

hr

March 19, 2025

```
[3]: import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
```

1 Loading Data

```
[4]: hr=pd.read_csv(r"C:\Mypythonfiles\Human_Resources.csv")
```

```
[5]: hr
```

```
[5]:      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
...  ...  ...  ...  ...  ...
1465   36         No  Travel_Frequently     884  Research & Development
1466   39         No    Travel_Rarely     613  Research & Development
1467   27         No    Travel_Rarely     155  Research & Development
1468   49         No  Travel_Frequently    1023      Sales
1469   34         No    Travel_Rarely     628  Research & Development
```

```
      DistanceFromHome  Education  EducationField  EmployeeCount \
0                      1          2  Life Sciences              1
1                      8          1  Life Sciences              1
2                      2          2          Other              1
3                      3          4  Life Sciences              1
4                      2          1          Medical              1
...  ...  ...  ...  ...
1465          23          2          Medical              1
1466           6          1          Medical              1
1467           4          3  Life Sciences              1
1468           2          3          Medical              1
1469           8          3          Medical              1
```

	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	\
0	1	...		1	80
1	2	...		4	80
2	4	...		2	80
3	5	...		3	80
4	7	...		4	80
...	
1465	2061	...		3	80
1466	2062	...		1	80
1467	2064	...		2	80
1468	2065	...		4	80
1469	2068	...		1	80

	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	\
0	0	8		0
1	1	10		3
2	0	7		3
3	0	8		3
4	1	6		3
...	
1465	1	17		3
1466	1	9		5
1467	1	6		0
1468	0	17		3
1469	0	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
...	
1465	3	5		2
1466	3	7		7
1467	3	6		2
1468	2	9		6
1469	4	4		3

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

1465	0	3
1466	1	7
1467	0	3
1468	0	8
1469	1	2

[1470 rows x 35 columns]

[6]: hr.head()

	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	

...	RelationshipSatisfaction	StandardHours	StockOptionLevel	\
0	...	1	80	0
1	...	4	80	1
2	...	2	80	0
3	...	3	80	0
4	...	4	80	1

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

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

[5 rows x 35 columns]

2 EDA Techniques

Checking the basic information

```
[7]: hr.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]: hr.isnull()
```

```
[8]:      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  ...  RelationshipSatisfaction  StandardHours  \
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

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
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

```

```

      WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
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

```

```

      YearsSinceLastPromotion  YearsWithCurrManager
0      False      False
1      False      False
2      False      False
3      False      False
4      False      False

```

```

...
1465      False      False
1466      False      False
1467      False      False
1468      False      False
1469      False      False

```

[1470 rows x 35 columns]

```
[10]: hr.dropna(inplace=True)
```

```
[11]: hr.isnull().sum()
```

```

[11]: Age      0
      Attrition      0
      BusinessTravel      0
      DailyRate      0
      Department      0
      DistanceFromHome      0
      Education      0
      EducationField      0

```

```

EmployeeCount      0
EmployeeNumber      0
EnvironmentSatisfaction  0
Gender              0
HourlyRate          0
JobInvolvement      0
JobLevel            0
JobRole             0
JobSatisfaction     0
MaritalStatus       0
MonthlyIncome       0
MonthlyRate         0
NumCompaniesWorked  0
Over18              0
OverTime            0
PercentSalaryHike   0
PerformanceRating   0
RelationshipSatisfaction  0
StandardHours       0
StockOptionLevel    0
TotalWorkingYears   0
TrainingTimesLastYear  0
WorkLifeBalance     0
YearsAtCompany      0
YearsInCurrentRole  0
YearsSinceLastPromotion  0
YearsWithCurrManager  0
dtype: int64

```

3 Summery Statistics

```
[12]: hr.describe(include="all")
```

```

[12]:
count      1470.000000      1470      1470      1470.000000
unique         NaN           2           3           NaN
top           NaN          No  Travel_Rarely           NaN
freq         NaN        1233        1043           NaN
mean        36.923810         NaN         NaN        802.485714
std          9.135373         NaN         NaN        403.509100
min         18.000000         NaN         NaN        102.000000
25%         30.000000         NaN         NaN        465.000000
50%         36.000000         NaN         NaN        802.000000
75%         43.000000         NaN         NaN       1157.000000
max         60.000000         NaN         NaN       1499.000000

