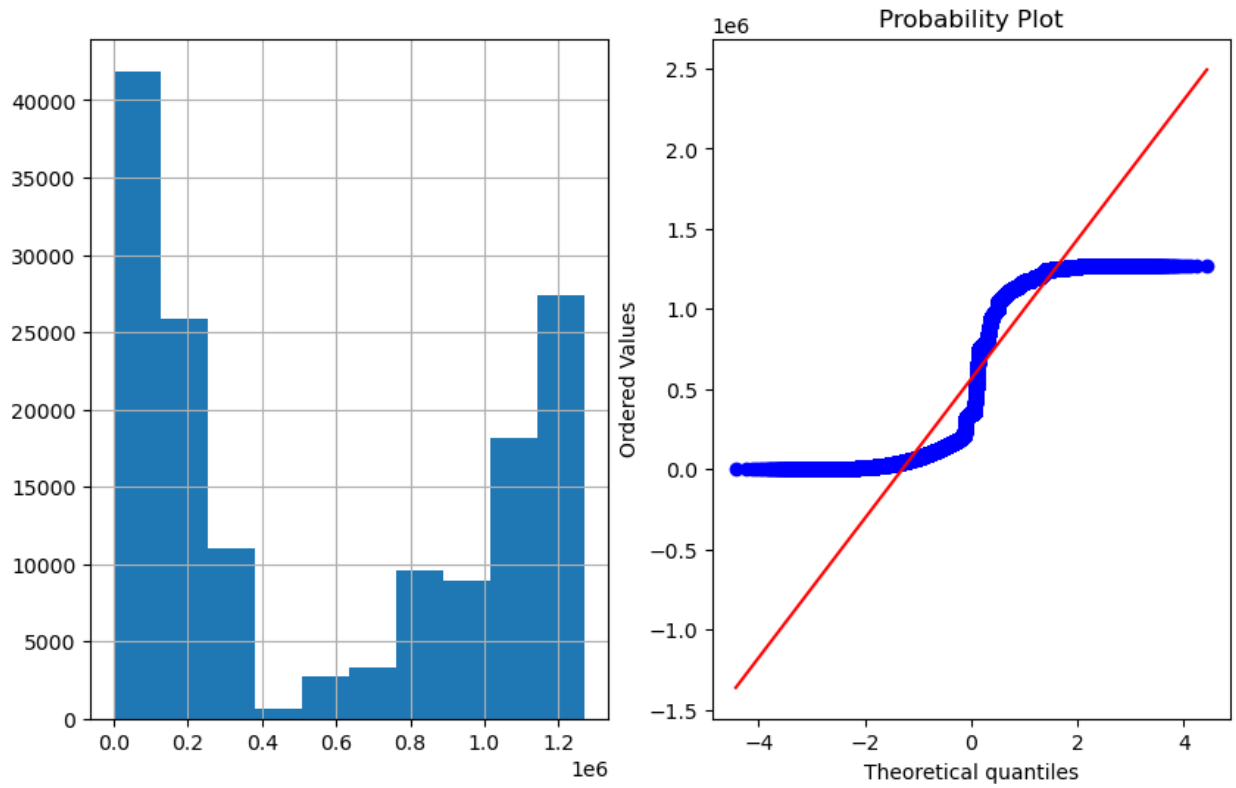

```
plot_data(numaric_data,'HRSPERWK')
```




```
plot_data(numaric_data,'ANNUAL')
```



```
plot_data(numaric_data,'STATENUMBER')
```




```
after_EDA_data.drop('AGENCYNAME',axis=1,inplace=True)

after_EDA_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149481 entries, 0 to 149480
Data columns (total 27 columns):
 #   Column                                                Non-Null Count
Dtype
---  ------                                                --------------
-----
 0   MI                                                    149481 non-null
int64
 1   CLASSCODE                                             149481 non-null
int64
 2   CLASSTITLE                                            149481 non-null
int64
 3   GENDER                                                149481 non-null
int32
 4   EMPLOYDATE                                            149481 non-null
datetime64[ns]
 5   EMPLOY_DAY                                            149481 non-null
int32
```

```
 6   ETHNICITY_ASIAN                                 149481 non-null
bool
 7   ETHNICITY_BLACK                                 149481 non-null
bool
 8   ETHNICITY_HISPANIC                              149481 non-null
bool
 9   ETHNICITY_OTHER                                 149481 non-null
bool
