----------Exploratory Data Analysis----------

In [2]:
```
import pandas as pd
import numpy as np
```

In [3]:
```
data = pd.read_csv('C:/Users/DELL/Desktop/Atomcamp Python/HR-Attrition.
```

In [4]:
```
data.head(10)
```

Out[4]:

	Index	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	I
0	1	41	Yes	Travel_Rarely	1102	Sales	1	
1	2	49	No	Travel_Frequently	279	Research & Development	8	
2	3	37	Yes	Travel_Rarely	1373	Research & Development	2	
3	4	33	No	Travel_Frequently	1392	Research & Development	3	
4	5	27	No	Travel_Rarely	591	Research & Development	2	
5	6	32	No	Travel_Frequently	1005	Research & Development	2	
6	7	59	No	Travel_Rarely	1324	Research & Development	3	
7	8	30	No	Travel_Rarely	1358	Research & Development	24	
8	9	38	No	Travel_Frequently	216	Research & Development	23	
9	10	36	No	Travel_Rarely	1299	Research & Development	27	

10 rows × 36 columns

In [5]:
```
data.head(10)
```

Out[5]:

	Index	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	I
0	1	41	Yes	Travel_Rarely	1102	Sales	1	
1	2	49	No	Travel_Frequently	279	Research & Development	8	
2	3	37	Yes	Travel_Rarely	1373	Research & Development	2	
3	4	33	No	Travel_Frequently	1392	Research & Development	3	
4	5	27	No	Travel_Rarely	591	Research & Development	2	
5	6	32	No	Travel_Frequently	1005	Research & Development	2	
6	7	59	No	Travel_Rarely	1324	Research & Development	3	
7	8	30	No	Travel_Rarely	1358	Research & Development	24	
8	9	38	No	Travel_Frequently	216	Research & Development	23	
9	10	36	No	Travel_Rarely	1299	Research & Development	27	

10 rows × 36 columns

In [6]:
```
data.tail(10)
```

Out[6]:

	Index	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHom
1460	1461	29	No	Travel_Rarely	468	Research & Development	2
1461	1462	50	Yes	Travel_Rarely	410	Sales	2
1462	1463	39	No	Travel_Rarely	722	Sales	2
1463	1464	31	No	Non-Travel	325	Research & Development	
1464	1465	26	No	Travel_Rarely	1167	Sales	
1465	1466	36	No	Travel_Frequently	884	Research & Development	2
1466	1467	39	No	Travel_Rarely	613	Research & Development	
1467	1468	27	No	Travel_Rarely	155	Research & Development	
1468	1469	49	No	Travel_Frequently	1023	Sales	
1469	1470	34	No	Travel_Rarely	628	Research & Development	

10 rows × 36 columns

In [8]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 36 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   Index                     1470 non-null   int64 
 1   Age                       1470 non-null   int64 
 2   Attrition                 1470 non-null   object
 3   BusinessTravel            1470 non-null   object
 4   DailyRate                 1470 non-null   int64 
 5   Department                1470 non-null   object
 6   DistanceFromHome          1470 non-null   int64 
 7   Education                 1470 non-null   int64 
 8   EducationField            1470 non-null   object
 9   EmployeeCount             1470 non-null   int64 
 10  EmployeeNumber            1470 non-null   int64 
 11  EnvironmentSatisfaction   1470 non-null   int64 
 12  Gender                    1470 non-null   object
 13  HourlyRate                1470 non-null   int64 
 14  JobInvolvement            1470 non-null   int64 
 15  JobLevel                  1470 non-null   int64 
 16  JobRole                   1470 non-null   object
 17  JobSatisfaction           1470 non-null   int64 
 18  MaritalStatus             1470 non-null   object
 19  MonthlyIncome             1470 non-null   int64 
 20  MonthlyRate               1470 non-null   int64 
 21  NumCompaniesWorked        1470 non-null   int64 
 22  Over18                    1470 non-null   object
 23  OverTime                  1470 non-null   object
 24  PercentSalaryHike         1470 non-null   int64 
 25  PerformanceRating         1470 non-null   int64 
 26  RelationshipSatisfaction  1470 non-null   int64 
 27  StandardHours             1470 non-null   int64 
 28  StockOptionLevel          1470 non-null   int64 
 29  TotalWorkingYears         1470 non-null   int64 
 30  TrainingTimesLastYear     1470 non-null   int64 
 31  WorkLifeBalance           1470 non-null   int64 
 32  YearsAtCompany            1470 non-null   int64 
 33  YearsInCurrentRole        1470 non-null   int64 
 34  YearsSinceLastPromotion   1470 non-null   int64 
 35  YearsWithCurrManager      1470 non-null   int64 
dtypes: int64(27), object(9)
memory usage: 413.6+ KB
```

In [11]:
```
data.shape
```

Out[11]:
```
(1470, 36)
```

In [14]:
```
data.describe().T
```

Out[14]:

	count	mean	std	min	25%	50%
Index	1470.0	735.500000	424.496761	1.0	368.25	735.5
Age	1470.0	36.923810	9.135373	18.0	30.00	36.0
DailyRate	1470.0	802.485714	403.509100	102.0	465.00	802.0
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00	7.0
Education	1470.0	2.912925	1.024165	1.0	2.00	3.0
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00	1.0
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25	1020.5
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00	3.0
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00	66.0
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00	3.0
JobLevel	1470.0	2.063946	1.106940	1.0	1.00	2.0
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.00	3.0
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.00	4919.0
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.00	14235.5
NumCompaniesWorked	1470.0	2.693197	2.498009	0.0	1.00	2.0
PercentSalaryHike	1470.0	15.209524	3.659938	11.0	12.00	14.0
PerformanceRating	1470.0	3.153741	0.360824	3.0	3.00	3.0
RelationshipSatisfaction	1470.0	2.712245	1.081209	1.0	2.00	3.0
StandardHours	1470.0	80.000000	0.000000	80.0	80.00	80.0
StockOptionLevel	1470.0	0.793878	0.852077	0.0	0.00	1.0
TotalWorkingYears	1470.0	11.279592	7.780782	0.0	6.00	10.0
TrainingTimesLastYear	1470.0	2.799320	1.289271	0.0	2.00	3.0
WorkLifeBalance	1470.0	2.761224	0.706476	1.0	2.00	3.0
YearsAtCompany	1470.0	7.008163	6.126525	0.0	3.00	5.0
YearsInCurrentRole	1470.0	4.229252	3.623137	0.0	2.00	3.0
YearsSinceLastPromotion	1470.0	2.187755	3.222430	0.0	0.00	1.0
YearsWithCurrManager	1470.0	4.123129	3.568136	0.0	2.00	3.0

In [15]:
```
data.describe(include='object')
```

Out[15]:

	Attrition	BusinessTravel	Department	EducationField	Gender	JobRole	Marita
count	1470	1470	1470	1470	1470	1470	
unique	2	3	3	6	2	9	
top	No	Travel_Rarely	Research & Development	Life Sciences	Male	Sales Executive	
freq	1233	1043	961	606	882	326	

In [17]:
```
data.describe(include=['bool','object'])
```

Out[17]:

	Attrition	BusinessTravel	Department	EducationField	Gender	JobRole	Marita
count	1470	1470	1470	1470	1470	1470	
unique	2	3	3	6	2	9	
top	No	Travel_Rarely	Research & Development	Life Sciences	Male	Sales Executive	
freq	1233	1043	961	606	882	326	

In [18]:
```
data.columns
```

Out[18]:
```
Index(['Index', 'Age', 'Attrition', 'BusinessTravel', 'DailyRate',
       'Department', 'DistanceFromHome', 'Education', 'EducationFiel
d',
       'EmployeeCount', 'EmployeeNumber', 'EnvironmentSatisfaction',
'Gender',
       'HourlyRate', 'JobInvolvement', 'JobLevel', 'JobRole',
       'JobSatisfaction', 'MaritalStatus', 'MonthlyIncome', 'MonthlyRa
te',
       'NumCompaniesWorked', 'Over18', 'OverTime', 'PercentSalaryHik
e',
       'PerformanceRating', 'RelationshipSatisfaction', 'StandardHour
s',
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYea
r',
       'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole',
       'YearsSinceLastPromotion', 'YearsWithCurrManager'],
      dtype='object')
```

In [19]:
```
data.dtypes
```

Out[19]:
```
Index                       int64
Age                         int64
Attrition                  object
BusinessTravel             object
DailyRate                   int64
Department                 object
DistanceFromHome            int64
Education                   int64
EducationField             object
EmployeeCount               int64
EmployeeNumber              int64
EnvironmentSatisfaction     int64
Gender                     object
HourlyRate                  int64
JobInvolvement              int64
JobLevel                    int64
JobRole                    object
JobSatisfaction             int64
MaritalStatus              object
MonthlyIncome               int64
MonthlyRate                 int64
NumCompaniesWorked          int64
Over18                     object
OverTime                   object
PercentSalaryHike           int64
PerformanceRating           int64
RelationshipSatisfaction    int64
StandardHours               int64
StockOptionLevel            int64
TotalWorkingYears           int64
TrainingTimesLastYear       int64
WorkLifeBalance             int64
YearsAtCompany              int64
YearsInCurrentRole          int64
YearsSinceLastPromotion     int64
YearsWithCurrManager        int64
dtype: object
```

In [20]:
```
data['EducationField'].unique()
```

Out[20]:
```
array(['Life Sciences', 'Other', 'Medical', 'Marketing',
       'Technical Degree', 'Human Resources'], dtype=object)
```

In [21]:
```
data['EducationField'].nunique()
```

Out[21]:
```
6
```

In [23]:
```
data['EducationField'].value_counts()
```

Out[23]:
```
EducationField
Life Sciences       606
Medical             464
Marketing           159
Technical Degree    132
Other                82
Human Resources      27
Name: count, dtype: int64
```

In [ ]: