```
import pandas as pd
 In [1]:
 In [2]: emp = pd.read_excel(r'C:\Users\soham\OneDrive\Desktop\Rawdata.xlsx')
 In [3]:
         emp
 Out[3]:
                           Domain
              Name
                                       Age
                                             Location
                                                          Salary
                                                                    Exp
                      Datascience#$ 34 years
          0
               Mike
                                              Mumbai
                                                         5^00#0
                                                                     2+
          1 Teddy^
                                      45' yr Bangalore
                                                                     <3
                            Testing
                                                       10%%000
          2
             Uma#r
                    Dataanalyst^^#
                                       NaN
                                                 NaN
                                                        1$5%000
                                                                  4> yrs
          3
                        Ana^^lytics
                                       NaN Hyderbad
               Jane
                                                         2000^0
                                                                    NaN
          4
             Uttam*
                           Statistics
                                      67-yr
                                                                 5+ year
                                                 NaN
                                                         30000-
          5
                               NLP
                                                                    10+
                Kim
                                       55yr
                                                 Delhi
                                                        6000^$0
 In [4]:
         emp.shape #total Lenth of dimension
 Out[4]: (6, 6)
 In [6]:
         len(emp)
 Out[6]: 6
 In [8]: emp.columns
 Out[8]: Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='object')
In [13]: emp.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
                       6 non-null
                                        object
         1
             Domain
         2
             Age
                       4 non-null
                                        object
         3
             Location 4 non-null
                                        object
         4
                       6 non-null
                                        object
             Salary
             Exp
                        5 non-null
                                        object
        dtypes: object(6)
        memory usage: 420.0+ bytes
In [15]: emp
```

```
Out[15]:
              Name
                            Domain
                                        Age
                                              Location
                                                          Salary
                                                                     Exp
          0
               Mike
                      Datascience#$ 34 years
                                               Mumbai
                                                          5^00#0
                                                                      2+
          1 Teddy^
                             Testing
                                       45' yr Bangalore
                                                        10%%000
                                                                      <3
             Uma#r Dataanalyst^^#
          2
                                       NaN
                                                  NaN
                                                        1$5%000
                                                                   4> yrs
          3
               Jane
                        Ana^^lytics
                                       NaN Hyderbad
                                                          2000^0
                                                                    NaN
          4
             Uttam*
                           Statistics
                                       67-yr
                                                  NaN
                                                          30000- 5+ year
          5
                Kim
                               NLP
                                        55yr
                                                  Delhi
                                                        6000^$0
                                                                     10+
In [17]:
         emp['Name']
Out[17]: 0
                 Mike
          1
               Teddy^
          2
                Uma#r
          3
                 Jane
          4
               Uttam*
                  Kim
          Name: Name, dtype: object
In [19]:
         emp['Domain']
Out[19]: 0
                Datascience#$
          1
                      Testing
               Dataanalyst^^#
          2
          3
                  Ana^^lytics
          4
                   Statistics
                           NLP
          Name: Domain, dtype: object
In [21]:
         emp['Age']
Out[21]:
          0
               34 years
                 45' yr
          1
          2
                    NaN
          3
                    NaN
          4
                  67-yr
          5
                   55yr
          Name: Age, dtype: object
In [23]:
         emp['Location']
Out[23]:
          0
                  Mumbai
          1
               Bangalore
          2
                     NaN
                Hyderbad
          3
          4
                     NaN
                   Delhi
          Name: Location, dtype: object
In [25]: emp['Salary']
```

```
Out[25]: 0
                 5^00#0
          1
               10%%000
               1$5%000
          3
                 2000^0
          4
                 30000-
               6000^$0
          Name: Salary, dtype: object
In [27]:
         emp['Exp']
Out[27]:
          1
                     <3
                4> yrs
          3
                    NaN
               5+ year
                    10+
          Name: Exp, dtype: object
          emp[['Name','Age']]
In [29]:
Out[29]:
              Name
                         Age
