In [6]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
dataset1=pd.read_csv("general_data.csv")
```

Out[6]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount
0	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1
1	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1
2	32	No	Travel_Frequently	Research & Development	17	4	Other	1
3	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1
4	32	No	Travel_Rarely	Research & Development	10	1	Medical	1

5 rows × 24 columns

In [2]:
```

```

Out[2]:
```
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
       'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',
       'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',
       'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
       'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],
      dtype='object')
```

In [7]:
```

```

Out[7]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCou
0	51	No	Travel_Rarely	Sales	6	2	Life Sciences	
1	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	
2	32	No	Travel_Frequently	Research & Development	17	4	Other	
3	38	No	Non-Travel	Research & Development	2	5	Life Sciences	
4	32	No	Travel_Rarely	Research & Development	10	1	Medical	
...	...	...	...	...	...	...	...	
4405	42	No	Travel_Rarely	Research & Development	5	4	Medical	
4406	29	No	Travel_Rarely	Research & Development	2	4	Medical	
4407	25	No	Travel_Rarely	Research & Development	25	2	Life Sciences	
4408	42	No	Travel_Rarely	Sales	18	2	Medical	
4409	40	No	Travel_Rarely	Research & Development	28	3	Medical	

4410 rows × 24 columns

In [9]:
```
dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompanie
                   'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','Ye
```

In [20]:
```
```

Out[20]:

	Age	DistanceFromHome	Education	MonthlyIncome	NumCompaniesWorked	PercentSalaryHike
count	3699.000000	3699.000000	3699.000000	3699.000000	3684.000000	3699.000000
mean	37.561233	9.227088	2.919708	65672.595296	2.648480	15.157340
std	8.885956	8.167978	1.025784	47472.814021	2.460537	3.634551
min	18.000000	1.000000	1.000000	10510.000000	0.000000	11.000000
25%	31.000000	2.000000	2.000000	29360.000000	1.000000	12.000000
50%	36.000000	7.000000	3.000000	49300.000000	2.000000	14.000000
75%	43.000000	14.000000	4.000000	86060.000000	4.000000	18.000000
max	60.000000	29.000000	5.000000	199990.000000	9.000000	25.000000

In [10]:
```
is_attrition = [dataset1['Attrition'] == 'Yes']
```

Out[10]:
```
[0       False
 1        True
 2       False
 3       False
 4       False
         ...
 4405    False
 4406    False
 4407    False
 4408    False
 4409    False
 Name: Attrition, Length: 4410, dtype: bool]
```

In [11]:
```
attrition_ds= dataset1.loc[dataset1['Attrition'] == 'Yes']
```

Out[11]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCou
1	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	
6	28	Yes	Travel_Rarely	Research & Development	11	2	Medical	
13	47	Yes	Non-Travel	Research & Development	1	1	Medical	
28	44	Yes	Travel_Frequently	Research & Development	1	2	Medical	
30	26	Yes	Travel_Rarely	Research & Development	4	3	Medical	
...	...	...	...	...	...	...	...	
4381	29	Yes	Travel_Rarely	Research & Development	7	1	Life Sciences	
4386	33	Yes	Travel_Rarely	Sales	11	4	Marketing	
4388	33	Yes	Travel_Rarely	Sales	1	3	Life Sciences	
4391	32	Yes	Travel_Rarely	Sales	23	1	Life Sciences	
4402	37	Yes	Travel_Frequently	Sales	2	3	Marketing	

711 rows x 24 columns

In [41]:

Out[41]:
```
0    Travel_Rarely
dtype: object
```

In [42]:

Out[42]:
```
711
```

In [12]:

Out[12]:
```
Travel_Rarely        468
Travel_Frequently    207
Non-Travel            36
Name: BusinessTravel, dtype: int64
```

Infrence1

People who travel rarely are more likely to leave the company..

In [48]:
```
no_attrition_ds= dataset1.loc[dataset1['Attrition'] == 'No']
```

Out[48]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount
0	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1
2	32	No	Travel_Frequently	Research & Development	17	4	Other	1
3	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1
4	32	No	Travel_Rarely	Research & Development	10	1	Medical	1
5	46	No	Travel_Rarely	Research & Development	8	3	Life Sciences	1
...	...	...	...	...	...	...	...	...
4405	42	No	Travel_Rarely	Research & Development	5	4	Medical	1
4406	29	No	Travel_Rarely	Research & Development	2	4	Medical	1

In [54]:

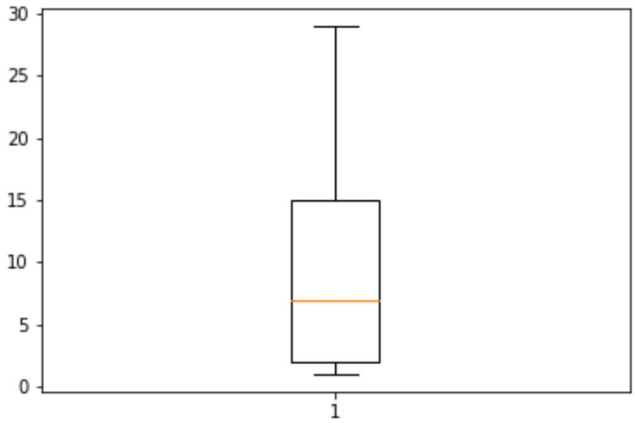
Out[54]:
```
Travel_Rarely        2661
Travel_Frequently     624
Non-Travel            414
Name: BusinessTravel, dtype: int64
```

In [5]:

Out[5]:
```
count    711.000000
mean       9.012658
std        7.772368
min        1.000000
25%        2.000000
50%        7.000000
75%       15.000000
max       29.000000
Name: DistanceFromHome, dtype: float64
```

In [9]:
```
attriton_distance_box_plot=attrition_ds['DistanceFromHome']
```

Out[9]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a76869d48>,
  <matplotlib.lines.Line2D at 0x19a76869f88>],
 'caps': [<matplotlib.lines.Line2D at 0x19a7686eac8>,
  <matplotlib.lines.Line2D at 0x19a76874fc8>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a76869648>],
 'medians': [<matplotlib.lines.Line2D at 0x19a7687fcc8>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a7687fec8>],
 'means': []}
```

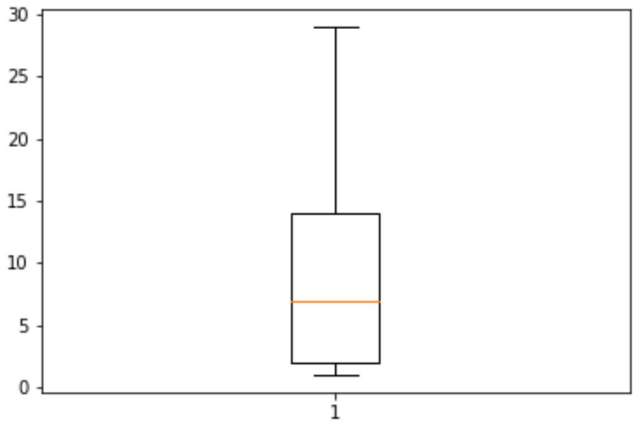


In [7]:

Out[7]:
```
count    3699.000000
mean        9.227088
std         8.167978
min         1.000000
25%         2.000000
50%         7.000000
75%        14.000000
max        29.000000
Name: DistanceFromHome, dtype: float64
```

In [11]:
```
no_attriton_distance_box_plot=no_attrition_ds['DistanceFromHome']
```

Out[11]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a769aa348>,
  <matplotlib.lines.Line2D at 0x19a76260e88>],
 'caps': [<matplotlib.lines.Line2D at 0x19a76425708>,
  <matplotlib.lines.Line2D at 0x19a7645ebc8>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a76473cc8>],
 'medians': [<matplotlib.lines.Line2D at 0x19a7624aa48>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a7640ef88>],
 'means': []}
```



Infrence2

Lookign at box plot of employess who left the company meadian is below mean. in other words employess who stays far away from company tend to leave the company.

In [13]:

Out[13]:
```
count      3699.000000
mean      65672.595296
std       47472.814021
min       10510.000000
25%       29360.000000
50%       49300.000000
75%       86060.000000
max      199990.000000
Name: MonthlyIncome, dtype: float64
```

In [14]:

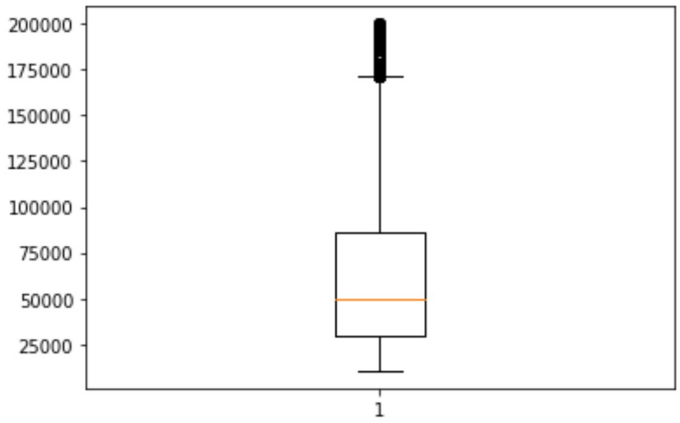
Out[14]:
```
count       711.000000
mean      61682.616034
std       44792.067695
min       10090.000000
25%       28440.000000
50%       49080.000000
75%       71040.000000
max      198590.000000
Name: MonthlyIncome, dtype: float64
```

In [16]:
```
no_attriton_distance_box_plot=no_attrition_ds['MonthlyIncome']

plt.boxplot(no_attriton_distance_box_plot)
```

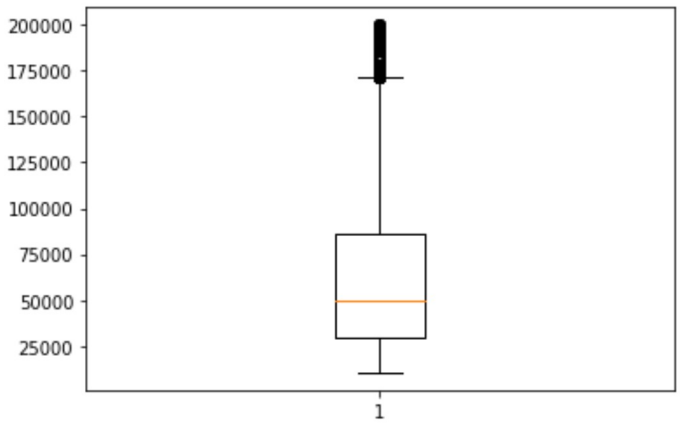
Out[16]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a76d0b248>,
  <matplotlib.lines.Line2D at 0x19a7657b4c8>],
 'caps': [<matplotlib.lines.Line2D at 0x19a76565a08>,
  <matplotlib.lines.Line2D at 0x19a7655e8c8>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a7657eb88>],
 'medians': [<matplotlib.lines.Line2D at 0x19a7655a248>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a76540348>],
```



In [17]:
```
attriton_distance_box_plot=attrition_ds['MonthlyIncome']
```

Out[17]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a76d34b48>,
  <matplotlib.lines.Line2D at 0x19a76d34688>],
 'caps': [<matplotlib.lines.Line2D at 0x19a76d2fbc8>,
  <matplotlib.lines.Line2D at 0x19a76d2c888>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a76d37948>],
 'medians': [<matplotlib.lines.Line2D at 0x19a76d2a988>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a766a82c8>],
 'means': []}
```



Infrence3

Salary is not normalised. Mean is less than median.

In [25]:
```
print(attrition_ds['Gender'].describe())
print("="*100)
print(attrition_ds['Gender'].value_counts())
print("="*100)
print(attrition_ds['Gender'].value_counts(normalize=True) * 100)
print("="*100)
print("="*100)
```

```
print(dataset1['Gender'].describe())
print("="*100)
print(dataset1['Gender'].value_counts())
print("="*100)
```

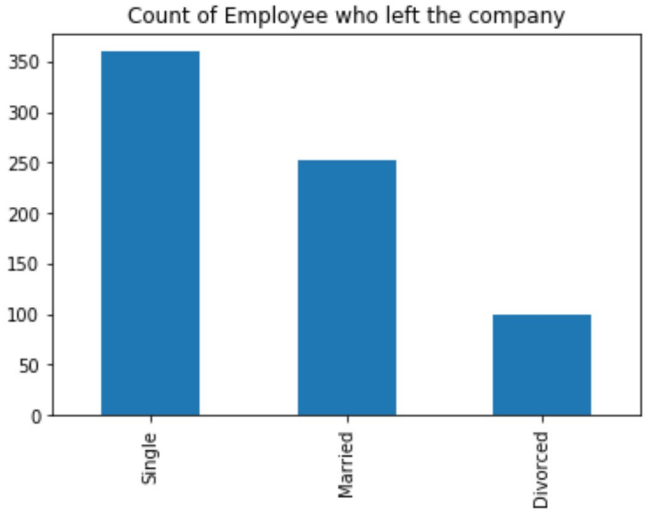
```
count      711
unique       2
top       Male
freq       441
Name: Gender, dtype: object
====================================================================================
==================
Male      441
Female    270
Name: Gender, dtype: int64
====================================================================================
==================
Male      62.025316
Female    37.974684
Name: Gender, dtype: float64
====================================================================================
==================
====================================================================================
==================
count     4410
unique       2
top       Male
freq      2646
Name: Gender, dtype: object
====================================================================================
==================
Male      2646
Female    1764
Name: Gender, dtype: int64
====================================================================================
==================
Male      60.0
Female    40.0
Name: Gender, dtype: float64
```

Infrence4

Gender Ration of company is 60:40 but the ration of people who leave company is 62:38 With this we can infer Male employess are more likely to leave than female employees.

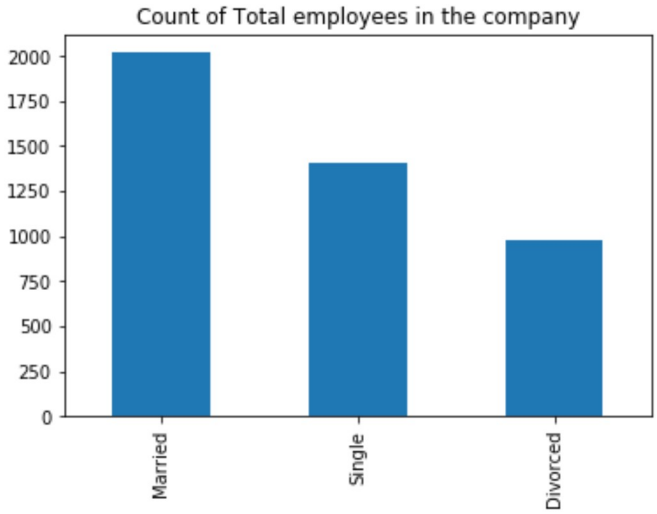
In [39]:
```
print(attrition_ds['MaritalStatus'].describe())
print("="*100)
print(attrition_ds['MaritalStatus'].value_counts())
print("="*100)
print(attrition_ds['MaritalStatus'].value_counts(normalize=True) * 100)
print("="*100)
attrition_ds['MaritalStatus'].value_counts().plot(kind='bar',title="Count of Employee
print("="*100)
print(dataset1['MaritalStatus'].describe())
print("="*100)
print(dataset1['MaritalStatus'].value_counts())
print("="*100)
print(dataset1['MaritalStatus'].value_counts(normalize=True) * 100)
```

```
count        711
unique         3
top       Single
freq         360
Name: MaritalStatus, dtype: object
================================================================================
==================
Single      360
Married     252
Divorced     99
Name: MaritalStatus, dtype: int64
================================================================================
==================
Single      50.632911
Married     35.443038
Divorced    13.924051
Name: MaritalStatus, dtype: float64
================================================================================
==================
================================================================================
==================
count        4410
unique          3
top       Married
freq         2019
Name: MaritalStatus, dtype: object
================================================================================
==================
Married     2019
Single      1410
Divorced     981
Name: MaritalStatus, dtype: int64
================================================================================
==================
Married     45.782313
Single      31.972789
```



In [38]:

Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x19a78571548>



Infrence5

Of the employee who leaves the company 50% are Single

In [43]:
```
print(attrition_ds['NumCompaniesWorked'].value_counts())
```

```
1.0    293
0.0     69
7.0     51
4.0     50
6.0     48
3.0     48
2.0     48
5.0     46
9.0     36
8.0     18
Name: NumCompaniesWorked, dtype: int64
AxesSubplot(0.125,0.125;0.775x0.755)
```

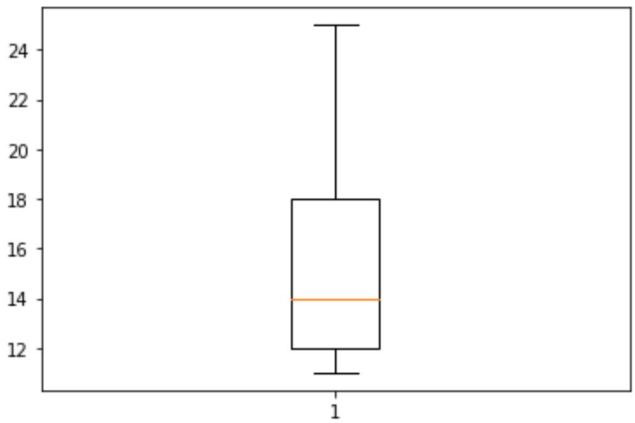
In [44]:
```
print(dataset1['NumCompaniesWorked'].value_counts())
```

```
1.0    1558
0.0     586
3.0     474
2.0     438
4.0     415
7.0     222
6.0     208
5.0     187
9.0     156
8.0     147
```

In [ ]:
```
# Infrence6
To be written
```

In [49]:
```
attrition_ds['PercentSalaryHike'].describe()
```

Out[49]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a7a1e6f48>,
  <matplotlib.lines.Line2D at 0x19a7a1e6fc8>],
 'caps': [<matplotlib.lines.Line2D at 0x19a7a1ebf08>,
  <matplotlib.lines.Line2D at 0x19a7a1ebfc8>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a7a1e6748>],
 'medians': [<matplotlib.lines.Line2D at 0x19a7a1effc8>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a7a1eff48>],
 'means': []}
```



Infrence 7

On Alanysing percentage hike, Mean lies below mean. So people who get less percentage hike are leaving the company.

In [50]:

```
Out[50]: count    4410.0
         mean        8.0
         std         0.0
         min         8.0
         25%         8.0
         50%         8.0
         75%         8.0
         max         8.0
         Name: StandardHours, dtype: float64
```

Infrence 8

Standar working hours is same for all employees so this cannot be parameter for one leaving company.

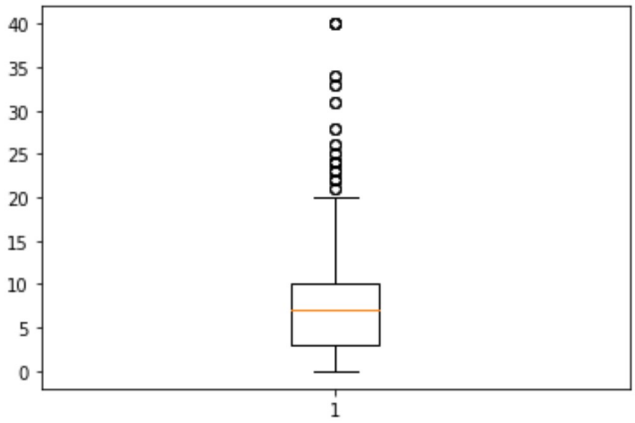
In [58]:

```
attrition_ds['TotalWorkingYears'].describe()
attrition_ds_1 = attrition_ds['TotalWorkingYears'].dropna()
print(attrition_ds_1.describe())
```

```
count    709.000000
mean       8.255289
std        7.164018
min        0.000000
25%        3.000000
50%        7.000000
75%       10.000000
max       40.000000
Name: TotalWorkingYears, dtype: float64
```

Out[58]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a7d841588>,
  <matplotlib.lines.Line2D at 0x19a7d841d48>],
 'caps': [<matplotlib.lines.Line2D at 0x19a7d841e08>,
  <matplotlib.lines.Line2D at 0x19a7d845cc8>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a7d83eac8>],
 'medians': [<matplotlib.lines.Line2D at 0x19a7d845b08>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a7d3ae448>],
 'means': []}
```



Infrence 9

Experince of employeess who leaves the company has more outliers Employees with total exp <5 leave the company frequntly

In [60]:
```
print(attrition_ds['TrainingTimesLastYear'].describe())
print("="*100)
```

```
count    711.000000
mean       2.654008
std        1.154834
min        0.000000
25%        2.000000
50%        3.000000
75%        3.000000
max        6.000000
Name: TrainingTimesLastYear, dtype: float64
====================================================================================================
count    3699.000000
mean        2.827251
std         1.311493
min         0.000000
25%         2.000000
50%         3.000000
75%         3.000000
max         6.000000
Name: TrainingTimesLastYear, dtype: float64
```

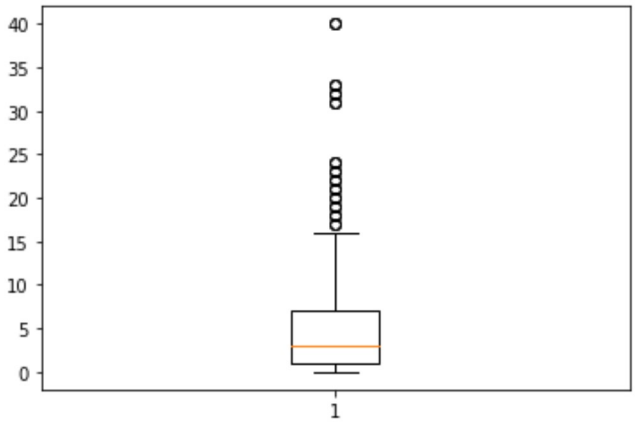
Infrence 10

mean Training time of Employee who left the company and staying in the company are

In [63]:
```
print(attrition_ds['YearsAtCompany'].describe())
print("="*100)
print(no_attrition_ds['YearsAtCompany'].describe())
```

```
count    711.000000
mean       5.130802
std        5.941598
min        0.000000
25%        1.000000
50%        3.000000
75%        7.000000
max       40.000000
Name: YearsAtCompany, dtype: float64
====================================================================================================
count    3699.000000
mean        7.369019
std         6.094649
min         0.000000
25%         3.000000
50%         6.000000
75%        10.000000
max        37.000000
Name: YearsAtCompany, dtype: float64
```

Out[63]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a7db4b808>,
  <matplotlib.lines.Line2D at 0x19a7db39888>],
 'caps': [<matplotlib.lines.Line2D at 0x19a7db2b0c8>,
  <matplotlib.lines.Line2D at 0x19a7db41788>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a7db39508>],
 'medians': [<matplotlib.lines.Line2D at 0x19a7dc2cfc8>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a7dc2ce48>],
 'means': []}
```



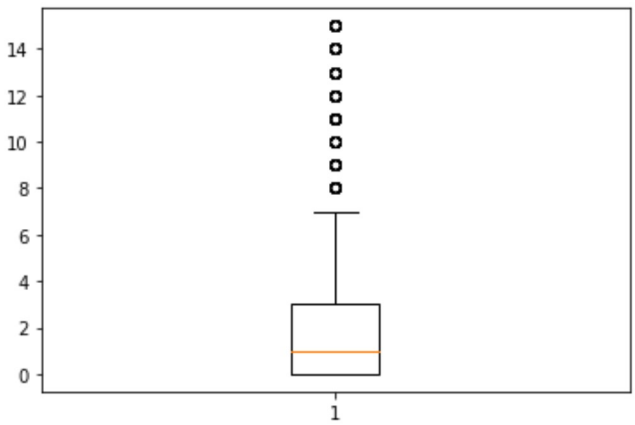
Infrence 12

The employess who left the company there are many outliers in terms of their experience in the companny also its mean is less than median so Employee who spend more time in the company are more likely to leave.

In [66]:
```
print(attrition_ds['YearsSinceLastPromotion'].describe())
print("="*100)
print(no_attrition_ds['YearsSinceLastPromotion'].describe())
```

```
count    711.000000
mean       1.945148
std        3.148633
min        0.000000
25%        0.000000
50%        1.000000
75%        2.000000
max       15.000000
Name: YearsSinceLastPromotion, dtype: float64
================================================================================
==================
count    3699.000000
mean        2.234388
std         3.233887
min         0.000000
25%         0.000000
50%         1.000000
```

Out[66]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x19a7e0a8048>,
  <matplotlib.lines.Line2D at 0x19a7e0a8088>],
 'caps': [<matplotlib.lines.Line2D at 0x19a7e0a34c8>,
  <matplotlib.lines.Line2D at 0x19a7e0a3508>],
 'boxes': [<matplotlib.lines.Line2D at 0x19a7e0a8a48>],
 'medians': [<matplotlib.lines.Line2D at 0x19a7e0a31c8>],
 'fliers': [<matplotlib.lines.Line2D at 0x19a7e0a0388>],
 'means': []}
```



In [75]:
```
print(attrition_ds['YearsWithCurrManager'].describe())
print("="*100)
print(no_attrition_ds['YearsWithCurrManager'].describe())
```

```
count    711.000000
mean       2.852321
std        3.138918
```

Infrence 13

With above data we can conculde that change in manager is one cause for employee leaving company.

In [78]:
```
print(attrition_ds['Age'].describe())
print("="*100)
```

```
count    711.000000
mean      33.607595
std        9.675693
min       18.000000
25%       28.000000
50%       32.000000
75%       39.000000
max       58.000000
Name: Age, dtype: float64
====================================================================================================
count    3699.000000
mean       37.561233
std         8.885956
min        18.000000
25%        31.000000
50%        36.000000
75%        43.000000
max        60.000000
Name: Age, dtype: float64
```

Infrence 14

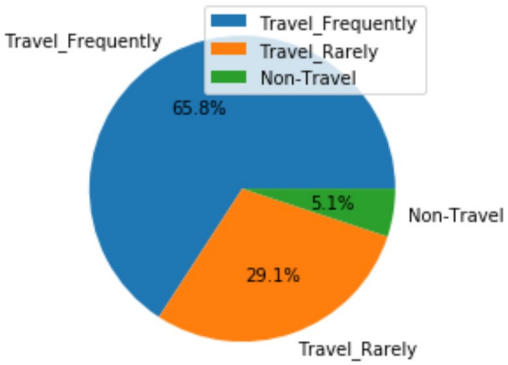
With Above Data we can conclude that People with less age tend to leave the company more.

In [23]:
```
rs = dataset1.groupby('Attrition').mean()
```

Out[23]:

Attrition	Age	DistanceFromHome	Education	EmployeeCount	EmployeeID	JobLevel	MonthlyIncome	Num
No	37.561233	9.227088	2.919708	1.0	2208.139497	2.068938	65672.595296	
Yes	33.607595	9.012658	2.877637	1.0	2191.767932	2.037975	61682.616034	

In [44]:
```
lbl = attrition_ds['BusinessTravel'].unique()
values = attrition_ds['BusinessTravel'].value_counts()
plt.pie(values,labels=lbl,autopct='%1.1f%%',)
plt.legend(lbl,loc=0)
```



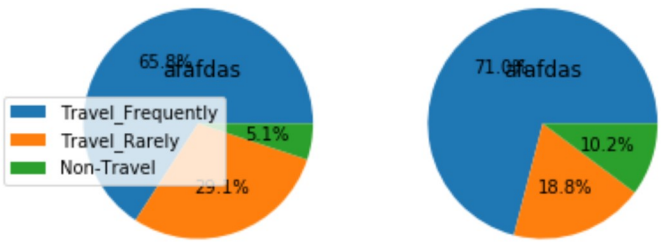
In [99]:
```
lbl = attrition_ds['BusinessTravel'].unique()
values = attrition_ds['BusinessTravel'].value_counts()
no_lbl = dataset1['BusinessTravel'].unique()
no_values = dataset1['BusinessTravel'].value_counts()

plt.subplot(1, 7, 1)
plt.pie(values, autopct='%1.1f%%',radius=4)
plt.title("afafdas",loc='left')
plt.legend(lbl,loc='best')

plt.subplot(1, 7, 5)
plt.title("afafdas")
plt.pie(no_values, autopct='%1.1f%%',radius=4)
#plt.legend(lbl,loc='center')

# plt.subplot(1, 7, 7)
# plt.legend(lbl,loc='center')

plt.show()
```



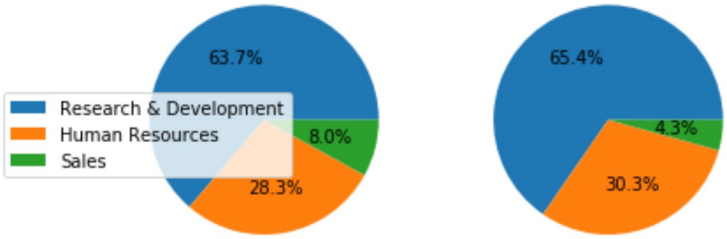
In [102]:
```
lbl = attrition_ds['Department'].unique()
values = attrition_ds['Department'].value_counts()
no_lbl = dataset1['Department'].unique()
no_values = dataset1['Department'].value_counts()

plt.subplot(1, 7, 1)
plt.pie(values, autopct='%1.1f%%',radius=4,)
plt.legend(lbl,loc='best')

plt.subplot(1, 7, 5)
plt.pie(no_values, autopct='%1.1f%%',radius=4)
#plt.legend(lbl,loc='center')

# plt.subplot(1, 7, 7)
```

```
# plt.legend(lbl,loc='center')
```



In [ ]: