```
In [1]: import pandas as pd
In [2]: | emp = pd.read_excel(r'Downloads\Rawdata.xlsx')
In [3]: emp
Out[3]:
              Name
                          Domain
                                     Age
                                           Location
                                                       Salary
                                                                 Exp
                    Datascience#$ 34 years
          0
               Mike
                                                       5^00#0
                                                                  2+
                                            Mumbai
             Teddy^
                          Testing
                                    45' yr Bangalore
                                                    10%%000
                                                                  <3
             Uma#r
                    Dataanalyst^^#
                                     NaN
                                               NaN
                                                     1$5%000
                                                               4> yrs
                       Ana^^lytics
                                     NaN
                                           Hyderbad
                                                       2000^0
               Jane
                                                                 NaN
                                                       30000- 5+ year
             Uttam*
                         Statistics
                                     67-yr
                                               NaN
               Kim
                             NLP
                                     55yr
                                               Delhi
                                                      6000^$0
                                                                 10+
In [4]: emp.shape
Out[4]: (6, 6)
In [5]: emp.columns
Out[5]: Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='objec
         t')
In [6]: len(emp.columns)
Out[6]: 6
In [7]: len(emp)
Out[7]: 6
In [8]: |emp.describe()
Out[8]:
                  Name
                              Domain
                                         Age Location
                                                        Salary Exp
                     6
                                   6
                                           4
                                                                 5
                                                     4
                                                            6
           count
                                   6
                                                                 5
          unique
                                                     4
                   Mike Datascience#$ 34 years
                                               Mumbai
                                                       5^00#0
                                                                2+
             top
            freq
                      1
                                           1
                                                     1
                                                            1
                                                                 1
```

```
In [9]: emp.columns
 Out[9]: Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='objec
          t')
In [10]: |emp[['Name', 'Domain']]
Out[10]:
               Name
                           Domain
                Mike
                     Datascience#$
              Teddy^
                           Testing
           1
                     Dataanalyst^^#
              Uma#r
           3
                Jane
                        Ana^^lytics
              Uttam*
                          Statistics
                Kim
                              NLP
           5
In [11]:
          emp[['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp']]
Out[11]:
               Name
                           Domain
                                      Age
                                            Location
                                                        Salary
                                                                  Exp
           0
                Mike
                     Datascience#$ 34 years
                                                        5^00#0
                                                                   2+
                                             Mumbai
           1
              Teddy<sup>^</sup>
                           Testing
                                     45' yr Bangalore
                                                     10%%000
                                                                   <3
                     Dataanalyst^^#
           2
              Uma#r
                                      NaN
                                                NaN
                                                      1$5%000
                                                                4> yrs
                Jane
                        Ana^^lytics
                                      NaN
                                            Hyderbad
                                                        2000^0
                                                                 NaN
              Uttam*
                          Statistics
                                      67-yr
                                                NaN
                                                        30000- 5+ year
                              NLP
                                      55yr
                                                       6000^$0
           5
                Kim
                                                Delhi
                                                                  10+
In [12]: emp.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
                                             object
           0
                Name
           1
                Domain
                           6 non-null
                                             object
           2
                Age
                           4 non-null
                                             object
           3
                Location
                           4 non-null
                                             object
           4
                           6 non-null
                Salary
                                             object
           5
                Exp
                           5 non-null
                                             object
          dtypes: object(6)
          memory usage: 416.0+ bytes
```

```
In [13]: emp['Name']
Out[13]: 0
                 Mike
          1
               Teddy^
          2
                Uma#r
          3
                  Jane
          4
               Uttam*
          5
                   Kim
          Name: Name, dtype: object
In [14]: | emp['Name'] = emp['Name'].str.replace(r'\W','')