```

	Department	DistanceFromHome	Education	EducationField	\
count	1470	1470.000000	1470.000000		1470
unique	3	NaN	NaN		6
top	Research & Development	NaN	NaN	Life Sciences	
freq	961	NaN	NaN		606
mean	NaN	9.192517	2.912925		NaN
std	NaN	8.106864	1.024165		NaN
min	NaN	1.000000	1.000000		NaN
25%	NaN	2.000000	2.000000		NaN
50%	NaN	7.000000	3.000000		NaN
75%	NaN	14.000000	4.000000		NaN
max	NaN	29.000000	5.000000		NaN

	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	\
count	1470.0	1470.000000	...	1470.000000	
unique	NaN	NaN	...	NaN	
top	NaN	NaN	...	NaN	
freq	NaN	NaN	...	NaN	
mean	1.0	1024.865306	...	2.712245	
std	0.0	602.024335	...	1.081209	
min	1.0	1.000000	...	1.000000	
25%	1.0	491.250000	...	2.000000	
50%	1.0	1020.500000	...	3.000000	
75%	1.0	1555.750000	...	4.000000	
max	1.0	2068.000000	...	4.000000	

	StandardHours	StockOptionLevel	TotalWorkingYears	\
count	1470.0	1470.000000	1470.000000	
unique	NaN	NaN	NaN	
top	NaN	NaN	NaN	
freq	NaN	NaN	NaN	
mean	80.0	0.793878	11.279592	
std	0.0	0.852077	7.780782	
min	80.0	0.000000	0.000000	
25%	80.0	0.000000	6.000000	
50%	80.0	1.000000	10.000000	
75%	80.0	1.000000	15.000000	
max	80.0	3.000000	40.000000	

	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	\
count	1470.000000	1470.000000	1470.000000	
unique	NaN	NaN	NaN	
top	NaN	NaN	NaN	
freq	NaN	NaN	NaN	
mean	2.799320	2.761224	7.008163	
std	1.289271	0.706476	6.126525	
min	0.000000	1.000000	0.000000	

25%	2.000000	2.000000	3.000000
50%	3.000000	3.000000	5.000000
75%	3.000000	3.000000	9.000000
max	6.000000	4.000000	40.000000

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
count	1470.000000	1470.000000	1470.000000
unique	NaN	NaN	NaN
top	NaN	NaN	NaN
freq	NaN	NaN	NaN
mean	4.229252	2.187755	4.123129
std	3.623137	3.222430	3.568136
min	0.000000	0.000000	0.000000
25%	2.000000	0.000000	2.000000
50%	3.000000	1.000000	3.000000
75%	7.000000	3.000000	7.000000
max	18.000000	15.000000	17.000000

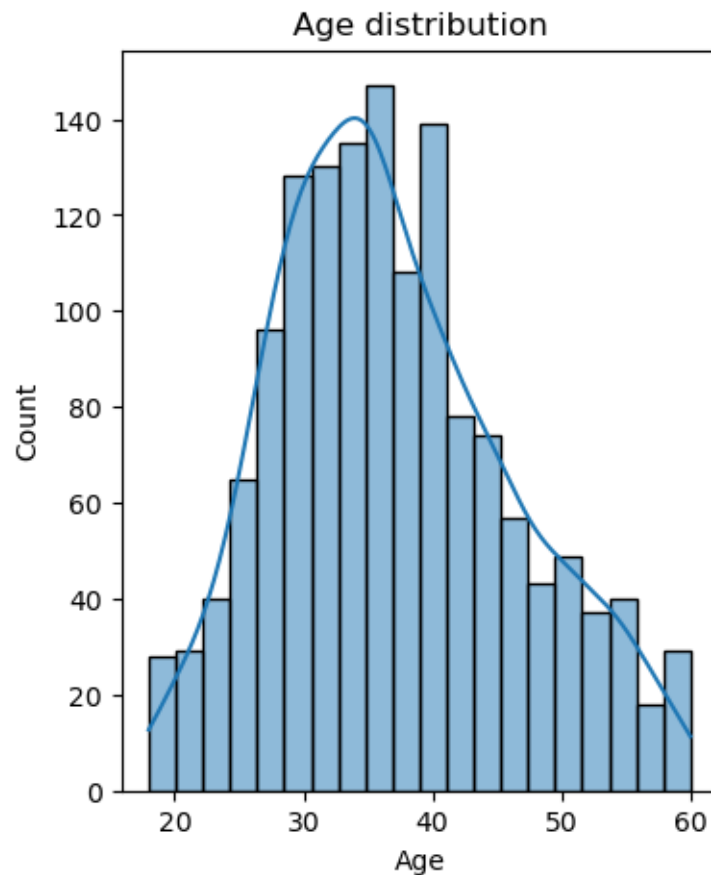
[11 rows x 35 columns]

