```
In [3]:
          import pandas as pd
          raw_data=pd.read_excel(r'C:\Users\user\Documents\Rawdata.xlsx')
 In [5]:
          raw data
 Out[5]:
              Name
                            Domain
                                              Location
                                        Age
                                                           Salary
                                                                      Exp
               Mike
                       Datascience#$ 34 years
                                               Mumbai
                                                          5^00#0
                                                                       2+
             Teddy^
                                       45' yr Bangalore
                             Testing
                                                        10%%000
                                                                       <3
          2
              Uma#r
                     Dataanalyst^^#
                                        NaN
                                                  NaN
                                                         1$5%000
                                                                    4> yrs
          3
                         Ana^^lytics
                                        NaN Hyderbad
               Jane
                                                          2000^0
                                                                     NaN
             Uttam*
                            Statistics
                                                           30000-
                                                                  5+ year
          4
                                       67-yr
                                                  NaN
          5
                Kim
                               NLP
                                        55yr
                                                  Delhi
                                                         6000^$0
                                                                      10+
          id(raw_data)
 In [7]:
 Out[7]:
          2258201618016
          raw_data.columns
 In [9]:
 Out[9]:
          Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='object')
In [11]:
          raw_data.shape
Out[11]: (6, 6)
          raw_data.head()
In [13]:
Out[13]:
              Name
                            Domain
                                        Age
                                               Location
                                                           Salary
                                                                      Exp
          0
               Mike
                       Datascience#$ 34 years
                                                          5^00#0
                                                                       2+
                                               Mumbai
             Teddy^
                                       45' yr Bangalore
                             Testing
                                                        10%%000
                                                                       <3
                     Dataanalyst^^#
          2
              Uma#r
                                        NaN
                                                  NaN
                                                         1$5%000
                                                                    4> yrs
                         Ana^^lytics
                                        NaN Hyderbad
          3
               Jane
                                                          2000^0
                                                                     NaN
                                                           30000-
                                                                  5+ year
             Uttam*
                            Statistics
                                       67-yr
                                                  NaN
In [15]:
          raw_data.tail()
```

```
Out[15]:
              Name
                            Domain
                                      Age
                                             Location
                                                         Salary
                                                                    Exp
          1 Teddy^
                                     45' yr
                             Testing
                                            Bangalore
                                                       10%%000
                                                                     <3
              Uma#r Dataanalyst^^#
          2
                                      NaN
                                                 NaN
                                                       1$5%000
                                                                  4> yrs
          3
                         Ana^^lytics
                                      NaN Hyderbad
                                                                    NaN
                Jane
                                                         2000^0
              Uttam*
                            Statistics
                                      67-yr
                                                 NaN
                                                         30000-
                                                                 5+ year
          5
                                                Delhi
                                                       6000^$0
                                                                    10+
                Kim
                                NLP
                                      55yr
```

In [17]: raw_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 6 columns):

#	Column	Non-Null Count	Dtype
0	Name	6 non-null	object
1	Domain	6 non-null	object
2	Age	4 non-null	object
3	Location	4 non-null	object
4	Salary	6 non-null	object
5	Exp	5 non-null	object
44		(6)	

dtypes: object(6)

memory usage: 420.0+ bytes

In [19]: raw_data.isnull()

Out[19]:

	Name	Domain	Age	Location	Salary	Ехр
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	True	True	False	False
3	False	False	True	False	False	True
4	False	False	False	True	False	False
5	False	False	False	False	False	False

In [21]: raw_data.isna() # same as isnull()

Out[21]:

	Name	Domain	Age	Location	Salary	Exp
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	True	True	False	False
3	False	False	True	False	False	True
4	False	False	False	True	False	False
5	False	False	False	False	False	False

```
raw data.isnull().sum()
In [23]:
                        0
Out[23]:
          Name
           Domain
                        0
           Age
                        2
                        2
           Location
           Salary
                        0
           Exp
                        1
           dtype: int64
In [25]:
          raw_data
Out[25]:
               Name
                             Domain
                                          Age
                                                 Location
                                                              Salary
                                                                         Exp
                Mike
                        Datascience#$
                                       34 years
                                                  Mumbai
                                                             5^00#0
                                                                          2+
              Teddy^
                                         45' yr Bangalore
                                                           10%%000
                                                                           <3
                              Testing
           2
              Uma#r
                      Dataanalyst^^#
                                          NaN
                                                     NaN
                                                            1$5%000
                                                                       4> yrs
           3
                          Ana^^lytics
                                          NaN
                                                Hyderbad
                                                             2000^0
                                                                         NaN
                Jane
           4
              Uttam*
                             Statistics
                                                              30000-
                                                                      5+ year
                                          67-yr
                                                     NaN
           5
                 Kim
                                 NLP
                                          55yr
                                                     Delhi
                                                            6000^$0
                                                                         10+
```

DATA CLEANING AND DATA CLEASING

```
In [28]:
         raw_data['Name']=raw_data['Name'].str.replace(r'\W','',regex=True) # \W capture
         raw_data['Name'] # '' This is the replacement string. Here, it's an empty string
Out[28]:
          0
                Mike
               Teddy
          1
          2
                Umar
          3
                Jane
          4
               Uttam
          5
                 Kim
          Name: Name, dtype: object
In [30]:
         raw data['Domain']
Out[30]:
          0
                Datascience#$
          1
                      Testing
          2
               Dataanalyst^^#
                  Ana^^lytics
          3
          4
                   Statistics
          5
                          NLP
          Name: Domain, dtype: object
         raw data['Domain']=raw data['Domain'].str.replace(r'\W','',regex=True)
In [32]:
         raw_data['Domain']
In [34]:
```

```
Out[34]: 0
               Datascience
          1
                   Testing
          2
               Dataanalyst
          3
                 Analytics
                Statistics
          4
          5
                       NLP
          Name: Domain, dtype: object
          raw_data['Age']=raw_data['Age'].str.replace(r'\W','',regex=True)
In [36]:
          raw_data['Age']
Out[36]:
          0
               34years
          1
                  45yr
          2
                   NaN
          3
                   NaN
          4
                  67yr
                  55yr
          Name: Age, dtype: object
          raw_data['Age']=raw_data['Age'].str.extract('(\d+)') # \d+: \d matches any digit
In [68]:
          raw_data['Age']
        <>:1: SyntaxWarning: invalid escape sequence '\d'
        <>:1: SyntaxWarning: invalid escape sequence '\d'
        C:\Users\user\AppData\Local\Temp\ipykernel_1852\2837976887.py:1: SyntaxWarning: i
        nvalid escape sequence '\d'
          raw_data['Age']=raw_data['Age'].str.extract('(\d+)') # \d+: \d matches any digi
        t (0-9), and the + means "one or more" digits.
                34
Out[68]: 0
          1
                45
          2
               NaN
               NaN
          3
          4
                67
                55
          Name: Age, dtype: object
In [70]:
          raw data
Out[70]:
             Name
                       Domain
                                Age
                                       Location
                                                Salary
                                                        Exp
                                                           2
          0
              Mike
                    Datascience
                                  34
                                        Mumbai
                                                  5000
          1
             Teddy
                        Testing
                                  45
                                      Bangalore
                                                 10000
                                                           3
          2
             Umar
                     Dataanalyst
                                NaN
                                           NaN
                                                 15000
                                                           4
          3
              Jane
                       Analytics
                                NaN
                                      Hyderbad
                                                 20000
                                                       NaN
          4
             Uttam
                       Statistics
                                  67
                                           NaN
                                                 30000
                                                           5
                           NLP
          5
               Kim
                                  55
                                          Delhi
                                                 60000
                                                          10
In [72]:
          raw_data['Location']=raw_data['Location'].str.replace(r'\W','',regex=True)
          raw data
```

```
Out[72]:
             Name
                        Domain
                                 Age
                                        Location
                                                 Salary
                                                          Exp
          0
              Mike Datascience
                                   34
                                         Mumbai
                                                   5000
                                                            2
              Teddy
                         Testing
                                   45 Bangalore
                                                  10000
                                                            3
          1
                     Dataanalyst
          2
              Umar
                                 NaN
                                            NaN
                                                  15000
                                                            4
          3
               Jane
                       Analytics
                                 NaN
                                       Hyderbad
                                                  20000 NaN
          4
             Uttam
                        Statistics
                                   67
                                            NaN
                                                  30000
                                                            5
          5
                Kim
                            NLP
                                   55
                                           Delhi
                                                  60000
                                                           10
In [74]:
          raw_data['Salary']=raw_data['Salary'].str.replace(r'\W','',regex=True)
          raw_data
Out[74]:
             Name
                        Domain
                                 Age
                                        Location Salary
                                                          Exp
          0
              Mike
                     Datascience
                                   34
                                         Mumbai
                                                   5000
                                                            2
              Teddy
                                       Bangalore
                                                  10000
                                                            3
          1
                         Testing
                                   45
          2
              Umar
                                                  15000
                     Dataanalyst
                                 NaN
                                            NaN
                                                            4
                                       Hyderbad
          3
               Jane
                       Analytics
                                 NaN
                                                  20000 NaN
                        Statistics
                                                  30000
          4
             Uttam
                                   67
                                            NaN
                                                            5
                            NLP
                                           Delhi
                                                  60000
          5
                Kim
                                   55
                                                           10
          raw_data['Exp']=raw_data['Exp'].str.extract('(\d+)')
In [76]:
          raw_data
         <>:1: SyntaxWarning: invalid escape sequence '\d'
         <>:1: SyntaxWarning: invalid escape sequence '\d'
         C:\Users\user\AppData\Local\Temp\ipykernel 1852\206955048.py:1: SyntaxWarning: in
        valid escape sequence '\d'
           raw_data['Exp']=raw_data['Exp'].str.extract('(\d+)')
Out[76]:
             Name
                        Domain
                                 Age
                                        Location
                                                  Salary
          0
              Mike
                     Datascience
                                                   5000
                                                            2
                                   34
                                         Mumbai
              Teddy
                         Testing
                                   45
                                       Bangalore
                                                  10000
                                                            3
          2
              Umar
                     Dataanalyst
                                 NaN
                                            NaN
                                                  15000
                                                            4
          3
               Jane
                       Analytics
                                 NaN
                                       Hyderbad
                                                  20000
          4
             Uttam
                        Statistics
                                   67
                                            NaN
                                                  30000
                                                            5
          5
                Kim
                            NLP
                                   55
                                           Delhi
                                                  60000
                                                           10
          clean data=raw data.copy()
In [78]:
In [80]:
          clean_data
```

Out[

80]:		Name	Domain	Age	Location	Salary	Ехр
			Datascience	34	Mumbai	5000	2
1 Teddy Testing 45 Bangalore		10000	3				
2 Umar D		Dataanalyst	NaN	NaN	15000	4	
	3	Jane	Analytics	NaN	Hyderbad	20000	NaN
		Uttam	Statistics	67	NaN	30000	5
	5	Kim	NLP	55	Delhi	60000	10

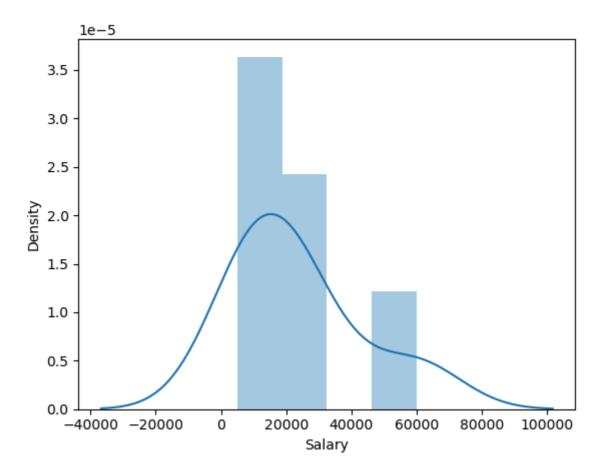
EDA Techniques

```
In [83]: clean_data.isnull().sum()
                                                                                                                0
Out[83]: Name
                                                   Domain
                                                                                                               2
                                                   Age
                                                   Location
                                                                                                               2
                                                   Salary
                                                   Exp
                                                                                                                1
                                                   dtype: int64
In [85]:
                                                clean_data['Age']
Out[85]:
                                                                                 34
                                                   1
                                                                                 45
                                                   2
                                                                           NaN
                                                   3
                                                                           NaN
                                                   4
                                                                                 67
                                                   Name: Age, dtype: object
                                             import numpy as np
In [87]:
                                                clean_data['Age']=clean_data['Age'].fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.mean(pd.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age']).fillna(np.to_numeric(clean_data['Age'])).fillna(np.to_numeric
In [89]:
                                                 clean_data['Age']
Out[89]:
                                                                                           34
                                                   1
                                                                                           45
                                                                          50.25
                                                   3
                                                                          50.25
                                                                                           67
                                                                                           55
                                                   Name: Age, dtype: object
                                                clean_data['Exp']=clean_data['Exp'].fillna(np.mean(pd.to_numeric(clean_data['Exp
                                                 clean_data['Exp']
```

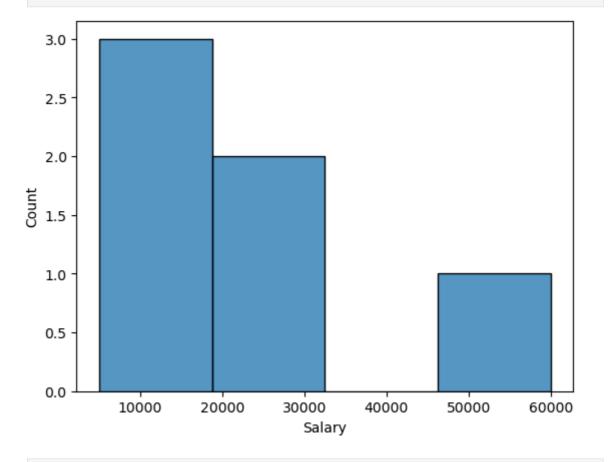
```
Out[91]:
                  2
           1
                  3
           2
                  4
           3
                4.8
           4
                  5
           5
                 10
           Name: Exp, dtype: object
          clean_data['Location'].isnull().sum()
Out[93]: 2
In [105...
           clean_data['Location']=clean_data['Location'].fillna(clean_data['Location'].mode
           clean_data['Location'] # The [0] is used to select the first mode from the Serie
Out[105...
                   Mumbai
           1
                Bangalore
           2
                 Hyderbad
           3
                 Hyderbad
           4
                 Hyderbad
           5
                     Delhi
           Name: Location, dtype: object
In [122...
           clean_data
Out[122...
              Name
                        Domain
                                         Location Salary Exp
                                  Age
                                                            2
           0
               Mike
                     Datascience
                                    34
                                          Mumbai
                                                    5000
              Teddy
                                                   10000
                                                            3
                         Testing
                                    45
                                        Bangalore
           2
                                 50.25
               Umar
                      Dataanalyst
                                        Bangalore
                                                   15000
                                                            4
           3
                                 50.25
               Jane
                        Analytics
                                        Hyderbad
                                                   20000
                                                           4.8
                                                            5
           4
              Uttam
                        Statistics
                                    67
                                        Bangalore
                                                   30000
           5
                Kim
                            NLP
                                    55
                                            Delhi
                                                   60000
                                                           10
In [124...
           clean data.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 6 entries, 0 to 5
         Data columns (total 6 columns):
              Column
                         Non-Null Count Dtype
          0
              Name
                         6 non-null
                                          object
          1
              Domain
                         6 non-null
                                          object
          2
                         6 non-null
                                          object
              Age
          3
               Location 6 non-null
                                          object
                         6 non-null
                                          object
               Salary
          5
               Exp
                         6 non-null
                                          object
         dtypes: object(6)
         memory usage: 420.0+ bytes
In [129...
           clean_data['Age']=clean_data['Age'].astype(int)
           clean_data['Age']
```

```
Out[129...
               34
          1
               45
          2
               50
          3
               50
          4
               67
          5
               55
          Name: Age, dtype: int32
In [131...
          clean_data['Salary']=clean_data['Salary'].astype(int)
          clean_data['Salary']
Out[131...
          0
                5000
          1
               10000
          2
               15000
          3
               20000
          4
               30000
          5
               60000
          Name: Salary, dtype: int32
In [133...
          clean_data['Exp']=clean_data['Exp'].astype(int)
          clean_data['Exp']
Out[133...
          0
                2
          1
                3
          2
                4
          3
                4
                5
          4
          5
               10
          Name: Exp, dtype: int32
In [135...
         clean_data.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 6 entries, 0 to 5
         Data columns (total 6 columns):
             Column
                      Non-Null Count Dtype
             -----
                        -----
          0
                                       object
             Name
                      6 non-null
          1 Domain 6 non-null
                                       object
          2
                       6 non-null
                                        int32
             Age
          3
             Location 6 non-null
                                        object
          4
                       6 non-null
                                        int32
              Salary
          5
              Exp
                        6 non-null
                                        int32
         dtypes: int32(3), object(3)
         memory usage: 348.0+ bytes
In [143...
          clean_data['Name']=clean_data['Name'].astype('category')
          clean_data['Location']=clean_data['Location'].astype('category')
          clean_data['Domain']=clean_data['Domain'].astype('category')
          clean data.info()
```

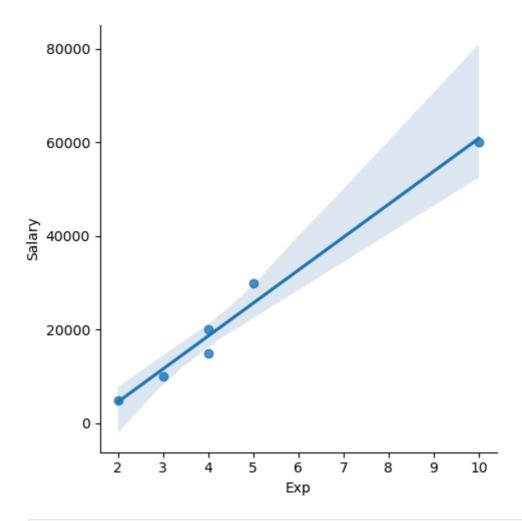
```
<class 'pandas.core.frame.DataFrame'>
         RangeIndex: 6 entries, 0 to 5
        Data columns (total 6 columns):
             Column
                     Non-Null Count Dtype
         ---
                      -----
                       6 non-null
            Name
         0
                                       category
                                    category
         1 Domain 6 non-null
         2 Age
                      6 non-null
                                      int32
         3 Location 6 non-null
                                       category
         4
             Salary
                       6 non-null
                                       int32
         5
                       6 non-null
                                       int32
             Exp
         dtypes: category(3), int32(3)
         memory usage: 866.0 bytes
In [146...
          clean_data
Out[146...
             Name
                      Domain Age Location Salary Exp
              Mike Datascience
          0
                                34
                                     Mumbai
                                               5000
                                                       2
          1
             Teddy
                       Testing
                                45 Bangalore 10000
                                                       3
          2
             Umar Dataanalyst
                                50 Bangalore
                                              15000
                                                       4
          3
              Jane
                      Analytics
                                50 Hyderbad 20000
                                                       4
          4
            Uttam
                      Statistics
                                67
                                    Bangalore 30000
                                                       5
          5
               Kim
                          NLP
                                55
                                        Delhi 60000
                                                      10
In [148...
          import os
          os.getcwd()
Out[148...
          'C:\\Users\\user'
          import matplotlib.pyplot as plt # visualization
In [150...
          import seaborn as sns
In [152...
          import warnings
          warnings.filterwarnings('ignore')
In [154...
          clean_data['Salary']
Out[154...
          0
                5000
          1
               10000
          2
               15000
          3
               20000
          4
               30000
          5
               60000
          Name: Salary, dtype: int32
In [156...
         vis1=sns.distplot(clean_data['Salary'])
```



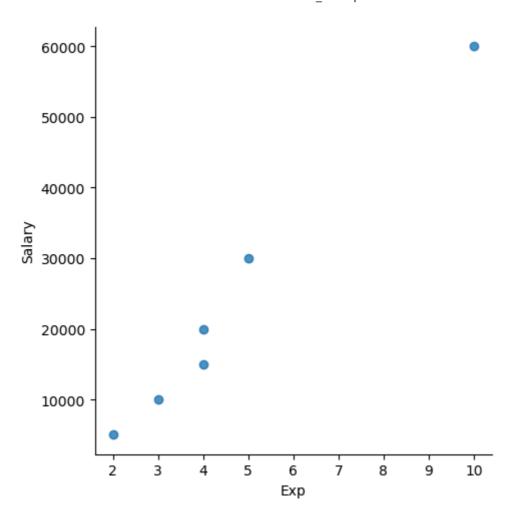
In [158... vis2=sns.histplot(clean_data['Salary'])



In [160... vis3=sns.lmplot(data=clean_data,x='Exp',y='Salary')



In [162... vis4=sns.lmplot(data=clean_data,x='Exp',y='Salary',fit_reg=False)



In [166... clean_data[:]

Out[166...

	Name	Domain	Age	Location	Salary	Ехр
0	Mike	Datascience	34	Mumbai	5000	2
1	Teddy	Testing	45	Bangalore	10000	3
2	Umar	Dataanalyst	50	Bangalore	15000	4
3	Jane	Analytics	50	Hyderbad	20000	4
4	Uttam	Statistics	67	Bangalore	30000	5
5	Kim	NLP	55	Delhi	60000	10

In [168... clean_data[0:6:2]

Out[168...