          0
               Mike
                     34 years
             Teddy^
                        45' yr
          2
              Uma#r
                         NaN
                         NaN
                Jane
             Uttam*
                        67-yr
          5
                         55yr
                Kim
          emp[['Name','Domain','Age','Location','Salary','Exp']]
Out[31]:
              Name
                            Domain
                                         Age
                                               Location
                                                            Salary
                                                                       Exp
          0
               Mike
                                                                        2+
                       Datascience#$
                                     34 years
                                                Mumbai
                                                           5^00#0
             Teddy^
                             Testing
                                        45' yr
                                              Bangalore
                                                         10%%000
                                                                        <3
                      Dataanalyst^^#
          2
              Uma#r
                                        NaN
                                                   NaN
                                                          1$5%000
                                                                     4> yrs
          3
                         Ana^^lytics
                                              Hyderbad
                                        NaN
                                                           2000^0
                Jane
                                                                      NaN
                                                                    5+ year
             Uttam*
                            Statistics
                                        67-yr
                                                   NaN
                                                            30000-
          5
                                NLP
                                         55yr
                                                   Delhi
                                                          6000^$0
                                                                       10+
                Kim
```

## Data cleansing of row formate#

```
In [34]: emp['Name']
```

```
Out[34]: 0
                Mike
             Teddy^
          1
          2
               Uma#r
          3
                 Jane
          4
              Uttam*
                  Kim
          Name: Name, dtype: object
In [36]: emp['Name'] = emp['Name'].str.replace(r'\W','',regex=True)
In [38]:
         emp['Name']
Out[38]: 0
               Mike
          1
              Teddy
          2
               Umar
          3
                Jane
              Uttam
          4
                 Kim
          Name: Name, dtype: object
        emp['Domain'] = emp['Domain'].str.replace(r'\W','',regex=True)
In [40]:
In [42]: emp['Domain']
              Datascience
Out[42]: 0
          1
                  Testing
          2
             Dataanalyst
          3
                Analytics
          4
                Statistics
          5
                       NLP
          Name: Domain, dtype: object
In [44]: emp['Age'] = emp['Age'].str.replace(r'\W','',regex=True)
In [46]: emp['Age']
Out[46]: 0
               34years
          1
                 45yr
          2
                  NaN
          3
                  NaN
          4
                  67yr
                 55yr
          Name: Age, dtype: object
In [48]: | emp['Age'] = emp['Age'].str.extract('(\d+)') #using age cancel year
        <>:1: SyntaxWarning: invalid escape sequence '\d'
        <>:1: SyntaxWarning: invalid escape sequence '\d'
        C:\Users\soham\AppData\Local\Temp\ipykernel_22004\4102034463.py:1: SyntaxWarning:
        invalid escape sequence '\d'
          emp['Age'] = emp['Age'].str.extract('(\d+)') #using age cancel year
In [50]: emp['Age']
```

```
Out[50]: 0
               34
               45
          1
          2
              NaN
          3
              NaN
              67
          4
          5
               55
          Name: Age, dtype: object
In [52]: emp['Location'] = emp['Location'].str.replace(r'\W','')
In [54]: emp['Location']
Out[54]: 0
                 Mumbai
          1
              Bangalore
          2
                     NaN
          3
              Hyderbad
                    NaN
          4
                  Delhi
          Name: Location, dtype: object
In [56]: emp['Salary'] = emp['Salary'].str.replace(r'\W','',regex=True)
In [58]: emp['Salary']
Out[58]: 0
               5000
              10000
          1
          2
              15000
          3
              20000
          4
              30000
              60000
          Name: Salary, dtype: object
In [60]: emp['Exp'] = emp['Exp'].str.replace(r'\W','',regex=True)
In [62]: emp['Exp'] = emp['Exp'].str.extract('(\d+)')
        <>:1: SyntaxWarning: invalid escape sequence '\d'
        <>:1: SyntaxWarning: invalid escape sequence '\d'
        C:\Users\soham\AppData\Local\Temp\ipykernel_22004\3836251810.py:1: SyntaxWarning:
        invalid escape sequence '\d'
          emp['Exp'] = emp['Exp'].str.extract('(\d+)')
In [64]: emp['Exp']
Out[64]: 0
                 2
          1
                 3
          2
                 4
          3
              NaN
          4
                 5
                10
          Name: Exp, dtype: object
         emp #cleaning of data
In [66]:
```

Out[66]:	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	67	NaN	30000	5
	5	Kim	NLP	55	Delhi	60000	10
In [68]:	cl	ean_dat	a = emp.copy	/()			
In [70]:	cl	ean_dat	a				
Out[70]:		Name	Domain	Age	Location	Salary	Ехр
	0	Mike	Datascience	34	Mumbai	5000	2
	1	Teddy	Testing	45	Mumbai       5000       2         Bangalore       10000       3         NaN       15000       4         Hyderbad       20000       NaN         NaN       30000       5         Delhi       60000       10         Location       Salary       Exp         Mumbai       5000       2         Bangalore       10000       3         NaN       15000       4         Hyderbad       20000       NaN         NaN       30000       5         Delhi       60000       10         Et #         Location       Salary       Exp         Mumbai       5000       2		
