	GO_STP_13248 Assessment 10
	Date - 12-06-2021
In [1]:	<pre>import numpy as np import pandas as pd import matplotlib.pyplot as plt import seaborn as sns</pre>
In [2]:	<pre>df = pd.read_table("salary.dat.txt", delim_whitespace= " ") df.head()</pre>
Out[2]:	sx         rk         yr         dg         yd         sl           0         male         full         25         doctorate         35         36350           1         male         full         13         doctorate         22         35350           2         male         full         7         doctorate         23         28200           3         female         full         7         doctorate         27         26775           4         male         full         19         masters         30         33696
In [3]:	<pre>df["sx"].unique()</pre>
Out[3]: In [4]:	<pre>array(['male', 'female'], dtype=object)  df["rk"].unique()</pre>
Out[4]:	array(['full', 'associate', 'assistant'], dtype=object)
	<pre>df['dg'].unique() array(['doctorate', 'masters'], dtype=object)</pre>
In [6]:	Pandas get_dummies() Function  df1 = pd.get_dummies(df["sx"])
Out[6]:	df1.head()  female male
	0       0       1         1       0       1         2       0       1
	3 1 0 4 0 1
In [7]:	<pre>df2 = pd.get_dummies(df["dg"]) df2.head()</pre>
Out[7]:	doctorate masters  0 1 0
	1       1       0         2       1       0         3       1       0         4       0       1
In [8]:	<pre>df3 = pd.get_dummies(df["rk"]) df3.head()</pre>
Out[8]:	assistant associate full
	0       0       0       1         1       0       0       1         2       0       0       1
	3 0 0 1 4 0 0 1
In [9]:	<pre>df_upd = pd.concat([df,df1, df2,df3], axis = 1)</pre>
In [10]:	df_upd.head() sx rk yr dg yd sl female male doctorate masters assistant associate full
Out[10]:	sx         rk         yr         dg         yd         st         female         masters         associate         full           0         male         full         25         doctorate         35         36350         0         1         1         0         0         1           1         male         full         13         doctorate         22         35350         0         1         1         0         0         1           2         male         full         7         doctorate         23         28200         0         1         1         0         0         1           3         female         full         7         doctorate         27         26775         1         0         1         0         0         1           4         male         full         19         masters         30         33696         0         1         0         0         1
In [11]:	<pre>df_upd = df_upd.drop(["sx", "rk", "dg"], axis = 1)</pre>
In [12]:	df_upd.head()
Out[12]:	yr         yd         sl         female         male         doctorate         masters         assistant         associate         full           0         25         35         36350         0         1         1         0         0         0         1
	1       13       22       35350       0       1       1       0       0       0       1         2       10       23       28200       0       1       1       0       0       0       1         3       7       27       26775       1       0       1       0       0       1         4       19       30       33696       0       1       0       0       0       1
In [13]:	<pre>One hot Encoding  from sklearn.preprocessing import LabelEncoder le = LabelEncoder() df.sx = le.fit_transform(df.sx) df.head()</pre>
Out[13]:	sx         rk         yr         dg         yd         sl           0         1         full         25         doctorate         3         35350           1         full         10         full         10         doctorate         22         35350           3         0         full         7         doctorate         27         26775           4         1         full         19         masters         30         33696
In [15]:	<pre>from sklearn.preprocessing import OneHotEncoder from sklearn.compose import ColumnTransformer ct = ColumnTransformer([('sx', OneHotEncoder(), [0])])</pre>
	<pre>newct_1 = ct.fit_transform(df) print(newct_1)  [[0. 1.]</pre>
	[0. 1.] [0. 1.] [1. 0.] [0. 1.] [0. 1.] [1. 0.]
	[0. 1.] [0. 1.] [0. 1.]
	[0. 1.] [0. 1.] [0. 1.] [0. 1.] [0. 1.]
	[0. 1.] [0. 1.] [0. 1.] [0. 1.] [0. 1.]
	[0. 1.] [0. 1.] [0. 1.] [0. 1.] [1. 0.]
	[0. 1.] [0. 1.] [0. 1.] [1. 0.]
	[0. 1.] [0. 1.] [1. 0.] [0. 1.]
	[0. 1.] [1. 0.] [1. 0.] [0. 1.]
	[1. 0.] [0. 1.] [0. 1.]
	[0. 1.] [0. 1.] [0. 1.] [1. 0.] [0. 1.]
	[0. 1.] [1. 0.] [1. 0.] [0. 1.] [1. 0.]
In [16]:	[1. 0.] [1. 0.]]
	<pre>from sklearn.preprocessing import LabelEncoder le = LabelEncoder() df.rk = le.fit_transform(df.rk) df.head()</pre>
Out[16]:	sx         rk         yr         dg         yd         sl           0         1         2         25         doctorate         35         36350
	1         1         2         13         doctorate         22         35350           2         1         2         10         doctorate         23         28200           3         0         2         7         doctorate         27         26775
	4 1 2 19 masters 30 33696
	<pre>from sklearn.preprocessing import LabelEncoder le = LabelEncoder() df.dg = le.fit_transform(df.dg) df.head()</pre>
Out[17]:	0         1         2         25         0         35         36350           1         1         2         13         0         22         35350           2         1         2         10         0         23         28200           3         0         2         7         0         27         26775
	4 1 2 19 1 30 33696  Thank you
In [ ]:	

Name - Debashis Saha