	Name	Domain	Age	Location	Salary	Ехр
0	Mike	Datascience	34	Mumbai	5000	2
2	Umar	Dataanalyst	50	Bangalore	15000	4
4	Uttam	Statistics	67	Bangalore	30000	5

In [170... clean_data[::-1]

```
Out[170...
              Name
                         Domain Age
                                       Location
                                                  Salary Exp
           5
                Kim
                             NLP
                                    55
                                            Delhi
                                                   60000
                                                           10
             Uttam
                        Statistics
                                    67
                                        Bangalore
                                                   30000
                                                             5
           3
                Jane
                        Analytics
                                    50
                                        Hyderbad
                                                   20000
                                                             4
           2
               Umar
                      Dataanalyst
                                    50
                                        Bangalore
                                                   15000
                                                             4
           1
               Teddy
                          Testing
                                    45
                                        Bangalore
                                                   10000
                                                             3
                Mike Datascience
                                    34
                                         Mumbai
                                                    5000
                                                             2
In [172...
           clean_data.columns
Out[172...
           Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='object')
           x_iv=clean_data[['Name','Domain','Age','Location','Salary','Exp']]
In [176...
           x_iv
Out[176...
                                       Location Salary Exp
              Name
                         Domain Age
           0
               Mike Datascience
                                    34
                                         Mumbai
                                                    5000
                                                             2
               Teddy
                          Testing
                                    45
                                        Bangalore
                                                   10000
                                                             3
           1
           2
               Umar
                      Dataanalyst
                                        Bangalore
                                                   15000
                                                             4
                                    50
           3
                Jane
                        Analytics
                                        Hyderbad
                                                   20000
                                    50
              Uttam
                         Statistics
                                        Bangalore
                                                   30000
                                                             5
           4
                                    67
           5
                 Kim
                             NLP
                                    55
                                            Delhi
                                                   60000
                                                            10
In [180...
           y_dv=clean_data['Salary']
           y_dv
Out[180...
           0
                  5000
           1
                 10000
           2
                 15000
           3
                 20000
           4
                 30000
                 60000
           Name: Salary, dtype: int32
```

In [182...

raw data

\cap	n+	Γ	1	Q	7
U	uс	L	+	O	۷.

	Name	Domain	Age	Location	Salary	Ехр
0	Mike	Datascience	34	Mumbai	5000	2
1	Teddy	Testing	45	Bangalore	10000	3
2	Umar	Dataanalyst	NaN	NaN	15000	4
3	Jane	Analytics	NaN	Hyderbad	20000	NaN
4	Uttam	Statistics	Statistics 67 NaN		30000	5
5	Kim	NLP	55	Delhi	60000	10

In [184...

clean_data

Out[184...

	Name	Domain	Age	Location	Salary	Ехр
0	Mike	ike Datascience		Mumbai	5000	2
1	Teddy	Testing	45	Bangalore	10000	3
2	Umar	Dataanalyst	50	Bangalore	15000	4
3	Jane	Analytics	50	Hyderbad	20000	4
4	Uttam	Statistics	67	67 Bangalore		5
5	Kim	NLP	55	Delhi	60000	10

In [186...

x_iv

Out[186...

	Name	Domain	Age	Location	Salary	Ехр
0	Mike	Datascience	34	Mumbai	5000	2
1	Teddy	Testing	45	Bangalore	10000	3
2	Umar	Dataanalyst	50	Bangalore	15000	4
3	Jane	Analytics	50	Hyderbad	20000	4
4	Uttam	Statistics	67	Bangalore	30000	5
5	Kim	NLP	55	Delhi	60000	10

In [188...

y_dv

Out[188...

- 0 5000
- 1 10000
- 2 15000
- 3 20000
- 4 30000
- 5 60000

Name: Salary, dtype: int32

In [192...

imputation=pd.get_dummies(clean_data) # variable creation
imputation

Out[192		Age	Salary	Ехр	Name_Jane	Name_Kim	Name_Mike	Name_Teddy	Name_Umar
	0	34	5000	2	False	False	True	False	False
	1	45	10000	3	False	False	False	True	False
	2	50	15000	4	False	False	False	False	True
	3	50	20000	4	True	False	False	False	False
	4	67	30000	5	False	False	False	False	False
	5	55	60000	10	False	True	False	False	False
	4								•
In []:									