Data Labelling

In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: # Read the dataset
df = pd.read_csv(r'F:\Technocolabs\WA_Fn-UseC_-HR-Employee-Attrition.csv')

In [3]: df

Out[3]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education
0	41	Yes	Travel_Rarely	1102	Sales	1	2
1	49	No	Travel_Frequently	279	Research & Development	8	1
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2
3	33	No	Travel_Frequently	1392	Research & Development	3	4
4	27	No	Travel_Rarely	591	Research & Development	2	1
1465	36	No	Travel_Frequently	884	Research & Development	23	2
1466	39	No	Travel_Rarely	613	Research & Development	6	1
1467	27	No	Travel_Rarely	155	Research & Development	4	3
1468	49	No	Travel_Frequently	1023	Sales	2	3
1469	34	No	Travel_Rarely	628	Research & Development	8	3

1470 rows × 35 columns

```
In [5]: import pandas as pd

# Load your dataset
# Replace 'your_dataset.csv' with the actual path to your dataset file
df = pd.read_csv(r'F:\Technocolabs\WA_Fn-UseC_-HR-Employee-Attrition.csv')

# Define your condition and label accordingly
def label_function(row):
    if row['Age'] > 30:
        return 'label_A'
    else:
        return 'label_B'

# Apply the label function to each row
df['label'] = df.apply(label_function, axis=1)

# Display the first few rows to verify the labeling
print(df.head())
```

	Age Attritio				DailyRate		Departmen	
0	41 Ye		Travel_Rarely		1102 279		Sale	
1		_	Travel_Frequently				& Developmen	
2	37 Ye		el_Rar	-	1373		& Developmen	
3		lo Travel_Fr		-	1392		& Developmen	
4	27 N	lo Trave	el_Rar	ely	591	Research 8	& Developmen	t
on	DistanceFrom	Home Educat	ion E	ducat	ionField	EmployeeCo	unt Employe	eNumb
er 0 1	`	1	2	Life	Sciences		1	
1 2		8	1	Life	Sciences		1	
2 4		2	2		Other		1	
3 5		3		Life	Sciences		1	
4 7		2	1		Medical		1	
	Standar	dHours Stock	(Optio	nLeve		orkingYears	\	
0	• • •	80			0	8		
1	• • •	80 80			1	10		
2		хи			0			
_	• • •					7		
3	•••	80			0	8		
3 4								
4	 TrainingTime	80 80	VorkLi	.feBa]	0	8	YearsInCurr	entRo
4 le 0	 TrainingTime	80 80	VorkLi	.feBa]	0	8	YearsInCurr	entRo
4 le	 TrainingTime	80 80 sLastYear W	WorkLi	.feBal	0 1 ance Years	8 6 SAtCompany	YearsInCurr	entRo
1e 0 4 1	 TrainingTime	80 80 sLastYear w	VorkLi	.feBa]	0 1 ance Years	8 6 SAtCompany 6	YearsInCurr	entRo
1e 0 4 1 7 2	 TrainingTime	80 80 sLastYear w 0 3	WorkLi	.feBal	0 1 ance Years 1 3	8 6 SAtCompany 6 10	YearsInCurr	entRo
1e 0 4 1 7 2 0 3	 TrainingTime	80 80 esLastYear w 0 3 3	VorkLi	.feBa]	0 1 ance Years 1 3	8 6 SAtCompany 6 10 0	YearsInCurr	entRo
1e 0 4 1 7 2 0 3 7 4 2	 TrainingTime	80 80 esLastYear w 0 3 3 3			0 1 ance Years 1 3 3	8 6 SAtCompany 6 10 0 8 2	YearsInCurr	entRo
1e 0 4 1 7 2 0 3 7 4 2	TrainingTime	80 80 esLastYear w 0 3 3 3			0 1 ance Years 1 3 3 3	8 6 6 6 10 0 8 2 1abel	YearsInCurr	entRo
1e 0 4 1 7 2 0 3 7 4 2	TrainingTime	80 80 **SLastYear W 0 3 3 3 3			0 1 ance Years 1 3 3 3 currManager	8 6 SAtCompany 6 10 0 8 2	YearsInCurr	entRo
1e 0 4 1 7 2 0 3 7 4 2	TrainingTime	80 80 **SLastYear W 0 3 3 3 3 **TPromotion 0			0 1 ance Years 1 3 3 3 currManager	8 6 SAtCompany 6 10 0 8 2 Salabel Salabel Salabel Salabel Salabel Salabel Salabel Salabel Salabel	YearsInCurr	entRo
1e 0 4 1 7 2 0 3 7 4 2	TrainingTime	80 80 **SLastYear W 0 3 3 3 3 **TPromotion 0 1			0 1 ance Years 1 3 3 3 currManager	8 6 SAtCompany 6 10 0 8 2 Label 6 label_A 7 label_A	YearsInCurr	entRo

[5 rows x 36 columns]

```
In [6]: # Define your conditions and labels accordingly
def label_function(row):
    if row['Age'] > 30 and row['Department'] == 'Sales':
        return 'label_A'
    # Example: If DailyRate is less than 500 and Education is greater than if
    elif row['DailyRate'] < 500 and row['Education'] > 3:
        return 'label_B'
    # Add more conditions and labels as needed
    else:
        return 'label_C' # Default label if none of the conditions are met

# Apply the label function to each row
df['label'] = df.apply(label_function, axis=1)

# Display the first few rows to verify the labeling
print(df.head())
```

0 1 2 3 4	Age 41 49 37 33 27	Attrition Yes No Yes No No	Travel_F Trav Travel_F	el_Ra reque el_Ra	rely ntly rely ntly	DailyRate 1102 279 1373 1392 591	Research Research	& De & De & De	velopment velopment	\
	Dist	tanceFromHo	me Educa	tion	Educa	tionField	EmployeeCo	unt	EmployeeN	lumb
er 0	\		1	2	Lifo	Sciences		1		
1			_	2	LITE	Sciences				
1			8	1	Life	Sciences		1		
2 2			2	2		Other		1		
4			2			<i>c</i> :				
3 5			3	4	L1†e	Sciences		1		
4 7			2	1		Medical		1		
0 1 2 3 4	••••	StandardH	ours Stoc 80 80 80 80 80	kOpti	onLeve	el TotalWo 0 1 0 0	orkingYears 8 10 7 8 6			
1.		iningTimesL	astYear	WorkL	ifeBa	lance Years	AtCompany	Yea	rsInCurrer	ntRo
le 0	\		0			1	6			
4 1			3			3	10			
7			5			5	10			
2 0			3			3	0			
3			3			3	8			
7 4			3			3	2			
2										
	Years	sSinceLastP	romotion	Year	sWith	CurrManager				
0 1			0 1				5 label_A 7 label_C			
2			0				label_C			
3			3				abel_C			
4			2				label_C			

