The Outline

1. **Introduction**
   - The project Goal
   - The key steps implemented

2. **Method and analysis**
   - Data preparation
   - The population sample
   - Data exploration and visualization
   - Pearson correlation Analysis
     - A Brief about the Method
     - Analyzing the distribution
     - Plotting the variables

3. **The result**

4. **Conclusion**

---

1. Introduction

The project Goal

Finding the relation between the emotional intelligence and the job performance of the teachers using the Pearson correlation method.

- H_0: there is no correlation between emotional intelligence and job performance
- H_1: there is a correlation between emotional intelligence and job performance

The key steps implemented

- Collect and prepare the data via surveys
- Process the data:
  Missing values
  Outliers
  Data calculations
- Explore the Data
- Perform in-depth Analysis using statistical software
- Communicate the results of the analysis

2. Method and analysis

Data preparation

The data was collected through two main surveys each consists of 10 questions related to two topics: emotional intelligence and job performance assessment. The candidate of the first assessment was the teachers in the school, where they had to answer questions related to their emotional intelligence using likert scale. On the other hand, the second group was the supervisors of which the teachers report to.

The population sample

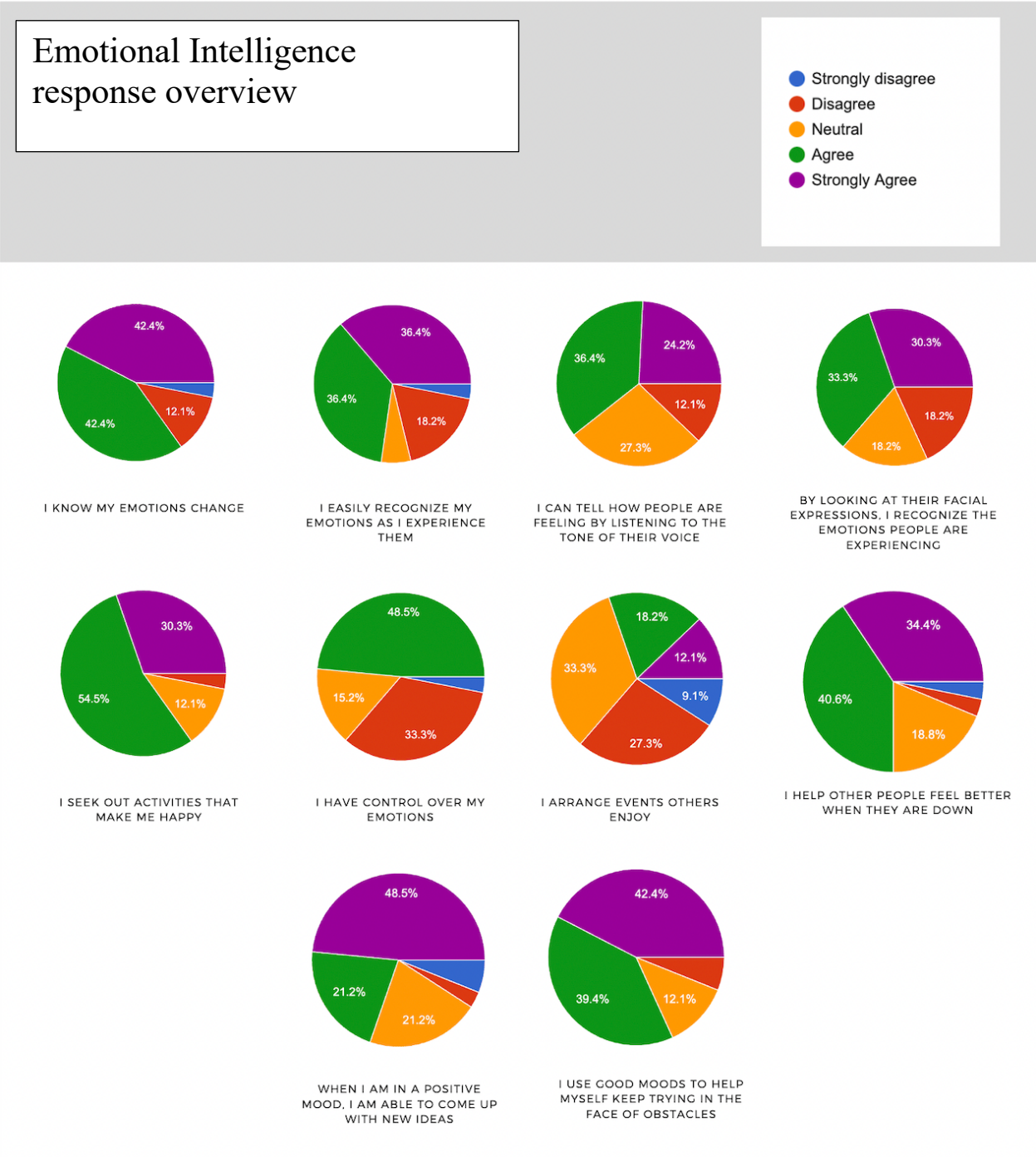
The data was collected from distinct group of 31 teachers and supervisor, note that some teachers in this group reports to the same supervisor.

Data exploration and visualization

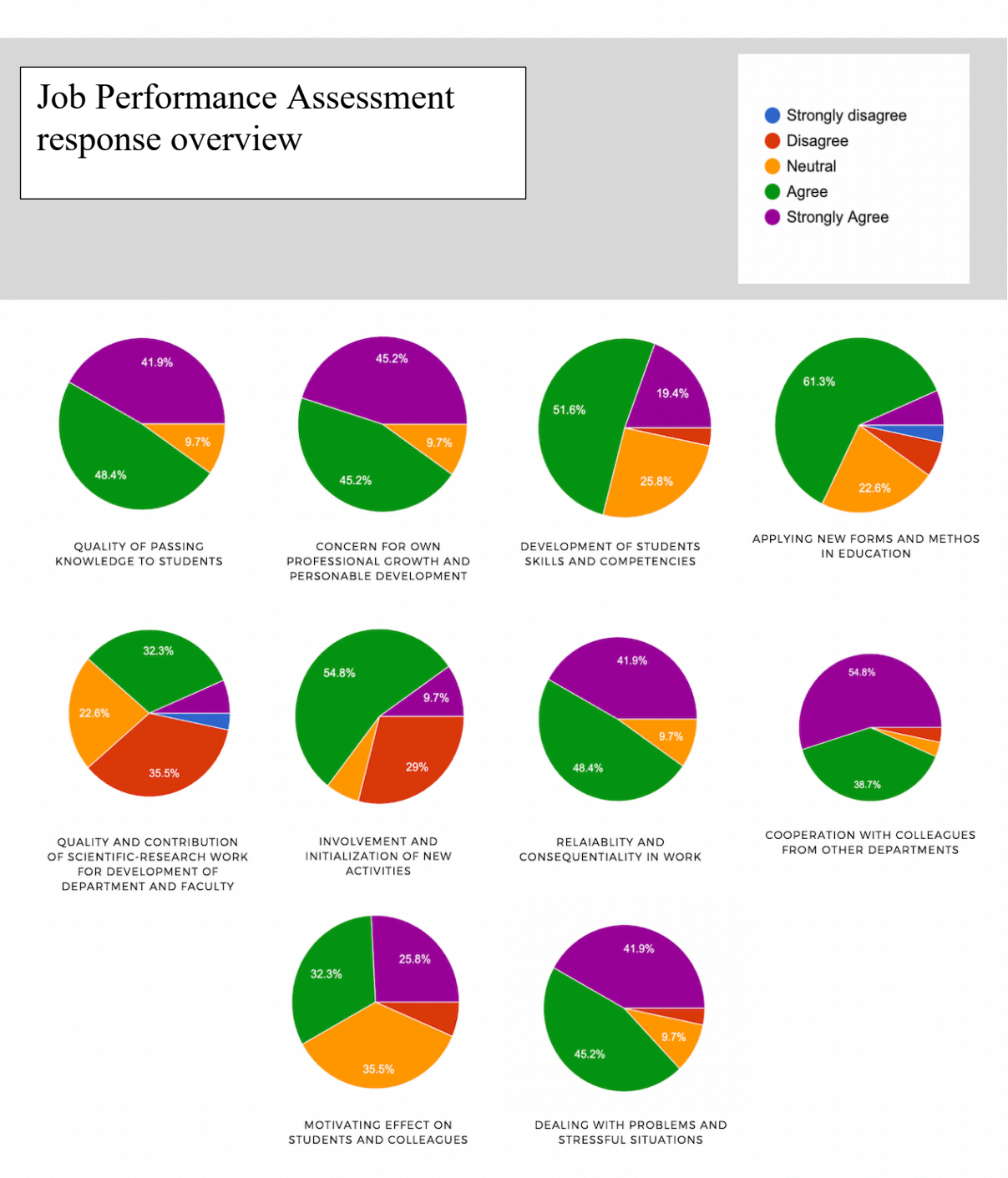
We are exploring the data using visualization tool (Python/seaborn library) and hypothesis testing to provide summary level insight into a dataset and uncover underlying patterns.

The main charts used in this project are the below:
Pie chart : to show data proportion
Density Chart: to observe and assess the distribution of our data
Scatter plot : to visualize the correlation between the variables of the bivariate data

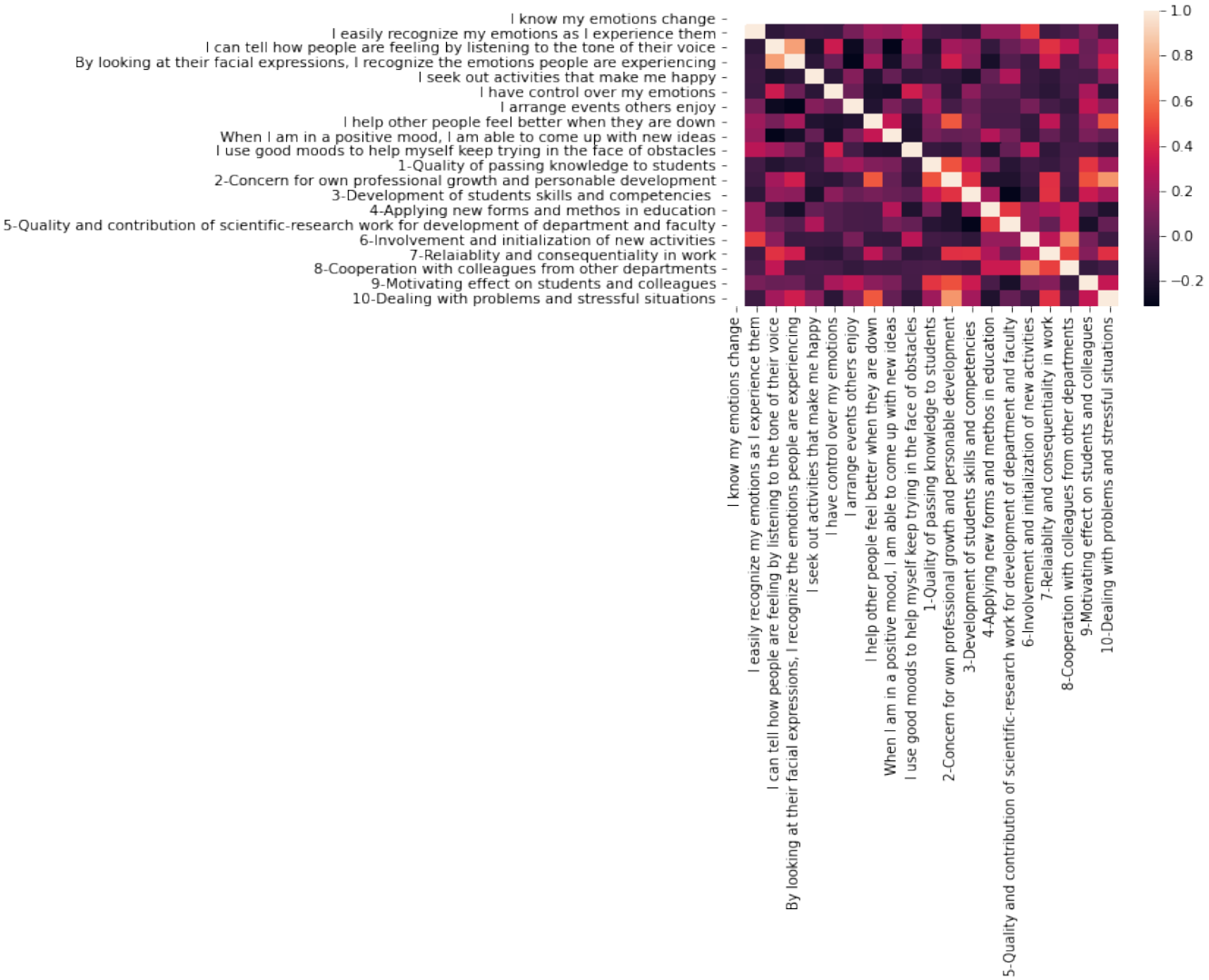
We can clearly observe that the positive response is dominant for most of the questions to the teachers related to their emotions, however once it comes to group activities we can notice that the answers where more conservative.



The overall pattern of responses of the supervisors is also positive as we can see below, which indicate that there could be a potential correlation between the responses of both groups



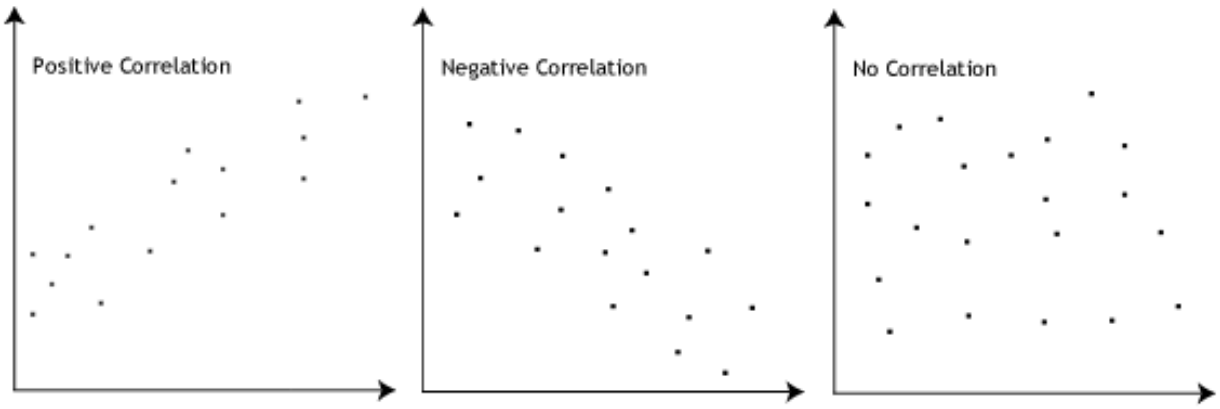
The below heatmap shows the relation of all the responses among each other for both groups, in order to show the correlation for each response in relation to all the different responses presented.



Pearson Correlation Analysis

A brief about the method

In brief the Pearson Correlation (r) measures the extent of correlation between two variables on a scale range from -1 to +1, r is calculated based on the mathematical equation below:



Regardless of the strength of the relation, To measure association or relationship between variables we use correlation coefficients to express the degree of association. There are three major status that can result from this test, r = 1 then it is a perfect Positive Correlation or r = -1 then it is a perfect negative Correlation or r = 0 then it implies that there is No Correlation between the variables. Below is a representation of the relations using scatter plots.

Pearson correlation formula:

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}}$$

mx and m_y are the means of x and y variables.
The p-value (significance level) of the correlation can be determined :

1. by using the correlation coefficient table for the degrees of freedom : $df = n - 2$, where n is the number of observation in x and y variables.
2. or by calculating the **t value** as follow:

Analyzing the distribution

The assumptions for Pearson correlation coefficient are as follows: level of measurement, related pairs, absence of outliers, normality of variables, linearity, and homoscedasticity.

Level of measurement refers to each variable. For a Pearson correlation, each variable should be continuous. If one or both of the variables are ordinal in measurement, then a other correlation method could be conducted instead.

Related pairs refers to the pairs of variables. Each participant or observation should have a pair of values.

Absence of outliers refers to not having outliers in either variable. Having an outlier can skew the results of the correlation by pulling the line of best fit formed by the correlation too far in one direction or another. Typically, an outlier is defined as a value that is 3.29 standard deviations from the mean, or a standardized value of less than ±3.29.

Linearity and homoscedasticity refer to the shape of the values formed by the scatterplot. For linearity, a “straight line” relationship between the variable should be formed. If a line were to be drawn between all the dots going from left to right, the line should be straight and not curved. Homoscedasticity refers to the distance between the points to that straight line. The shape of the scatterplot should be tube-like in shape. If the shape is cone-like, then homoskedasticity would not be met.

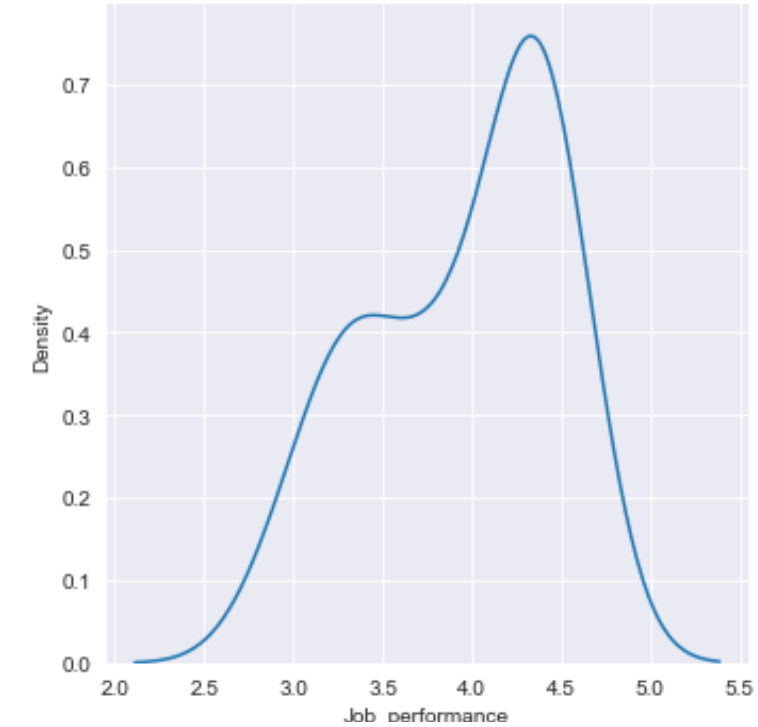
Plotting the Variables

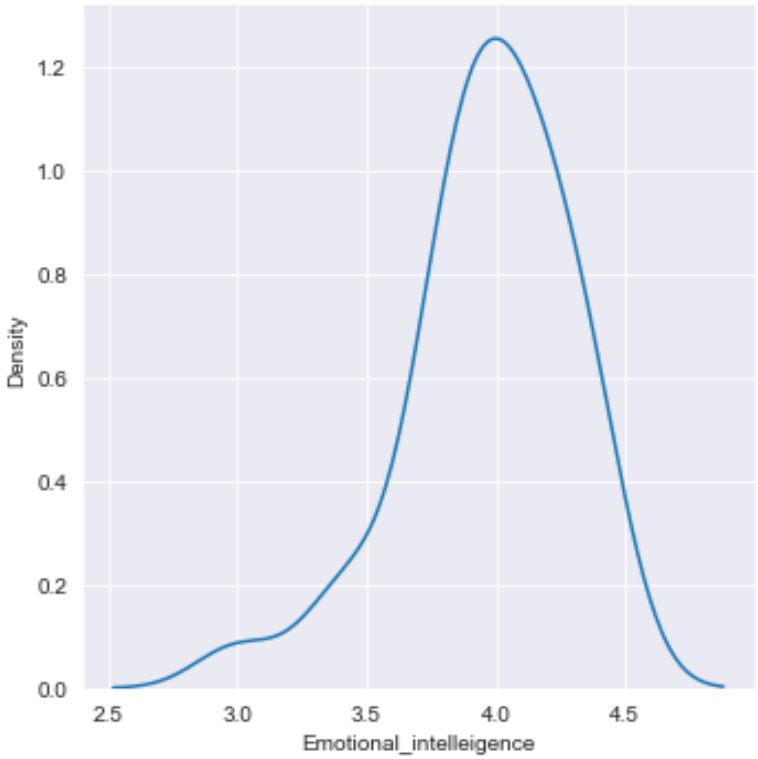
The data was tested and the abnormality was treated by removing the outliers, our data fits all the assumptions for Pearson correlation, and that can be shown by plotting the respective charts which will reveal various insights of the variables and their correlations.

In general Pearson correlation does not assume normality, even though it can violate the tests of significance If the data are badly skewed, bimodal, or otherwise violate the assumptions of the general linear model then it is better to use another ranked correlation method (non- parametric) in order to stay free from any unexpected correlations obtained from Pearson correlation analysis.

We can check the normality of the data by plotting a density graph which will show clearly the distribution of the data, hence we can conclude the normality.

The normal distribution is the most common type of distribution, is the proper term for a probability bell curve. That shows how the data is being distributed where most of the observations are clustered in the central peak.





In our graphs we notice that both curves are skewed to the left and that is due to the tendency of the positive response for both the teachers and the supervisor.

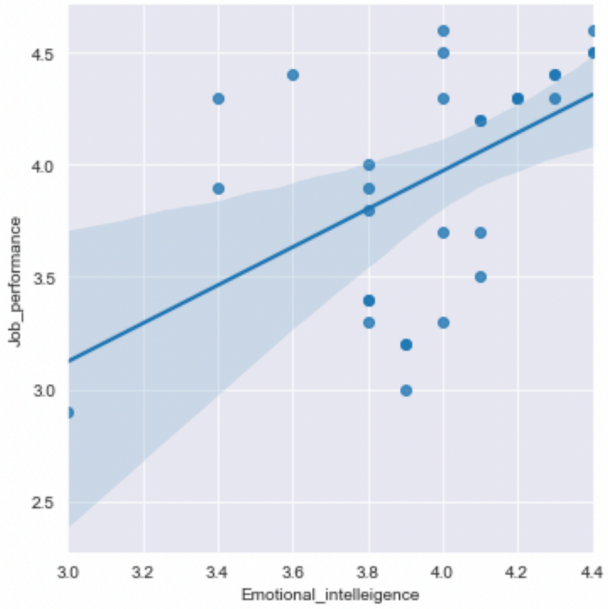
The distribution appear to be sufficient to apply Pearson correlation method and it doesn't violate any assumptions

The hypotheses:
H0: There is no correlation between emotional intelligence and job performance
H1: There is a correlation between emotional intelligence and job performance

α = 0.005 – is the significance level which indicates that the risk of concluding that a correlation exists
P-value= 0.002826

We can see that the trend of the data is positive, the higher the emotional intelligence the higher is the job performance.

```
In [54]: ax=sns.lmplot(x="Emotional_intelleigence", y="Job_performance", data=df)
```




```
In [55]: corr(df0['Emotional_intelleigence'], df0['Job_performance'])
```

Out[55]:

	n	r	CI95%	p-val	BF10	power
pearson	31	0.518186	[0.2, 0.74]	0.002826	15.795	0.868223

3. The result

The result of the correlation function as presented above shows that the p-value is less than 0.005, which indicates that we can reject the null hypothesis and conclude that there is a significant positive correlation between emotional intelligence and the job performance of the teacher

4. The conclusion

We can conclude that