39hlhgox8

February 20, 2025

```
import pandas as pd
[177]:
[179]: pd.__version__
[179]: '2.2.2'
       emp = pd.read_excel(r'C:\Users\mikim\Downloads\Rawdata.xlsx')
[183]:
       emp
[183]:
             Name
                            Domain
                                          Age
                                                 Location
                                                             Salary
                                                                          Exp
             Mike
                                                             5^00#0
                                                                           2+
                    Datascience#$
                                     34 years
                                                   Mumbai
       1
          Teddv^
                                       45' yr
                                                                           <3
                           Testing
                                                Bangalore
                                                            10%%000
                   Dataanalyst^^#
       2
            Uma#r
                                          NaN
                                                      NaN
                                                            1$5%000
                                                                       4> yrs
                                                 Hyderbad
                       Ana^^lytics
                                                             2000^0
       3
             Jane
                                          \mathtt{NaN}
                                                                          NaN
       4
          Uttam*
                        Statistics
                                        67-yr
                                                      NaN
                                                             30000-
                                                                     5+ year
       5
              Kim
                               NI.P
                                         55yr
                                                    Delhi
                                                            6000^$0
                                                                          10+
[185]:
       id(emp)
[185]: 2352588227120
       emp.columns
[187]:
[187]: Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='object')
[189]:
       emp.shape
[189]: (6, 6)
       emp.head()
[191]:
                                                             Salary
             Name
                            Domain
                                          Age
                                                 Location
                                                                          Exp
             Mike
                    Datascience#$
                                     34 years
                                                   Mumbai
                                                             5^00#0
                                                                           2+
                                       45' yr
         Teddy^
                                               Bangalore
                                                            10%%000
       1
                           Testing
                                                                           <3
                   Dataanalyst^^#
       2
            Uma#r
                                          NaN
                                                      NaN
                                                            1$5%000
                                                                       4> yrs
       3
                       Ana^^lytics
                                                 Hyderbad
                                                             2000^0
             Jane
                                          \mathtt{NaN}
                                                                          NaN
       4 Uttam*
                        Statistics
                                        67-yr
                                                      NaN
                                                             30000-
                                                                     5+ year
```

```
emp.tail()
[193]:
[193]:
            Name
                            Domain
                                              Location
                                                          Salary
                                                                       Exp
                                       Age
          Teddy^
                                    45' yr
                                                         10%%000
                                                                        <3
       1
                          Testing
                                             Bangalore
       2
           Uma#r
                   Dataanalyst^^#
                                       NaN
                                                   NaN
                                                         1$5%000
                                                                   4> yrs
       3
                      Ana^^lytics
             Jane
                                       NaN
                                              Hyderbad
                                                          2000^0
                                                                       NaN
       4
          Uttam*
                       Statistics
                                     67-yr
                                                   NaN
                                                          30000-
                                                                  5+ year
       5
             Kim
                               NLP
                                      55yr
                                                 Delhi
                                                         6000^$0
                                                                       10+
[195]:
      emp.info()
       <class 'pandas.core.frame.DataFrame'>
      RangeIndex: 6 entries, 0 to 5
      Data columns (total 6 columns):
            Column
                      Non-Null Count
                                        Dtype
                                        ____
        0
            Name
                      6 non-null
                                        object
        1
            Domain
                      6 non-null
                                        object
        2
            Age
                      4 non-null
                                        object
        3
            Location 4 non-null
                                        object
        4
            Salary
                      6 non-null
                                        object
        5
                       5 non-null
            Exp
                                        object
      dtypes: object(6)
      memory usage: 420.0+ bytes
[197]:
       emp
[197]:
            Name
                            Domain
                                                Location
                                                            Salary
                                          Age
                                                                         Exp
            Mike
                    Datascience#$
                                    34 years
                                                  Mumbai
                                                            5^00#0
                                                                          2+
       0
                                      45' yr
       1
          Teddy^
                          Testing
                                               Bangalore
                                                           10%%000
                                                                          <3
       2
           Uma#r
                   Dataanalyst^^#
                                         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
                                                           6000^$0
             Kim
                               NLP
                                        55yr
                                                   Delhi
                                                                         10+
[199]:
       emp.isnull()
[199]:
                  Domain
                                  Location
           Name
                             Age
                                             Salary
                                                       Exp
                   False
         False
                          False
                                     False
                                              False
                                                     False
         False
                   False
                          False
                                     False
                                              False
                                                    False
       1
       2 False
                   False
                            True
                                      True
                                              False
                                                     False
                   False
                            True
                                              False
                                                      True
       3 False
                                     False
       4 False
                   False
                          False
                                      True
                                              False
                                                    False
          False
                   False
                          False
                                     False
                                              False
                                                     False
[201]: emp.isna() #isnull and isna both are same
```

```
[201]:
           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
[203]: emp.isnull().sum()
[203]: Name
                   0
       Domain
                   0
                   2
       Age
       Location
                   2
       Salary
       Exp
       dtype: int64
      Data Cleaning
[206]: emp['Name']
[206]: 0
              Mike
       1
            Teddy^
       2
             Uma#r
       3
              Jane
       4
            Uttam*
       5
               Kim
       Name: Name, dtype: object
[208]: emp['Name'] = emp['Name'].str.replace(r'\W','', regex=True) #non word character
[210]: emp['Name']
[210]: 0
             Mike
       1
            Teddy
       2
             Umar
       3
             Jane
       4
            Uttam
              Kim
       5
       Name: Name, dtype: object
[212]:
       emp
[212]:
           Name
                         Domain
                                       Age
                                             Location
                                                         Salary
                                                                     Exp
                  Datascience#$
                                 34 years
                                               Mumbai
                                                         5^00#0
                                                                      2+
       0
           Mike
          Teddy
                         Testing
                                    45' yr
                                            Bangalore
                                                        10%%000
                                                                      <3
```

```
4> yrs
       2
            Umar
                  Dataanalyst^^#
                                         {\tt NaN}
                                                     {\tt NaN}
                                                           1$5%000
                     Ana^^lytics
       3
            Jane
                                         NaN
                                                Hyderbad
                                                            2000^0
                                                                         {\tt NaN}
                                                            30000-
       4
          Uttam
                       Statistics
                                       67-yr
                                                     NaN
                                                                     5+ year
             Kim
                              NLP
                                        55yr
                                                   Delhi
                                                           6000^$0
                                                                         10+
       emp['Domain']
[214]:
[214]: 0
              Datascience#$
                    Testing
       1
             Dataanalyst^^#
       2
       3
                Ana^^lytics
       4
                 Statistics
                         NLP
       Name: Domain, dtype: object
[216]: emp['Domain'] = emp['Domain'].str.replace(r'\W','', regex=True)
[218]: emp['Domain']
[218]: 0
             Datascience
       1
                 Testing
       2
             Dataanalyst
       3
               Analytics
       4
              Statistics
       5
                     NI.P
       Name: Domain, dtype: object
[220]:
       emp
[220]:
            Name
                        Domain
                                      Age
                                            Location
                                                         Salary
                                                                      Exp
                  Datascience
                                                         5^00#0
           Mike
                                34 years
                                               Mumbai
                                                                       2+
       1
          Teddy
                       Testing
                                   45' yr
                                           Bangalore
                                                        10%%000
                                                                       <3
       2
            Umar
                  Dataanalyst
                                      NaN
                                                  NaN
                                                       1$5%000
                                                                   4> yrs
       3
            Jane
                    Analytics
                                      NaN
                                             Hyderbad
                                                         2000^0
                                                                      NaN
       4
          Uttam
                   Statistics
                                                         30000-
                                    67-yr
                                                  NaN
                                                                  5+ year
       5
             Kim
                           NLP
                                                       6000^$0
                                     55yr
                                                Delhi
                                                                      10+
       emp['Age'] = emp['Age'].str.replace(r'\W','', regex=True)
[222]:
[224]:
       emp['Age']
[224]: 0
             34years
       1
                45yr
       2
                 NaN
       3
                 NaN
       4
                67yr
       5
                55yr
```

```
Name: Age, dtype: object
[226]: emp['Age'] = emp['Age'].str.extract('(\d+)')
      <>:1: SyntaxWarning: invalid escape sequence '\d'
      <>:1: SyntaxWarning: invalid escape sequence '\d'
      C:\Users\mikim\AppData\Local\Temp\ipykernel_3648\1884116463.py:1: SyntaxWarning:
      invalid escape sequence '\d'
        emp['Age'] = emp['Age'].str.extract('(\d+)')
[228]: emp['Age']
[228]: 0
             34
             45
       1
       2
            NaN
       3
            NaN
       4
             67
       5
             55
       Name: Age, dtype: object
[230]: emp
[230]:
                       Domain
                                     Location
           Name
                               Age
                                                 Salary
                                                              Exp
       0
           Mike
                 Datascience
                                34
                                       Mumbai
                                                 5^00#0
                                                               2+
         Teddy
                                                10%%000
       1
                      Testing
                                45
                                    Bangalore
                                                               <3
       2
           Umar
                 Dataanalyst
                                                1$5%000
                               NaN
                                           NaN
                                                           4> yrs
       3
           Jane
                   Analytics
                               NaN
                                     Hyderbad
                                                 2000^0
                                                              NaN
       4
         Uttam
                  Statistics
                                                 30000-
                                67
                                           NaN
                                                          5+ year
                                                6000^$0
            Kim
                          NLP
                                55
                                        Delhi
                                                              10+
[232]:
       emp['Location'] = emp['Location'].str.replace(r'\W','', regex=True)
       emp['Location']
[234]:
[234]: 0
               Mumbai
       1
            Bangalore
       2
                  NaN
       3
             Hyderbad
       4
                  NaN
       5
                Delhi
       Name: Location, dtype: object
       emp['Salary'] = emp['Salary'].str.replace(r'\W','', regex=True)
[236]:
[238]:
       emp['Salary']
```

```
[238]: 0
             5000
             10000
       1
       2
            15000
       3
            20000
       4
            30000
       5
            60000
       Name: Salary, dtype: object
[240]:
       emp
[240]:
           Name
                       Domain
                                Age
                                      Location Salary
                                                             Exp
       0
           Mike
                 Datascience
                                 34
                                        Mumbai
                                                  5000
                                                              2+
                                                 10000
       1
          Teddy
                      Testing
                                 45
                                     Bangalore
                                                              <3
       2
           Umar
                  Dataanalyst
                                                 15000
                                NaN
                                            NaN
                                                          4> yrs
       3
           Jane
                    Analytics
                                NaN
                                                 20000
                                                             NaN
                                      Hyderbad
          Uttam
                   Statistics
                                 67
                                            NaN
                                                 30000
                                                         5+ year
       5
            Kim
                          NLP
                                 55
                                          Delhi
                                                 60000
                                                             10+
[244]: emp['Exp'] = emp['Exp'].str.extract('(\d+)')
      <>:1: SyntaxWarning: invalid escape sequence '\d'
      <>:1: SyntaxWarning: invalid escape sequence '\d'
      C:\Users\mikim\AppData\Local\Temp\ipykernel_3648\3836251810.py:1: SyntaxWarning:
      invalid escape sequence '\d'
         emp['Exp'] = emp['Exp'].str.extract('(\d+)')
[246]: emp['Exp']
[246]: 0
               2
       1
               3
       2
               4
       3
            NaN
       4
               5
       5
              10
       Name: Exp, dtype: object
[248]:
       emp
[248]:
                                Age
           Name
                       Domain
                                      Location Salary
                                                         Exp
                                                  5000
           Mike
                  Datascience
                                 34
                                        Mumbai
                                                           3
       1
          Teddy
                      Testing
                                 45
                                     Bangalore
                                                 10000
       2
           Umar
                  Dataanalyst
                                NaN
                                            NaN
                                                 15000
       3
           Jane
                    Analytics
                                NaN
                                      Hyderbad
                                                 20000
                                                         NaN
       4
          Uttam
                   Statistics
                                 67
                                            NaN
                                                 30000
                                                           5
       5
            Kim
                          NLP
                                 55
                                                 60000
                                                          10
                                          Delhi
[250]: clean_data = emp.copy()
```

```
[252]: clean_data
[252]:
           Name
                       Domain
                              Age
                                      Location Salary
                                                        Exp
           Mike
                 Datascience
                                 34
                                        Mumbai
                                                  5000
                                     Bangalore
                                                10000
                                                          3
       1
          Teddy
                      Testing
                                 45
       2
           Umar
                  Dataanalyst
                                                15000
                                                          4
                               NaN
                                           NaN
       3
           Jane
                    Analytics
                               NaN
                                      Hyderbad
                                                20000
                                                        NaN
                   Statistics
       4
          Uttam
                                 67
                                           NaN
                                                30000
                                                          5
       5
                          NLP
                                 55
                                         Delhi
                                                60000
                                                         10
            Kim
[254]:
      clean_data['Age']
[254]: 0
             34
             45
       1
       2
            NaN
       3
            NaN
       4
             67
       5
             55
       Name: Age, dtype: object
[256]: import numpy as np
[258]: clean_data['Age'] = clean_data['Age'].fillna(np.mean(pd.
        →to_numeric(clean_data['Age'])))
[260]: clean_data['Age']
[260]: 0
               34
               45
       1
       2
            50.25
            50.25
       3
       4
               67
               55
       5
       Name: Age, dtype: object
[262]: clean_data['Location'] = clean_data['Location'].fillna(clean_data['Location'].
         →mode()[0])
[264]: clean_data['Location']
[264]: 0
               Mumbai
       1
            Bangalore
       2
            Bangalore
       3
             Hyderbad
       4
            Bangalore
                Delhi
       Name: Location, dtype: object
```

```
[266]: clean_data['Exp'] = clean_data['Exp'].fillna(np.mean(pd.
        →to_numeric(clean_data['Exp'])))
[268]: clean_data['Exp']
[268]: 0
              2
       1
              3
       2
              4
       3
            4.8
       4
              5
       5
             10
       Name: Exp, dtype: object
[270]: clean_data
[270]:
                                                        Exp
           Name
                      Domain
                                 Age
                                       Location Salary
           Mike
                 Datascience
                                  34
                                         Mumbai
                                                  5000
                                                           2
       0
                     Testing
                                  45
                                      Bangalore 10000
                                                           3
       1 Teddy
       2
           Umar
                 Dataanalyst 50.25
                                      Bangalore
                                                 15000
                                                           4
       3
           Jane
                   Analytics 50.25
                                       Hyderbad
                                                 20000
                                                        4.8
                  Statistics
                                      Bangalore
        Uttam
                                  67
                                                 30000
                                                           5
                                          Delhi
       5
            Kim
                         NLP
                                  55
                                                 60000
                                                         10
[272]: 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
       0
           Name
                                      object
                      6 non-null
       1
           Domain
                                      object
       2
                      6 non-null
           Age
                                      object
           Location 6 non-null
                                      object
           Salary
                      6 non-null
                                      object
       5
           Exp
                      6 non-null
                                      object
      dtypes: object(6)
      memory usage: 420.0+ bytes
[274]: clean_data['Age'] = clean_data['Age'].astype(int)
[276]:
      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
       1
           Domain
                                      object
       2
                     6 non-null
                                      int32
           Age
       3
           Location 6 non-null
                                      object
       4
           Salary
                     6 non-null
                                      object
                     6 non-null
           Exp
                                      object
      dtypes: int32(1), object(5)
      memory usage: 396.0+ bytes
      clean_data['Salary'] = clean_data['Salary'].astype(int)
[278]:
[284]:
       clean_data['Exp'] = clean_data['Exp'].astype(int)
[286]:
       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 6 non-null
                                      object
       4
           Salary
                     6 non-null
                                      int32
                     6 non-null
           Exp
                                      int32
      dtypes: int32(3), object(3)
      memory usage: 348.0+ bytes
[288]: | clean_data['Name'] = clean_data['Name'].astype('category')
       clean_data['Domain'] = clean_data['Domain'].astype('category')
       clean_data['Location'] = clean_data['Location'].astype('category')
[290]: 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
                                      category
           Name
       1
           Domain
                     6 non-null
                                      category
       2
           Age
                     6 non-null
                                      int32
       3
           Location 6 non-null
                                      category
       4
           Salary
                     6 non-null
                                      int32
                     6 non-null
           Exp
                                      int32
      dtypes: category(3), int32(3)
      memory usage: 866.0 bytes
```

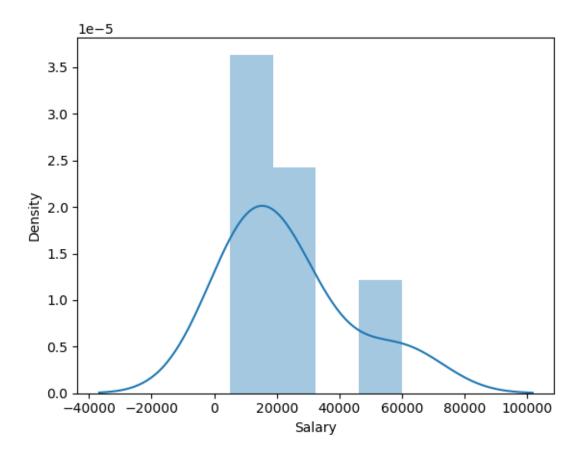
0

Name

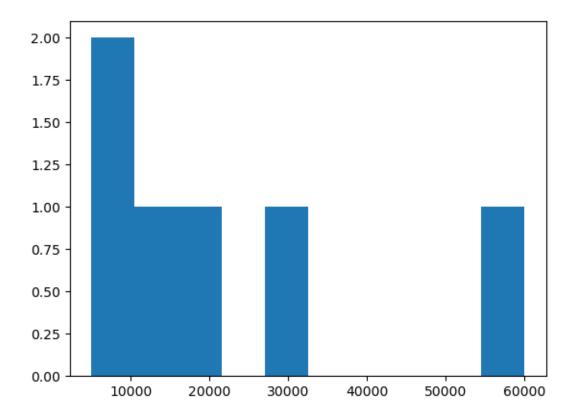
6 non-null

object

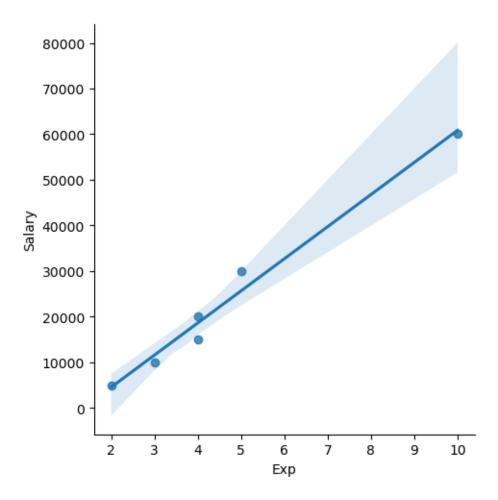
```
[292]: clean_data
[292]:
           Name
                                      Location Salary
                       Domain Age
                                                        Exp
           Mike
                 Datascience
                                34
                                        Mumbai
                                                  5000
                                                           2
                                                           3
       1
          Teddy
                      Testing
                                45
                                    Bangalore
                                                 10000
           Umar
                 Dataanalyst
                                    Bangalore
                                                 15000
                                                           4
       2
                                50
       3
           Jane
                    Analytics
                                50
                                      Hyderbad
                                                 20000
       4
          Uttam
                   Statistics
                                67
                                     Bangalore
                                                 30000
                                                           5
       5
                          NLP
                                55
                                         Delhi
                                                 60000
            Kim
                                                          10
       clean_data.to_csv('clean_data.csv')
[294]:
[298]: import os
       os.getcwd() #from the os give the saved current working directly
[298]: 'C:\\Users\\mikim'
[300]: clean_data
[300]:
           Name
                       Domain
                               Age
                                      Location
                                                Salary
                                                         Exp
           Mike
                 Datascience
                                        Mumbai
                                                  5000
                                                           2
                                34
          Teddy
                      Testing
                                     Bangalore
                                                 10000
                                                           3
       1
                                45
                 Dataanalyst
       2
           Umar
                                50
                                    Bangalore
                                                 15000
                                                           4
       3
           Jane
                   Analytics
                                50
                                     Hyderbad
                                                 20000
                                                           4
       4
         Uttam
                  Statistics
                                67
                                    Bangalore
                                                 30000
                                                           5
                          NLP
                                         Delhi
                                                 60000
       5
            Kim
                                55
                                                          10
[302]: import matplotlib.pyplot as plt
       import seaborn as sns
[303]: import warnings
       warnings.filterwarnings('ignore')
[304]: clean_data['Salary']
[304]: 0
             5000
       1
            10000
       2
            15000
       3
            20000
       4
            30000
            60000
       5
       Name: Salary, dtype: int32
[308]: vis1 = sns.distplot(clean_data['Salary'])
       plt.show()
```

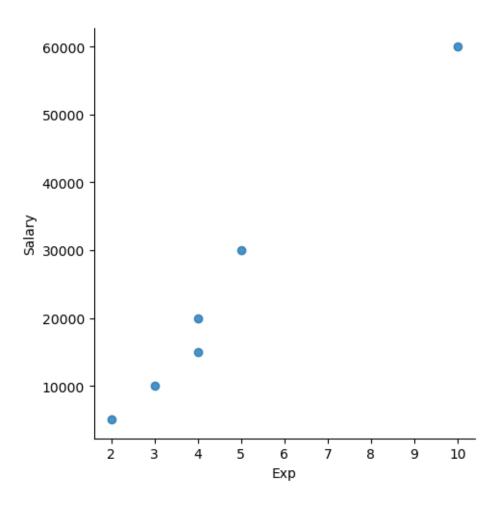


```
[310]: vis2 = plt.hist(clean_data['Salary'])
plt.show()
```



```
[312]: vis4= sns.lmplot(data=clean_data, x='Exp', y='Salary')
```





```
[316]:
       clean_data
                                                          Exp
[316]:
           Name
                       Domain
                                Age
                                      Location
                                                 Salary
           Mike
                                 34
                                         Mumbai
                                                   5000
                                                            2
       0
                  Datascience
                                     Bangalore
                                                            3
       1
          Teddy
                      Testing
                                 45
                                                  10000
                                     Bangalore
       2
           Umar
                  Dataanalyst
                                                  15000
                                 50
       3
           Jane
                    Analytics
                                 50
                                      Hyderbad
                                                  20000
                                                            4
       4
          Uttam
                   Statistics
                                 67
                                     Bangalore
                                                  30000
                                                            5
       5
                          NLP
                                 55
                                          Delhi
                                                  60000
            {\tt Kim}
                                                           10
[318]: X_iv = clean_data[['Name','Domain','Age','Location','Exp']]
[320]: X_iv #independent variable
[320]:
           Name
                       Domain
                                Age
                                      Location Exp
           Mike
                 Datascience
                                 34
                                         Mumbai
                                                   2
       1
          Teddy
                      Testing
                                 45
                                     Bangalore
                                                   3
                 Dataanalyst
                                     Bangalore
                                                   4
           Umar
                                 50
```

```
3
            Jane
                    Analytics
                                  50
                                       Hyderbad
                                                     4
                                                     5
       4
         Uttam
                   Statistics
                                  67
                                      Bangalore
       5
             Kim
                           NLP
                                  55
                                           Delhi
                                                    10
[324]: y_dv = clean_data['Salary']
      y_dv #dependent variable
[326]:
[326]: 0
              5000
       1
             10000
       2
             15000
       3
             20000
       4
             30000
       5
             60000
       Name: Salary, dtype: int32
[328]: clean_data
[328]:
            Name
                        Domain
                                Age
                                       Location
                                                  Salary
                                                           Exp
           Mike
                  Datascience
                                          Mumbai
                                                     5000
                                                              2
                                  34
       1
          Teddy
                       Testing
                                  45
                                      Bangalore
                                                    10000
                                                              3
       2
            Umar
                  Dataanalyst
                                      Bangalore
                                                    15000
                                                              4
                                  50
       3
            Jane
                    Analytics
                                  50
                                       Hyderbad
                                                    20000
                                                             4
                   Statistics
       4
          Uttam
                                  67
                                      Bangalore
                                                    30000
                                                             5
       5
             Kim
                           NLP
                                           Delhi
                                                    60000
                                  55
                                                             10
       imputation = pd.get_dummies(clean_data,dtype=int)
[330]:
[332]:
       imputation
[332]:
           Age
                Salary
                         Exp
                              Name_Jane
                                           Name_Kim
                                                      Name_Mike
                                                                  Name_Teddy
                                                                               Name_Umar
       0
            34
                  5000
                           2
                                       0
                                                  0
                                                                            0
                                                                                        0
                                                               1
       1
            45
                 10000
                           3
                                       0
                                                  0
                                                               0
                                                                            1
                                                                                        0
       2
            50
                 15000
                           4
                                       0
                                                  0
                                                               0
                                                                            0
                                                                                        1
       3
            50
                           4
                                                  0
                                                                            0
                                                                                        0
                 20000
                                       1
                                                               0
                           5
       4
            67
                 30000
                                       0
                                                  0
                                                               0
                                                                            0
                                                                                        0
       5
                 60000
            55
                          10
                                       0
                                                                                        0
                                            Domain_Dataanalyst
                                                                  Domain_Datascience
          Name_Uttam
                       Domain_Analytics
       0
                    0
                                        0
                                                               0
                                                                                     1
                    0
                                        0
                                                               0
                                                                                     0
       1
       2
                     0
                                        0
                                                               1
                                                                                     0
       3
                     0
                                        1
                                                               0
                                                                                     0
                                                                                     0
       4
                     1
                                        0
                                                               0
       5
                     0
                                        0
                                                               0
                                                                                     0
                        Domain_Statistics Domain_Testing Location_Bangalore
          Domain_NLP
```

```
0
                                        0
       1
                    0
                                        0
                                                         1
                                                                              1
                    0
       2
                                        0
                                                         0
                                                                              1
       3
                    0
                                        0
                                                         0
                                                                              0
       4
                    0
                                        1
                                                         0
                                                                              1
       5
                    1
                                        0
                                                                              0
                                                         0
                          Location_Hyderbad Location_Mumbai
          Location_Delhi
       0
       1
                        0
                                            0
                                                              0
       2
                        0
                                            0
                                                              0
       3
                        0
                                            1
                                                              0
       4
                        0
                                            0
                                                              0
       5
                                            0
                                                              0
                        1
[334]: clean_data
[334]:
           Name
                       Domain
                               Age
                                      Location
                                                Salary Exp
                                        Mumbai
                                                   5000
                                                           2
           Mike
                 Datascience
                                34
       1 Teddy
                      Testing
                                     Bangalore
                                                  10000
                                                           3
                                45
                 Dataanalyst
           Umar
       2
                                50
                                     Bangalore
                                                  15000
       3
           Jane
                   Analytics
                                50
                                      Hyderbad
                                                  20000
                                                           4
       4
         Uttam
                   Statistics
                                67
                                     Bangalore
                                                  30000
                                                           5
       5
            Kim
                          NLP
                                55
                                         Delhi
                                                  60000
                                                          10
[336]: len(clean data)
[336]: 6
[338]:
       imputation.columns
[338]: Index(['Age', 'Salary', 'Exp', 'Name_Jane', 'Name_Kim', 'Name_Mike',
               'Name_Teddy', 'Name_Umar', 'Name_Uttam', 'Domain_Analytics',
               'Domain_Dataanalyst', 'Domain_Datascience', 'Domain_NLP',
               'Domain_Statistics', 'Domain_Testing', 'Location_Bangalore',
               'Location_Delhi', 'Location_Hyderbad', 'Location_Mumbai'],
             dtype='object')
[340]: len(imputation.columns)
[340]: 19
  []:
```