	2	Mike Datas Teddy T Umar Datas Jane An Uttam Sta Kim  ean_data = er ean_data  Mame Datas Teddy T Umar Datas Jane An Uttam Sta Kim  Missing value ean_data  Name Datas Teddy T Teddy T Teddy T The Company of the Company	Dataanalyst	NaN	NaN	15000	4
	3	Jane	Analytics	e 34 Mumbai 5000 2 g 45 Bangalore 10000 3 et NaN			
	4	Uttam	Statistics	67	NaN	30000	5
	5	Kim	NLP	55	Delhi	60000	10
In [72]:	#	Missing	values tred	atment	#		
In [74]:	cl	ean_dat	a				
Out[74]:		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	67	NaN	30000	5
	5	Kim	NLP	55	Delhi	60000	10

In [76]: 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
             Domain
         1
                    6 non-null
                                       object
         2
                       4 non-null
                                       object
           Age
            Location 4 non-null
                                       object
         3
         4
             Salary
                       6 non-null
                                       object
         5
                       5 non-null
                                       object
             Exp
        dtypes: object(6)
        memory usage: 420.0+ bytes
In [78]: clean_data.head(2)
Out[78]:
            Name
                      Domain Age
                                     Location Salary Exp
          0
             Mike Datascience
                                34
                                      Mumbai
                                                5000
                                                        2
             Teddy
                       Testing
                                 45
                                    Bangalore
                                               10000
                                                        3
In [80]:
         import numpy as np
         import pandas as pd
In [82]:
         clean_data['Age']
Out[82]: 0
                34
          1
                45
          2
               NaN
          3
              NaN
          4
                67
          5
                55
          Name: Age, dtype: object
In [84]:
         clean_data['Age'] = clean_data['Age'].fillna(np.mean(pd.to_numeric(clean_data['
In [86]:
         import numpy as np
         import pandas as pd
In [88]:
         clean_data
Out[88]:
             Name
                      Domain
                                Age
                                      Location Salary
                                                       Ехр
          0
             Mike
                                                 5000
                                                          2
                   Datascience
                                 34
                                       Mumbai
             Teddy
                        Testing
                                 45
                                     Bangalore
                                                10000
                                                          3
          2
             Umar
                                                         4
                    Dataanalyst
                               50.25
                                          NaN
                                                15000
          3
              Jane
                      Analytics
                               50.25
                                      Hyderbad
                                                20000 NaN
                                                          5
          4
            Uttam
                      Statistics
                                 67
                                          NaN
                                                30000
                          NLP
                                  55
               Kim
                                          Delhi
                                                60000
                                                         10
In [90]:
         clean_data['Age']
```

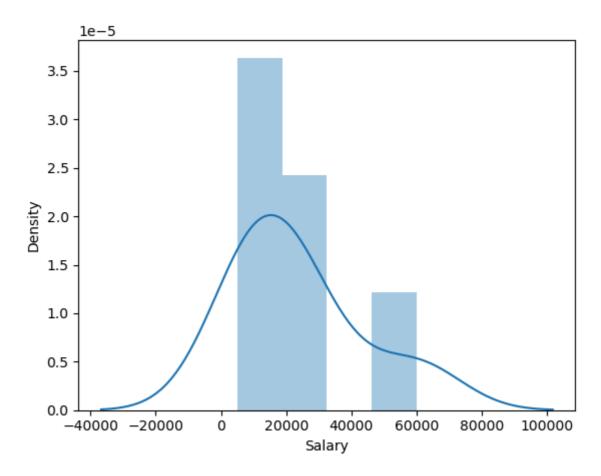
```
Out[90]: 0
                   34
                   45
           1
                50.25
           2
           3
                50.25
           4
                   67
           5
                   55
           Name: Age, dtype: object
          clean_data['Exp'] = clean_data['Exp'].fillna(np.mean(pd.to_numeric(clean data['
In [94]:
          clean_data['Exp']
Out[94]:
           0
                  2
           1
                  3
           2
                  4
           3
                4.8
           4
                  5
                 10
           Name: Exp, dtype: object
In [96]:
          clean_data
Out[96]:
              Name
                        Domain
                                         Location Salary Exp
                                  Age
           0
               Mike Datascience
                                    34
                                         Mumbai
                                                    5000
                                                            2
              Teddy
                         Testing
                                        Bangalore
                                                   10000
           1
                                    45
                                                            3
           2
              Umar
                     Dataanalyst 50.25
                                            NaN
                                                   15000
                                                            4
               Jane
                        Analytics 50.25
                                        Hyderbad
                                                   20000
           3
                                                          4.8
           4
              Uttam
                        Statistics
                                            NaN
                                                   30000
                                                            5
                                    67
           5
                Kim
                            NLP
                                    55
                                            Delhi
                                                  60000
                                                           10
          clean_data['Location'] = clean_data['Location'].fillna(clean_data['Location'].m
In [98]:
In [100...
           clean_data['Location']
Out[100...
           0
                   Mumbai
           1
                Bangalore
           2
                Bangalore
           3
                 Hyderbad
           4
                Bangalore
                    Delhi
           Name: Location, dtype: object
In [102...
           clean_data
```

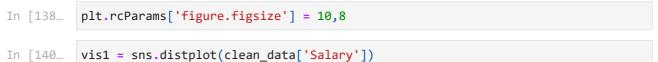
```
Out[102...
              Name
                         Domain
                                   Age
                                         Location Salary Exp
           0
               Mike Datascience
                                    34
                                          Mumbai
                                                    5000
                                                             2
               Teddy
                          Testing
                                    45
                                        Bangalore
                                                   10000
                                                             3
           1
           2
               Umar
                      Dataanalyst
                                  50.25
                                        Bangalore
                                                   15000
                                                             4
           3
                Jane
                        Analytics
                                  50.25
                                         Hyderbad
                                                   20000
                                                           4.8
           4
              Uttam
                        Statistics
                                        Bangalore
                                                   30000
                                                             5
                                    67
           5
                Kim
                            NLP
                                    55
                                             Delhi
                                                   60000
                                                            10
           clean_data['Age'] = clean_data['Age'].astype(int)
In [104...
           clean_data['Salary'] = clean_data['Salary'].astype(int)
In [106...
In [108...
           clean_data['Name'] = clean_data['Name'].astype('category')
           clean_data['Location'] = clean_data['Location'].astype('category')
In [110...
           clean_data['Domain'] = clean_data['Domain'].astype('category')
In [112...
In [114...
           clean_data['Exp'] = clean_data['Exp'].astype(int)
In [116...
           clean_data
Out[116...
              Name
                         Domain Age
                                        Location Salary Exp
                                                            2
           0
               Mike
                     Datascience
                                   34
                                         Mumbai
                                                   5000
              Teddy
                                                  10000
                                                            3
           1
                          Testing
                                   45
                                       Bangalore
           2
               Umar
                      Dataanalyst
                                       Bangalore
                                                   15000
                                                            4
                                   50
           3
                        Analytics
                                       Hyderbad
                                                  20000
                Jane
                                   50
                                                            4
                                                            5
              Uttam
                        Statistics
                                                   30000
           4
                                   67
                                       Bangalore
           5
                            NLP
                                   55
                                                  60000
                Kim
                                            Delhi
                                                           10
In [118...
           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
                                           category
          1
               Domain
                         6 non-null
                                           category
          2
                         6 non-null
                                           int32
               Location 6 non-null
          3
                                           category
          4
               Salary
                         6 non-null
                                           int32
          5
               Exp
                          6 non-null
                                           int32
          dtypes: category(3), int32(3)
         memory usage: 866.0 bytes
```

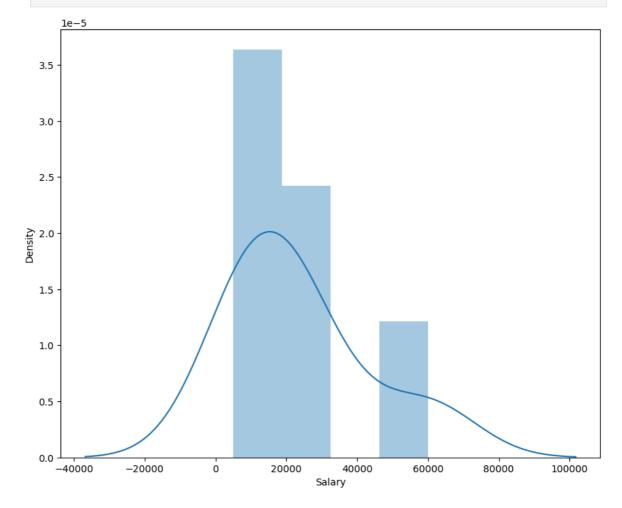
```
In [120...
           clean_data
Out[120...
              Name
                         Domain
                                  Age
                                        Location Salary Exp
                                                             2
           0
               Mike
                      Datascience
                                    34
                                         Mumbai
                                                    5000
               Teddy
                          Testing
                                   45
                                        Bangalore
                                                   10000
                                                             3
           2
               Umar
                      Dataanalyst
                                    50
                                        Bangalore
                                                   15000
                                                            4
           3
                Jane
                        Analytics
                                    50
                                        Hyderbad
                                                  20000
                                                            4
                         Statistics
                                        Bangalore
                                                   30000
                                                             5
           4
              Uttam
                                    67
           5
                             NLP
                 Kim
                                    55
                                            Delhi
                                                   60000
                                                            10
           clean_data.to_csv('clean_data.csv')
In [122...
           import os
In [124...
           os.getcwd()
Out[124...
           'C:\\Users\\soham'
In [126...
           clean_data.columns
Out[126...
           Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='object')
In [128...
           import matplotlib.pyplot as plt
           import seaborn as sns
In [129...
           import warnings
           warnings.filterwarnings('ignore')
In [130...
           clean_data['Salary']
           0
Out[130...
                  5000
           1
                 10000
           2
                 15000
           3
                 20000
           4
                 30000
                 60000
           Name: Salary, dtype: int32
```

vis1 = sns.distplot(clean\_data['Salary'])

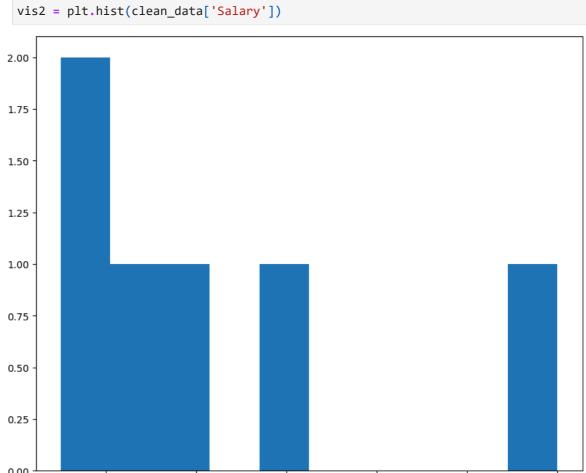
In [134...

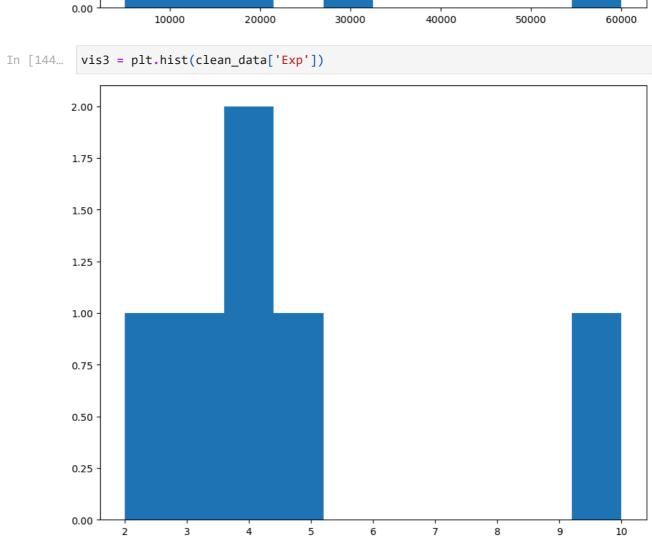




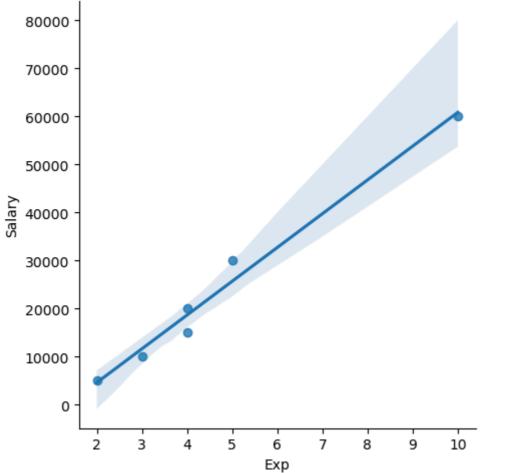


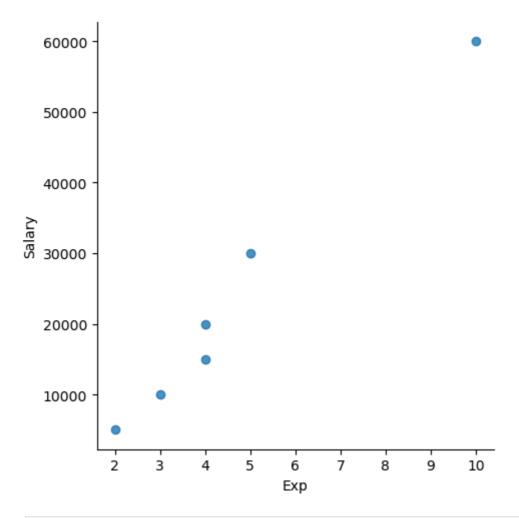
In [142...

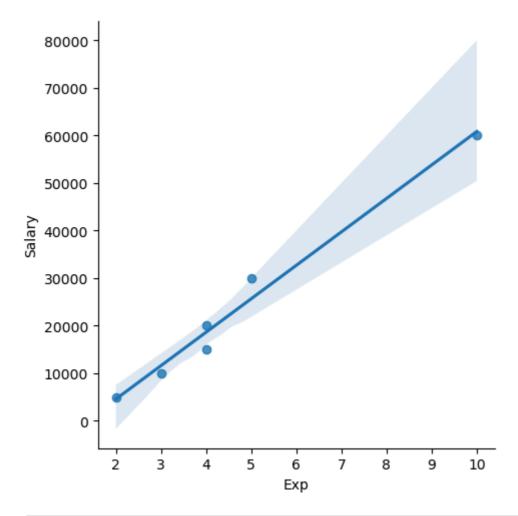












In [154... clean\_data

Out[154...

	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 [156...

clean\_data[:]

Out[156		Name	Domain	Age	Location	Salary	Ехр
	0	Mike	Datascience	34	Mumbai	5000	2
	1	Teddy	Testing	45	Bangalore	10000	3
	<b>0</b> Mike Datascience 34 Mumbai 5000 2						
	3	Name         Domain         Age         Location         Salary         Exp           Name         Domain         Age         Location         Salary         Exp           Numar         Domain         Age         Location         Salary         Exp           Name         Domain         Age         Location         Salary         Exp           Location_data[2:]         Location         Salary         Exp           Name         Domain         Age         Location         Salary         Exp           Location_data[2:]         Location         Salary         Exp           Location_data[2:]         Location         Salary         Exp           Location_data[2:]         Location         Salary         Exp           Location_data[2:]         Location         Salary         Exp           Location_data[0:1]         Location         Salary         Exp           Location_data[0:1]         Location         Salary         Exp           Location_data[0:3]         Location         Salary         Exp					
	4	Mike   Datascience   34   Mumbai   5000   2     Teddy   Testing   45   Bangalore   10000   3     Umar   Dataanalyst   50   Bangalore   15000   4     Jane   Analytics   50   Hyderbad   20000   4     Uttam   Statistics   67   Bangalore   30000   5     Kim   NLP   55   Delhi   60000   10     Delan_data[:2]					
	5	Kim	NLP	55	Delhi	60000	10
In [158	clo	ean_dat	a[:2]				
Out[158		Name	Domain	Age	Location	Salary	Ехр
	0	Mike	Datascience	34	Mumbai	5000	2
	1	Teddy	Testing	45	Bangalore	10000	3
T- [100	- 1		- [2.1				
In [160	C10	ean_dat	a[2:]				
Out[160		Name	Domain	Age	Location	Salary	Ехр
	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 [162	clo	ean_dat	a[0:1]				
Out[162		Name	Domain	Age	Location	Salary	Ехр
	0	Mike	Datascience	34	Mumbai	5000	2
in [164	clo	ean_dat	a[0:3]				
out[164		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
To [170	61.	oon dot					
