ASSIGNMENT

## FINDING CORRELATION BTW VARIABLES

In [27]:

## import pandas as pd

In [28]:

URL = 'https://raw.githubusercontent.com/LetsUpgrade/AI-ML-July-2020/master/Day-7/Assignm ent/general\_data.csv'

data = pd.read\_csv(URL)

In [29]:

data Out[29]:

**Age Attrition BusinessTravel Department DistanceFromHome Education EducationField EmployeeCount Employ**

**0** 51 No Travel\_Rarely Sales 6 2 Life Sciences 1

**1** 31 Yes Travel\_Frequently Research &

Development

10 1 Life Sciences 1

**3** 38 No Non-Travel Research &

Development

2 5 Life Sciences 1

**...** ... ... ... ... ... ... ... ...

**4405** 42 No Travel\_Rarely Research &

Development

**4406** 29 No Travel\_Rarely Research &

Development

5 4 Medical 1

2 4 Medical 1

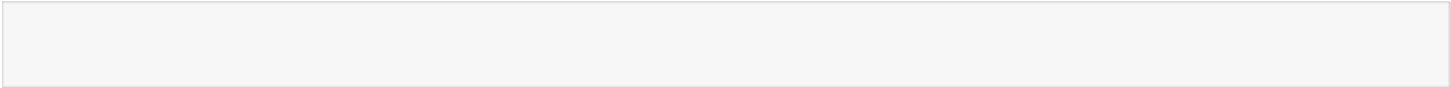
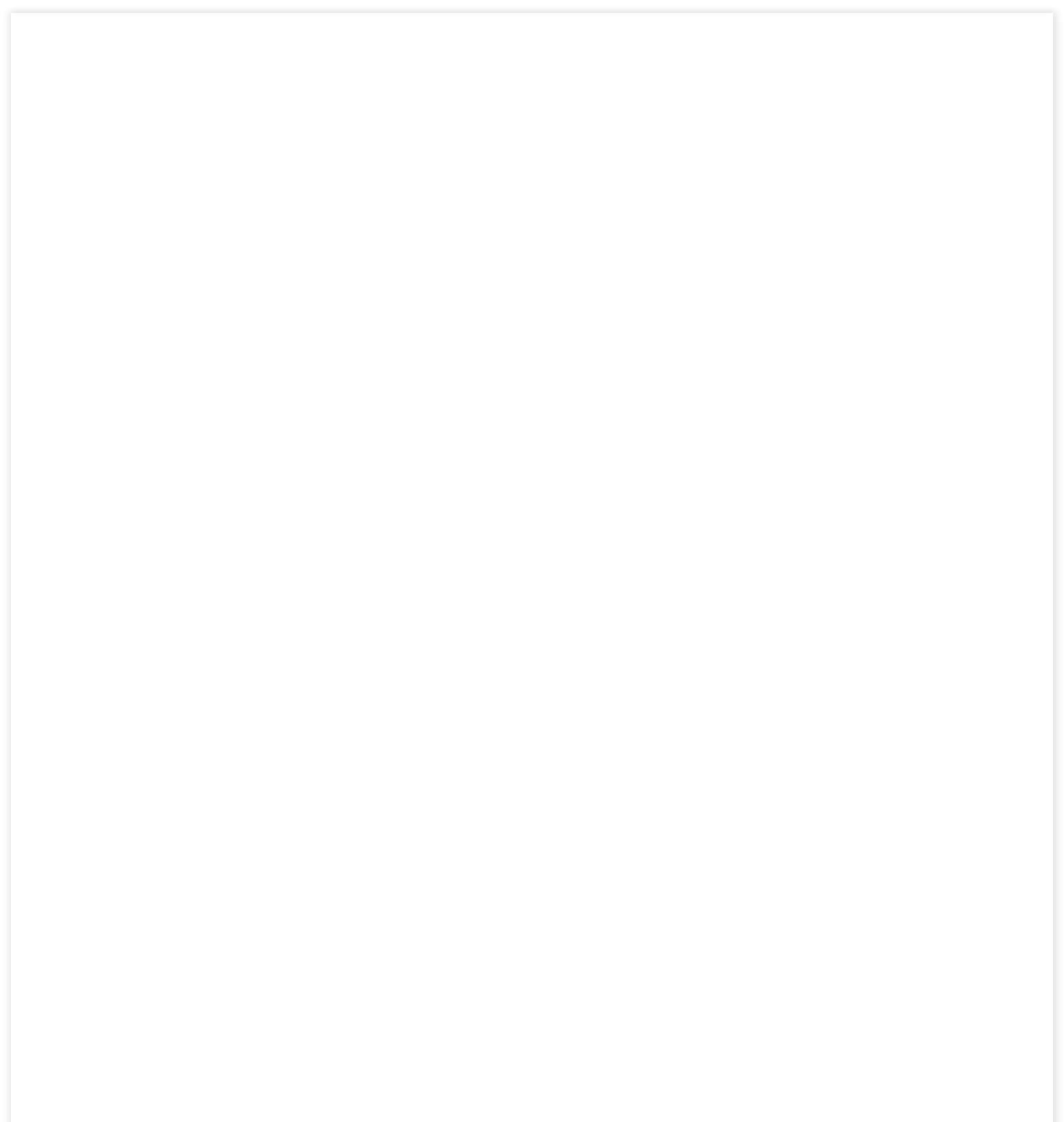
**4408** 42 No Travel\_Rarely Sales 18 2 Medical 1

**4409** 40 No Travel\_Rarely Research &

Development

28 3 Medical 1

4410 rows × 24 columns



**2** 32

No Travel\_Frequently

Research &

Development

17

4

Other

1

**4** 32

No

Travel\_Rarely

Research &

Development

10

1

Medical

1

**4407** 25

No

Travel\_Rarely

Research &

Development

25

2 Life Sciences

1

In [30]:

**from scipy.stats import** pearsonr

In [31]:

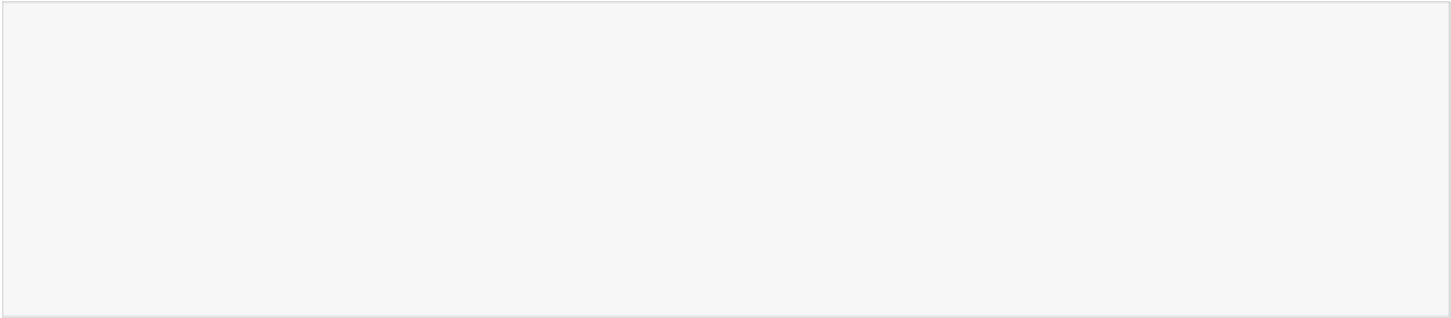
## import matplotlib.pyplot as plt

In [32]:

data['Attrition'].replace(to\_replace=("No","Yes"),value=(0,1),inplace=**True**)

# Attrition and Age

In [47]:



stats,p = pearsonr(data.Attrition,data.Age) print("Correlation: ",stats,"**\n** P-value: ",p) **if** stats == 0 :

print("**\n\n**No Correlation") **elif** stats > 0 :

print("**\n\n**Positive correlation") **else**:

print("**\n\n**Negative correlation") **if** p<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.Age)

Correlation: -0.15920500686577949 P-value: 1.996801615887198e-26

Negative correlation Reject null hypothesis

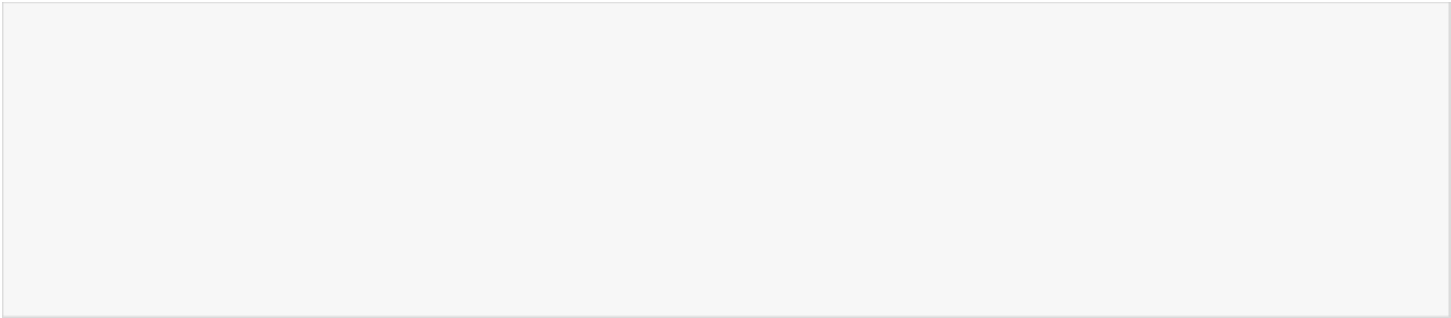
Out[47]:

<matplotlib.collections.PathCollection at 0x7fa453a1fa58>



# Attrition and Monthlyincome

In [48]:



stats1,p1 = pearsonr(data.Attrition,data.MonthlyIncome) print("Correlation: ",stats1,"**\n** P-value: ",p1)

**if** stats1 == 0 : print("**\n\n**No Correlation")

**elif** stats1 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p1<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.MonthlyIncome)

Correlation: -0.03117628169811501

P-value: 0.03842748490600132

Negative correlation Reject null hypothesis

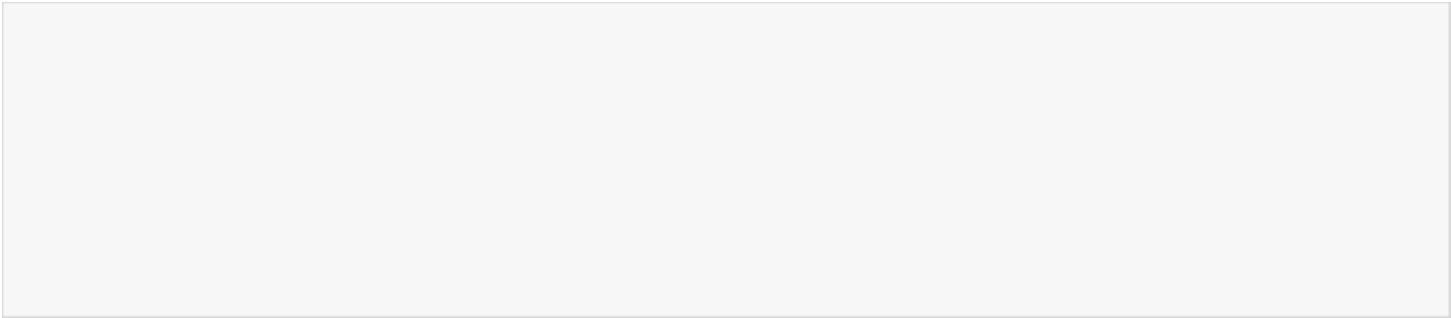
Out[48]:

<matplotlib.collections.PathCollection at 0x7fa453982588>



# Attrition and DistanceFromHome

In [49]:



stats2,p2 = pearsonr(data.Attrition,data.DistanceFromHome) print("Correlation: ",stats2,"**\n** P-value: ")

**if** stats2 == 0 : print("**\n\n**No Correlation")

**elif** stats2 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p2<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.DistanceFromHome)

Correlation: -0.009730141010179692 P-value:

Negative correlation Accept null hypothesis

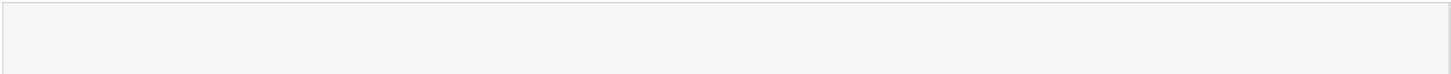
Out[49]:

<matplotlib.collections.PathCollection at 0x7fa45396a240>



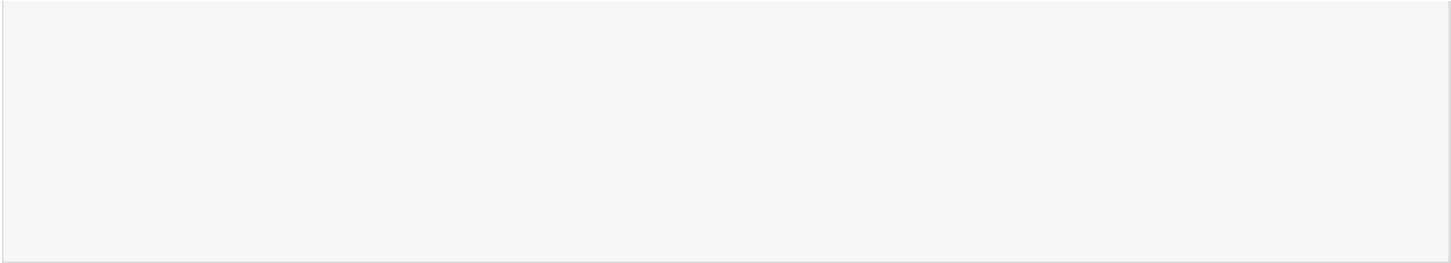
# Attrition and Joblevel

In [50]:



stats3,p3 = pearsonr(data.Attrition,data.JobLevel) print("Correlation: ",stats3,"**\n** P-value: ",p3)





**if** stats3 == 0 : print("**\n\n**No Correlation")

**elif** stats3 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p3<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.JobLevel)

Correlation: -0.010289713287494923

P-value: 0.4945171727183978

Negative correlation Accept null hypothesis

Out[50]:

<matplotlib.collections.PathCollection at 0x7fa4538cc470>



# Attrition and BusinessTravel

In [37]:



data['BusinessTravel']

Out[37]:

1. Travel\_Rarely
2. Travel\_Frequently
3. Travel\_Frequently
4. Non-Travel
5. Travel\_Rarely

...

|  |  |  |
| --- | --- | --- |
| 4405 | Travel\_Rarely |  |
| 4406 | Travel\_Rarely |  |
| 4407 | Travel\_Rarely |  |
| 4408 | Travel\_Rarely |  |
| 4409 | Travel\_Rarely |  |
| Name: | BusinessTravel, Length: 4410, dtype: | object |

In [38]:



data['BusinessTravel'].unique().tolist()

Out[38]:

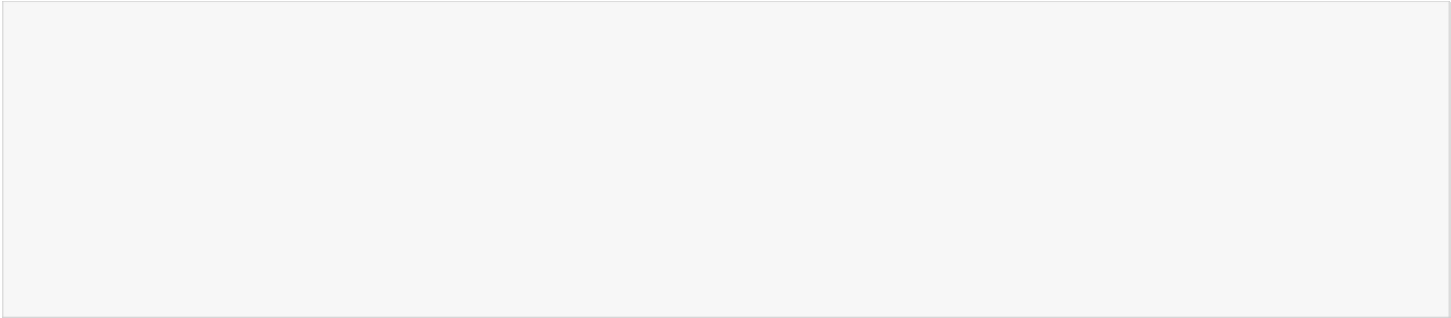
['Travel\_Rarely', 'Travel\_Frequently', 'Non-Travel'] In [39]:



data["BusinessTravel"].replace(to\_replace=("Travel\_Rarely" ,"Travel\_Frequently", "Non-Tra vel"),value=(0,1,2),inplace=**True**)



In [51]:



stats4,p4 = pearsonr(data.Attrition,data.BusinessTravel) print("Correlation: ",stats4,"**\n** P-value: ",p4)

**if** stats4 == 0 : print("**\n\n**No Correlation")

**elif** stats4 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p4<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.BusinessTravel)

Correlation: -7.377694602216824e-05 P-value: 0.9960919945440154

Negative correlation Accept null hypothesis

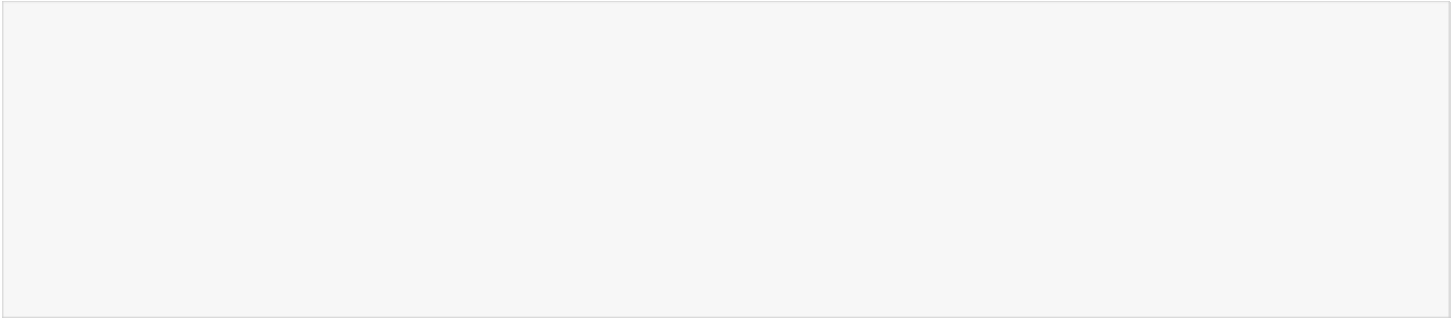
Out[51]:

<matplotlib.collections.PathCollection at 0x7fa4538b3160>



# Attrition and PercentSalaryHike

In [52]:



stats5,p5 = pearsonr(data.Attrition,data.PercentSalaryHike) print("Correlation: ",stats5,"**\n** P-value: ",p5)

**if** stats5 == 0 : print("**\n\n**No Correlation")

**elif** stats5 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p5<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.PercentSalaryHike)

Correlation: 0.0325325948910535

P-value: 0.030743386433355353

Positive correlation Reject null hypothesis

Out[52]:

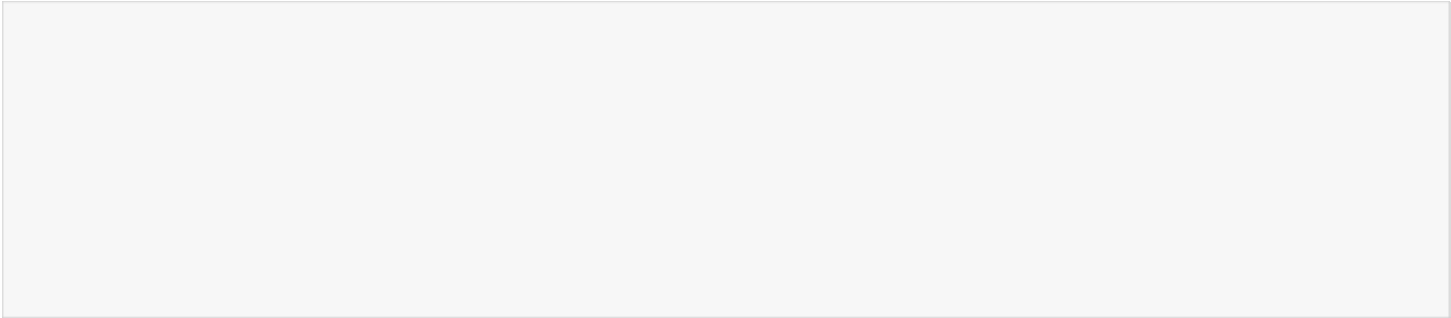
<matplotlib.collections.PathCollection at 0x7fa45380edd8>





# Attrition and StockOptionLevel

In [53]:



stats6,p6 = pearsonr(data.Attrition,data.StockOptionLevel) print("Correlation: ",stats6,"**\n** P-value: ",p6)

**if** stats6 == 0 : print("**\n\n**No Correlation")

**elif** stats6 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p6<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.StockOptionLevel)

Correlation: -0.006838852403261526

P-value: 0.6498072937475723

Negative correlation Accept null hypothesis

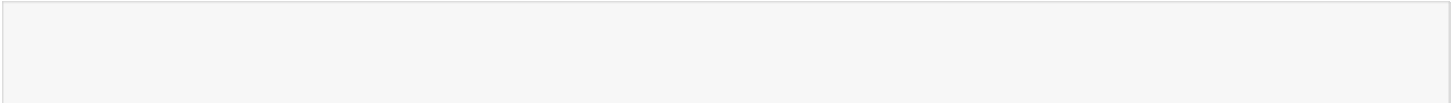
Out[53]:

<matplotlib.collections.PathCollection at 0x7fa45377f0b8>



# Attrition and TrainingTimesLastYear

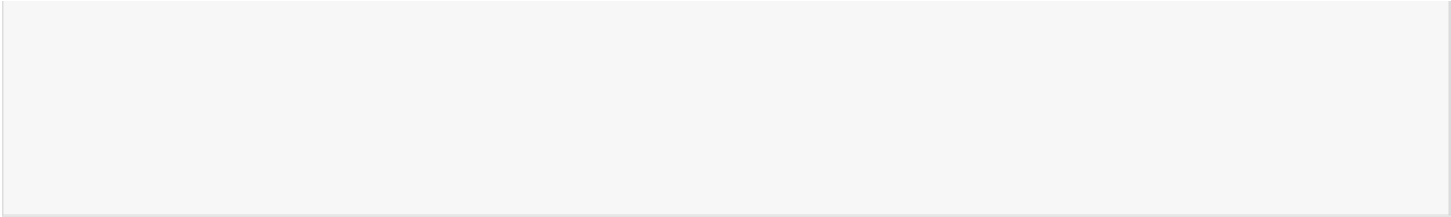
In [54]:



stats7,p7 = pearsonr(data.Attrition,data.TrainingTimesLastYear) print("Correlation: ",stats7,"**\n** P-value: ",p7)

**if** stats7 == 0 : print("**\n\n**No Correlation")





**elif** stats7 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p7<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.TrainingTimesLastYear)

Correlation: -0.049430576244254974

P-value: 0.0010247061915365072

Negative correlation Reject null hypothesis

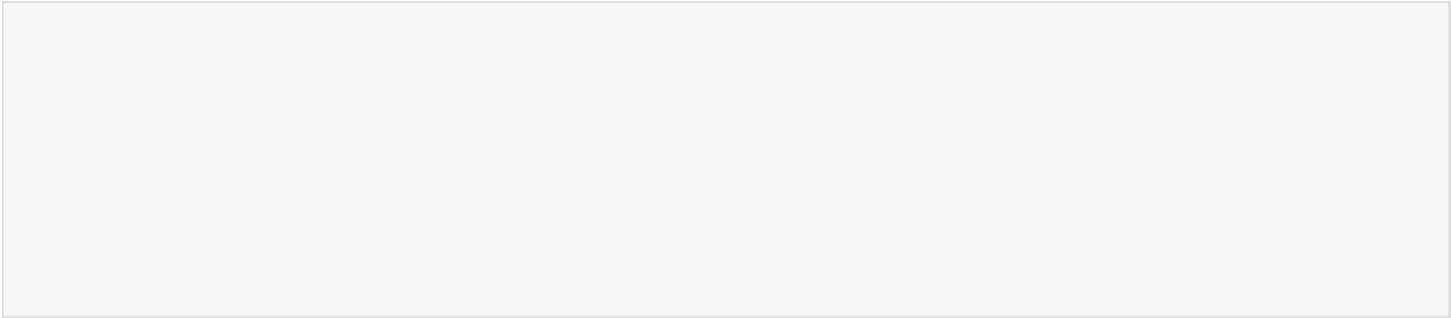
Out[54]:

<matplotlib.collections.PathCollection at 0x7fa45375b320>



# Attrition and YearsSinceLastPromotion

In [55]:



stats8,p8 = pearsonr(data.Attrition,data.YearsSinceLastPromotion) print("Correlation: ",stats8,"**\n** P-value: ",p8)

**if** stats8 == 0 : print("**\n\n**No Correlation")

**elif** stats8 > 0 : print("**\n\n**Positive correlation")

**else**:

print("**\n\n**Negative correlation") **if** p8<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.YearsSinceLastPromotion)

Correlation: -0.03301877514258437

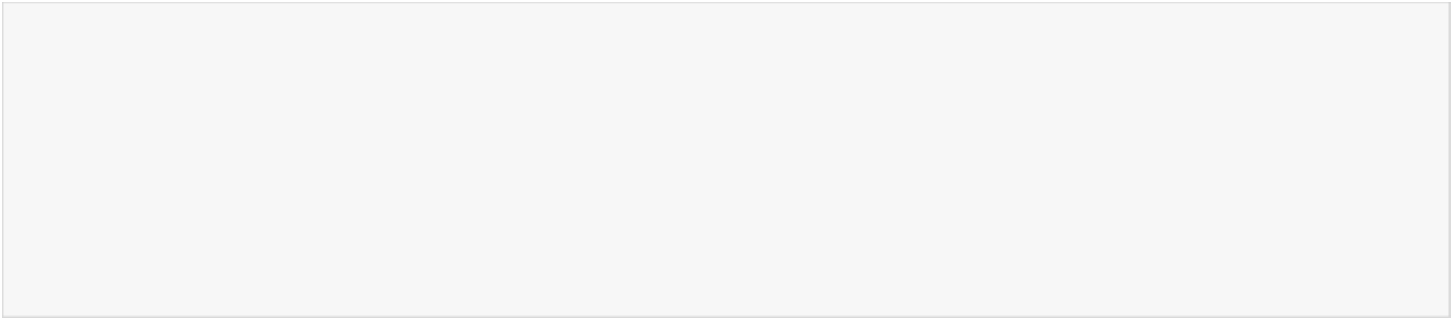
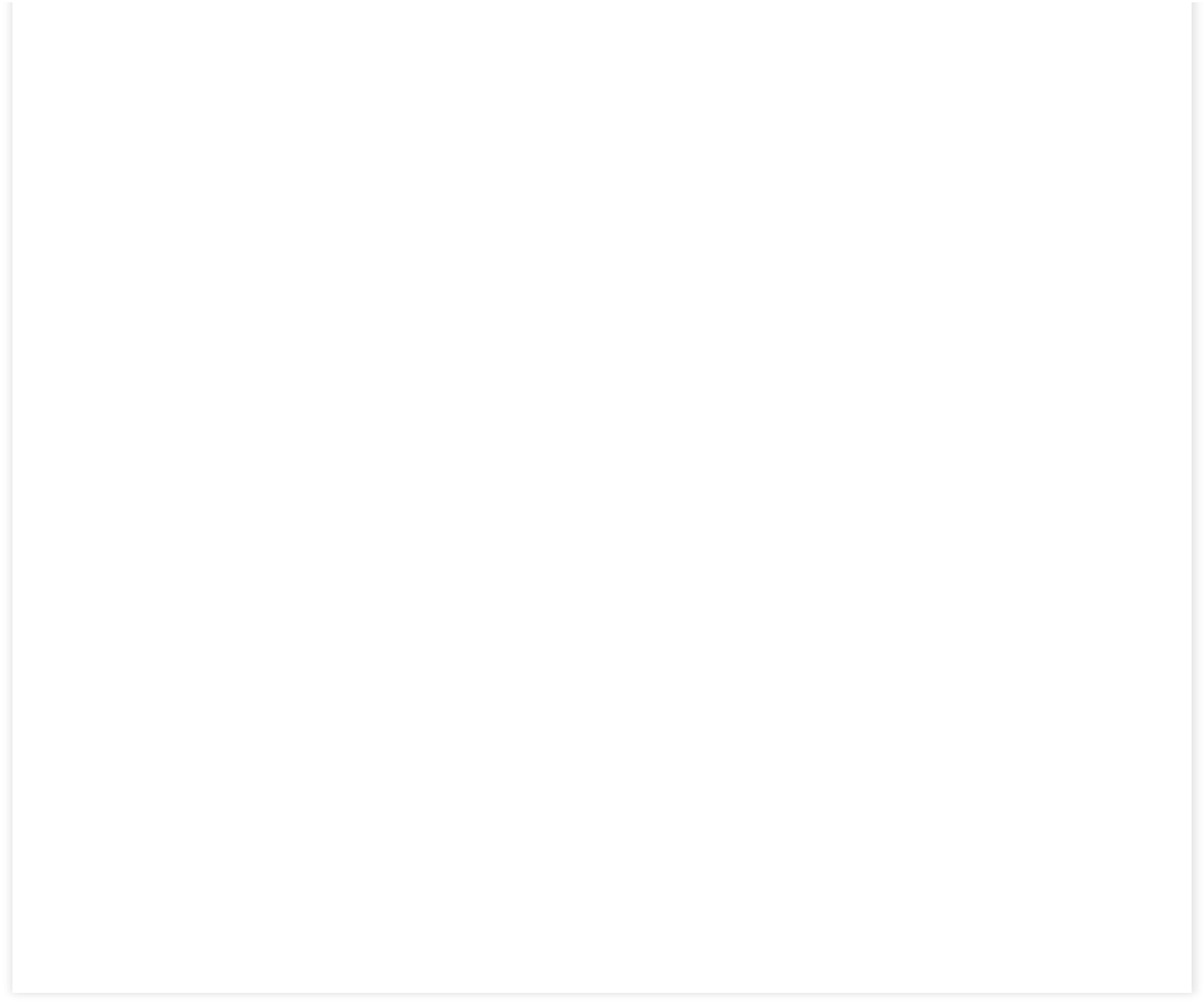
P-value: 0.028330336189396753

Negative correlation Reject null hypothesis

Out[55]:

<matplotlib.collections.PathCollection at 0x7fa454236438>





# Attrition and YearsWithCurrManager

In [56]:

stats9,p9 = pearsonr(data.Attrition,data.YearsWithCurrManager) print("Correlation: ",stats9,"**\n** P-value: ",p9)

**if** stats9 == 0 : print("**\n\n**No Correlation")

**elif** stats9 > 0 : print("**\n\n**Positive correlation")

## else:

print("**\n\n**Negative correlation") **if** p9<0.05:

print("Reject null hypothesis") **else**:

print("Accept null hypothesis") plt.scatter(data.Attrition,data.YearsWithCurrManager)

Correlation: -0.15619931590162806 P-value: 1.7339322652900218e-25

Negative correlation Reject null hypothesis

Out[56]:

<matplotlib.collections.PathCollection at 0x7fa4543d9208>