          C:\Users\Admin\AppData\Local\Temp\ipykernel_8372\389424325.py:1: FutureWarni
          ng: The default value of regex will change from True to False in a future ve
          rsion.
            emp['Name'] = emp['Name'].str.replace(r'\W','')
In [15]: emp['Name']
Out[15]:
          0
                Mike
               Teddy
          1
          2
                Umar
          3
                Jane
          4
               Uttam
          5
                  Kim
          Name: Name, dtype: object
In [16]: | emp['Domain'] = emp['Domain'].str.replace(r'\W','')
          C:\Users\Admin\AppData\Local\Temp\ipykernel_8372\2360087947.py:1: FutureWarn
          ing: The default value of regex will change from True to False in a future v
          ersion.
            emp['Domain'] = emp['Domain'].str.replace(r'\W','')
In [17]:
          emp
Out[17]:
             Name
                       Domain
                                       Location
                                  Age
                                                   Salary
                                                            Exp
              Mike
                   Datascience
                              34 years
                                        Mumbai
                                                  5^00#0
                                                             2+
           1
             Teddy
                        Testing
                                 45' yr Bangalore
                                                10%%000
                                                             <3
              Umar
                    Dataanalyst
                                  NaN
                                           NaN
                                                 1$5%000
                                                           4> yrs
           3
              Jane
                      Analytics
                                  NaN
                                       Hyderbad
                                                  2000^0
                                                            NaN
                                                   30000- 5+ year
             Uttam
                      Statistics
                                 67-yr
           4
                                           NaN
                          NLP
           5
               Kim
                                  55yr
                                           Delhi
                                                 6000^$0
                                                             10+
```

```
In [18]: emp['Location'] = emp['Location'].str.replace(r'\W','')
          C:\Users\Admin\AppData\Local\Temp\ipykernel_8372\3886403992.py:1: FutureWarn
          ing: The default value of regex will change from True to False in a future v
          ersion.
             emp['Location'] = emp['Location'].str.replace(r'\W','')
In [19]:
          emp
Out[19]:
              Name
                        Domain
                                   Age
                                         Location
                                                     Salary
                                                               Exp
           0
                    Datascience
                                34 years
                                                    5^00#0
               Mike
                                          Mumbai
                                                                2+
           1
              Teddy
                        Testing
                                  45' yr
                                        Bangalore
                                                  10%%000
                                                                <3
           2
              Umar
                     Dataanalyst
                                             NaN
                                                   1$5%000
                                                             4> yrs
                                   NaN
           3
               Jane
                       Analytics
                                   NaN
                                         Hyderbad
                                                    2000^0
                                                              NaN
                                                     30000- 5+ year
           4
              Uttam
                       Statistics
                                  67-yr
                                             NaN
                           NLP
                                                   6000^$0
           5
                Kim
                                   55yr
                                            Delhi
                                                               10+
          emp['Age'] = emp['Age'].str.replace(r'\W','')
In [20]:
          C:\Users\Admin\AppData\Local\Temp\ipykernel 8372\3358378917.py:1: FutureWarn
          ing: The default value of regex will change from True to False in a future v
          ersion.
             emp['Age'] = emp['Age'].str.replace(r'\W','')
In [21]:
          emp
Out[21]:
              Name
                        Domain
                                   Age
                                        Location
                                                    Salary
                                                              Exp
           0
               Mike
                    Datascience
                                34years
                                          Mumbai
                                                    5^00#0
                                                               2+
                                                 10%%000
           1
              Teddy
                        Testing
                                   45yr
                                        Bangalore
                                                               <3
                                                  1$5%000
           2
              Umar
                     Dataanalyst
                                  NaN
                                            NaN
                                                            4> yrs
               Jane
                       Analytics
                                  NaN
                                        Hyderbad
                                                    2000^0
                                                              NaN
              Uttam
                       Statistics
                                   67yr
                                            NaN
                                                    30000- 5+ year
                           NLP
                                   55yr
                                                   6000^$0
                                                              10+
           5
                Kim
                                            Delhi
```

emp['Age'] = emp['Age'].str.extract('(\d+)')

In [22]:

```
In [23]: emp
Out[23]:
               Name
                         Domain
                                 Age
                                        Location
                                                    Salary
                                                              Exp
                                                    5^00#0
                                                                2+
            0
                Mike
                     Datascience
                                   34
                                         Mumbai
            1
               Teddy
                          Testing
                                   45
                                       Bangalore
                                                 10%%000
                                                                <3
            2
               Umar
                      Dataanalyst
                                 NaN
                                            NaN
                                                  1$5%000
                                                             4> yrs
                                                    2000^0
                        Analytics
                                 NaN
                                       Hyderbad
            3
                Jane
                                                              NaN
                                                    30000- 5+ year
               Uttam
                        Statistics
                                   67
                                            NaN
                            NLP
                                   55
            5
                 Kim
                                           Delhi
                                                   6000^$0
                                                               10+
           emp['Salary'] = emp['Salary'].str.replace(r'\W','')
In [24]:
           C:\Users\Admin\AppData\Local\Temp\ipykernel 8372\1304150360.py:1: FutureWarn
           ing: The default value of regex will change from True to False in a future v
           ersion.
             emp['Salary'] = emp['Salary'].str.replace(r'\W','')
In [25]:
           emp
Out[25]:
               Name
                         Domain
                                  Age
                                        Location
                                                 Salary
                                                           Exp
            0
                Mike
                     Datascience
                                   34
                                         Mumbai
                                                   5000
                                                             2+
            1
               Teddy
                          Testing
                                   45
                                       Bangalore
                                                  10000
                                                             <3
            2
                      Dataanalyst
                                            NaN
                                                  15000
               Umar
                                 NaN
                                                          4> yrs
            3
                Jane
                        Analytics
                                 NaN
                                       Hyderbad
                                                 20000
                                                           NaN
            4
               Uttam
                        Statistics
                                   67
                                            NaN
                                                 30000
                                                        5+ year
                            NLP
            5
                 Kim
                                   55
                                           Delhi
                                                 60000
                                                            10+
           emp['Exp'] = emp['Exp'].str.extract('(\d+)')
In [26]:
In [27]:
           emp
Out[27]:
                                                 Salary
                                                         Exp
               Name
                         Domain
                                        Location
                                 Age
            0
                Mike
                     Datascience
                                   34
                                         Mumbai
                                                   5000
                                                           2
                                       Bangalore
                                                  10000
                                                           3
            1
               Teddy
                          Testing
                                   45
               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 [28]: | clean_data = emp.copy()
In [29]:
          emp
Out[29]:
                                                      Exp
              Name
                        Domain Age
                                      Location Salary
               Mike Datascience
                                 34
                                                5000
                                                        2
           0
                                       Mumbai
           1
              Teddy
                        Testing
                                 45
                                     Bangalore
                                               10000
                                                        3
                    Dataanalyst NaN
                                               15000
              Umar
                                         NaN
                                                        4
                                     Hyderbad
               Jane
                       Analytics NaN
                                               20000 NaN
                       Statistics
                                               30000
              Uttam
                                 67
                                         NaN
                                                        5
                          NLP
                                               60000
           5
                Kim
                                 55
                                         Delhi
                                                       10
In [30]:
          clean data
Out[30]:
              Name
                       Domain Age
                                     Location Salary
                                                      Exp
                                                        2
           0
              Mike Datascience
                                 34
                                       Mumbai
                                                5000
           1
              Teddy
                        Testing
                                 45
                                     Bangalore
                                               10000
                                                        3
              Umar
                    Dataanalyst
                               NaN
                                         NaN
                                               15000
           3
                       Analytics
                               NaN
                                     Hyderbad
                                               20000 NaN
               Jane
              Uttam
                       Statistics
                                 67
                                         NaN
                                               30000
                                                        5
               Kim
                          NLP
                                         Delhi
                                               60000
                                                       10
                                 55
In [31]:
          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
                           6 non-null
                Name
                                             object
           1
                Domain
                           6 non-null
                                             object
           2
                           4 non-null
                                             object
                Age
           3
                Location 4 non-null
                                             object
           4
                Salary
                           6 non-null
                                             object
           5
                           5 non-null
                                             object
                Exp
          dtypes: object(6)
          memory usage: 416.0+ bytes
In [32]: import numpy as np
```