In [172	CI	ean_dat	a				

ut[172		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
174	Х	= clean	_data.drop(	['Sala	ary'],axis	=1)	
176	cl	ean_dat	a				
76		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
[178	Х						
[178		Name	Domain	Age	Location	Ехр	
	0	Mike	Datascience	34	Mumbai	2	
	1	Teddy	Testing	45	Bangalore	3	
	2	Umar	Dataanalyst	50	Bangalore	4	
	3	Jane	Analytics	50	Hyderbad	4	
	4	Uttam	Statistics	67	Bangalore	5	
	5	Kim	NLP	55	Delhi	10	
[180	Χ.	columns					
t[180	Ir	ndex(['N	lame', 'Doma	in',	'Age', 'Lo	cation'	, 'Ex
[182	cl	ean_dat	a.columns				
t[182	Ir	ndex(['N	lame', 'Doma	in',	'Age', 'Lo	cation'	, 'Sa
[186	c1	ean_dat	a				
11 LT00"	CI	can_uat	u				

Out[186		Name		Domain	Age	Location	Salary	Ехр
	0	Mike	Data	science	34	Mumbai	5000	2
	1	Teddy		Testing	45	Bangalore	10000	3
	2	Umar	Data	aanalyst	50	Bangalore	15000	4
	3	Jane	А	nalytics	50	Hyderbad	20000	4
	4	Uttam	S	tatistics	67	Bangalore	30000	5
	5	Kim		NLP	55	Delhi	60000	10
In [188	У	= clean	_data	a.drop(	'Name	e','Age','l	ocation	n'],ax
In [190	у							
Out[190		Dor	main	Salary	Ехр			
	0	Datasci	ence	5000	2			
	1	Tes	sting	10000	3			
	2	Dataan	alyst	15000	4			
	3		lytics	20000	4			
	4		istics	30000	5			
	5		NLP	60000	10			
In [199	cl	ean_dat	а					
ut[199		Name		Domain	Age	Location	Salary	Ехр
	0	Mike	Data	science	34	Mumbai	5000	2
	1	Teddy		Testing	45	Bangalore	10000	3
	2	Umar	Data	aanalyst	50	Bangalore	15000	4
	3	Jane	А	nalytics	50	Hyderbad	20000	4
	4	Uttam	S	tatistics	67	Bangalore	30000	5
	5	Kim		NLP	55	Delhi	60000	10
In [201	х							

ut[201		Name	<u>.</u> D	omain	Age	Locati	on Exp		
-	0	Mike		science	34	Muml			
	1	Teddy		Testing	45	Bangalo			
	2	Umai		analyst		Bangalo			
	3	Jane		nalytics		Hyderb			
	4	Uttam		tatistics		Bangalo			
	5	Kim		NLP			elhi 10		
203	У								
203		Do	omain	Salary	Ехр				
	0	Datas	cience	5000	2	-			
	1	Т	esting	10000	3				
	2	Dataa	ınalyst	15000	4				
	3	An	alytics	20000	4				
	4	Sta	atistics	30000	5				
	5		NLP	60000	10				
[207	im	putati	.on = p	od.get_	dummie	es(clear	n_data)		
209	im	putati	.on						
209		Age	Salary	Ехр	Name <sub>.</sub>	_Jane N	Name_Kim	Name_Mike	Name_Teddy
	0	34	5000	2		False	False	True	False
	1	45	10000	3		False	False	False	True
	2	50	15000	4		False	False	False	False
	3	50	20000	4		True	False	False	False
	4	67	30000	5		False	False	False	False
	5	55	60000	10		False	True	False	False
	4					)			
	-								