 10  ETHNICITY_WHITE                                 149481 non-null
bool
 11  STATUS_CRP - CLASSIFIED REGULAR PART-TIME       149481 non-null
bool
 12  STATUS_CTF - CLASSIFIED TEMPORARY FULL-TIME     149481 non-null
bool
 13  STATUS_CTP - CLASSIFIED TEMPORARY FULL-TIME     149481 non-null
bool
 14  STATUS_CTP - CLASSIFIED TEMPORARY PART-TIME     149481 non-null
bool
 15  STATUS_ERF - EXEMPT REGULAR FULL-TIME           149481 non-null
bool
 16  STATUS_ERP - EXEMPT REGULAR PART-TIME           149481 non-null
bool
 17  STATUS_URF - UNCLASSIFIED REGULAR FULL-TIME     149481 non-null
bool
 18  STATUS_URP - UNCLASSIFIED REGULAR PART-TIME     149481 non-null
bool
 19  STATUS_UTF - UNCLASSIFIED TEMPORARY FULL-TIME   149481 non-null
bool
 20  STATUS_UTP - UNCLASSIFIED TEMPORARY PART-TIME   149481 non-null
bool
 21  AGENCY                                          149481 non-null
int64
 22  HRLYRATE                                        149481 non-null
float64
 23  HRSPERWK                                        149481 non-null
float64
 24  MONTHLY                                         149481 non-null
float64
 25  ANNUAL                                          149481 non-null
float64
 26  STATENUMBER                                     149481 non-null
int64
dtypes: bool(15), datetime64[ns](1), float64(4), int32(2), int64(5)
memory usage: 14.7 MB
```

Modeling

```
X=after_EDA_data.iloc[:,:-1]
```

```
X

        MI  CLASSCODE  CLASSTITLE  GENDER EMPLOYDATE  EMPLOY_DAY  \
0       11        324         324       1 1988-02-18          18
1       11         47          47       1 2015-02-01           1
2       17        324         324       1 2020-02-01           1
3       17         47          47       1 2018-09-01           1
4        6       9267        9267       0 2020-06-29          29
...     ..        ...         ...     ...        ...         ...
149476  22        385         385       1 2017-10-30          30
149477  13         48          48       0 2015-07-13          13
149478  20        221         221       1 2012-10-15          15
149479  14       1083        1082       1 1989-09-22          22
149480  22        221         221       0 2012-02-16          16

        ETHNICITY_ASIAN            ETHNICITY_BLACK            \
0                         False                       False
1                         False                       False
2                         False                       False
3                         False                       False
4                         False                       False
...                         ...                         ...
149476                    False                       False
149477                    False                       False
149478                    False                       False
149479                    False                       False
149480                    False                       False

        ETHNICITY_HISPANIC         ETHNICITY_OTHER            ...  \
0                         False                       False  ...
1                         False                       False  ...
2                         False                       False  ...
3                         False                       False  ...
4                          True                       False  ...
...                         ...                         ...  ...
149476                    False                       False  ...
149477                    False                       False  ...
149478                    False                       False  ...
149479                    False                       False  ...
149480                    False                       False  ...

        STATUS_ERP - EXEMPT REGULAR PART-TIME            \
0                                                 False
1                                                 False
2                                                 False
3                                                 False
4                                                 False
...                                                 ...
149476                                            False
149477                                            False
```

```
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_URF - UNCLASSIFIED REGULAR FULL-TIME      \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_URP - UNCLASSIFIED REGULAR PART-TIME      \
0                                                        True
1                                                       False
2                                                        True
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_UTF - UNCLASSIFIED TEMPORARY FULL-TIME    \
0                                                       False
1                                                       False
2                                                       False
3                                                       False
4                                                       False
...                                                       ...
149476                                                  False
149477                                                  False
149478                                                  False
149479                                                  False
149480                                                  False

        STATUS_UTP - UNCLASSIFIED TEMPORARY PART-TIME    AGENCY
HRLYRATE  \
0                                                       False     241
75.96150
1                                                       False     212
81.04454
2                                                       False     241
```

```
75.96150
3                                                    False     212
81.04453
4                                                    False     696
0.00000
...                                                    ...     ...
...
149476                                               False     809
0.00000
149477                                               False     809
0.00000
149478                                               False     809
12.93000
149479                                               False     809
0.00000
149480                                               False     809
11.74000

        HRSPERWK  MONTHLY     ANNUAL
0           29.0  9545.82  114549.84
1            4.0  1404.77   16857.24
2           29.0  9545.82  114549.84
3            4.0  1404.77   16857.24
4           40.0  3284.27   39411.24
...          ...      ...        ...
149476      40.0  2899.00   34788.00
149477      40.0  5500.00   66000.00
149478      20.0  1120.60   13447.20
149479      40.0  5744.16   68929.92
149480      20.0  1017.46   12209.52

[149481 rows x 26 columns]

y=after_EDA_data.iloc[:,-1]

y

0         127717
1         127717
2          59115
3          59115
4         165030
           ...
149476    770781
149477    847431
149478     34266
149479    123490
149480    103583
Name: STATENUMBER, Length: 149481, dtype: int64
```

```
X['EMPLOY_DAYS'] = (pd.Timestamp('today') -
pd.to_datetime(X['EMPLOYDATE'], errors='coerce')).dt.days

X = X.drop(columns=['EMPLOYDATE'])# then drop the original datetime

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaled_data_X = scaler.fit_transform(X)

X.isnull().sum()
```

```
MI                                                  0
CLASSCODE                                           0
CLASSTITLE                                          0
GENDER                                              0
EMPLOY_DAY                                          0
ETHNICITY_ASIAN                                     0
ETHNICITY_BLACK                                     0
ETHNICITY_HISPANIC                                  0
ETHNICITY_OTHER                                     0
ETHNICITY_WHITE                                     0
STATUS_CRP - CLASSIFIED REGULAR PART-TIME           0
STATUS_CTF - CLASSIFIED TEMPORARY FULL-TIME         0
STATUS_CTP - CLASSIFIED TEMPORARY FULL-TIME         0
STATUS_CTP - CLASSIFIED TEMPORARY PART-TIME         0
STATUS_ERF - EXEMPT REGULAR FULL-TIME               0
STATUS_ERP - EXEMPT REGULAR PART-TIME               0
STATUS_URF - UNCLASSIFIED REGULAR FULL-TIME         0
STATUS_URP - UNCLASSIFIED REGULAR PART-TIME         0
STATUS_UTF - UNCLASSIFIED TEMPORARY FULL-TIME       0
STATUS_UTP - UNCLASSIFIED TEMPORARY PART-TIME       0
AGENCY                                              0
HRLYRATE                                            0
HRSPERWK                                            0
MONTHLY                                             0
ANNUAL                                              0
EMPLOY_DAYS                                         0
dtype: int64
```

