**Problem Statement**- A large company named XYZ, employs, at any given point of time, around 4000 employees. However, every year, around 15% of its employees leave the company and need to be replaced with the talent pool available in the job market. The management believes that this level of attrition (employees leaving, either on their own or because they got fired) is bad for the company, because of the following reasons -

The former employees’ projects get delayed, which makes it difficult to meet timelines, resulting in a reputation loss among consumers and partners.

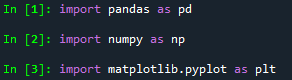
A sizeable department has to be maintained, for the purposes of recruiting new talent. More often than not, the new employees have to be trained for the job and/or given time to acclimatise themselves to the company.

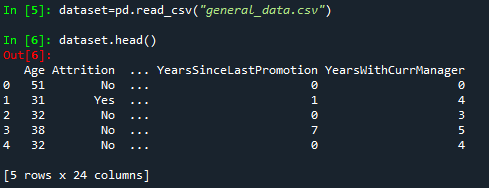
Hence, the management has contracted an HR analytics firm to understand what factors they should focus on, in order to curb attrition. In other words, they want to know what changes they should make to their workplace, in order to get most of their employees to stay. Also, they want to know which of these variables is most important and needs to be addressed right away.

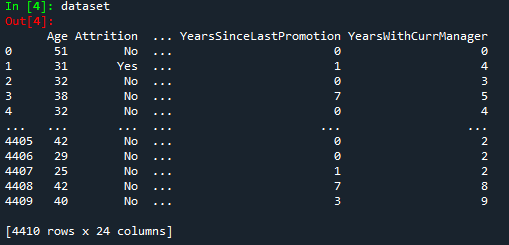
Since you are one of the star analysts at the firm, this project has been given to you.

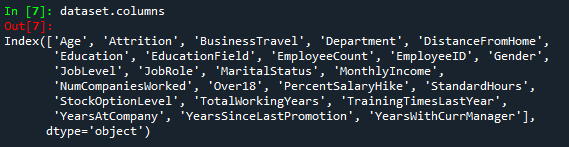
Goal of the case study You are required to model the probability of attrition. The results thus obtained will be used by the management to understand what changes they should make to their workplace, in order to get most of their employees to stay.

**Step 1- Launching**

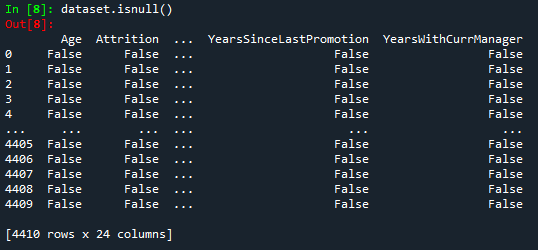




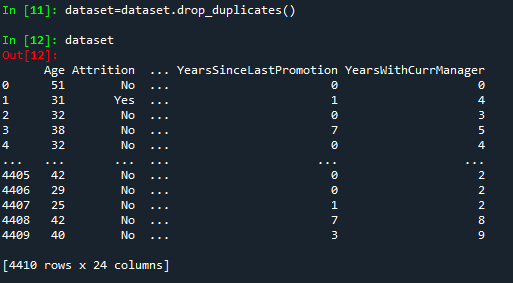
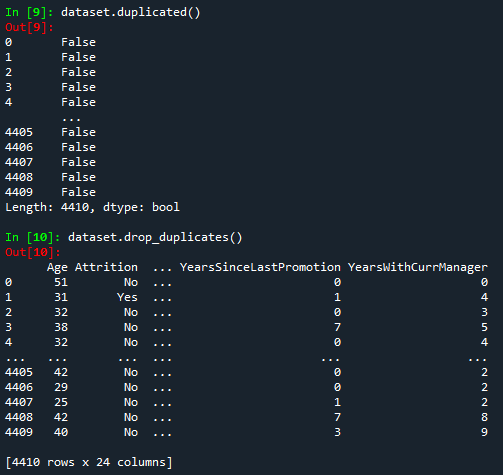
****



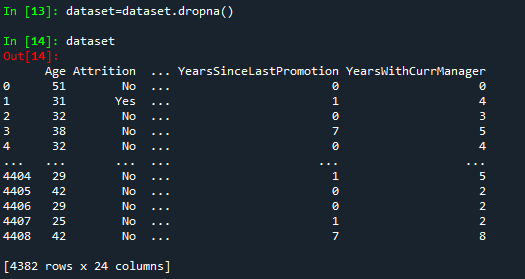
**Step 2- Data Treatment**



Checked and Deleted Duplicate Values (here no duplicates found)

****

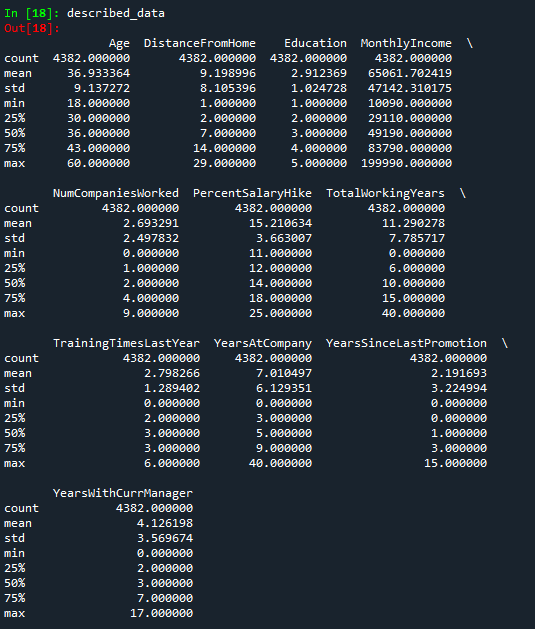
Deleted Null Values



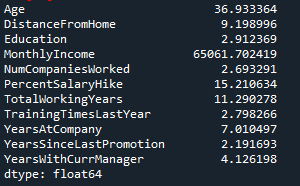
**Step 3- Univariate Analysis**

**C:\Users\Chaitanya\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot (33).png**

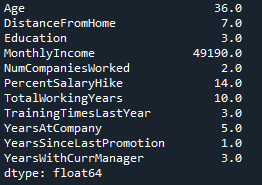
Described Data-



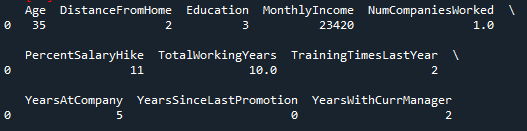
Mean-



Median-



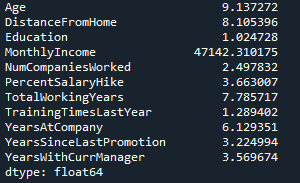
Mode-



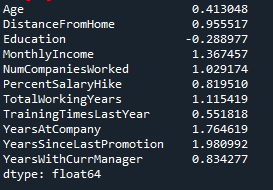
Variance-



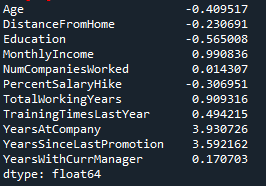
Std Deviation-



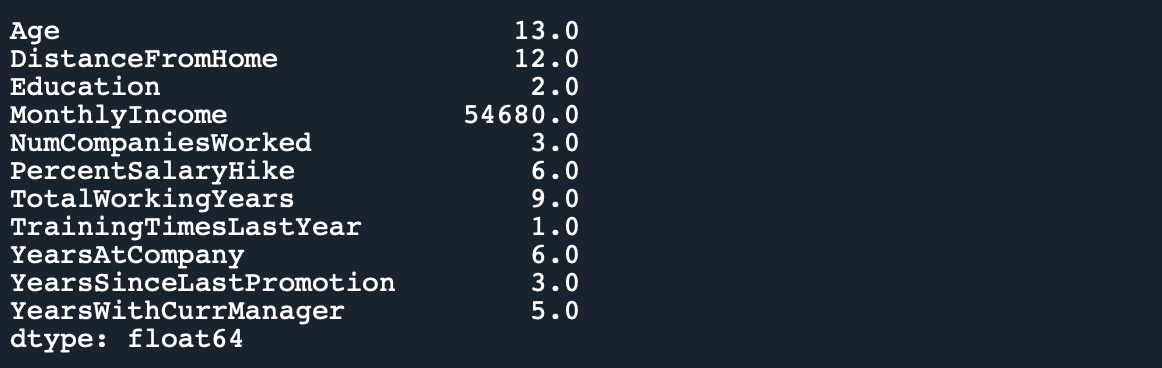
Skewness-



Kurtosis-



IQR-

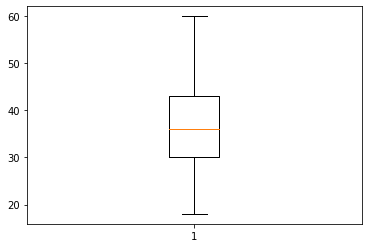
****

**Step 4- Inference from the Analysis**

* All the above variables are positively skewed (mean > median) except Education which is negatively skewed.
* Age, DistanceFromHome, Education and PercentSalaryHike are platykurtic in nature while all the other variables are leptokurtic.
* The MonthlyIncome’s IQR is at 54K suggesting companywide attrition across all income bands.
* Age forms a near normal distribution with 13 years of IQR.

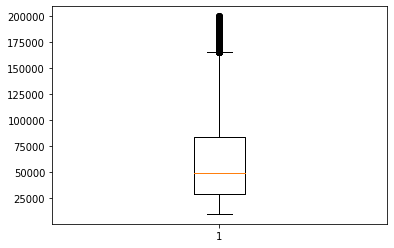
**Step 5- Outliers**

Age-

****

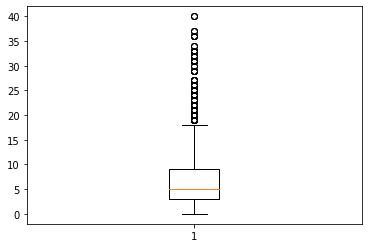
Age is normally distributed without any outliers.

Monthly Income-



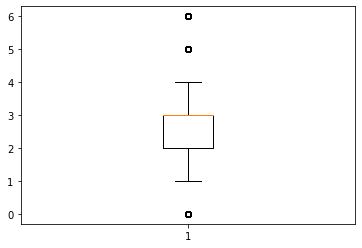
Monthly Income is positively skewed with several outliers.

Years at Company-



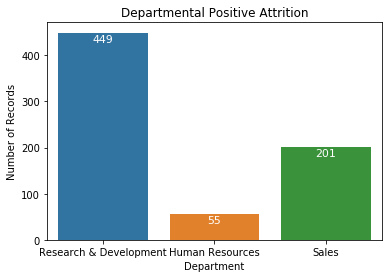
Years at Company is positively skewed with several outliers.

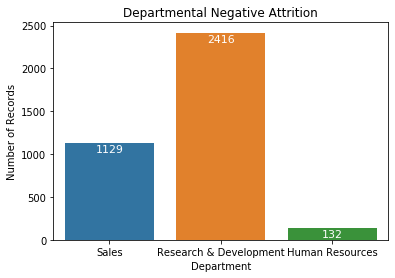
Training Times Last Year-



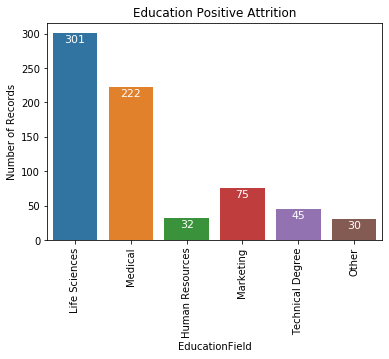
Training Times Last Year is negatively skewed with some outliers.

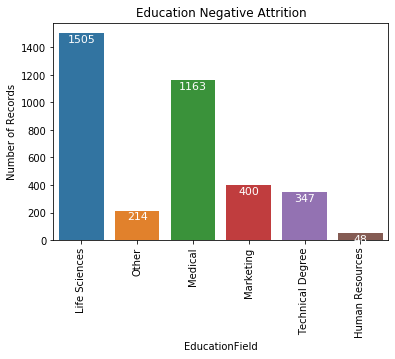
**Step 6- Visualisation**





Visualisation of each department with positive and negative attrition.



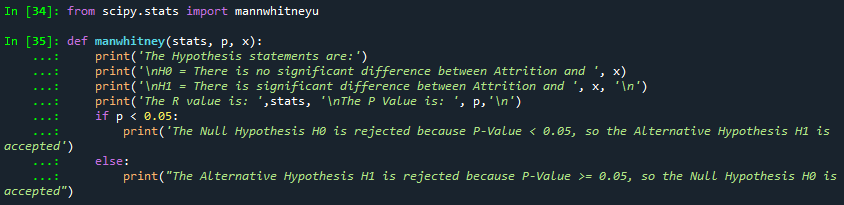


Visualisation of each department with positive and negative attrition.

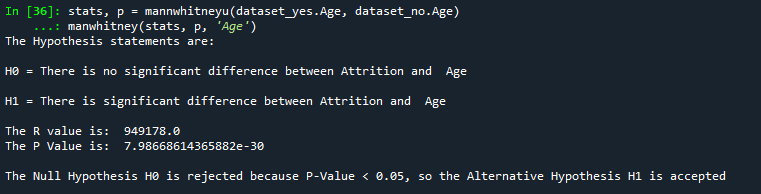
### **Step 7 - Statistical Tests**

Mann-Whitney Test-

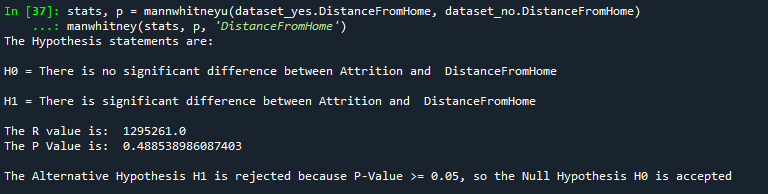
Imported mannwhitneyu and defined a function for hypothesis testing.



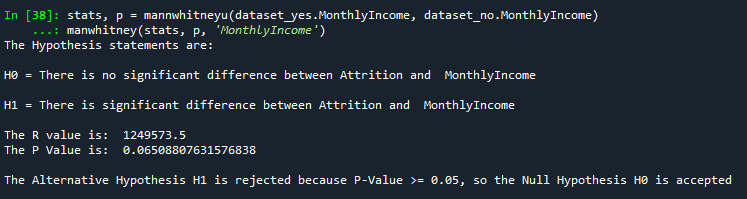
1. Attrition and Age-



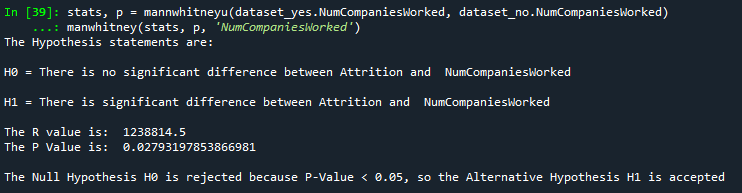
1. Attrition and Distance from home-



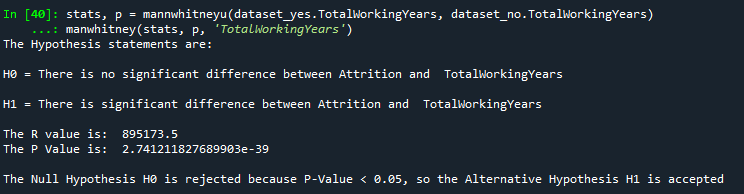
1. Attrition and Monthly Income-



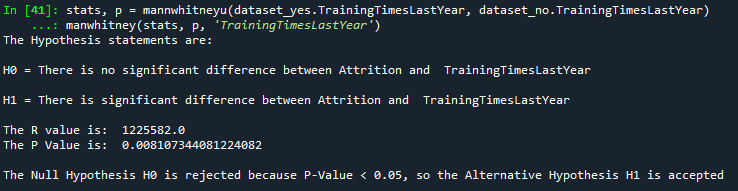
1. Attrition and Number of companies worked-



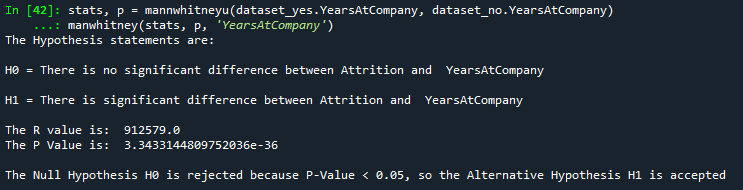
1. Attrition and Total working years-



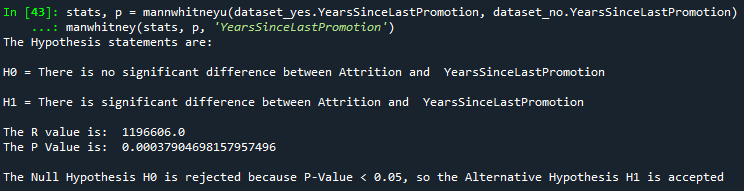
1. Attrition and Training times last year-



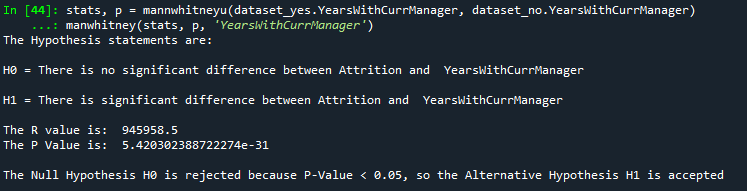
1. Attrition and Years at company-



1. Attrition and Years since last promotion-

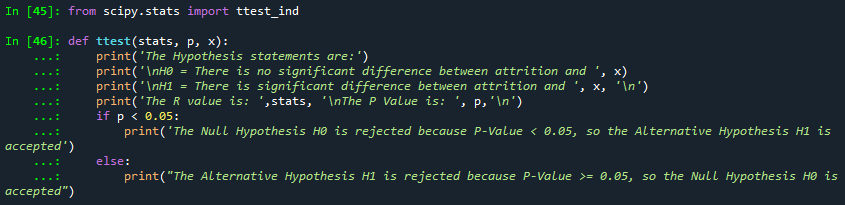


1. Attrition and Years with current manager-

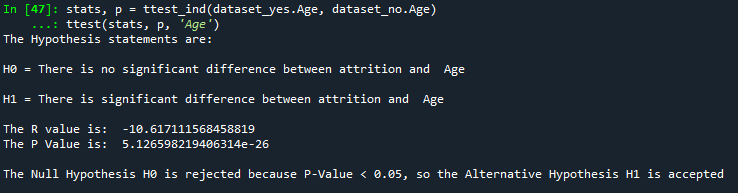


Separate T Test-

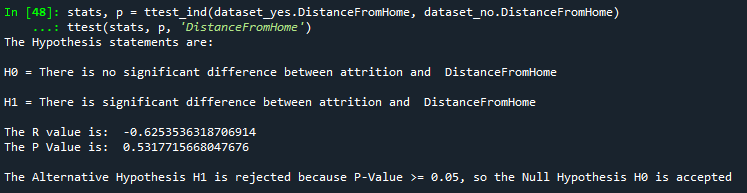
Imported ttest\_ind and defined a function for hypothesis testing.



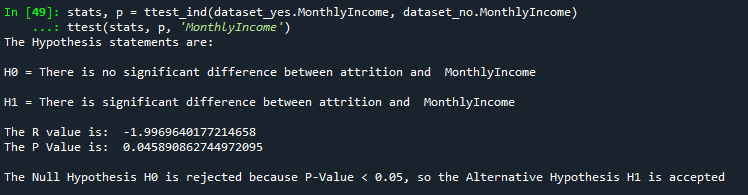
1. Attrition and Age-



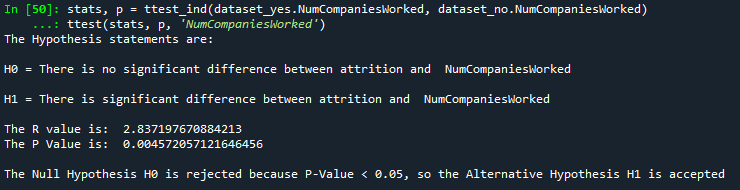
1. Attrition and Distance from home-



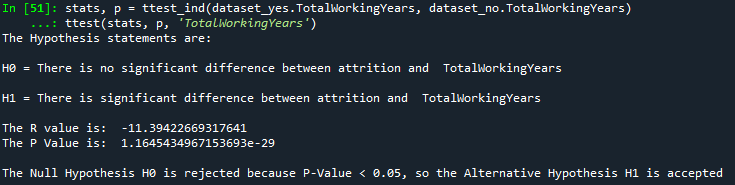
1. Attrition and Monthly Income-



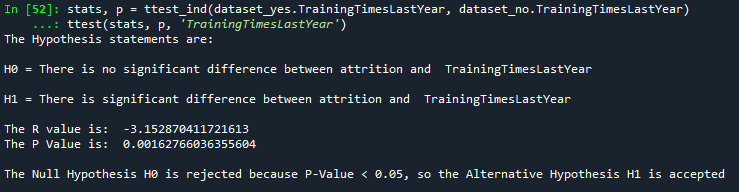
1. Attrition and Number of companies worked-



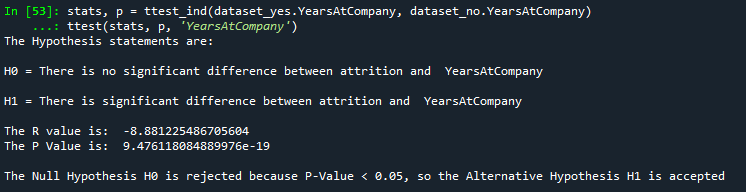
1. Attrition and Total working years-



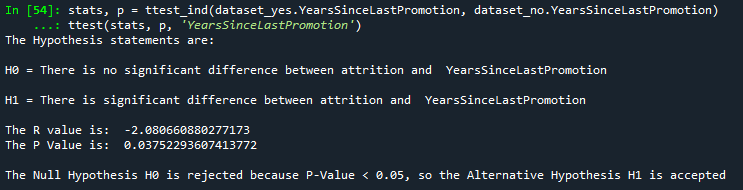
1. Attrition and Training times last year-



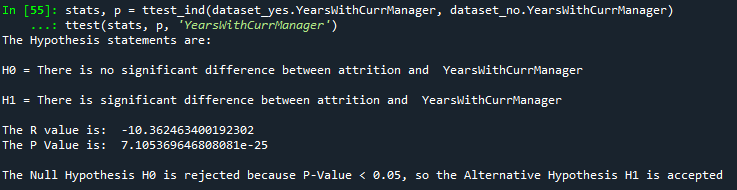
1. Attrition and Years at company-



1. Attrition and Years since last promotion-

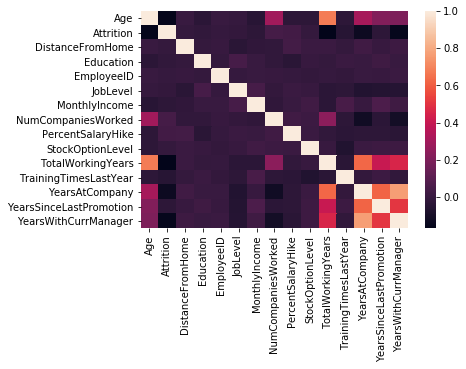


1. Attrition and Years with current manager-



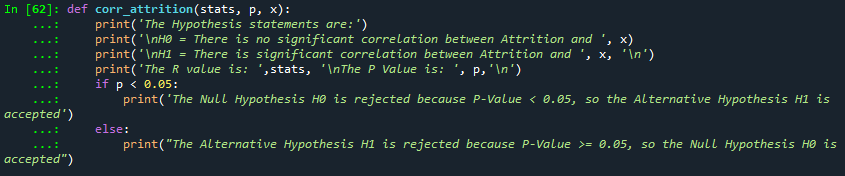
### **Step 8 – Unsupervised Learning – Correlation Analysis**

Standard Hours and Employee Count has no impact in correlation as they have same values in all records. So, we remove the columns and plot a graph to have an idea on correlation of variables.

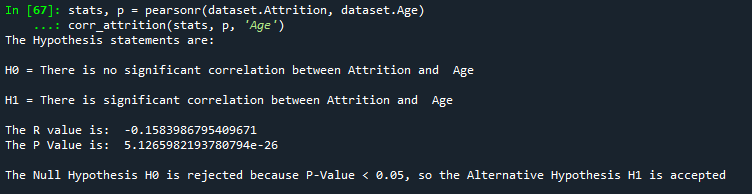


Imported pearsonr and defined a function for calculating the correlation.

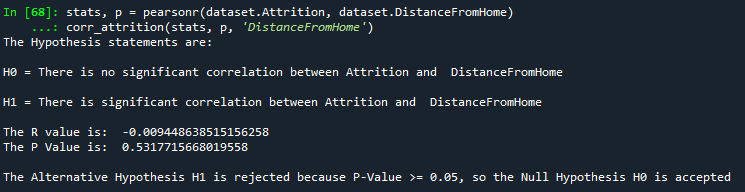




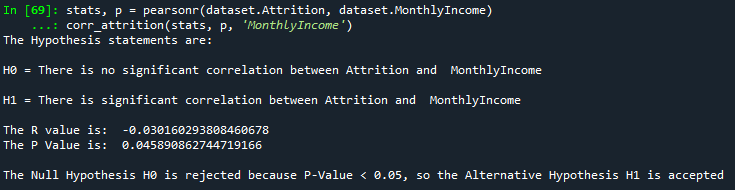
* Attrition and Age Correlation-



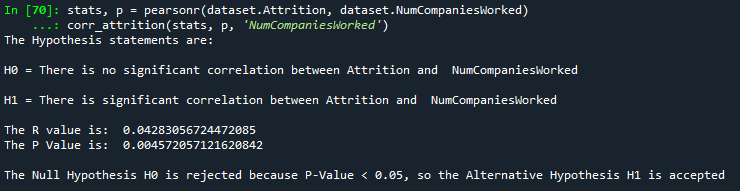
* Attrition and Distance from home Correlation-



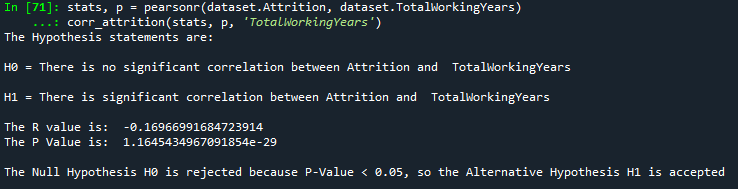
* Attrition and Monthly Income Correlation-



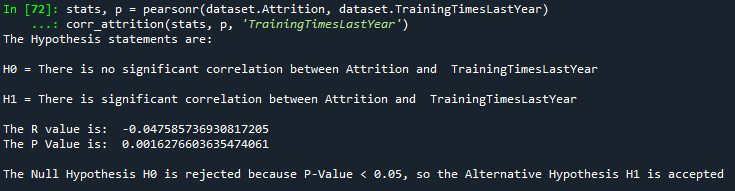
* Attrition and Number of companies worked Correlation-



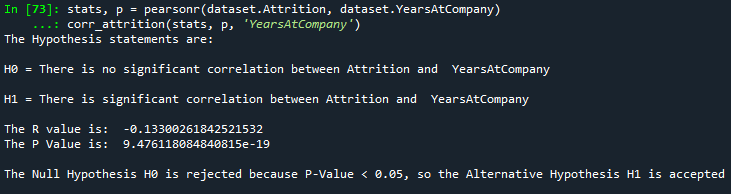
* Attrition and Total working years Correlation-



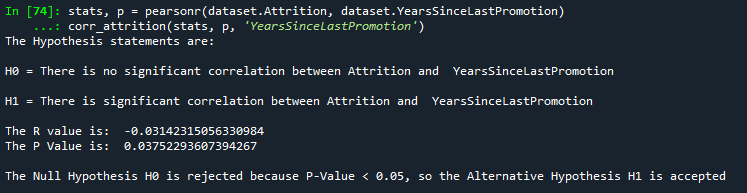
* Attrition and Training times last year Correlation-



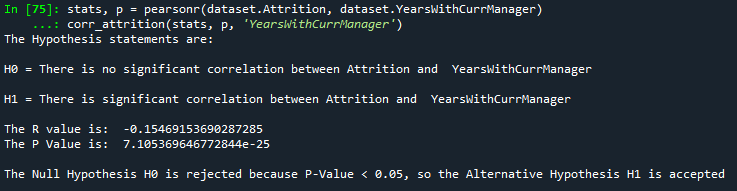
* Attrition and Years at company Correlation-



* Attrition and Years since last promotion Correlation-



* Attrition and Years with current manager Correlation-



### **Inference from above Analysis-**

1. **Attrition and Age**- As r = -0.1583, there is low negative correlation between Attrition and Age. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and Age.
2. **Attrition and Distance from home**- As r = -0.0094, there is low negative correlation between Attrition and DistanceFromHome. As the P value is > 0.05, the null hypothesis is accepted, so there is no significant correlation between Attrition and DistanceFromHome.
3. **Attrition and Monthly Income-** As r = -0.0301, there is low negative correlation between Attrition and Monthly Income. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and Monthly Income.
4. **Attrition and Num of companies worked**: As r = 0.0428, there is low positive correlation between Attrition and Num of companies worked. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and Num of companies worked.
5. **Attrition and Total working years-** As r = -0.1696, there is low negative correlation between Attrition and TotalWorkingYears. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and TotalWorkingYears.
6. **Attrition and Training times last year**- As r = -0.0475, there is low negative correlation between Attrition and TrainingTimesLastYear. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and TrainingTimesLastYear.
7. **Attrition and Years at company**- As r = -0.1330, there is low negative correlation between Attrition and YearsAtCompany. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and YearsAtCompany.
8. **Attrition and Years since last promotion**- As r = -0.0314, there is low negative correlation between Attrition and YearsSinceLastPromotion. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and YearsSinceLastPromotion.
9. **Attrition and Years with current manager**- As r = -0.1546, there is low negative correlation between Attrition and YearsWithCurrManager. As the P value is < 0.05, the null hypothesis is rejected, so there is significant correlation between Attrition and YearsWithCurrManager.

**Conclusion-**

From the above Statistical Tests and Correlation Analysis, we can say that there are few factors which are related to Attrition and rest have no significance for Attrition. From the above Inferences, we can conclude that the Company has to make following changes to reduce the number of Attrition in the Company.

1. Hire middle aged employees having age of approximately 36 years old or above.
2. Reduce the number of business trips of employees.
3. Other departments should be more familiar with HR department.
4. All departments should have more skilled and promising employees.
5. Creating familiar environment with employees so they don't leave the job.
6. Hiring well experienced employees and train appropriate skills to the freshers.
7. Hire more married people who are aware of the responsibilities.
8. Don’t change the manager at a frequent interval.