M504_Indiviual_Final_Project_GH1019253

March 17, 2022

1 Final Assessment for M504 (A.I and Application)

[1]:
```python
from IPython.display import Image
Image("GISMA_LOGO.png",width = 200, height = 200)
```

[1]:

1.0.1 IBM Employees Survey on the Attrition Dataset

2 Table of the Content

2.0.1 1) Introduction

2.0.2 2) Background & need for study:

2.0.3 3) Scope & Objectives(Business Questions):

2.0.4 4) Approach & methodology:

2.0.5 5) Importing Libraries

2.0.6 6) Importing Dataset Using OS Libraries

2.0.7 7) Loading Dataset

2.0.8 8) Data Description and filteration

8.1) Checking for Null values in dataset

8.2)Visualizing Data Insights

8.3) Here we are Checking for the duplicated column if it is there any)

8.4) Checking for Unique values in each column

2.0.9 9) Data Cleansing

2.0.10 10) Plotting Histogram

2.0.11 11) Data Findings on intial analysis

2.0.12 12) Dropping the Columns we don't need

2.0.13 13) Statistics and Insight in detail

13.1) Lets see what the statistics tell us and compare them

13.1.1) Employees who left the company

13.1.2) Employee who stayed with the company

13.2) Answer to the Business Questions

2.0.14 14) Plotting the heatmap to find correlations

14.1) Analysis of Heatmap

2.0.15 15) Feature Selection and Target Mapping (Addressing the findings)

15.1) Age vs Atttrition

15.2) JobRole vs Attrition

15.3) Marital Status vs Attrition

15.4) JobInvolvement vs Attrition

15.5) JobLevel vs Attrition

15.6) Distance from Home vs Attrition

15.7) Years with Current Manager vs Attrition

15.8) Total working years vs Attrition

15.9) Monthly Income vs JobRole

15.10) Monthly Income vs JobRole vs Attrition

2.0.16 16) Conclusion

1) Introduction My Clients company is concerned about high level of Attrition rate and due to which their integrity can be on the high risk, They have hired me as a Data analyst to figure out what are the facts that affects the attrition rate and to bring the insights of data to the company directors so they can come up with the solution and try to make decision which can benefit their company.The Dataset we are going to use is provided by IBM on Kaggle(www.kaggle.com).

If we discuss about the type of data we have, it's based on numerical and categorical data, we have 1500 entries approximately. We have 35 unique column and 1470 rows.It's actually based on the survey done by the IBM itself. We will be working on this dataset to find out the insights we will try to map the different variable on the target variable to find out facts thats affecting the attrition rate.

3 2) Background & need for study:

My Client's organization is confronting an imposing challenge of hiring and holding the talented employees while simultaneously experiencing difficulty to manage loss through attrition.Losing employees and representatives are bringing about execution misfortunes which can have long haul adverse consequence on organizations particularly in the event that the withdrawing ability leaves holes in its execution capacity and human asset working which incorporates lost efficiency as well as potentially loss of work group agreement and social generosity. The progress of any organization relies to a great extent upon the employees, the workers are considered as the foundation of any company.Thats why they employed me as a Data analyst to figure out why this occurring and what are the causes and how might we conquer this situtaion. This study was principally attempted to distinguish the level of worker's disposition, the disappointment factors they face in the organization and why they are leaving the company. When the levels of employees's mentality are recognized and the factors that appearing in the company then administration would be able to make an important move to decrease whittling down level(Attrition).

4 3) Scope & Objectives(Business Questions):

1. Does age affects the tendency of leaving the company?
2. Does the MonthlyIncome affects the Attrition rate?
3. Does the Distance from Home to work affects the Employees to leave the company ?
4. Do Environment Satisfaction matters to make an employee stay at the company ?
5. The position on which Attrition is higher or lower?
6. How does the level of involvement at work affect the Attrition?
7. Do martial status affects the Attrition level ?
8. Does JobRole have any tendency in the Attrition level?
9. What are the wage distribution between different job position?
10. Does the length of stay as a Current Manager influence the departure of employees?
11. Is there a relationship between total working time in the company and Attrition?

5 4) Approach & methodology:

The dataset is handed by IBM on Kaggle. We are going to import the dataset using OS library and then visualize it and look for the missing values or null values in dataset, if we find any we will deal with it by either dropping that rows or we manipulate it using some ML algorithms,then we visualize it and look for the unique values in the each column, moving further we will do some data cleansing and try to convert the categorical data into numerical data, after doing all of these step we will plot some histogram to visualize what are the insights so far we have figured out.After that I am gonna map the values against the target label to find the answers of the questions i have mentioned above and on the bases of the answers i will conclude my findings.

6 5) Importing Libraries

Here I have imported some major libraries which i will be using most frequently.

[2]:
```
# To Ignore the warnings
import warnings
warnings.filterwarnings('always')
warnings.filterwarnings('ignore')


import hvplot


import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import hvplot.pandas
```

```
import missingno as msno

%matplotlib inline
sns.set_style("whitegrid")
plt.style.use("Solarize_Light2")

pd.set_option("display.float_format", "{:.2f}".format)
pd.set_option("display.max_columns", 80)
pd.set_option("display.max_rows", 80)
```

7 6) Importing Dataset Using OS Libraries

Dataset = HR_Employee_data.csv

[3]:
```
import os
for dirname, _, filenames in os.walk('/data/'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

[4]:
```
!pip install hvplot
```

```
Requirement already satisfied: hvplot in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (0.7.3)
Requirement already satisfied: bokeh>=1.0.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from hvplot)
(2.4.2)
Requirement already satisfied: colorcet>=2 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from hvplot)
(3.0.0)
Requirement already satisfied: pandas in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from hvplot)
(1.3.5)
Requirement already satisfied: holoviews>=1.11.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from hvplot)
(1.14.8)
Requirement already satisfied: numpy>=1.15 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from hvplot)
(1.21.2)

WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
```

```
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)


Requirement already satisfied: pillow>=7.1.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
bokeh>=1.0.0->hvplot) (8.4.0)
Requirement already satisfied: packaging>=16.8 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
bokeh>=1.0.0->hvplot) (21.3)
Requirement already satisfied: typing-extensions>=3.10.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
bokeh>=1.0.0->hvplot) (3.10.0.2)
Requirement already satisfied: PyYAML>=3.10 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
bokeh>=1.0.0->hvplot) (6.0)
Requirement already satisfied: tornado>=5.1 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
bokeh>=1.0.0->hvplot) (6.1)
Requirement already satisfied: Jinja2>=2.9 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
bokeh>=1.0.0->hvplot) (2.11.3)
Requirement already satisfied: param>=1.7.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
colorcet>=2->hvplot) (1.12.0)
Requirement already satisfied: pyct>=0.4.4 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
colorcet>=2->hvplot) (0.4.8)
Requirement already satisfied: panel>=0.8.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
holoviews>=1.11.0->hvplot) (0.12.6)
Requirement already satisfied: pyviz-comms>=0.7.4 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
holoviews>=1.11.0->hvplot) (2.1.0)
Requirement already satisfied: pytz>=2017.3 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
pandas->hvplot) (2021.3)
Requirement already satisfied: python-dateutil>=2.7.3 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
pandas->hvplot) (2.8.2)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
Jinja2>=2.9->bokeh>=1.0.0->hvplot) (1.1.1)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
packaging>=16.8->bokeh>=1.0.0->hvplot) (3.0.4)
Requirement already satisfied: requests in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
```

```
panel>=0.8.0->holoviews>=1.11.0->hvplot) (2.27.1)
Requirement already satisfied: tqdm>=4.48.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
panel>=0.8.0->holoviews>=1.11.0->hvplot) (4.62.3)
Requirement already satisfied: bleach in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
panel>=0.8.0->holoviews>=1.11.0->hvplot) (4.1.0)
Requirement already satisfied: markdown in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
panel>=0.8.0->holoviews>=1.11.0->hvplot) (3.3.6)
Requirement already satisfied: six>=1.5 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from python-
dateutil>=2.7.3->pandas->hvplot) (1.16.0)
Requirement already satisfied: colorama in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
tqdm>=4.48.0->panel>=0.8.0->holoviews>=1.11.0->hvplot) (0.4.4)
Requirement already satisfied: webencodings in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
bleach->panel>=0.8.0->holoviews>=1.11.0->hvplot) (0.5.1)
Requirement already satisfied: importlib-metadata>=4.4 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
markdown->panel>=0.8.0->holoviews>=1.11.0->hvplot) (4.10.0)
Requirement already satisfied: idna<4,>=2.5 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
requests->panel>=0.8.0->holoviews>=1.11.0->hvplot) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
requests->panel>=0.8.0->holoviews>=1.11.0->hvplot) (2021.10.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
requests->panel>=0.8.0->holoviews>=1.11.0->hvplot) (1.26.7)
Requirement already satisfied: charset-normalizer~=2.0.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
requests->panel>=0.8.0->holoviews>=1.11.0->hvplot) (2.0.4)
Requirement already satisfied: zipp>=0.5 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
importlib-metadata>=4.4->markdown->panel>=0.8.0->holoviews>=1.11.0->hvplot)
(3.7.0)
```

[5]:
```
!pip install missingno
```

```
Requirement already satisfied: missingno in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (0.5.0)
```

```
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
```

```
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy
(c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages)


Requirement already satisfied: scipy in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
missingno) (1.7.3)
Requirement already satisfied: numpy in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
missingno) (1.21.2)
Requirement already satisfied: seaborn in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
missingno) (0.11.2)
Requirement already satisfied: matplotlib in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
missingno) (3.5.0)
Requirement already satisfied: packaging>=20.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
matplotlib->missingno) (21.3)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
matplotlib->missingno) (1.3.1)
Requirement already satisfied: cycler>=0.10 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
matplotlib->missingno) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
matplotlib->missingno) (4.25.0)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
matplotlib->missingno) (2.8.2)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
matplotlib->missingno) (8.4.0)
Requirement already satisfied: pyparsing>=2.2.1 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
matplotlib->missingno) (3.0.4)
Requirement already satisfied: pandas>=0.23 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
seaborn->missingno) (1.3.5)
Requirement already satisfied: pytz>=2017.3 in
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from
pandas>=0.23->seaborn->missingno) (2021.3)
Requirement already satisfied: six>=1.5 in
```

```
c:\users\hursh\appdata\local\continuum\anaconda3\lib\site-packages (from python-
dateutil>=2.7->matplotlib->missingno) (1.16.0)
```

8 7) Loading Dataset

[6]:
```python
#Loading the dataset using pandas csv read

employee_data_set = pd.read_csv("data/HR_Employee_data.csv")
```

[7]:
```python
# Viewing the first few lines of the dataset using Head() function

employee_data_set.head()
```

[7]:
```
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   EnvironmentSatisfaction  Gender  HourlyRate  JobInvolvement  JobLevel  \
0                        2  Female          94               3         2
1                        3    Male          61               2         2
2                        4    Male          92               2         1
3                        4  Female          56               3         1
4                        1    Male          40               3         1

                 JobRole  JobSatisfaction MaritalStatus  MonthlyIncome  \
0        Sales Executive                4        Single           5993
1     Research Scientist                2       Married           5130
2  Laboratory Technician                3        Single           2090
3     Research Scientist                3       Married           2909
4  Laboratory Technician                2       Married           3468

   MonthlyRate  NumCompaniesWorked Over18 OverTime  PercentSalaryHike  \
0        19479                   8      Y      Yes                 11
1        24907                   1      Y       No                 23
2         2396                   6      Y      Yes                 15
3        23159                   1      Y      Yes                 11
```

```
4       16632                 9      Y       No                 12

   PerformanceRating  RelationshipSatisfaction  StandardHours  \
0                  3                         1             80
1                  4                         4             80
2                  3                         2             80
3                  3                         3             80
4                  3                         4             80

   StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                 0                  8                      0
1                 1                 10                      3
2                 0                  7                      3
3                 0                  8                      3
4                 1                  6                      3

   WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
0                1               6                   4
1                3              10                   7
2                3               0                   0
3                3               8                   7
4                3               2                   2

   YearsSinceLastPromotion  YearsWithCurrManager
0                        0                     5
1                        1                     7
2                        0                     0
3                        3                     0
4                        2                     2
```

[8]:

```python
# checking if there is any null or Nan values in the given dataset.
employee_data_set.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
```

```
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

8.0.1 Data Description

- The Employee dataset has 35 features in total (columns), 26 of which are numeric and 9 are categorical, in addition to 1470 rows.
- The Employee dataset does not contain any null values

9 8) Data Description and filteration

Here we are going to check the dataset for the null values or missing values or duplicated values and try to get rid of them.

9.1 8.1) Checking for Null values in dataset

[9]:
```
employee_data_set.isnull()
```

[9]:
```
     Age  Attrition  BusinessTravel  DailyRate  Department  \
0  False      False           False      False       False
1  False      False           False      False       False
```

```
2     False      False          False      False       False
3     False      False          False      False       False
4     False      False          False      False       False
…     …        …            …        …         …
1465  False      False          False      False       False
1466  False      False          False      False       False
1467  False      False          False      False       False
1468  False      False          False      False       False
1469  False      False          False      False       False

      DistanceFromHome  Education  EducationField  EmployeeCount  \
0                False      False           False          False
1                False      False           False          False
2                False      False           False          False
3                False      False           False          False
4                False      False           False          False
…                …        …             …            …
1465             False      False           False          False
1466             False      False           False          False
1467             False      False           False          False
1468             False      False           False          False
1469             False      False           False          False

      EmployeeNumber  EnvironmentSatisfaction  Gender  HourlyRate  \
0              False                    False   False       False
1              False                    False   False       False
2              False                    False   False       False
3              False                    False   False       False
4              False                    False   False       False
…              …                      …     …         …
1465           False                    False   False       False
1466           False                    False   False       False
1467           False                    False   False       False
1468           False                    False   False       False
1469           False                    False   False       False

      JobInvolvement  JobLevel  JobRole  JobSatisfaction  MaritalStatus  \
0              False     False    False            False          False
1              False     False    False            False          False
2              False     False    False            False          False
3              False     False    False            False          False
4              False     False    False            False          False
…              …       …      …              …            …
1465           False     False    False            False          False
1466           False     False    False            False          False
1467           False     False    False            False          False
1468           False     False    False            False          False
```

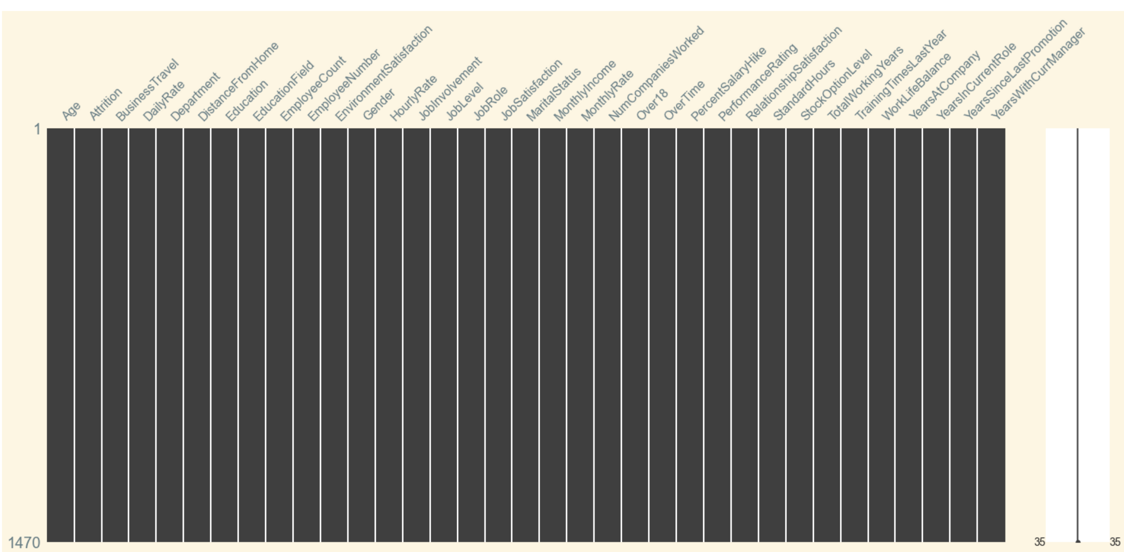
```
1469          False     False    False           False          False

      MonthlyIncome  MonthlyRate  NumCompaniesWorked  Over18  OverTime  \
0             False        False               False   False     False
1             False        False               False   False     False
2             False        False               False   False     False
3             False        False               False   False     False
4             False        False               False   False     False
…               …            …                   …       …         …
1465          False        False               False   False     False
1466          False        False               False   False     False
1467          False        False               False   False     False
1468          False        False               False   False     False
1469          False        False               False   False     False

      PercentSalaryHike  PerformanceRating  RelationshipSatisfaction  \
0                 False              False                     False
1                 False              False                     False
2                 False              False                     False
3                 False              False                     False
4                 False              False                     False
…                   …                  …                         …
1465              False              False                     False
1466              False              False                     False
1467              False              False                     False
1468              False              False                     False
1469              False              False                     False

      StandardHours  StockOptionLevel  TotalWorkingYears  \
0             False             False              False
1             False             False              False
2             False             False              False
3             False             False              False
4             False             False              False
…               …                 …                  …
1465          False             False              False
1466          False             False              False
1467          False             False              False
1468          False             False              False
1469          False             False              False

      TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany  \
0                     False            False           False
1                     False            False           False
2                     False            False           False
3                     False            False           False
4                     False            False           False
```

```
…                         …                 …                …
1465                      False             False            False
1466                      False             False            False
1467                      False             False            False
1468                      False             False            False
1469                      False             False            False

      YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  False                    False                 False
1                  False                    False                 False
2                  False                    False                 False
3                  False                    False                 False
4                  False                    False                 False
…                  …                        …                     …
1465               False                    False                 False
1466               False                    False                 False
1467               False                    False                 False
1468               False                    False                 False
1469               False                    False                 False

[1470 rows x 35 columns]
```