Feature importance

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(X,y,test_size=0.2,rando
m_state=20)
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

```
((119584, 26), (29897, 26), (119584,), (29897,))
```

```
from sklearn.tree import DecisionTreeRegressor,ExtraTreeRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import Lasso, Ridge, ElasticNet
from sklearn.ensemble import
GradientBoostingRegressor,AdaBoostRegressor,RandomForestRegressor
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import median_absolute_error,r2_score
from sklearn.model_selection import GridSearchCV,RandomizedSearchCV
# from xgboost import XGBRegressor
```

DecisionTree

```
model=DecisionTreeRegressor()
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('DECISIONTREE REGRESSOR',model.score(x_train,y_train)*100)
print('DECISIONTREE REGRESSOR',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
DECISIONTREE REGRESSOR 98.35008772671401
DECISIONTREE REGRESSOR 36.628417974501005
Mean Squard Error IS :      145982600079.8433
Mean Absolute Error Is :    213099.79538311213
Median Absolute Error Is :  65247.0
```

```
from sklearn.metrics import
precision_recall_curve,r2_score,confusion_matrix,classification_report
print(r2_score(y_test,y_predict))
```

```
0.36628417974501004
```

Hyper parameter tunning for DecisionTree

```
from scipy.stats import randint

# Setup the parameters and distributions to sample from: param_dist
param_dist = {"max_depth": [3,4,5,6],
              "max_features": randint(1, 9),
              "min_samples_leaf": randint(1, 9)}
```

```
# Instantiate a Decision Tree regressor: tree
tree = DecisionTreeRegressor()

# Instantiate the RandomizedSearchCV object: tree_cv
tree_cv = RandomizedSearchCV(tree, param_dist, cv = 5)

# Fit it to the data
tree_cv.fit(X, y)

# Print the tuned parameters and score
print("Tuned Decision Tree Parameters:
{}".format(tree_cv.best_params_))
print("Best score is {}".format(tree_cv.best_score_))
```

```
Tuned Decision Tree Parameters: {'max_depth': 6, 'max_features': 8,
'min_samples_leaf': 5}
Best score is 0.5458840869933715
```

K-Nearest Neighbor(KNN)

```
model=KNeighborsRegressor(n_neighbors=25,leaf_size=8,metric=
'minkowski',weights='distance',p=2)
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('KNeighborsRegressor',model.score(x_train,y_train)*100)
print('KNeighborsRegressor',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
KNeighborsRegressor 98.34997022248467
KNeighborsRegressor 61.47231028449816
Mean Squard Error IS :      88752278860.1885
Mean Absolute Error Is :    196492.16444850984
Median Absolute Error Is :  99718.46227279367
```

Hyper parameter tunning for KNN

```
#List Hyperparameters that we want to tune.
n_neighbors=list(range(1,30))
weights=['uniform','distance']
leaf_size=list(range(1,50))
algoritham=['auto', 'ball_tree', 'kd_tree', 'brute']
metric=['minkowski','Euclidean']

#Convert to dictionary
kNN_grid={'n_neighbors':n_neighbors,'weights':weights,'leaf_size':leaf
```

```
_size,'metric':metric}

#Use GridSearch
random_cv =
RandomizedSearchCV(estimator=model,param_distributions=kNN_grid,cv=5,r
andom_state=42,n_jobs=1,scoring='r2')

#Fit the model
random_cv.fit(x_train,y_train)
```

```
RandomizedSearchCV(cv=5,
                   estimator=KNeighborsRegressor(leaf_size=8,
n_neighbors=25,
                                                 weights='distance'),
                   n_jobs=1,
                   param_distributions={'leaf_size': [1, 2, 3, 4, 5,
6, 7, 8, 9,
                                                      10, 11, 12, 13,
14, 15,
                                                      16, 17, 18, 19,
20, 21,
                                                      22, 23, 24, 25,
26, 27,
                                                      28, 29,
30, ...],
                                        'metric': ['minkowski',
'Euclidean'],
                                        'n_neighbors': [1, 2, 3, 4, 5,
6, 7, 8,
                                                        9, 10, 11, 12,
13, 14,
                                                        15, 16, 17,
18, 19, 20,
                                                        21, 22, 23,
24, 25, 26,
                                                        27, 28, 29],
                                        'weights': ['uniform',
'distance']},
                   random_state=42, scoring='r2')
```

```
#Print The value of best Hyperparameters
random_cv.best_estimator_
```

```
KNeighborsRegressor(leaf_size=8, n_neighbors=25)
```

```
random_cv.best_params_
```

```
{'weights': 'uniform',
 'n_neighbors': 25,
 'metric': 'minkowski',
 'leaf_size': 8}
```

```
random_cv.best_score_
```

```
0.6147754028465494
```

```
random_cv.fit(x_test, y_test)
```

```
RandomizedSearchCV(cv=5,
                   estimator=KNeighborsRegressor(leaf_size=8,
n_neighbors=25,
                                                 weights='distance'),
                   n_jobs=1,
                   param_distributions={'leaf_size': [1, 2, 3, 4, 5,
6, 7, 8, 9,
                                                      10, 11, 12, 13,
14, 15,
                                                      16, 17, 18, 19,
20, 21,
                                                      22, 23, 24, 25,
26, 27,
                                                      28, 29,
30, ...],
                                        'metric': ['minkowski',
'Euclidean'],
                                        'n_neighbors': [1, 2, 3, 4, 5,
6, 7, 8,
                                                        9, 10, 11, 12,
13, 14,
                                                        15, 16, 17,
18, 19, 20,
                                                        21, 22, 23,
24, 25, 26,
                                                        27, 28, 29],
                                        'weights': ['uniform',
'distance']},
                   random_state=42, scoring='r2')
```

```
random_cv.best_score_
```

```
0.5975754890719027
```

AdaBoost

```
model=AdaBoostRegressor(learning_rate=0.01, loss = 'exponential',
n_estimators=100)
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('AdaBoostRegressor',model.score(x_train,y_train)*100)
print('AdaBoostRegressor',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
```

```
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
AdaBoostRegressor 57.98342868423586
AdaBoostRegressor 58.51762138842924
Mean Squard Error IS :      95558691982.42406
Mean Absolute Error Is :    230333.64569465606
Median Absolute Error Is :  177159.9018097294
```

Hyper parameter tunning for AdaBoost

```
#List Hyperparameters that we want to tune.
param_dist = {
 'n_estimators': [50, 100],
 'learning_rate' : [0.01,0.05,0.1,0.3,1],
 'loss' : ['linear', 'square', 'exponential']
 }

#Convert to dictionary
pre_gs_inst = RandomizedSearchCV(estimator=model,param_distributions =
param_dist,cv=3,n_iter = 10,n_jobs=-1)

#Fit the model
pre_gs_inst.fit(x_train, y_train)
```

```
RandomizedSearchCV(cv=3,
                   estimator=AdaBoostRegressor(learning_rate=0.01,
                                               loss='exponential',
                                               n_estimators=100),
                   n_jobs=-1,
                   param_distributions={'learning_rate': [0.01, 0.05,
0.1, 0.3,
                                                          1],
                                        'loss': ['linear', 'square',
                                                 'exponential'],
                                        'n_estimators': [50, 100]})
```

```
#Print The value of best Hyperparameters
pre_gs_inst.best_params_
```

```
{'n_estimators': 100, 'loss': 'exponential', 'learning_rate': 0.01}
```

```
pre_gs_inst.best_score_
```

```
0.579500775199341
```

```
pre_gs_inst.fit(x_test, y_test)
```

```
RandomizedSearchCV(cv=3,
                   estimator=AdaBoostRegressor(learning_rate=0.01,
                                               loss='exponential',
                                               n_estimators=100),
                   n_jobs=-1,
                   param_distributions={'learning_rate': [0.01, 0.05,
0.1, 0.3,
                                                          1],
                                        'loss': ['linear', 'square',
                                                 'exponential'],
                                        'n_estimators': [50, 100]})

pre_gs_inst.best_score_

0.5875590009668653
```

GradientBoostingRegressor

```
model=GradientBoostingRegressor(learning_rate=0.01,
loss='squared_error', n_estimators=100)
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('GradientBoostingRegressor',model.score(x_train,y_train)*100)
print('GradientBoostingRegressor',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))

GradientBoostingRegressor 50.791387120749775
GradientBoostingRegressor 51.14993044914311
Mean Squard Error IS :      112530884336.22891
Mean Absolute Error Is :    293681.8137052154
Median Absolute Error Is :  299686.6427504248
```

Hyper parameter tunning for GradientBoostingRegressor

```
from scipy.stats import loguniform

### Hyper parameter tunning for GradientBoostingRegressor
param_dist = {
    "n_estimators": [1, 2, 5, 10, 20, 50, 100, 200, 500],
    "max_leaf_nodes": [2, 5, 10, 20, 50, 100],
    "learning_rate": loguniform(0.01, 1),
    "loss": ['squared_error', 'absolute_error', 'huber', 'quantile']
}
```

```
# Instantiate the RandomizedSearchCV object: tree_cv
pre_gbr_inst =
RandomizedSearchCV(estimator=model,param_distributions=param_dist,cv=5
,random_state=42,n_jobs=1,scoring='r2')

#Fit the model
pre_gbr_inst.fit(x_train,y_train)
```

```
RandomizedSearchCV(cv=5,

estimator=GradientBoostingRegressor(learning_rate=0.01),
                   n_jobs=1,
                   param_distributions={'learning_rate':
<scipy.stats._distn_infrastructure.rv_continuous_frozen object at
0x000001D8448BE330>,
                                        'loss': ['squared_error',
                                                 'absolute_error',
'huber',
                                                 'quantile'],
                                        'max_leaf_nodes': [2, 5, 10,
20, 50,
                                                           100],
                                        'n_estimators': [1, 2, 5, 10,
20, 50,
                                                         100, 200,
500]},
                   random_state=42, scoring='r2')
```

```
# print best parameter after tuning
pre_gbr_inst.best_params_
```

```
{'learning_rate': 0.5246634533625282,
 'loss': 'huber',
 'max_leaf_nodes': 50,
 'n_estimators': 500}
```

```
pre_gbr_inst.best_score_
```

```
0.6450114858817925
```

```
pre_gbr_inst.fit(x_test, y_test)
```

```
pre_gbr_inst.best_score_
```

```
0.6486435938003832
```

Support Vector Regression (SVR)

```
#kernel='rbf', degree=3, gamma='scale',tol=0.001, C=1.0, epsilon=0.1,
shrinking=True, cache_size=200, verbose=False, max_iter=-1
model=SVR()
```

```
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('SVR',model.score(x_train,y_train)*100)
print('SVR',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

Hyper parameter tunning for Support Vector Regression (SVR)

```
# defining parameter range
param_grid = {'C': [0.1, 1, 10, 100, 1000],
              'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
              'kernel': ['linear', 'poly', 'rbf', 'sigmoid',
'precomputed']}

from sklearn.model_selection import GridSearchCV

# Instantiate the Support Vector Regression (SVR)
pre_svr_inst = GridSearchCV(SVR(), param_grid, refit = True, verbose =
3, cv=2, random_state=42, n_jobs=-1, scoring='r2')

# fitting the model for grid search
pre_svr_inst.fit(x_train, y_train)

# print best parameter after tuning
print(pre_svr_inst.best_params_)

# print how our model looks after hyper-parameter tuning
print(pre_svr_inst.best_estimator_)

pre_svr_inst.best_score_

pre_svr_inst.fit(x_test, y_test)

pre_svr_inst.best_score_
```

Nu Support Vector Regression (NuSVR)

```
from sklearn.svm import NuSVR
model=NuSVR(nu=0.5,C=1.0, cache_size=200, coef0=0.0, degree=3,
gamma='scale', kernel='rbf',
            max_iter=-1, shrinking=True, tol=0.001, verbose=False)
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('NuSVR',model.score(x_train,y_train)*100)
print('NuSVR',model.score(x_test,y_test)*100)
```

```
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
NuSVR -0.33910596880928345
NuSVR -0.27632562306028063
Mean Squard Error IS :      230996264777.12326
Mean Absolute Error Is :    458791.22380199796
Median Absolute Error Is :  493572.40015789215
```

Linear Regression

```
model=LinearRegression()
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('LinearRegression',model.score(x_train,y_train)*100)
print('LinearRegression',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
LinearRegression 40.33476065532401
LinearRegression 40.66609242685354
Mean Squard Error IS :      136681424442.58035
Mean Absolute Error Is :    324514.2211189139
Median Absolute Error Is :  331282.1605035942
```

ElasticNet

```
model=ElasticNet()
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('ElasticNet',model.score(x_train,y_train)*100)
print('ElasticNet',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
ElasticNet 39.73993059669861
ElasticNet 40.211585319361355
```

```
Mean Squard Error IS :      137728425751.1445
Mean Absolute Error Is :    326811.0385257531
Median Absolute Error Is :  335241.4139910855
```

Lasso

```
model=Lasso()
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('Lasso',model.score(x_train,y_train)*100)
print('Lasso',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
Lasso 40.3341465291015
Lasso 40.66683296688186
Mean Squard Error IS :      136679718536.63101
Mean Absolute Error Is :    324512.2332136873
Median Absolute Error Is :  331263.8391224691
```

Ridge

```
model=Ridge()
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('Ridge',model.score(x_train,y_train)*100)
print('Ridge',model.score(x_test,y_test)*100)
print('Mean Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
Ridge 40.3341742269261
Ridge 40.666086310326364
Mean Squard Error IS :      136681438532.59537
Mean Absolute Error Is :    324517.01562923816
Median Absolute Error Is :  331294.00476229243
```

Random Forest

```
model=RandomForestRegressor(random_state=42)
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
```

```
print('LinearRegression',model.score(x_train,y_train)*100)
print('LinearRegression',model.score(x_test,y_test)*100)
print('Mena Squard Error IS :     ',mean_squared_error(y_test,
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
LinearRegression 93.81843862040414
LinearRegression 64.32672035487738
Mena Squard Error IS :      82176867761.8766
Mean Absolute Error Is :    190329.92190750537
Median Absolute Error Is :  98896.88500000001
```

Hyper parameter tunning for Random Forest

```
random_grid = {
    'bootstrap': [True],
    'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, None],
    'max_features': ['auto', 'sqrt'],
    'min_samples_leaf': [1, 2, 4],
    'min_samples_split': [2, 5, 10],
    'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800,
2000]}

# Using the random grid and searching for best hyperparameters
rf = RandomForestRegressor() #creating base model

random_cv_rf = RandomizedSearchCV(estimator = rf, param_distributions
= random_grid, n_iter = 100,
                                  cv = 2, verbose=2, random_state=42,
n_jobs = -1)

random_cv_rf.fit(x_train,y_train)
```

```
Fitting 2 folds for each of 100 candidates, totalling 200 fits
```

```
# print best parameter after tuning
print(random_cv_rf.best_params_)

# print how our model looks after hyper-parameter tuning
#print(random_cv_rf.best_estimator_)

random_cv_rf.best_score_

random_cv_rf.fit(x_test, y_test)

random_cv_rf.best_score_
```

Hyper parameter tunning for Random Forest using GridSearchCV

```
from sklearn.model_selection import GridSearchCV

# Create the parameter grid based on the results of random search

param_grid = {'bootstrap': [True],
              'max_depth': [80, 90, 100, 110],
              'max_features': [2, 3],
              'min_samples_leaf': [3, 4, 5],
              'min_samples_split': [8, 10, 12],
              'n_estimators': [100, 200, 300, 1000]
             }

# Create a based model

rf = RandomForestRegressor()

# Instantiate the grid search model

grid_search_rf = GridSearchCV(estimator = rf, param_grid = param_grid,

                              cv = 3, n_jobs = -1, verbose = 2)

# Fit the grid search to the data

grid_search_rf.fit(x_train,y_train)

# print best parameter after tuning
print(grid_search_rf.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_search_rf.best_estimator_)

grid_search_rf.best_score_

grid_search_rf.fit(x_test, y_test)

grid_search_rf.best_score_
```

XGBRegressor

```
# !pip install xgboost

from xgboost import XGBRegressor

model=XGBRegressor()
model.fit(x_train,y_train)
y_predict=model.predict(x_test)
print('XGBRegressor',model.score(x_train,y_train)*100)
print('XGBRegressor',model.score(x_test,y_test)*100)
print('Mena Squard Error IS :     ',mean_squared_error(y_test,
```

```
y_predict))
print('Mean Absolute Error Is :   ',mean_absolute_error(y_test,
y_predict))
print('Median Absolute Error Is : ',median_absolute_error(y_test,
y_predict))
```

```
XGBRegressor 71.09030556504347
XGBRegressor 67.54413003448789
Mena Squard Error IS :      74765251773.45758
Mean Absolute Error Is :    187117.56211084806
Median Absolute Error Is :  114142.5625
```

Hyper parameter tunning for XGBRegressor

```
hyper_grid = {
    'n_estimators': [100, 400, 800],
    'max_depth': [3, 6, 9],
    'learning_rate': [0.05, 0.1, 0.20],
    'min_child_weight': [1, 10, 100]
    }

# Create a based model

XGBR = XGBRegressor()

# Instantiate the grid search model

grid_xgbr = RandomizedSearchCV(XGBR, hyper_grid, scoring="accuracy",
n_iter = 500, cv=4)

# Fit the grid search to the data

grid_xgbr.fit(x_train,y_train)

# print best parameter after tuning
print(grid_xgbr.best_params_)

# print how our model looks after hyper-parameter tuning
#print(grid_xgbr.best_estimator_)

grid_xgbr.best_score_

grid_xgbr.fit(x_test, y_test)

grid_xgbr.best_score_
```

Conclusion

The dataset has been examined, the model has been created, and the outcomes have been forecasted using the test data in accordance with the stated objectives. In the project at hand,

we made an effort to apply every method and the hyper parameter, however because of the hyper parameter, the project is taking longer to complete.