Employee_attribution_Predication

August 3, 2023

[70]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
```

[71]:
```
emp_data = pd.read_csv("1673873196_hr_comma_sep.csv")
```

[72]:
```
emp_data.head()
```

[72]:

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	sales	salary
0	0.38	0.53	2	157	3	0	1	0	sales	low
1	0.80	0.86	5	262	6	0	1	0	sales	medium
2	0.11	0.88	7	272	4	0	1	0	sales	medium
3	0.72	0.87	5	223	5	0	1	0	sales	low
4	0.37	0.52	2	159	3	0	1	0	sales	low

[60]:
```
emp_data.tail()
```

[60]:

	satisfaction_level	last_evaluation	number_project
14994	0.40	0.57	2
14995	0.37	0.48	2
14996	0.37	0.53	2
14997	0.11	0.96	6
14998	0.37	0.52	2

```
       average_montly_hours  time_spend_company  Work_accident  left  \
14994                   151                   3              0     1
14995                   160                   3              0     1
14996                   143                   3              0     1
14997                   280                   4              0     1
14998                   158                   3              0     1

       promotion_last_5years    sales salary
14994                      0  support    low
14995                      0  support    low
14996                      0  support    low
14997                      0  support    low
14998                      0  support    low
```

[61]:
```
emp_data.shape
```

[61]:
```
(14999, 10)
```

[62]:
```
emp_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14999 entries, 0 to 14998
Data columns (total 10 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   satisfaction_level     14999 non-null  float64
 1   last_evaluation        14999 non-null  float64
 2   number_project         14999 non-null  int64
 3   average_montly_hours   14999 non-null  int64
 4   time_spend_company     14999 non-null  int64
 5   Work_accident          14999 non-null  int64
 6   left                   14999 non-null  int64
 7   promotion_last_5years  14999 non-null  int64
 8   sales                  14999 non-null  object
 9   salary                 14999 non-null  object
dtypes: float64(2), int64(6), object(2)
memory usage: 1.1+ MB
```

[17]:
```
# Data Preprocessing:
# identify and treating the missing values.
```

[63]:
```
emp_data.isnull().sum()
```

[63]:
```
satisfaction_level       0
last_evaluation          0
number_project           0
average_montly_hours     0
```

```
time_spend_company       0
Work_accident            0
left                     0
promotion_last_5years    0
sales                    0
salary                   0
dtype: int64
```

[ ]:
```
# Their is many way to treatment of the missing value but in this case
```

[22]:
```
# check Duplicate records of the data
# overfitting of the model
```

[64]:
```
emp_data[emp_data.duplicated()]
```

[64]:
```
       satisfaction_level  last_evaluation  number_project  \
396                  0.46             0.57               2
866                  0.41             0.46               2
1317                 0.37             0.51               2
1368                 0.41             0.52               2
1461                 0.42             0.53               2
…                     …                …                …
14994                0.40             0.57               2
14995                0.37             0.48               2
14996                0.37             0.53               2
14997                0.11             0.96               6
14998                0.37             0.52               2

       average_montly_hours  time_spend_company  Work_accident  left  \
396                     139                   3              0     1
866                     128                   3              0     1
1317                    127                   3              0     1
1368                    132                   3              0     1
1461                    142                   3              0     1
…                       …                     …              …    …
14994                   151                   3              0     1
14995                   160                   3              0     1
14996                   143                   3              0     1
14997                   280                   4              0     1
14998                   158                   3              0     1

       promotion_last_5years       sales  salary
396                        0       sales     low
866                        0  accounting     low
1317                       0       sales  medium
1368                       0       RandD     low
1461                       0       sales     low
```

```
…                       …          …      …
14994                       0     support     low
14995                       0     support     low
14996                       0     support     low
14997                       0     support     low
14998                       0     support     low

[3008 rows x 10 columns]
```

[65]:
```
emp_data.drop_duplicates(keep="first")
```

[65]:
```
       satisfaction_level  last_evaluation  number_project  \
0                    0.38             0.53               2
1                    0.80             0.86               5
2                    0.11             0.88               7
3                    0.72             0.87               5
4                    0.37             0.52               2
…                     …                …                …
11995                0.90             0.55               3
11996                0.74             0.95               5
11997                0.85             0.54               3
11998                0.33             0.65               3
11999                0.50             0.73               4

       average_montly_hours  time_spend_company  Work_accident  left  \
0                       157                   3              0     1
1                       262                   6              0     1
2                       272                   4              0     1
3                       223                   5              0     1
4                       159                   3              0     1
…                        …                   …              …    …
11995                   259                  10              1     0
11996                   266                  10              0     0
11997                   185                  10              0     0
11998                   172                  10              0     0
11999                   180                   3              0     0

       promotion_last_5years       sales  salary
0                          0       sales     low
1                          0       sales  medium
2                          0       sales  medium
3                          0       sales     low
4                          0       sales     low
…                          …          …      …
11995                      1  management    high
11996                      1  management    high
11997                      1  management    high
```

```
11998                     1   marketing    high
11999                     0          IT     low

[11991 rows x 10 columns]
```

[66]:
```
# removing the Duplocate
data=emp_data.drop_duplicates(keep="first")
```

[67]:
```
data.shape
```

[67]:
```
(11991, 10)
```

[38]:
```
# correlation Matrix

data.corr()

# Positive correlation
# Negative correlation
# No correlation
```

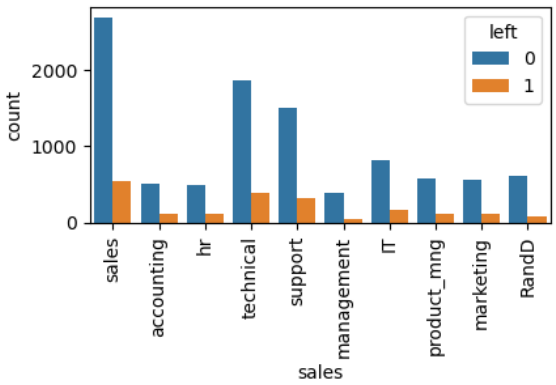
[38]:
```
                      satisfaction_level  last_evaluation  number_project  \
satisfaction_level              1.000000         0.095186       -0.133246
last_evaluation                 0.095186         1.000000        0.270256
number_project                 -0.133246         0.270256        1.000000
average_montly_hours           -0.006252         0.264678        0.331516
time_spend_company             -0.152915         0.096829        0.188837
Work_accident                   0.039940        -0.005695       -0.005612
left                           -0.350558         0.013520        0.030928
promotion_last_5years           0.019789        -0.007206       -0.000544

                      average_montly_hours  time_spend_company  \
satisfaction_level               -0.006252           -0.152915
last_evaluation                   0.264678            0.096829
number_project                    0.331516            0.188837
average_montly_hours              1.000000            0.102875
time_spend_company                0.102875            1.000000
Work_accident                    -0.012860            0.000003
left                              0.070409            0.173295
promotion_last_5years            -0.004964            0.056828

                      Work_accident      left  promotion_last_5years
satisfaction_level         0.039940 -0.350558               0.019789
last_evaluation           -0.005695  0.013520              -0.007206
number_project            -0.005612  0.030928              -0.000544
average_montly_hours      -0.012860  0.070409              -0.004964
time_spend_company         0.000003  0.173295               0.056828
Work_accident              1.000000 -0.125436               0.029852
```

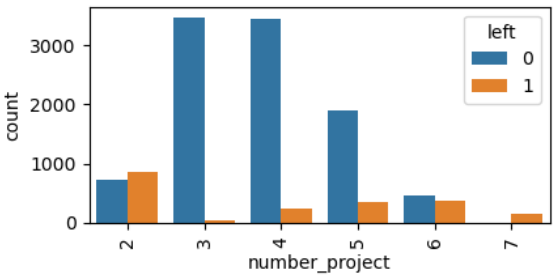
```
left                      -0.125436  1.000000             -0.044657
promotion_last_5years      0.029852 -0.044657              1.000000
```

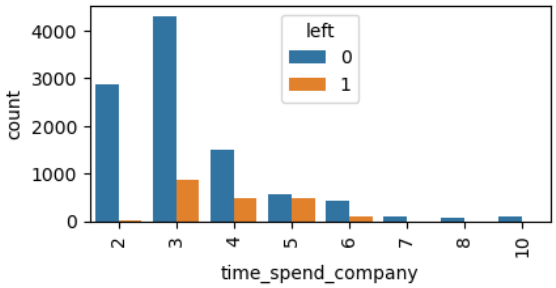
Data Visualtion, analysing different features

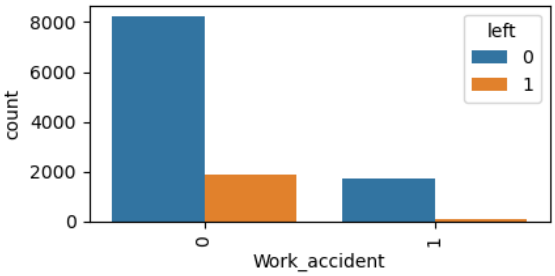
[68]:
```
frt = ["sales","number_project","time_spend_company","Work_accident","salary"]
```

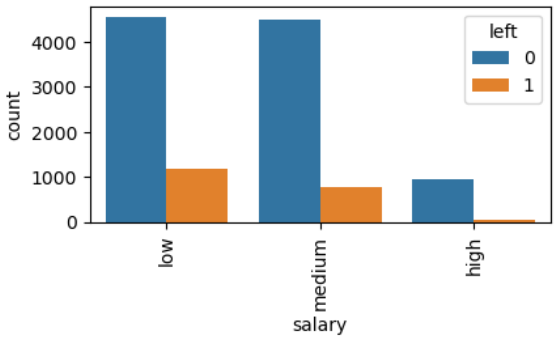
[69]:
```
fig=plt.subplots(figsize=(10,15))

for p,q in enumerate(frt):
    plt.subplot(4,2,p+1)
    plt.subplots_adjust(hspace=1)
    plt.xticks(rotation=90)
    sns.countplot(x=q,data=data,hue="left")
```











[51]:
```
# attrition rate: number of employee left/ number of total employee
```

[73]:
```
# Understanding
# people with low project as well as people having more project or having more␣
 ↪project that are leaving the company
# Years of experiance: 3 - 6 years
# salary .......> Low,Medium
# Promotion: Likely, quit ....> Havent received promotation
# Time with Company: After 3-6 years crucial the employee, affection with␣
 ↪organisation
```

```
# Number of project: if the opportunities are less or if the employee is␣
 ↪overburdened, more change of the employee to quit the job
# salary : incentive based system to be introduced
```