Results Age: * the minimum age of employee is 18, maximum age is 60 and average age is 36.92 * the majority of age is fall between 30 and 43. Attrition: * the frequency of attrition of an employee is 1233. BusinessTravel: * The frequency of BusinessTravel of an employee is 1043. DailyRate: * the DailyRate minimum of employee is 102, maximum DailyRate is 1499 and average age is 802.4 . * the majority of DailyRate is fall between 465 and 1157. NaN-> non-numeric values are not applicable.

4 Visualization

5 Histogram

```
[13]: plt.figure(figsize=(4,5))
sns.histplot(hr["Age"],bins=20,kde=True)
plt.title("Age distribution")
plt.show()
```



- The range of the age is fall between 18 and 60.
- There is positive skew.
- no outlier is exist.
- The average age is fall between 30 and 40
- there are no employee above the age of 60.

6 Heatmap

```
[21]: hr1=hr.select_dtypes(include=["number"])
      hr1.head()
```

```
[21]:
```

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	\
0	41	1102	1	2	1	1	
1	49	279	8	1	1	2	
2	37	1373	2	2	1	4	
3	33	1392	3	4	1	5	
4	27	591	2	1	1	7	

	EnvironmentSatisfaction	HourlyRate	JobInvolvement	JobLevel	...	\
0	2	94	3	2	...	
1	3	61	2	2	...	
2	4	92	2	1	...	
3	4	56	3	1	...	
4	1	40	3	1	...	

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

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

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

[5 rows x 26 columns]

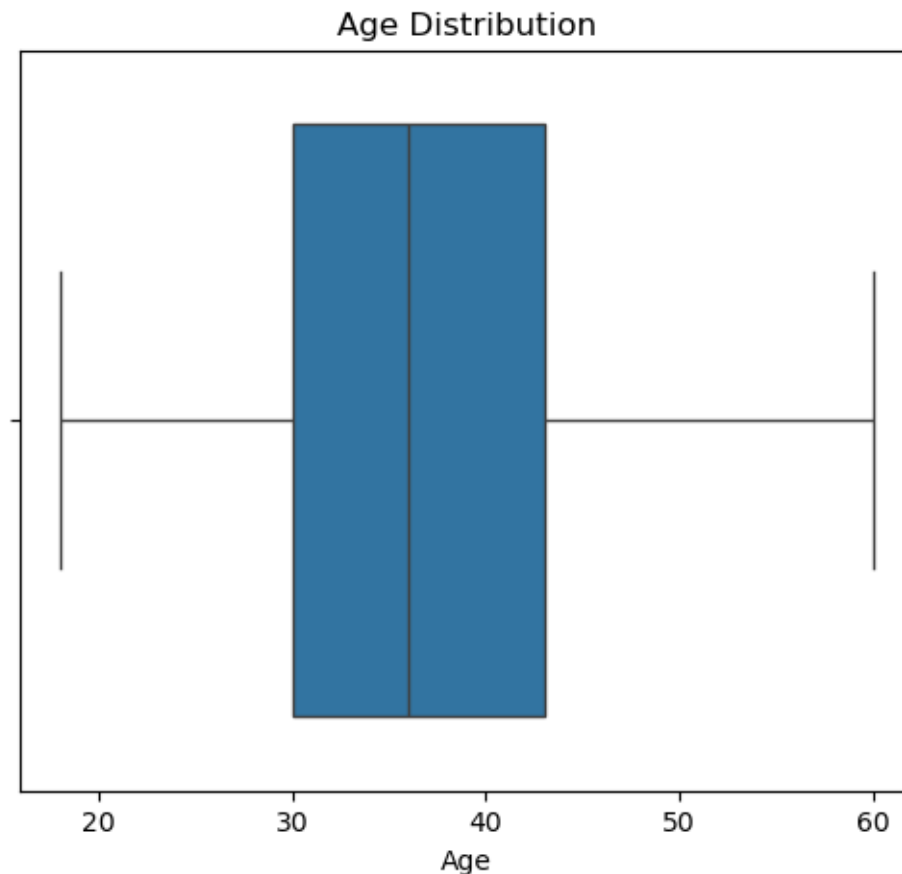
```
[22]: plt.figure(figsize=(10,20))
sns.heatmap(hr1.corr(),cmap="viridis",annot=True,linewidth=0.5)
plt.title("Correletion Matrix ")
plt.show()
```



Results: * The above graph shows the co-relation of an employee entries. * Age will affects the employee entries. * there is a good co-rekation between total working years and monthly income. * there is not co-relation between work life balance and years in current role.

7 Boxplot

```
[16]: plt.figure(figsize=(6,5))
sns.boxplot(x=hr["Age"])
plt.title("Age Distribution")
plt.show()
```

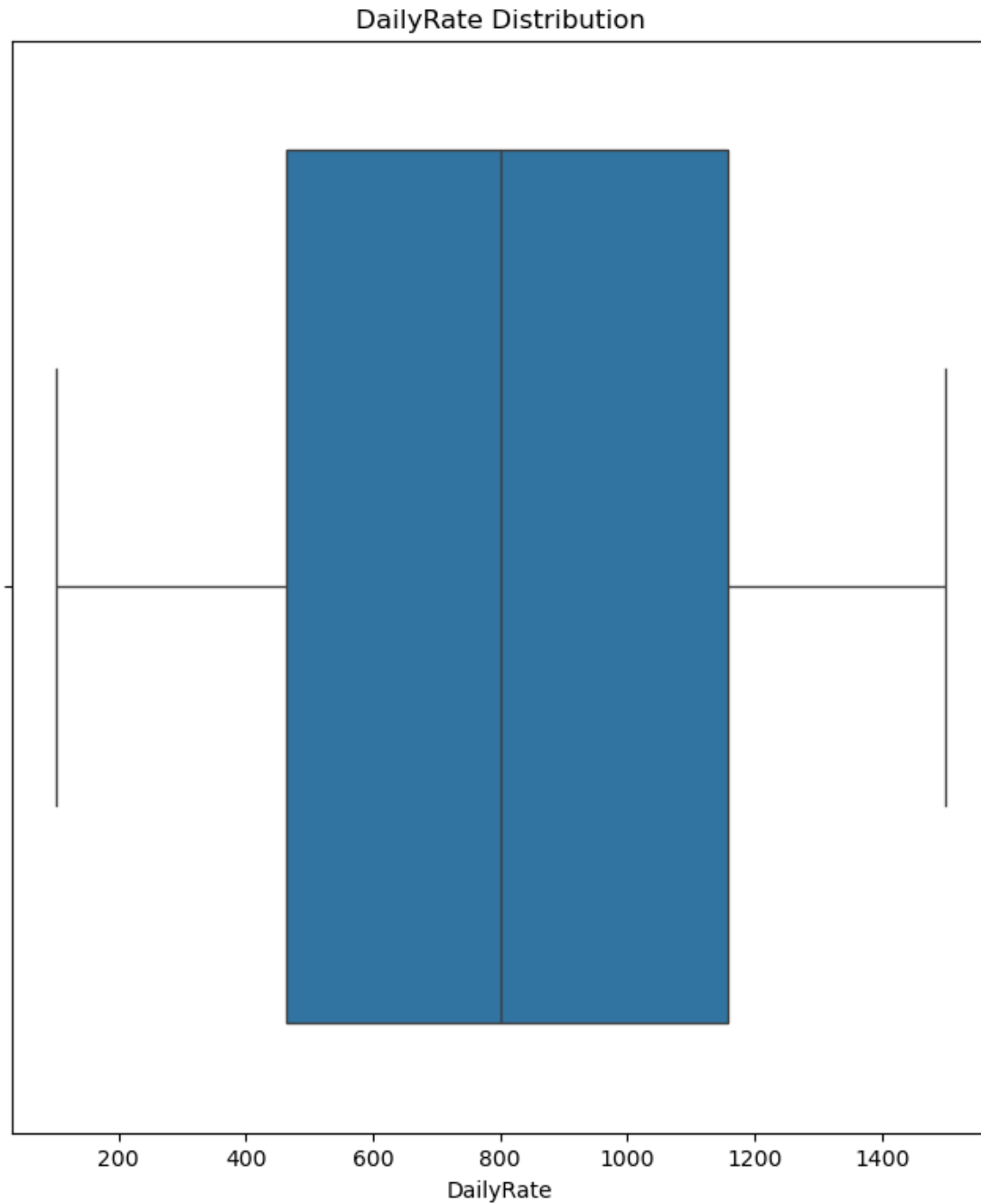


Results: *No abnormal outerlier exist.

*The avearage age is betwenn 30 and 40.

- the lower limit is at the age of 20 and upper limit is at the age of 60

```
[17]: plt.figure(figsize=(8,9))
sns.boxplot(x=hr["DailyRate"])
plt.title("DailyRate Distribution")
plt.show()
```



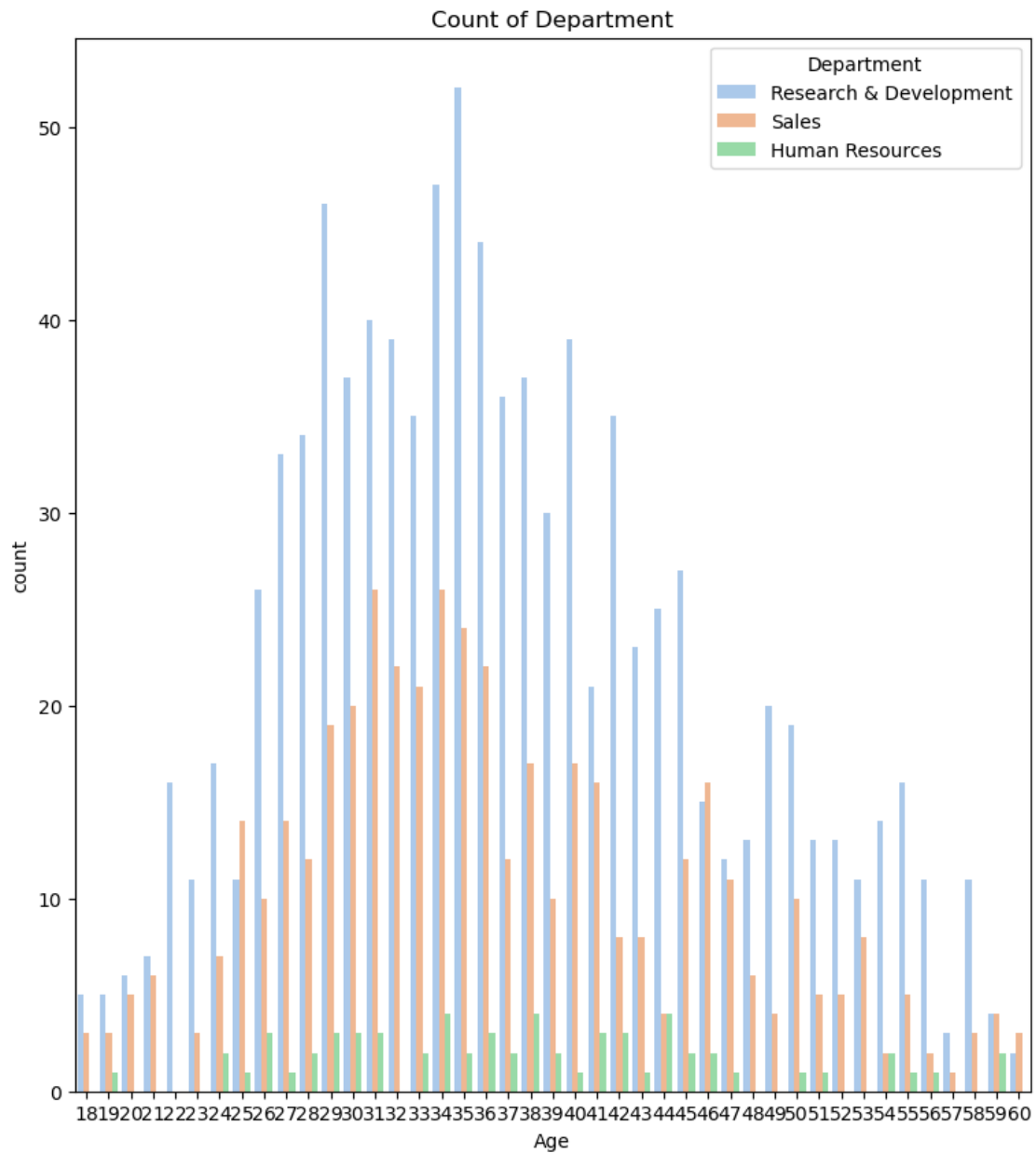
Results: *No abnormal outerlier exist.

*The average DailyRate is between 800.

- the lower limit is at 450 and upper limit is at the age of 1200

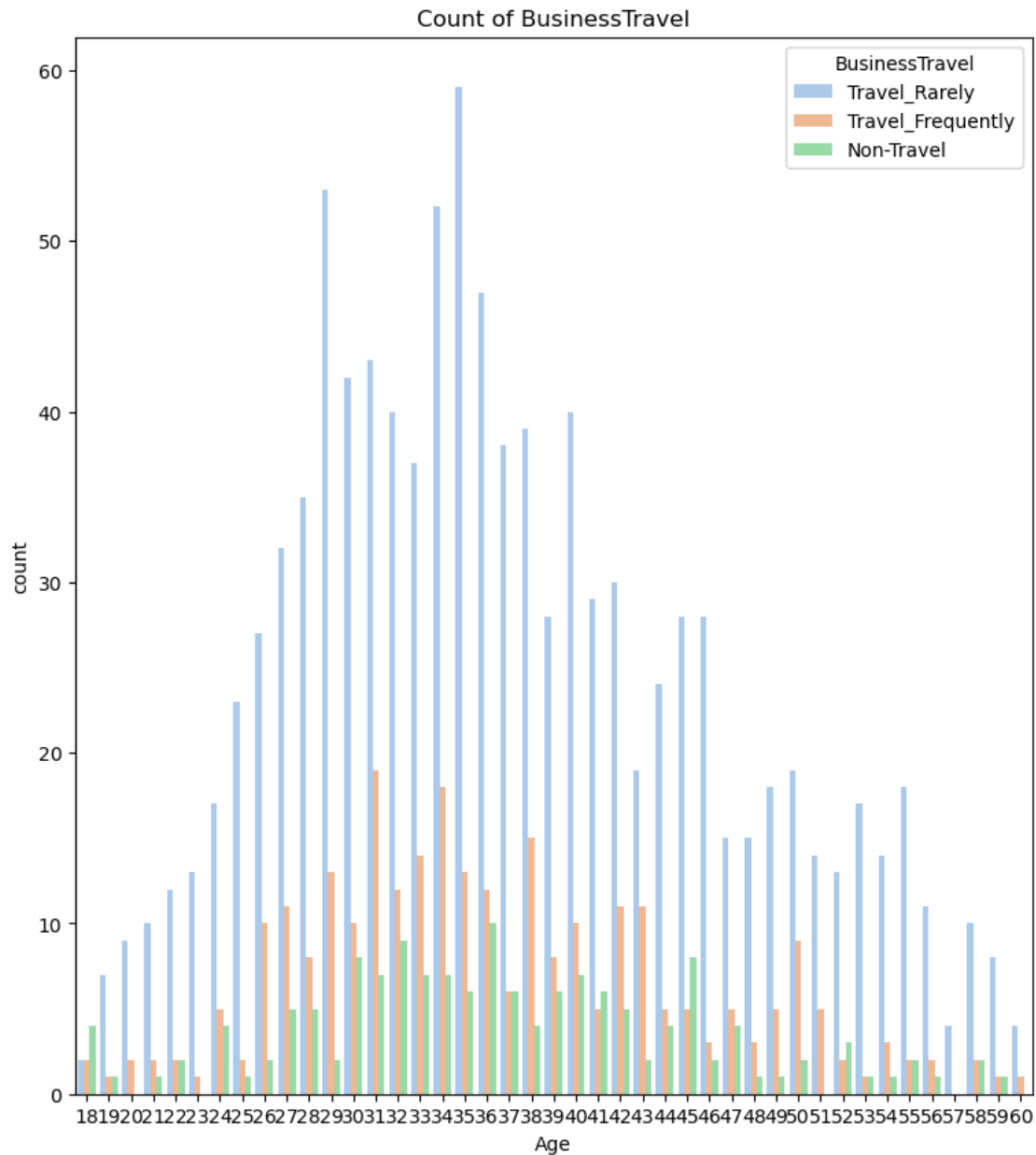
8 Countplot

```
[18]: plt.figure(figsize=(9,10))
sns.countplot(x=hr["Age"],palette="pastel",hue=hr["Department"])
plt.title("Count of Department")
plt.show()
```



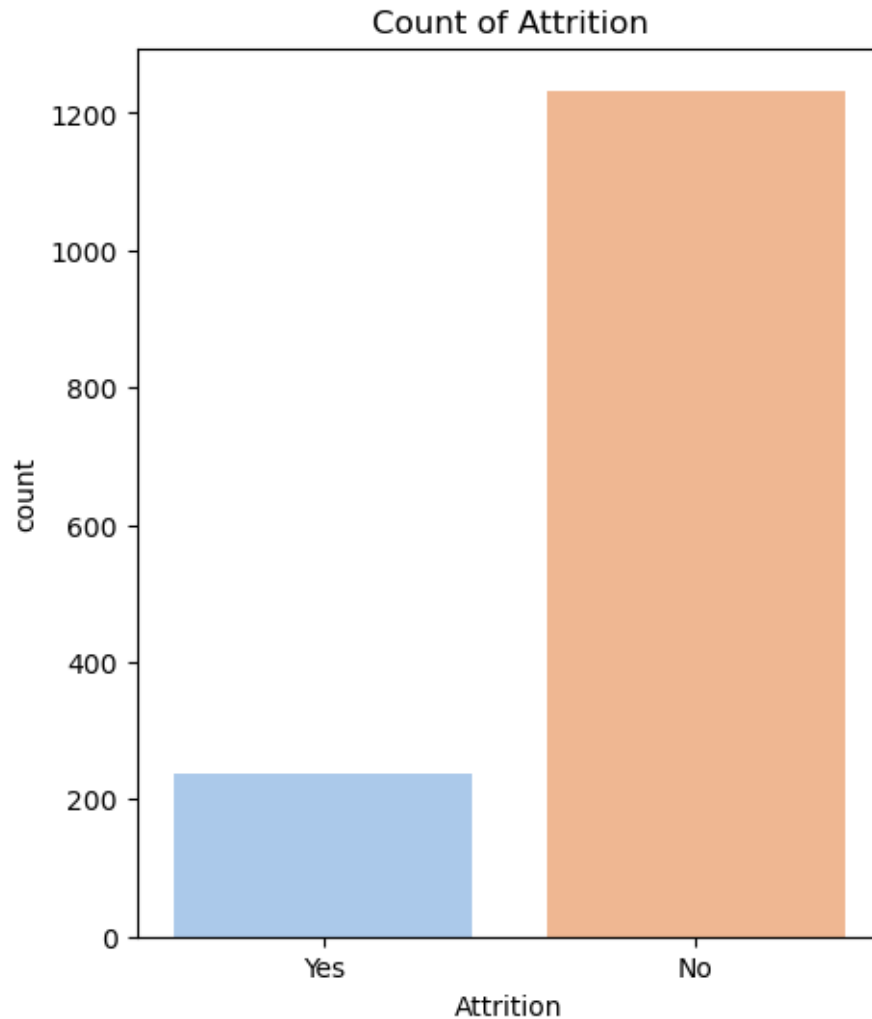
Results: * The above graph shows the count of employees of different department with respect to age. * Human Resource department has the lowest count across the data. * Research and Development has the highest count across the data. * The highest count is pf the age 35.

```
[19]: plt.figure(figsize=(9,10))
sns.countplot(x=hr["Age"],palette="pastel",hue=hr["BusinessTravel"])
plt.title("Count of BusinessTravel")
plt.show()
```



Results: * The above graph represents the count of the employees business travel with respect to the age. * Travel frequency seems to decrease with age, with noticeable drop after 40s. * The highest count for business travel appears to be in the age of range 30-35. * The non-travel category is generally has a higher count across most age groups compared to the other categories.

```
[20]: plt.figure(figsize=(5,6))
sns.countplot(x=hr["Attrition"],palette="pastel",hue=hr["Attrition"])
plt.title("Count of Attrition")
plt.show()
```



Results: * The above graph represents the count of employees stayed with company and left the company. * Approximately 200 employees have left the company(Attrition=Yes). * Around 1200 employees have stayed with the company(Attrition=No). * The chart indicates that there is low attrition rate within the company.

```
[2]: import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[3]: hr=pd.read_csv(r"C:\Mypythonfiles\Human_Resources.csv")
```

```
[4]: hr
```

```
[4]:      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
...  ...
1465   36      No      Travel_Frequently      884  Research & Development
1466   39      No      Travel_Rarely      613  Research & Development
1467   27      No      Travel_Rarely      155  Research & Development
1468   49      No      Travel_Frequently      1023      Sales
1469   34      No      Travel_Rarely      628  Research & Development
```

```
      DistanceFromHome  Education  EducationField  EmployeeCount \
0                      1          2  Life Sciences          1
1                      8          1  Life Sciences          1
2                      2          2      Other          1
3                      3          4  Life Sciences          1
4                      2          1      Medical          1
...
1465                  23          2      Medical          1
1466                   6          1      Medical          1
1467                   4          3  Life Sciences          1
1468                   2          3      Medical          1
1469                   8          3      Medical          1
```

```
      EmployeeNumber  ...  RelationshipSatisfaction  StandardHours \
0                   1  ...                        1              80
1                   2  ...                        4              80
2                   4  ...                        2              80
3                   5  ...                        3              80
4                   7  ...                        4              80
...
1465              2061  ...                        3              80
1466              2062  ...                        1              80
1467              2064  ...                        2              80
1468              2065  ...                        4              80
1469              2068  ...                        1              80
```

	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	\
0	0	8	0	
1	1	10	3	
2	0	7	3	
3	0	8	3	
4	1	6	3	
...	
1465	1	17	3	
1466	1	9	5	
1467	1	6	0	
1468	0	17	3	
1469	0	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	
...	
1465	3	5	2	
1466	3	7	7	
1467	3	6	2	
1468	2	9	6	
1469	4	4	3	

	YearsSinceLastPromotion	YearsWithCurrManager
0	0	5
1	1	7
2	0	0
3	3	0
4	2	2
...
1465	0	3
1466	1	7
1467	0	3
1468	0	8
1469	1	2

[1470 rows x 35 columns]

```
[5]: hr.groupby("EducationField")["Education"].mean()
```

```
[5]: EducationField
Human Resources    3.111111
Life Sciences      2.929043
```

```
Marketing      3.125786
Medical        2.803879
Other           3.073171
Technical Degree 2.825758
Name: Education, dtype: float64
```

```
[6]: hr.groupby("Department")["TotalWorkingYears"].mean()
```

```
[6]: Department
Human Resources      11.555556
Research & Development 11.342352
Sales                11.105381
Name: TotalWorkingYears, dtype: float64
```

9 Aggregation

```
[7]: hr.groupby("Attrition").agg({"Age": ["count", "mean"]})
```

```
[7]:           Age
      count    mean
Attrition
No       1233  37.561233
Yes        237  33.607595
```

10 Segmentation

```
[11]: hr2=hr[(hr["Attrition"]=="Yes")]
      hr2["YearsAtCompany"].mean()
```

```
[11]: 5.1308016877637135
```

```
[ ]:
```