9.2 8.2)Visualizing Data Insights

[10]:
```python
# Using Missining number matrix to visualize
msno.matrix(employee_data_set)
```

[10]:
```
<AxesSubplot:>
```



8.3) Here we are Checking for the duplicated column if it is there any

[11]:
```
employee_data_set[employee_data_set.duplicated(keep='first')].shape
```

[11]:
```
(0, 35)
```

[12]:
```
employee_data_set.describe()
```

[12]:
```
           Age  DailyRate  DistanceFromHome  Education  EmployeeCount  \
count  1470.00    1470.00           1470.00    1470.00        1470.00
mean     36.92     802.49              9.19       2.91           1.00
std       9.14     403.51              8.11       1.02           0.00
min      18.00     102.00              1.00       1.00           1.00
25%      30.00     465.00              2.00       2.00           1.00
50%      36.00     802.00              7.00       3.00           1.00
75%      43.00    1157.00             14.00       4.00           1.00
max      60.00    1499.00             29.00       5.00           1.00

       EmployeeNumber  EnvironmentSatisfaction  HourlyRate  JobInvolvement  \
count         1470.00                  1470.00     1470.00         1470.00
mean          1024.87                     2.72       65.89            2.73
std            602.02                     1.09       20.33            0.71
min              1.00                     1.00       30.00            1.00
25%            491.25                     2.00       48.00            2.00
50%           1020.50                     3.00       66.00            3.00
75%           1555.75                     4.00       83.75            3.00
max           2068.00                     4.00      100.00            4.00

       JobLevel  JobSatisfaction  MonthlyIncome  MonthlyRate  \
count   1470.00          1470.00        1470.00      1470.00
mean       2.06             2.73        6502.93     14313.10
std        1.11             1.10        4707.96      7117.79
min        1.00             1.00        1009.00      2094.00
25%        1.00             2.00        2911.00      8047.00
50%        2.00             3.00        4919.00     14235.50
75%        3.00             4.00        8379.00     20461.50
max        5.00             4.00       19999.00     26999.00

       NumCompaniesWorked  PercentSalaryHike  PerformanceRating  \
count             1470.00            1470.00            1470.00
mean                 2.69              15.21               3.15
std                  2.50               3.66               0.36
min                  0.00              11.00               3.00
25%                  1.00              12.00               3.00
50%                  2.00              14.00               3.00
75%                  4.00              18.00               3.00
max                  9.00              25.00               4.00
```

	RelationshipSatisfaction	StandardHours	StockOptionLevel	\
count	1470.00	1470.00	1470.00	
mean	2.71	80.00	0.79	
std	1.08	0.00	0.85	
min	1.00	80.00	0.00	
25%	2.00	80.00	0.00	
50%	3.00	80.00	1.00	
75%	4.00	80.00	1.00	
max	4.00	80.00	3.00	

	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	\
count	1470.00	1470.00	1470.00	
mean	11.28	2.80	2.76	
std	7.78	1.29	0.71	
min	0.00	0.00	1.00	
25%	6.00	2.00	2.00	
50%	10.00	3.00	3.00	
75%	15.00	3.00	3.00	
max	40.00	6.00	4.00	

	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	\
count	1470.00	1470.00	1470.00	
mean	7.01	4.23	2.19	
std	6.13	3.62	3.22	
min	0.00	0.00	0.00	
25%	3.00	2.00	0.00	
50%	5.00	3.00	1.00	
75%	9.00	7.00	3.00	
max	40.00	18.00	15.00	

	YearsWithCurrManager
count	1470.00
mean	4.12
std	3.57
min	0.00
25%	2.00
50%	3.00
75%	7.00
max	17.00

9.3 8.4) Checking for Unique values in each column

[13]:
```
for column in employee_data_set.columns:
    print(f"{column}:Unique values in No = {employee_data_set[column].
↪nunique()}")
    print("*************************************************")
```

```
Age:Unique values in No = 43
*************************************************
Attrition:Unique values in No = 2
*************************************************
BusinessTravel:Unique values in No = 3
*************************************************
DailyRate:Unique values in No = 886
*************************************************
Department:Unique values in No = 3
*************************************************
DistanceFromHome:Unique values in No = 29
*************************************************
Education:Unique values in No = 5
*************************************************
EducationField:Unique values in No = 6
*************************************************
EmployeeCount:Unique values in No = 1
*************************************************
EmployeeNumber:Unique values in No = 1470
*************************************************
EnvironmentSatisfaction:Unique values in No = 4
*************************************************
Gender:Unique values in No = 2
*************************************************
HourlyRate:Unique values in No = 71
*************************************************
JobInvolvement:Unique values in No = 4
*************************************************
JobLevel:Unique values in No = 5
*************************************************
JobRole:Unique values in No = 9
*************************************************
JobSatisfaction:Unique values in No = 4
*************************************************
MaritalStatus:Unique values in No = 3
*************************************************
MonthlyIncome:Unique values in No = 1349
*************************************************
MonthlyRate:Unique values in No = 1427
*************************************************
NumCompaniesWorked:Unique values in No = 10
*************************************************
```

```
Over18:Unique values in No = 1
**************************************************
OverTime:Unique values in No = 2
**************************************************
PercentSalaryHike:Unique values in No = 15
**************************************************
PerformanceRating:Unique values in No = 2
**************************************************
RelationshipSatisfaction:Unique values in No = 4
**************************************************
StandardHours:Unique values in No = 1
**************************************************
StockOptionLevel:Unique values in No = 4
**************************************************
TotalWorkingYears:Unique values in No = 40
**************************************************
TrainingTimesLastYear:Unique values in No = 7
**************************************************
WorkLifeBalance:Unique values in No = 4
**************************************************
YearsAtCompany:Unique values in No = 37
**************************************************
YearsInCurrentRole:Unique values in No = 19
**************************************************
YearsSinceLastPromotion:Unique values in No = 16
**************************************************
YearsWithCurrManager:Unique values in No = 18
**************************************************
```

[14]:
```
employee_data_set['Attrition'].value_counts()
```

[14]:
```
No     1233
Yes     237
Name: Attrition, dtype: int64
```

[15]:
```
employee_data_set['OverTime'].value_counts()
```

[15]:
```
No     1054
Yes     416
Name: OverTime, dtype: int64
```

[16]:
```
employee_data_set['Over18'].value_counts()
```

[16]:
```
Y    1470
Name: Over18, dtype: int64
```

10 9) Data Cleansing

[17]:
```
employee_data_set['Attrition'] = employee_data_set['Attrition'].apply(lambda x:␣
↪1 if x == 'Yes' else 0)
employee_data_set['OverTime'] = employee_data_set['OverTime'].apply(lambda x: 1␣
↪if x == 'Yes' else 0)
employee_data_set['Over18'] = employee_data_set['Over18'].apply(lambda x: 1 if␣
↪x == 'Y' else 0)
```

Why we are doing this, - So here we are converting "Attrition" column into boolean which means that actually the "Attrition" column contain "Yes" for those employees who quit and "No" for those who stayed at the company, we have used lambda funtion to keep the code concise and converted yes and no into "1" and "0" - Same we have done for the "overtime" column - The same methodology is applied for "Over18" column

[18]:
```
employee_data_set.head()
```

[18]:
```
   Age  Attrition     BusinessTravel  DailyRate              Department  \
0   41          1      Travel_Rarely       1102                   Sales
1   49          0  Travel_Frequently        279  Research & Development
2   37          1      Travel_Rarely       1373  Research & Development
3   33          0  Travel_Frequently       1392  Research & Development
4   27          0      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   EnvironmentSatisfaction  Gender  HourlyRate  JobInvolvement  JobLevel  \
0                        2  Female          94               3         2
1                        3    Male          61               2         2
2                        4    Male          92               2         1
3                        4  Female          56               3         1
4                        1    Male          40               3         1

                 JobRole  JobSatisfaction MaritalStatus  MonthlyIncome  \
0        Sales Executive                4        Single           5993
1     Research Scientist                2       Married           5130
2  Laboratory Technician                3        Single           2090
3     Research Scientist                3       Married           2909
4  Laboratory Technician                2       Married           3468

   MonthlyRate  NumCompaniesWorked  Over18  OverTime  PercentSalaryHike  \
0        19479                   8       1         1                 11
```

```
1        24907                     1       1         0                 23
2         2396                     6       1         1                 15
3        23159                     1       1         1                 11
4        16632                     9       1         0                 12

   PerformanceRating  RelationshipSatisfaction  StandardHours  \
0                  3                         1             80
1                  4                         4             80
2                  3                         2             80
3                  3                         3             80
4                  3                         4             80

   StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                 0                  8                      0
1                 1                 10                      3
2                 0                  7                      3
3                 0                  8                      3
4                 1                  6                      3

   WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
0                1               6                   4
1                3              10                   7
2                3               0                   0
3                3               8                   7
4                3               2                   2

   YearsSinceLastPromotion  YearsWithCurrManager
0                        0                     5
1                        1                     7
2                        0                     0
3                        3                     0
4                        2                     2
```

11 10) Plotting Histogram

[19]:
```
# Here we are plotting the histogram of each feature in the given dataset to␣
 ↪visualize its distribution

employee_data_set.hist(bins = 30, figsize = (20,20), color = 'orange');
```

12 11) Data Findings on intial analysis

12.0.1 What we have Analyze so far

- Our a large portion of the Employees lies between 27 to 40 age bunch.
- A large portion of our Employees is located(lives close by) from workplace.
- The Majority of our workers have level 3 of Education.
- The vast majority of the workers are the part of our organization for under 10 years

So as we probably aware now that the data columns are not exactly helpful for us i.e “EmployeeCount” , “Standardhours”, “Over18” in light of the fact that it will be something very similar starting with one worker then onto the next and assuming we talk about “Over18” so it’s definitley be “Yes” for all since all of the Employee will be more than 18, Moreover we don’t require the “EmployeeNumber” so lets dispose of them.

13 12) Dropping the Columns we don't need

[20]:
```
employee_data_set.drop(['EmployeeCount', 'StandardHours', 'Over18',␣
↪'EmployeeNumber'], axis=1, inplace=True)
```

[21]:
```
employee_data_set.head()
```

[21]:
```
   Age  Attrition     BusinessTravel  DailyRate              Department  \
0   41          1      Travel_Rarely       1102                   Sales
1   49          0  Travel_Frequently        279  Research & Development
2   37          1      Travel_Rarely       1373  Research & Development
3   33          0  Travel_Frequently       1392  Research & Development
4   27          0      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EnvironmentSatisfaction  \
0                 1          2  Life Sciences                        2
1                 8          1  Life Sciences                        3
2                 2          2          Other                        4
3                 3          4  Life Sciences                        4
4                 2          1        Medical                        1

   Gender  HourlyRate  JobInvolvement  JobLevel                JobRole  \
0  Female          94               3         2        Sales Executive
1    Male          61               2         2     Research Scientist
2    Male          92               2         1  Laboratory Technician
3  Female          56               3         1     Research Scientist
4    Male          40               3         1  Laboratory Technician

   JobSatisfaction MaritalStatus  MonthlyIncome  MonthlyRate  \
0                4        Single           5993        19479
1                2       Married           5130        24907
2                3        Single           2090         2396
3                3       Married           2909        23159
4                2       Married           3468        16632

   NumCompaniesWorked  OverTime  PercentSalaryHike  PerformanceRating  \
0                   8         1                 11                  3
1                   1         0                 23                  4
2                   6         1                 15                  3
3                   1         1                 11                  3
4                   9         0                 12                  3

   RelationshipSatisfaction  StockOptionLevel  TotalWorkingYears  \
0                         1                 0                  8
1                         4                 1                 10
2                         2                 0                  7
3                         3                 0                  8
```

```
4                          4                  1                   6

   TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
0                      0                1               6                   4
1                      3                3              10                   7
2                      3                3               0                   0
3                      3                3               8                   7
4                      3                3               2                   2

   YearsSinceLastPromotion  YearsWithCurrManager
0                        0                     5
1                        1                     7
2                        0                     0
3                        3                     0
4                        2                     2
```

- So now we are left with 31 columns

14 13) Statistics and Insight in detail

[22]:
```
# So now it's time to check how many employees stayed and how many left the␣
↪company
Employee_left_df = employee_data_set[employee_data_set['Attrition'] == 1]␣
↪#Checking for the employee who left
Employee_stayed_df = employee_data_set[employee_data_set['Attrition'] == 0]␣
↪#Checking for those who stayed


#Lets have a overview and count the statistic we got.
print("Total Employees of company=", len(employee_data_set))

print("Employees who left:", len(Employee_left_df))
print(f"Employees who left in Percentage: {1.*len(Employee_left_df)/
↪len(employee_data_set)*100.0:.2f}%")
print("Employees who did not leave or (stayed) =", len(Employee_stayed_df))
print(f"Employees who did not leave (stayed) : {1.*len(Employee_stayed_df)/
↪len(employee_data_set)*100.0:.2f}%")
```

```
Total Employees of company= 1470
Employees who left: 237
Employees who left in Percentage: 16.12%
Employees who did not leave or (stayed) = 1233
Employees who did not leave (stayed) : 83.88%
```

14.1 13.1) Lets see what the statistics tell us and compare them

14.1.1 13.1.1) Employees who left the company

[23]:
```
Employee_left_df.describe()
```

[23]:
```
          Age  Attrition  DailyRate  DistanceFromHome  Education  \
count  237.00     237.00     237.00            237.00     237.00
mean    33.61       1.00     750.36             10.63       2.84
std      9.69       0.00     401.90              8.45       1.01
min     18.00       1.00     103.00              1.00       1.00
25%     28.00       1.00     408.00              3.00       2.00
50%     32.00       1.00     699.00              9.00       3.00
75%     39.00       1.00    1092.00             17.00       4.00
max     58.00       1.00    1496.00             29.00       5.00

       EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
count                   237.00      237.00          237.00    237.00
mean                      2.46       65.57            2.52      1.64
std                       1.17       20.10            0.77      0.94
min                       1.00       31.00            1.00      1.00
25%                       1.00       50.00            2.00      1.00
50%                       3.00       66.00            3.00      1.00
75%                       4.00       84.00            3.00      2.00
max                       4.00      100.00            4.00      5.00

       JobSatisfaction  MonthlyIncome  MonthlyRate  NumCompaniesWorked  \
count           237.00         237.00       237.00              237.00
mean              2.47        4787.09     14559.31                2.94
std               1.12        3640.21      7208.15                2.68
min               1.00        1009.00      2326.00                0.00
25%               1.00        2373.00      8870.00                1.00
50%               3.00        3202.00     14618.00                1.00
75%               3.00        5916.00     21081.00                5.00
max               4.00       19859.00     26999.00                9.00

       OverTime  PercentSalaryHike  PerformanceRating  \
count    237.00             237.00             237.00
mean       0.54              15.10               3.16
std        0.50               3.77               0.36
min        0.00              11.00               3.00
25%        0.00              12.00               3.00
50%        1.00              14.00               3.00
75%        1.00              17.00               3.00
max        1.00              25.00               4.00

       RelationshipSatisfaction  StockOptionLevel  TotalWorkingYears  \
count                    237.00            237.00             237.00
```

mean	2.60	0.53	8.24
std	1.13	0.86	7.17
min	1.00	0.00	0.00
25%	2.00	0.00	3.00
50%	3.00	0.00	7.00
75%	4.00	1.00	10.00
max	4.00	3.00	40.00

	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	\
count	237.00	237.00	237.00	
mean	2.62	2.66	5.13	
std	1.25	0.82	5.95	
min	0.00	1.00	0.00	
25%	2.00	2.00	1.00	
50%	2.00	3.00	3.00	
75%	3.00	3.00	7.00	
max	6.00	4.00	40.00	

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
count	237.00	237.00	237.00
mean	2.90	1.95	2.85
std	3.17	3.15	3.14
min	0.00	0.00	0.00
25%	0.00	0.00	0.00
50%	2.00	1.00	2.00
75%	4.00	2.00	5.00
max	15.00	15.00	14.00

14.1.2 13.1.2) Employee who stayed with the company

[24]:
```
Employee_stayed_df.describe()
```

[24]:

	Age	Attrition	DailyRate	DistanceFromHome	Education	\
count	1233.00	1233.00	1233.00	1233.00	1233.00	
mean	37.56	0.00	812.50	8.92	2.93	
std	8.89	0.00	403.21	8.01	1.03	
min	18.00	0.00	102.00	1.00	1.00	
25%	31.00	0.00	477.00	2.00	2.00	
50%	36.00	0.00	817.00	7.00	3.00	
75%	43.00	0.00	1176.00	13.00	4.00	
max	60.00	0.00	1499.00	29.00	5.00	

	EnvironmentSatisfaction	HourlyRate	JobInvolvement	JobLevel	\
count	1233.00	1233.00	1233.00	1233.00	
mean	2.77	65.95	2.77	2.15	
std	1.07	20.38	0.69	1.12	
min	1.00	30.00	1.00	1.00	

25%	2.00	48.00	2.00	1.00
50%	3.00	66.00	3.00	2.00
75%	4.00	83.00	3.00	3.00
max	4.00	100.00	4.00	5.00

	JobSatisfaction	MonthlyIncome	MonthlyRate	NumCompaniesWorked \
count	1233.00	1233.00	1233.00	1233.00
mean	2.78	6832.74	14265.78	2.65
std	1.09	4818.21	7102.26	2.46
min	1.00	1051.00	2094.00	0.00
25%	2.00	3211.00	7973.00	1.00
50%	3.00	5204.00	14120.00	2.00
75%	4.00	8834.00	20364.00	4.00
max	4.00	19999.00	26997.00	9.00

	OverTime	PercentSalaryHike	PerformanceRating \
count	1233.00	1233.00	1233.00
mean	0.23	15.23	3.15
std	0.42	3.64	0.36
min	0.00	11.00	3.00
25%	0.00	12.00	3.00
50%	0.00	14.00	3.00
75%	0.00	18.00	3.00
max	1.00	25.00	4.00

	RelationshipSatisfaction	StockOptionLevel	TotalWorkingYears \
count	1233.00	1233.00	1233.00
mean	2.73	0.85	11.86
std	1.07	0.84	7.76
min	1.00	0.00	0.00
25%	2.00	0.00	6.00
50%	3.00	1.00	10.00
75%	4.00	1.00	16.00
max	4.00	3.00	38.00

	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany \
count	1233.00	1233.00	1233.00
mean	2.83	2.78	7.37
std	1.29	0.68	6.10
min	0.00	1.00	0.00
25%	2.00	2.00	3.00
50%	3.00	3.00	6.00
75%	3.00	3.00	10.00
max	6.00	4.00	37.00

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
count	1233.00	1233.00	1233.00

mean	4.48	2.23	4.37
std	3.65	3.23	3.59
min	0.00	0.00	0.00
25%	2.00	0.00	2.00
50%	3.00	1.00	3.00
75%	7.00	3.00	7.00
max	18.00	15.00	17.00

14.2 13.2) Answer to the Business Questions.....

14.2.1 Looking at the insights we have investigated up until this point we are contrasting std and mean of the workers who remained and left the organization

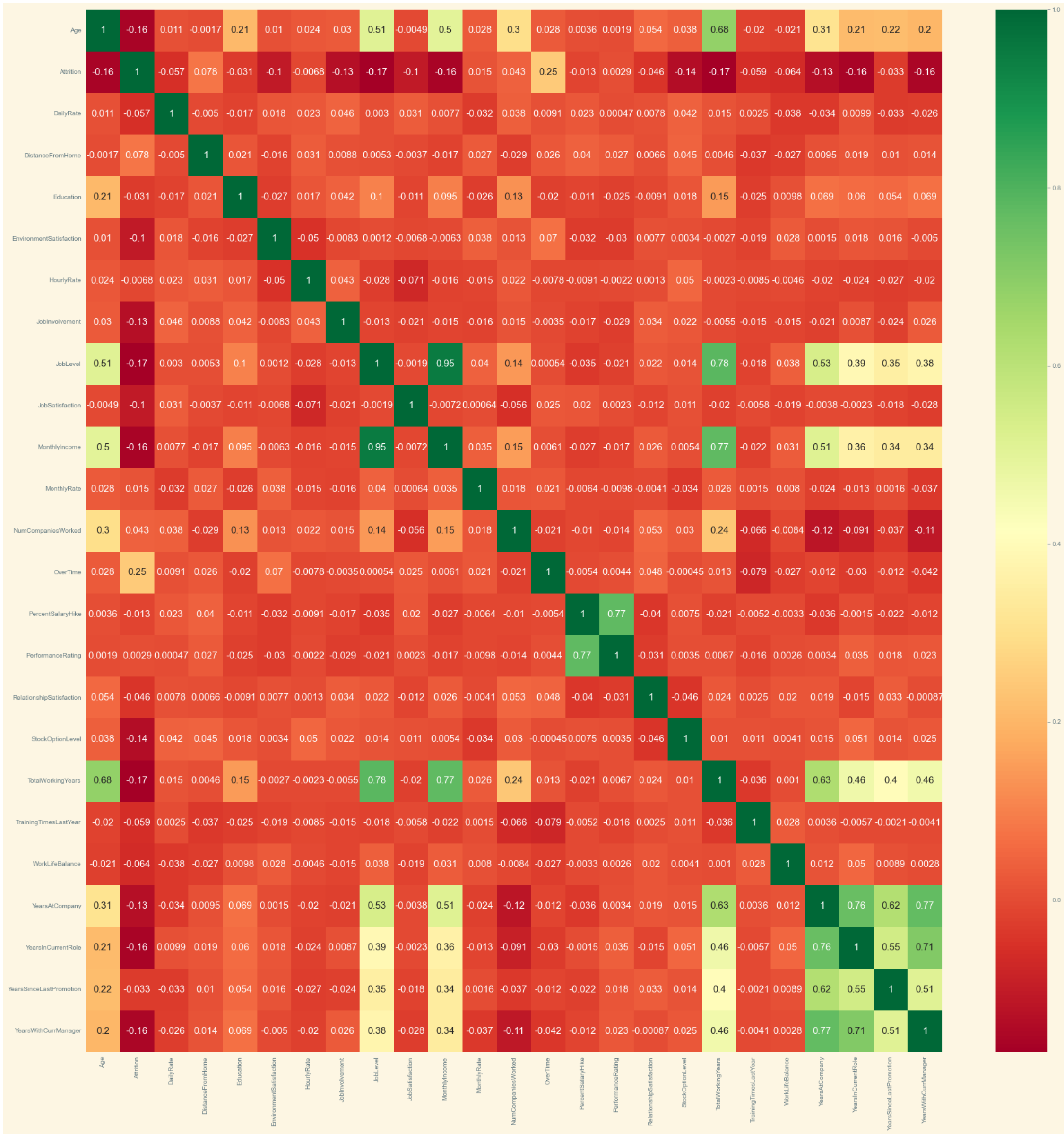
1. Age: The Insights state that the mean age of the employees who remained with the company is higher than who left. (37.5 x 33.6)
2. DailyRate: Rate of employees who remained with the organization is higher as compare to those who left. (812 x 750)
3. DistanceFromHome: Those worker who resides nearer to work place remained. (8.9km x 10.6km)
4. EnvironmentSatisfaction and JobSatisfaction: The Insights tell us that those employees who remained with our organization are more staisfied in general with their job.
5. StockOptionLevel: We have noticed that those employee who stayed have more higher stock level.

15 14) Plotting the heatmap to find correlations

Having the insight of those features which are corelated and find out the outliers and get rid of them.

[25]:
```
# Lets have the insight of those features which are corelated and find out the␣
↪outliers and get rid of them.

plt.figure(figsize= (30, 30))
sns.heatmap(employee_data_set.corr(), annot= True, cmap= "RdYlGn", annot_kws=␣
↪{"size":15})
```

[25]:
```
<AxesSubplot:>
```



16 14.1) Analysis of Heatmap

From the Correlation table we have discovered that month to month pay(Monthly Income) is profoundly connected with job level. Be that as it may, everyday rate, hourly rate and month to month rate are scarcely connected with anything. We will involve month to month pay(monthlyIncome) in later examination as an estimation of pay.

A few different ends we get from the correlation table:

1. Job level and complete working years are profoundly related.
2. Monthly Income and all out working years are exceptionally related.

3. Performance rating and percentage salary hike are highly correlated.
4. Years in current job and years at organization are profoundly connected.
5. Years with current Manager and years at organization are profoundly correlated.

17 15) Feature Selection and Target Mapping(Addressing the findings)

17.0.1 Finding some more correlations on some specific variables by plotting the feature against the "Target" variable

The feature Selection is one of the most important and cruicial steps of the preprocessing phase as the features which we choose directly effects the model performance. While some of the features seem to be less useful in terms of the context; others seem to equally useful. The better features we use the better our model will perform.

17.1 15.1) Age vs Atttrition

[26]:
```
# lets figure out the correlation between's kin who left the organization for␣
↪certain particular factors, for example, 'Age', 'JobRole', 'MaritalStatus',␣
↪'JobInvolvement' and 'JobLevel'

employee_data_set.hvplot.hist(y= 'Age', by= 'Attrition', subplots= False,␣
↪width= 700, height= 450, bins= 30)
```

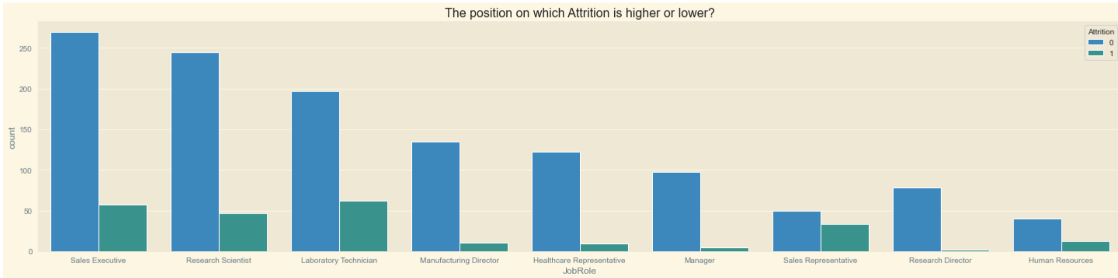
[26]:
```
:NdOverlay   [Attrition]
   :Histogram   [Age]   (Age_count)
```

Here we are utilizing blue tone to address the employees who remained with the organization, and involving orange for the people who left the organization.

- The Majority of the workers who left the organization is concentrated contrasted with the people who remained and they depend on age 31.
- Between 18 to 21 years old are concentrated the majority of the employees that leave proportionally the amount that remains.
- After the 31's, as age increases, there is a lessening in the quantity of workers who left the organization.

17.2 15.2) JobRole vs Attrition

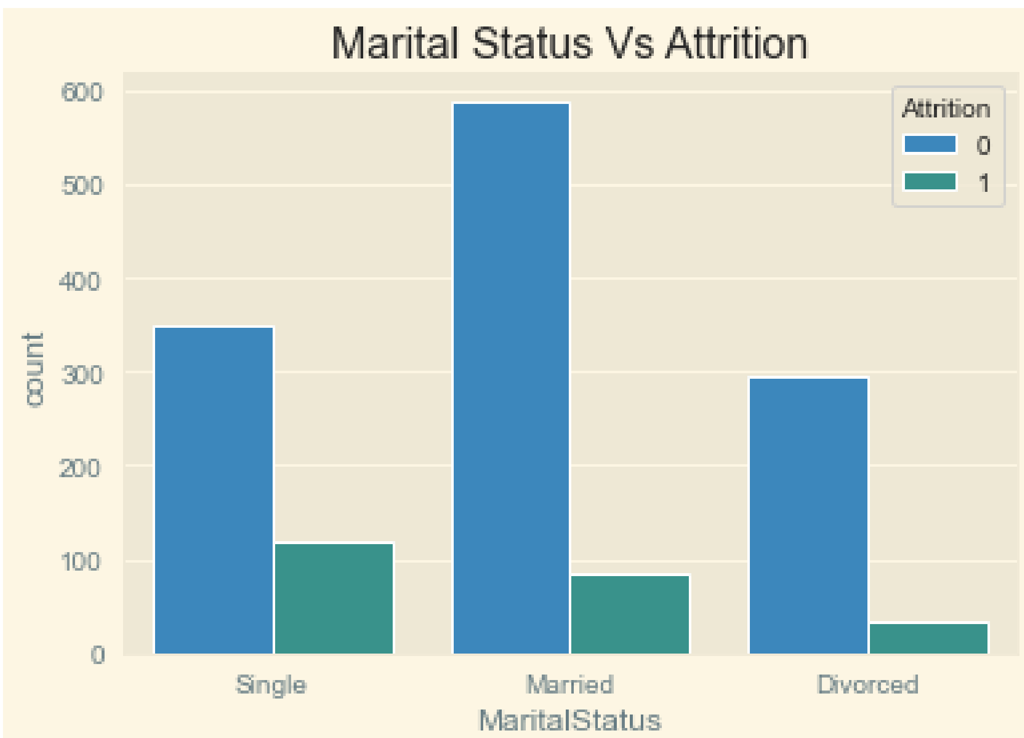
[27]:
```
plt.figure(figsize=[25,25])

plt.subplot(411)

sns.countplot(x = 'JobRole', hue = 'Attrition', data = employee_data_set)
plt.title("The position on which Attrition is higher or lower?");
```



Note (Blue is representing those who stayed and green is representing those who left) - We observed that almost half of the workers in Sale Representative team left the company and we have seen least attrition on the Research Director post.

17.3 15.3) Marital Status vs Attrition

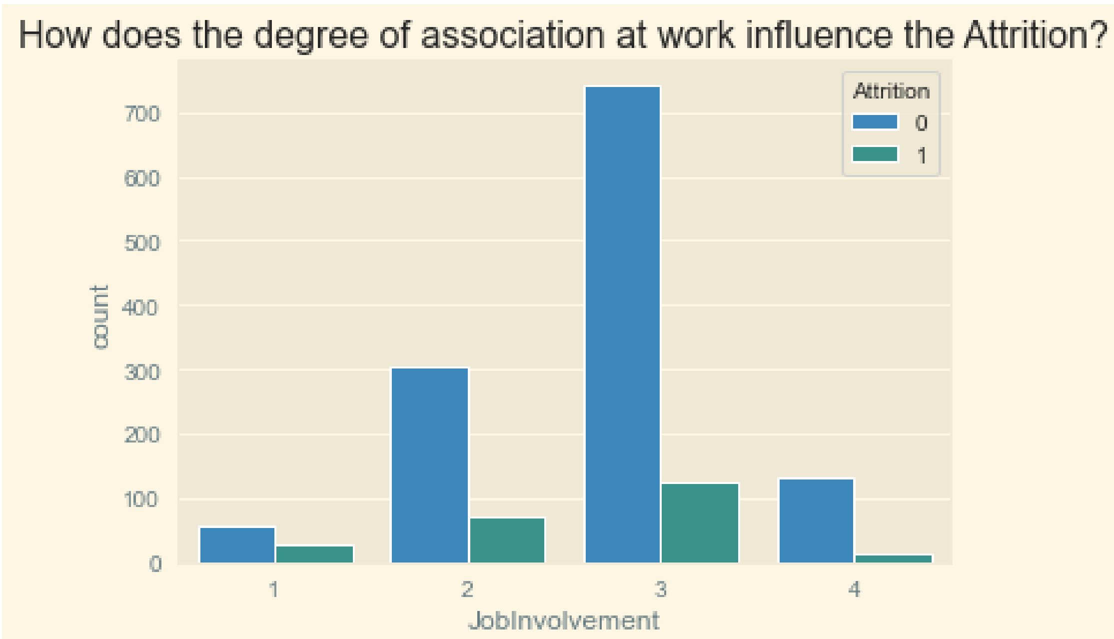
[28]:
```
sns.countplot(x = 'MaritalStatus', hue = 'Attrition', data = employee_data_set);
plt.title("Marital Status Vs Attrition");
```



Note (Blue is representing those who stayed and green is representing those who left) - Single Employees will in general leave the organization more than married ones and divorced ones.

17.4 15.4) JobInvolvement vs Attrition

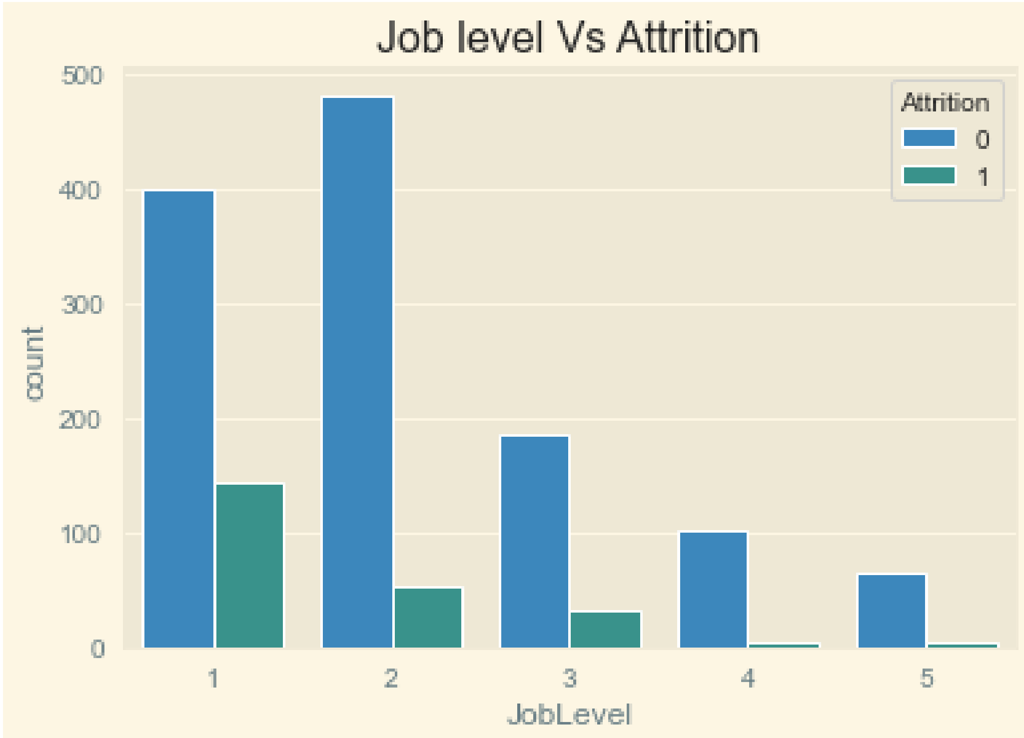
[29]:
```
sns.countplot(x = 'JobInvolvement', hue = 'Attrition', data =␣
  ↪employee_data_set);
plt.title("How does the degree of association at work influence the Attrition?
  ↪");
```



Note (Blue is representing those who stayed and green is representing those who left) - The less the employees are get invloved the more they tends to leave the job.

17.5 15.5) JobLevel vs Attrition

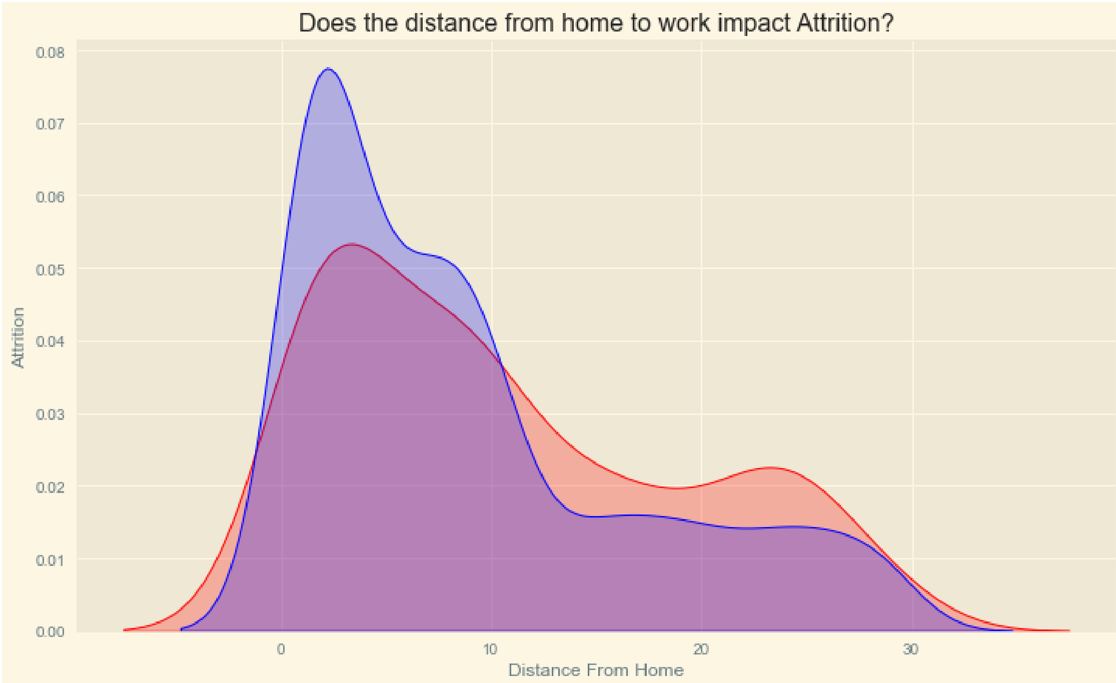
[30]:
```
sns.countplot(x = 'JobLevel', hue = 'Attrition', data = employee_data_set)
plt.title("Job level Vs Attrition");
```



Note (Blue is representing those who stayed and green is representing those who left) - The less the job level (less experienced) the more the employee leaves the company

17.6 15.6) Distance from Home vs Attrition

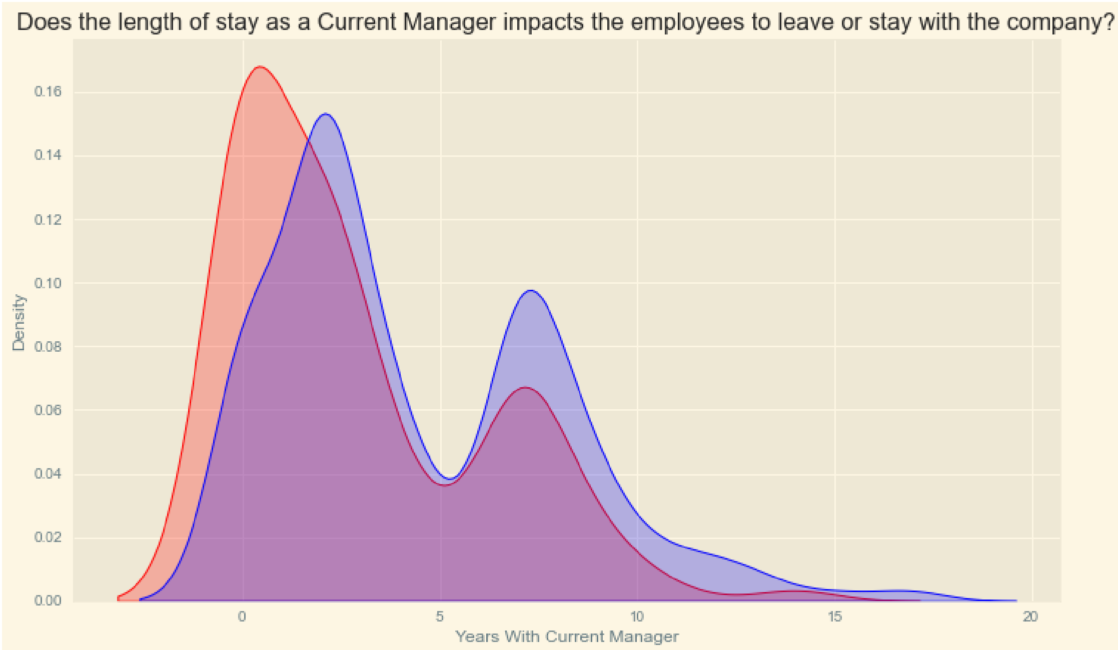
[31]:
```
# We will utilize KDE (Kernel Density Estimate) to envision the likelihood␣
↪density of a continuous variable.


plt.figure(figsize=(12,7))
sns.kdeplot(Employee_left_df['DistanceFromHome'], label = 'Employees who left',␣
↪shade = True, color = 'r')
sns.kdeplot(Employee_stayed_df['DistanceFromHome'], label = 'Employees who␣
↪Stayed', shade = True, color = 'b')
plt.xlabel('Distance From Home');
plt.ylabel('Attrition');
plt.title("Does the distance from home to work impact Attrition?");
```



- As the distance from home to workplace expands, the number of worker who tends to leave is higher.

17.7 15.7) Years with Current Manager vs Attrition

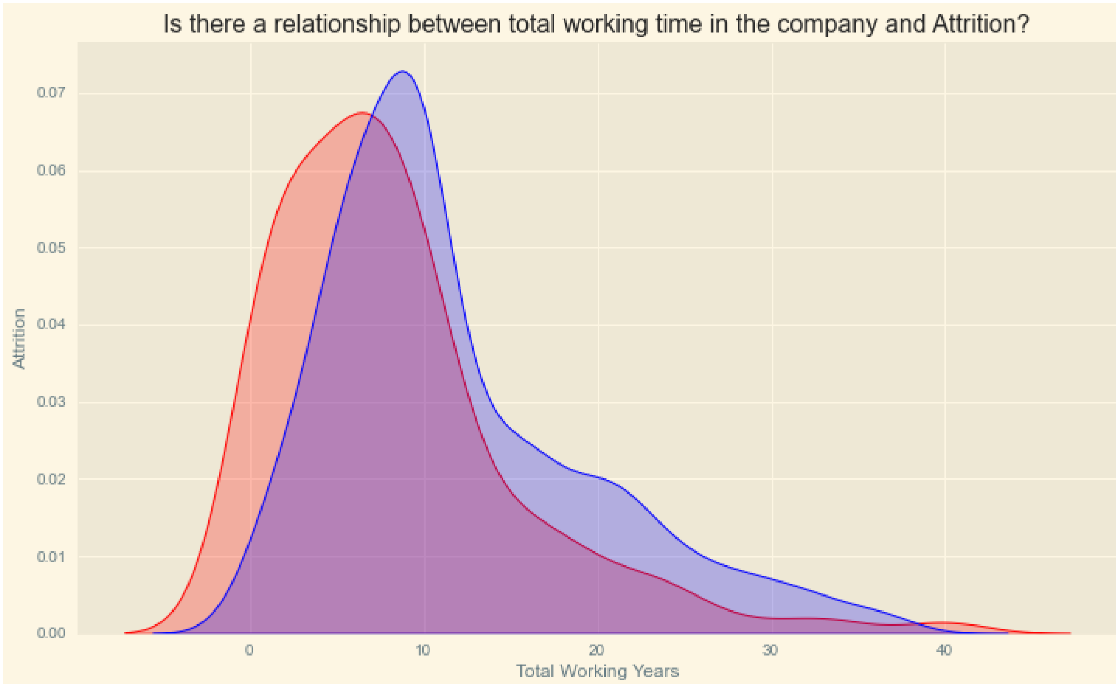
[32]:
```
plt.figure(figsize=(12,7))
sns.kdeplot(Employee_left_df['YearsWithCurrManager'], label = 'Employees who␣
↪left', shade = True, color = 'r')
sns.kdeplot(Employee_stayed_df['YearsWithCurrManager'], label = 'Employees who␣
↪Stayed', shade = True, color = 'b')
plt.xlabel('Years With Current Manager');
plt.title("Does the length of stay as a Current Manager impacts the employees␣
↪to leave or stay with the company?");
```



The more limited the time as a Current Manager, the more noteworthy the inclination for employees to leave.

17.8 15.8) Total working years vs Attrition

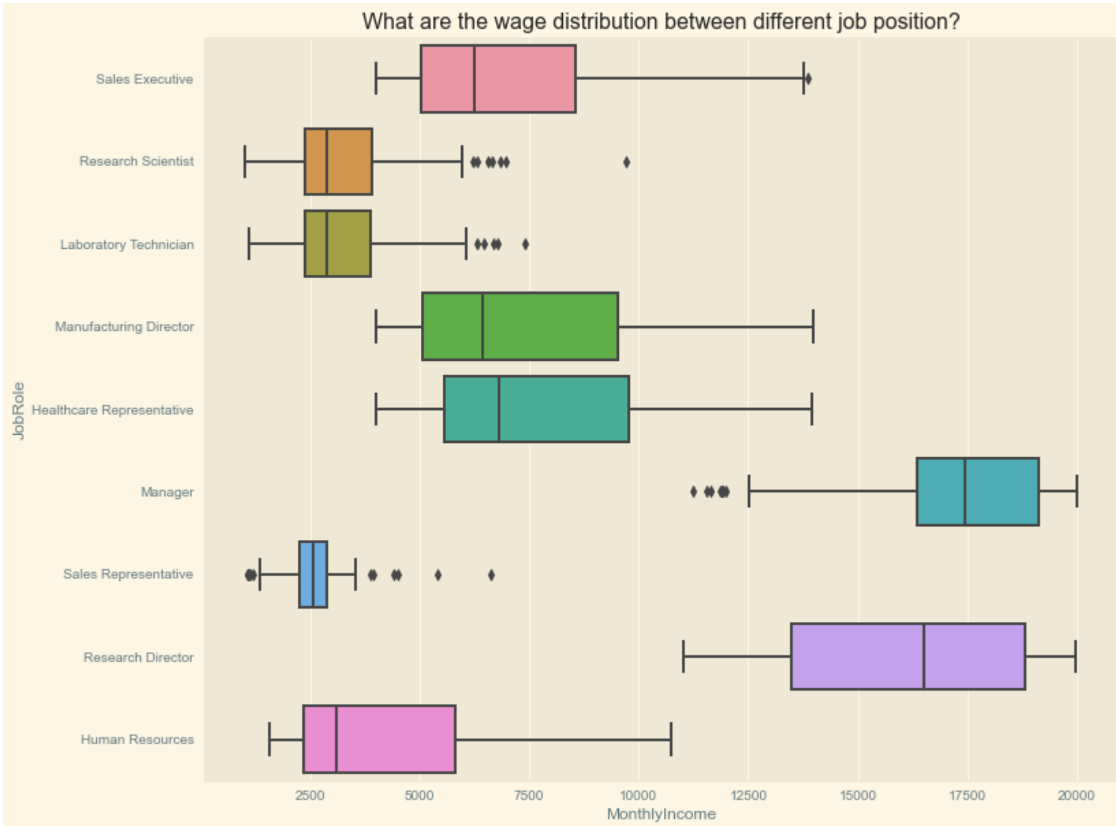
[33]:
```
plt.figure(figsize=(12,7))
sns.kdeplot(Employee_left_df['TotalWorkingYears'], shade = True, label =␣
 ↪'Employees who left', color = 'r')
sns.kdeplot(Employee_stayed_df['TotalWorkingYears'], shade = True, label =␣
 ↪'Employees who Stayed', color = 'b')
plt.xlabel('Total Working Years');
plt.ylabel('Attrition');
plt.title("Is there a relationship between total working time in the company␣
 ↪and Attrition?");
```



The basic time frame that Employees generally will quite often surrender is to around 7 years working at the organization. From that point they will generally remain.

17.9 15.9) Monthly Income vs JobRole

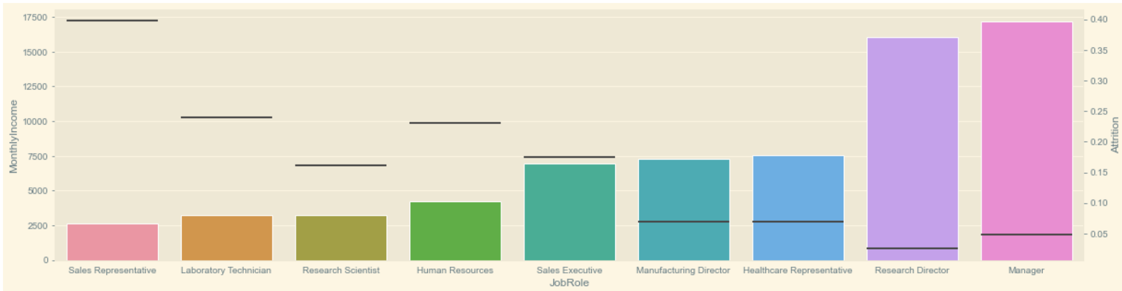
[34]:
```
plt.figure(figsize=(12, 10))
sns.boxplot(x = 'MonthlyIncome', y = 'JobRole', data = employee_data_set);
plt.title("What are the wage distribution between different job position?");
```



- Sales Executive are paid more than Research scientist and Laboratory Technician and Sales Representative are the least paid.
- While Research Director and Manager are best paid among all of the Employees.

17.10 15.10) Monthly Income vs JobRole vs Attrition

[35]:
```
role_income = employee_data_set.groupby('JobRole',␣
↪as_index=False)[['MonthlyIncome', 'Attrition']].mean().sort_values(
    by=['MonthlyIncome'])
role_income


fig, ax = plt.subplots(figsize=(20,5))
sns.barplot(x= 'JobRole', y= 'MonthlyIncome', data= role_income, ax= ax)
ax2 = ax.twinx()
sns.boxplot(x= 'JobRole', y= 'Attrition', data= role_income, ax= ax2)
ax2.grid(None)
plt.close(2)
plt.close(3)
```



Here the black bar is representing Attrition level and colored bars are showing monthly Income Increase The plot is showing a general pattern that as month to month pay increases, the steady loss rate goes down. In any case, we see the exemption for HR. Contrasted with other non-administrative jobs, human resource role is having higher month to month pay yet it's whittling down rate(Attrition) is shockingly higher than other jobs roles. To explore expected reasons, we check out at years since last advancement and occupation fulfillment.

18 Conclusion(Recommendations & actionable insights):

Employees are the backbone of every organization and if the company is not able to hold their employees it will be directly propotional to the loss in their business and integrity in the job market.Weakening is an issue that impacts all organizations, regardless of topography, industry and size of the organization. Employees leaving the company prompts huge expenses for a business, including the expense of business interruption, recruiting new staff and preparing new staff. All things considered, there is incredible financial matter in understanding the drivers of and limiting staff whittling down.So from all of this pipelining and data exploration, what we have explored and learned is that, The employees who left the company was from the Sales Representative department and the least Attrition was found on the role of Research director and Managers. Overall approximately 17% of the employees left the company and 83% employees stayed with the company.Well we have found that Distance from home to work place matters alot and the job involvements is the other factor that impacts the overall attrition rate.From all of the knowledge we gain now we are at the point where we can tell a story to the board of directors that what's going wrong and which department facing the most rate of Attrition.