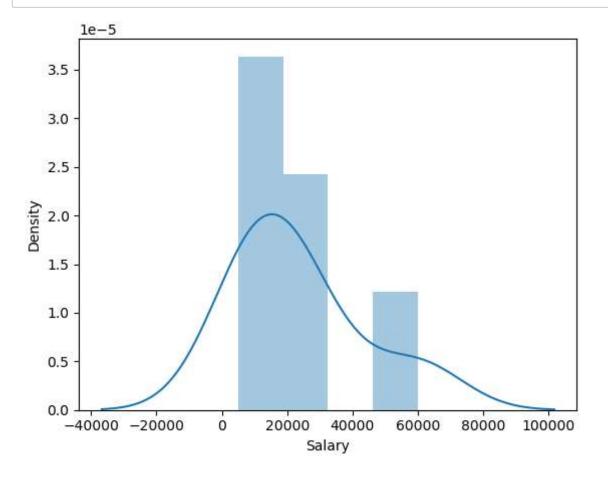
```
In [33]: clean_data
Out[33]:
                                         Name
                                                                     Domain Age
                                                                                                             Location Salary
                                           Mike Datascience
                                                                                                                                           5000
                                                                                                                                                                  2
                                                                                                 34
                                                                                                                Mumbai
                                 1
                                         Teddy
                                                                       Testing
                                                                                                45
                                                                                                           Bangalore
                                                                                                                                        10000
                                                                                                                                                                  3
                                                                                                                                        15000
                                         Umar
                                                            Dataanalyst
                                                                                           NaN
                                                                                                                        NaN
                                                                                                                                                                  4
                                                                                                            Hyderbad
                                           Jane
                                                                   Analytics
                                                                                            NaN
                                                                                                                                        20000 NaN
                                         Uttam
                                                                   Statistics
                                                                                                 67
                                                                                                                        NaN
                                                                                                                                        30000
                                                                                                                                                                  5
                                                                             NLP
                                                                                                                       Delhi
                                                                                                                                        60000
                                 5
                                              Kim
                                                                                                 55
                                                                                                                                                                10
                              clean_data["Age"]=clean_data["Age"].fillna(np.mean(pd.to_numeric
In [35]:
                                                                                                                                                                                            (clean_data["Age"]) ))
In [36]: | clean data['Age']
Out[36]: 0
                                                        34
                              1
                                                        45
                              2
                                              50.25
                                              50.25
                              3
                                                        67
                              4
                              5
                                                        55
                              Name: Age, dtype: object
In [37]: | clean_data['Age'] = clean_data['Age'].astype(int)
In [38]: | 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
                                                                                                                                   int32
                                              Age
                                 3
                                              Location 4 non-null
                                                                                                                                  object
                                 4
                                                                               6 non-null
                                              Salary
                                                                                                                                  object
                                                                               5 non-null
                                                                                                                                  object
                                              Exp
                              dtypes: int32(1), object(5)
                              memory usage: 392.0+ bytes
In [39]: clean_data['Exp']=clean_data['Exp'].fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.mean(pd.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp']).fillna(np.to_numeric(clean_data['Exp
```

```
In [40]: clean_data['Exp']
                2
Out[40]: 0
         1
                3
         2
                4
         3
              4.8
         4
                5
         5
               10
         Name: Exp, dtype: object
In [41]: | clean data['Exp'] = clean data['Exp'].astype(int)
In [42]: | clean_data['Location'] = clean_data['Location'].fillna(np.mode(pd.to_numeric(
         AttributeError
                                                    Traceback (most recent call last)
         Cell In[42], line 1
         ----> 1 clean data['Location'] = clean data['Location'].fillna(np.mode(pd.to
         _numeric(clean_data['Location'])))
         File ~\anaconda3\lib\site-packages\numpy\__init__.py:311, in __getattr__(att
         r)
                      from .testing import Tester
             308
                      return Tester
             309
         --> 311 raise AttributeError("module {!r} has no attribute "
                                       "{!r}".format(__name__, attr))
         AttributeError: module 'numpy' has no attribute 'mode'
In [44]: clean_data['Location'] = clean_data['Location'].fillna(clean_data['Location'].
In [45]: clean_data['Location']
Out[45]: 0
                 Mumbai
              Bangalore
         1
         2
              Bangalore
               Hyderbad
         3
         4
              Bangalore
                  Delhi
         Name: Location, dtype: object
```

```
In [46]: | clean_data
Out[46]:
                                     Location Salary Exp
              Name
                       Domain Age
                                               5000
              Mike Datascience
                                34
                                      Mumbai
                                                       2
              Teddy
                        Testing
                                45
                                    Bangalore
                                              10000
                                                       3
                    Dataanalyst
                                    Bangalore
                                              15000
              Umar
                                50
                                                       4
                                    Hyderbad
                                              20000
              Jane
                      Analytics
                                50
                                                       4
              Uttam
                      Statistics
                                67
                                    Bangalore
                                              30000
                                                       5
                          NLP
                                55
                                        Delhi
                                              60000
                                                      10
               Kim
          clean_data.to_csv('clean_data.csv')
In [47]:
In [48]:
          import os
          os.getcwd()
Out[48]: 'C:\\Users\\Admin'
          import matplotlib.pyplot as plt
In [49]:
          import seaborn as sns
In [50]: |%matplotlib inline
          import warnings
          warnings.filterwarnings('ignore')
In [51]: | clean_data['Salary']
Out[51]: 0
                 5000
          1
               10000
          2
               15000
          3
                20000
          4
                30000
          5
                60000
```

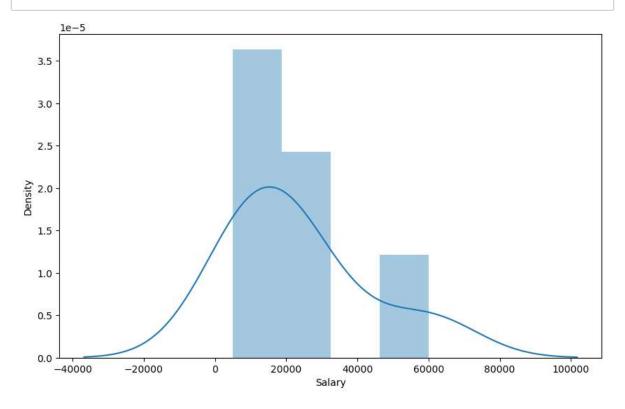
Name: Salary, dtype: object

```
In [52]: vis1 = sns.distplot(clean_data['Salary'])
```

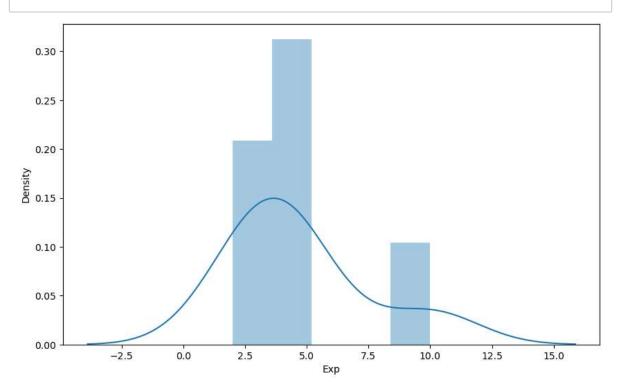


In [55]: plt.rcParams['figure.figsize']=10,6

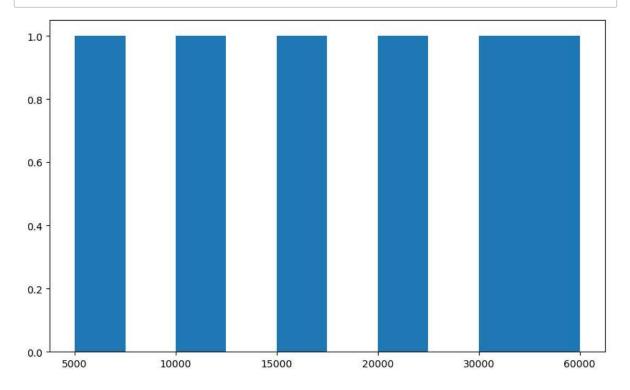
In [56]: vis1 = sns.distplot(clean_data['Salary'])

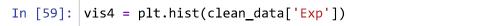


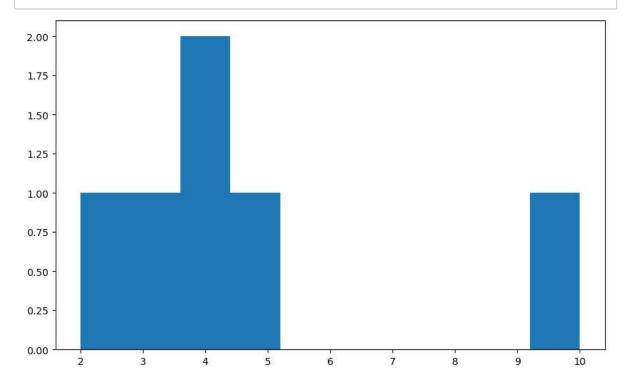
In [57]: vis2 = sns.distplot(clean_data['Exp'])



In [58]: vis3 = plt.hist(clean_data['Salary'])

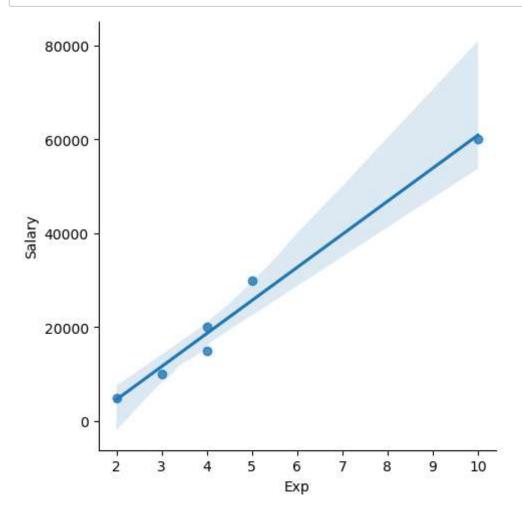




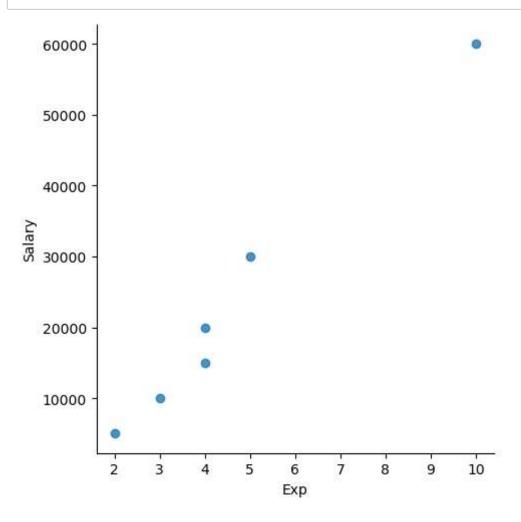


```
In [60]: clean_data.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 6 entries, 0 to 5
         Data columns (total 6 columns):
                       Non-Null Count Dtype
              Column
              -----
                        -----
              Name
                        6 non-null
                                       object
          0
          1
              Domain
                       6 non-null
                                       object
          2
              Age
                       6 non-null
                                        int32
          3
              Location 6 non-null
                                       object
          4
                        6 non-null
                                       object
              Salary
          5
              Exp
                        6 non-null
                                        int32
         dtypes: int32(2), object(4)
         memory usage: 368.0+ bytes
In [61]: | clean_data.Name = clean_data.Name.astype('category')
In [62]: | clean_data.Domain = clean_data.Domain.astype('category')
In [63]: | clean_data.Location = clean_data.Location.astype('category')
In [64]: | clean_data['Salary'] = clean_data['Salary'].astype(int)
In [65]: | clean_data.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 6 entries, 0 to 5
         Data columns (total 6 columns):
                       Non-Null Count Dtype
              Column
              -----
                        -----
                                       ____
          0
                        6 non-null
              Name
                                       category
              Domain
                        6 non-null
          1
                                       category
                       6 non-null
          2
              Age
                                       int32
          3
              Location 6 non-null
                                       category
          4
                       6 non-null
              Salary
                                       int32
          5
              Exp
                        6 non-null
                                       int32
         dtypes: category(3), int32(3)
         memory usage: 862.0 bytes
```

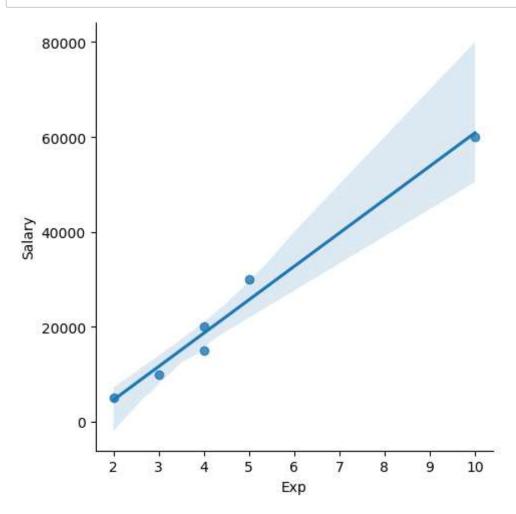
In [66]: vis6 = sns.lmplot(data = clean_data, x = "Exp", y='Salary')



In [67]: vis6 = sns.lmplot(data = clean_data, x = "Exp", y='Salary', fit_reg = False)



In [68]: vis6 = sns.lmplot(data = clean_data, x = "Exp", y='Salary', fit_reg = True)



In [69]: clean_data[0:5:2]

Out[69]:

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

In [70]: clean_data

Out[70]:

	Name	Domain	Age	Location	Salary	Ехр
() 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
	1 Uttam	Statistics	67	Bangalore	30000	5
;	5 Kim	NLP	55	De l hi	60000	10

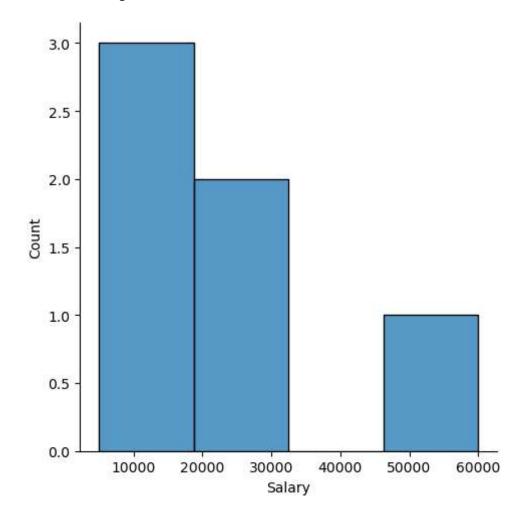
```
In [71]: emp
Out[71]:
               Name
                         Domain Age
                                        Location
                                                 Salary
                                                         Exp
                                                           2
                     Datascience
                                                   5000
            0
                Mike
                                   34
                                         Mumbai
            1
               Teddy
                          Testing
                                   45
                                       Bangalore
                                                 10000
                                                           3
                                                 15000
            2
               Umar
                      Dataanalyst NaN
                                            NaN
                                                           4
                                       Hyderbad
                                                 20000 NaN
            3
                Jane
                        Analytics
                                 NaN
               Uttam
                        Statistics
                                   67
                                            NaN
                                                 30000
                                                           5
                            NLP
                                   55
                                           Delhi
                                                 60000
                                                          10
            5
                 Kim
In [72]: x_iv = clean_data.drop(['Salary'], axis=1)
In [73]: x_iv
Out[73]:
                                       Location Exp
               Name
                         Domain Age
            0
                Mike
                     Datascience
                                   34
                                         Mumbai
                                                   2
            1
               Teddy
                                       Bangalore
                                                   3
                          Testing
                                   45
                      Dataanalyst
                                       Bangalore
               Umar
                                   50
                                                   4
            3
                Jane
                        Analytics
                                   50
                                       Hyderbad
                                                   4
               Uttam
                        Statistics
                                       Bangalore
                                                   5
            5
                 Kim
                            NLP
                                   55
                                           Delhi
                                                  10
In [74]: y_dv = clean_data.drop(['Name','Domain','Age','Location','Exp'],axis = 1)
In [75]: y_dv
Out[75]:
               Salary
            0
                5000
            1
               10000
               15000
            2
            3
               20000
               30000
```

60000

In [76]: | clean_data Out[76]: Salary Exp Name Domain Age Location Mike Datascience Mumbai Teddy Testing Bangalore Umar Dataanalyst Bangalore Analytics Hyderbad Jane Uttam Statistics Bangalore Kim NLP Delhi In [77]: imputation = pd.get_dummies(clean_data) In [78]: imputation Out[78]: Name_Jane Name_Kim Name_Mike Name_Teddy Name_Umar Name_Utt Age Salary Exp

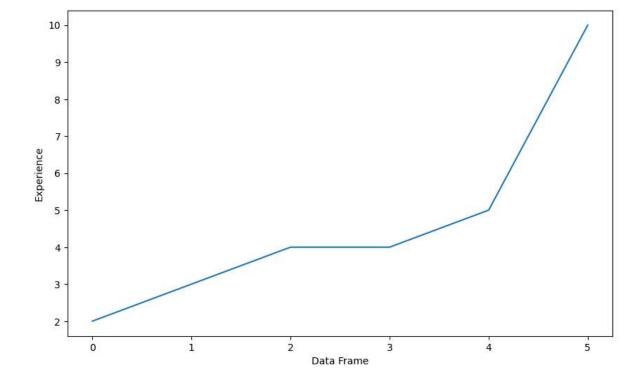
```
In [80]: sns.displot(clean_data['Salary'])
```

Out[80]: <seaborn.axisgrid.FacetGrid at 0x297bebbe200>



```
In [81]: plt.plot(clean_data['Exp'])
   plt.xlabel("Data Frame")
   plt.ylabel("Experience")
```

Out[81]: Text(0, 0.5, 'Experience')



```
In [ ]:
```