[5 rows x 36 columns]

```
In [7]: # Define your conditions and labels accordingly
        def label_attrition(row):
            if row['Attrition'] == 'Yes':
                return 'High Attrition'
            else:
                return 'Low Attrition'
        def label_income(row):
            if row['MonthlyIncome'] > 5000:
                return 'High Income'
            else:
                return 'Low Income'
        # Apply the label functions to each row
        df['Attrition_Label'] = df.apply(label_attrition, axis=1)
        df['Income_Label'] = df.apply(label_income, axis=1)
        # Display the first few rows to verify the labeling
        print(df.head())
```

```
Age Attrition
                      BusinessTravel DailyRate
                                                                 Department \
0
    41
              Yes
                       Travel Rarely
                                             1102
                                                                      Sales
1
    49
                                              279
                                                   Research & Development
               No
                   Travel_Frequently
2
    37
              Yes
                       Travel_Rarely
                                             1373
                                                    Research & Development
3
    33
               No
                   Travel_Frequently
                                             1392
                                                    Research & Development
                       Travel_Rarely
                                              591
                                                    Research & Development
4
    27
               No
   DistanceFromHome
                      Education EducationField
                                                  EmployeeCount
                                                                   EmployeeNumb
   \
er
                               2 Life Sciences
                                                                1
0
                   1
1
1
                   8
                                  Life Sciences
                                                                1
2
2
                   2
                               2
                                           Other
                                                                1
4
3
                                  Life Sciences
                   3
                                                                1
5
4
                   2
                               1
                                         Medical
                                                                1
7
        TotalWorkingYears TrainingTimesLastYear
                                                     WorkLifeBalance
0
                          8
                                                  0
                                                                    1
                         10
                                                  3
1
                                                                    3
   . . .
2
                          7
                                                  3
                                                                    3
   . . .
3
                          8
                                                  3
                                                                    3
   . . .
                                                  3
                          6
                                                                    3
4
   . . .
   YearsAtCompany
                    YearsInCurrentRole YearsSinceLastPromotion
0
                 6
                                       4
                                                                 0
1
                10
                                       7
                                                                 1
2
                                       0
                                                                 0
                 0
                                       7
                 8
                                                                 3
3
                 2
                                       2
                                                                 2
4
   YearsWithCurrManager
                             label
                                    Attrition_Label
                                                       Income_Label
0
                       5
                          label_A
                                     High Attrition
                                                        High Income
1
                       7
                           label C
                                      Low Attrition
                                                        High Income
2
                       0
                           label C
                                      High Attrition
                                                         Low Income
3
                       0
                           label C
                                      Low Attrition
                                                         Low Income
4
                       2
                           label_C
                                       Low Attrition
                                                         Low Income
```

[5 rows x 38 columns]

```
In [8]: # Define your conditions and labels accordingly
        def label_attrition(row):
            if row['Attrition'] == 'Yes':
                return 'High Attrition'
            else:
                return 'Low Attrition'
        def label_income(row):
            if row['MonthlyIncome'] > 5000:
                return 'High Income'
            else:
                return 'Low Income'
        # Apply the label functions to each row
        df['Attrition_Label'] = df.apply(label_attrition, axis=1)
        df['Income_Label'] = df.apply(label_income, axis=1)
        # Display the first few rows to verify the labeling
        print(df.head())
```

```
Age Attrition
                      BusinessTravel DailyRate
                                                                Department \
0
    41
             Yes
                       Travel Rarely
                                             1102
                                                                     Sales
1
    49
                                             279
                                                  Research & Development
              No
                   Travel_Frequently
2
    37
             Yes
                       Travel_Rarely
                                             1373 Research & Development
3
    33
               No
                   Travel_Frequently
                                             1392
                                                   Research & Development
                       Travel_Rarely
                                              591
                                                   Research & Development
4
    27
               No
                      Education EducationField EmployeeCount EmployeeNumb
   DistanceFromHome
   \
er
                               2 Life Sciences
0
                   1
                                                               1
1
1
                   8
                                 Life Sciences
                                                               1
2
2
                   2
                               2
                                          Other
                                                               1
4
3
                                  Life Sciences
                   3
                                                               1
5
4
                   2
                              1
                                        Medical
                                                               1
7
        TotalWorkingYears TrainingTimesLastYear
                                                    WorkLifeBalance
0
                         8
                                                 0
                                                                   1
                        10
                                                 3
1
                                                                   3
   . . .
2
                         7
                                                 3
                                                                   3
   . . .
3
                         8
                                                 3
                                                                   3
                                                 3
                         6
                                                                   3
4
   . . .
                    YearsInCurrentRole YearsSinceLastPromotion
   YearsAtCompany
0
                 6
                                      4
                                                                0
1
                10
                                      7
                                                                1
2
                                      0
                                                                0
                 0
                                      7
                 8
                                                                3
3
                                      2
                                                                2
4
                 2
                            label
   YearsWithCurrManager
                                   Attrition_Label Income_Label
0
                       5
                          label_A
                                     High Attrition
                                                       High Income
1
                       7
                          label C
                                      Low Attrition
                                                       High Income
2
                       0
                          label C
                                     High Attrition
                                                        Low Income
3
                       0
                          label C
                                     Low Attrition
                                                        Low Income
4
                       2
                          label_C
                                      Low Attrition
                                                        Low Income
[5 rows x 38 columns]
```

In []: