**Objective :**

This analysis is done to understand the factors which are responsible for employees of a given company to leave prematurely. This will help the company to take preventive steps to retain productive and valuable employees.

**Obtaining and Understanding the Data :**

*# Import required libraries for Data Analysis*

**import** numpy **as** np  
**import** pandas **as** pd  
pd.set\_option(**'display.max\_columns'**, 500)  
pd.set\_option(**'display.max\_rows'**, 500 )

*# Read the csv file and store the data in a dataframe*dataset = pd.read\_csv(**"Employee\_Details.csv"**)  
  
*# Get count of rows and columns*dataset.shape  
  
*# View top few rows of data*dataset.head()  
  
*# Datatype of each column*dataset.dtypes

**Data Insight :**

There are total 14999 rows and 10 columns in this data set. There are 2 categorical features (namely salary and job\_role) and 8 quantitative features.

**Cleaning the Data :**

*# Rename columns for better readability*dataset.columns = [**"satisfaction\_level"**,**"last\_evaluation"**,**"project\_count"**,

**"average\_monthly\_hours"**,**"exp\_in\_company"**,**"work\_accident"**,**"left\_company"**,**"promotion\_last\_5years"**,**"job\_role"**,**"salary"**]  
  
*# Change datatype of categorical data*dataset.job\_role = dataset.job\_role.astype(**"category"**)  
dataset.salary = dataset.salary.astype(**"category"**)  
dataset.dtypes  
  
*# Check for columns with missing values*dataset.isnull().any()  
  
*# Replace missing values with zero*dataset.satisfaction\_level.fillna(0)

**Data Insight :**

# satisfaction\_level has null values which are replaced with 0 and columns are renamed for better readability.

**Exploring the Data :**

*# Summary of numeric columns*dataset.describe()

satisfaction\_level last\_evaluation project\_count average\_monthly\_hours exp\_in\_company work\_accident

count 14999.00 14999.00 14999.000 14999.000000 14999.000000 14999.000000

mean 0.612834 0.716102 3.803054 201.050337 3.498233 0.144610

std 0.248631 0.171169 1.232592 49.943099 1.460136 0.351719

min 0.090000 0.360000 2.000000 96.000000 2.000000 0.000000

25% 0.440000 0.560000 3.000000 156.000000 3.000000 0.000000

50% 0.640000 0.720000 4.000000 200.000000 3.000000 0.000000

75% 0.820000 0.870000 5.000000 245.000000 4.000000 0.000000

max 1.000000 1.000000 7.000000 310.000000 10.000000 1.000000

left\_company promotion\_last\_5years

count 14999.000000 14999.000000

mean 0.238083 0.021268

std 0.425924 0.144281

min 0.000000 0.000000

25% 0.000000 0.000000

50% 0.000000 0.000000

75% 0.000000 0.000000

max 1.000000 1.000000

*#Summary of categorical columns*dataset.salary.describe()  
dataset.job\_role.describe()

Output :

count 14999

unique 3

top low

freq 7316

Name: salary, dtype: object

count 14999

unique 10

top sales

freq 4140

Name: job\_role, dtype: object

*# Percentage of employee who left the company*perc\_employee\_left = np.sum(dataset.left\_company/len(dataset.left\_company))\*100  
print(**'Percentage of employee left = {0} %'**.format(round(perc\_employee\_left)))

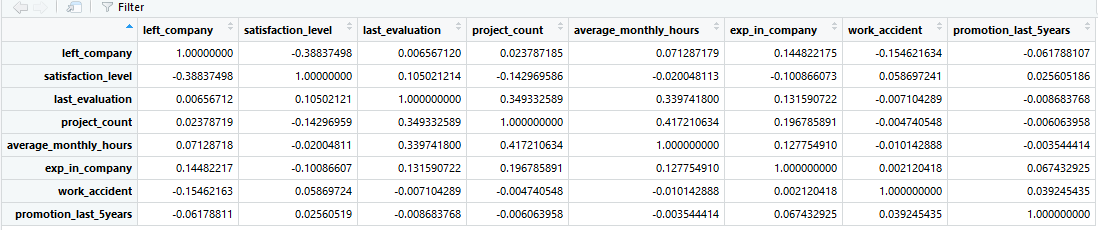
**Data Insight :**

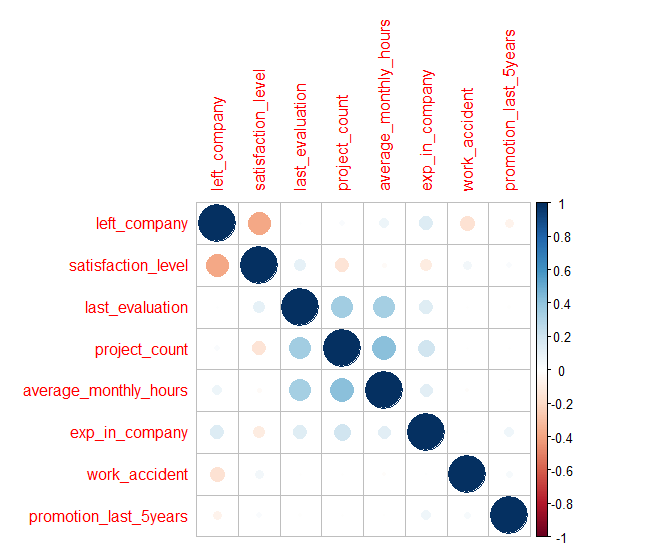
About 24% people left the company.

**Correlation among features :**

*# Correlation Matrix*corr = dataset.corr()  
print(corr)  
sns.heatmap(corr,xticklabels=corr.columns.values,yticklabels=corr.columns.values)

Output :





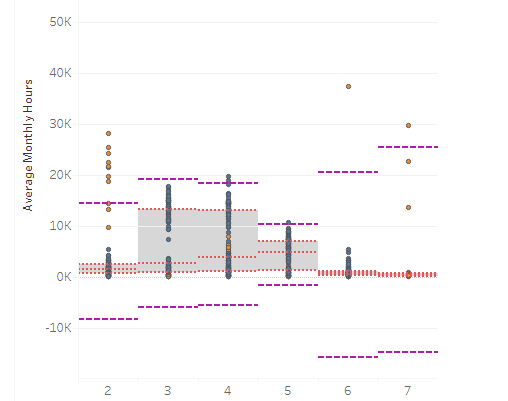
**Data Insight :**

**Positively correlated features :** Project\_count and average\_monthly\_hours are positively correlated to last\_evalulation. This means employees who spend more time on projects and allocated more projects and are evaluated higher.

Project count is also positively correlated to average\_monthly\_hours which means employees having more number of projects spend more hours working.

**Negatively correlated features :** Left\_company and satisfaction\_level are highly negatively correlated. This means less satisfied employees tend to leave a company more.

**Project count vs average monthly hours :**



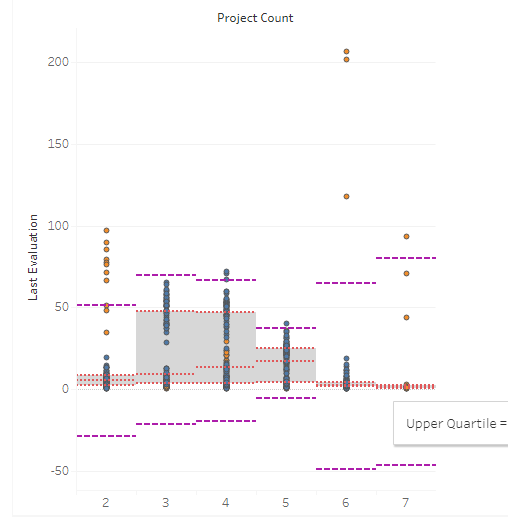
**Link to Tableau public :**

https://public.tableau.com/profile/mehak.narang6693#!/vizhome/EmployeeRetentionAnalysis/ProjectCountVsMonthlyHours

**Data Insight :**

For employees who left the company, average monthly working hours increase with the increase in number of projects.

**Project count vs Evaluation:**



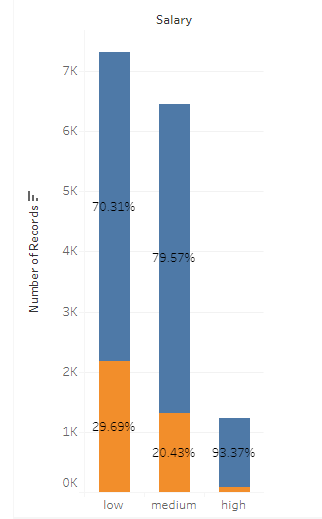
**Link to Tableau public :**

https://public.tableau.com/profile/mehak.narang6693#!/vizhome/EmployeeRetentionAnalysis/ProjectCountVsEvaluation

**Data Insight :**

More employees left the company who are evaluated higher and have done more number of projects.

**Employee Salary vs Left\_Company :**



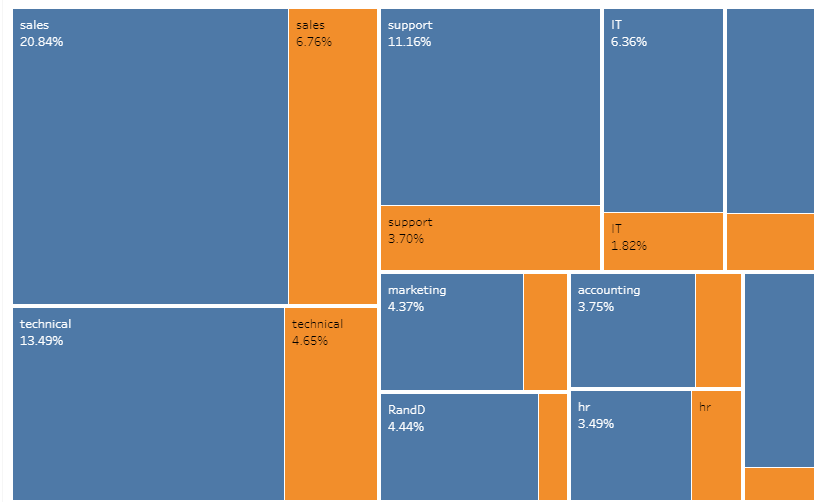
**Link to Tableau public :**

https://public.tableau.com/profile/mehak.narang6693#!/vizhome/EmployeeRetentionAnalysis/SalaryDistribution

**Data Insight :**

Employees with low and medium salary tend to leave the company. Very few employee with high salary left.

**Employee Department vs Left\_Company :**



**Link to Tableau public :**

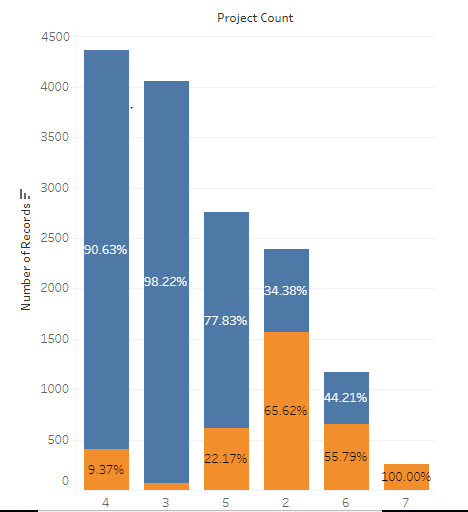
https://public.tableau.com/profile/mehak.narang6693#!/vizhome/EmployeeRetentionAnalysis/DepartmentDistribution

**Data Insight :**

Employee from Sales, technical and Support department tend to leave the company.

Employee from Management and Product Management department tend to stay in the company for a longer duration.

**Project count vs Left\_Company :**



**Link to Tableau public :**

https://public.tableau.com/profile/mehak.narang6693#!/vizhome/EmployeeRetentionAnalysis/ProjectCountDistribution

**Data Insight :**

All the employees with Project Count = 7 left the company.

More than 50% of the employee with Project Count = 2, 6 left the company.

**Measuring Satisfaction Level among employees :**

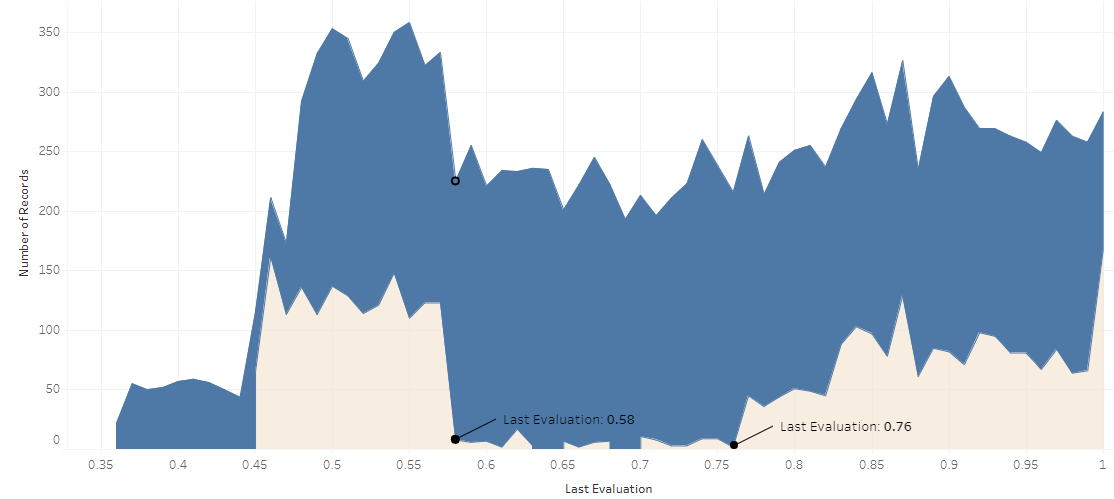
*# Comparing satisfaction among all employees against satisfaction among employees who left the company*perc\_employee\_satisfaction = sum(dataset.satisfaction\_level/len(dataset.satisfaction\_level))\*100  
print(**"Percentage satisfaction among all employees : {0} %"**.format(round(perc\_employee\_satisfaction)))  
  
employee\_left = dataset[dataset.left\_company==1]  
perc\_employee\_left = sum(employee\_left.satisfaction\_level/len(employee\_left.satisfaction\_level))\*100  
print(**"Percentage satisfaction among employees who left the company : {0} %"**.format(round(perc\_employee\_left)))

**Data Insight :**

Employee having least satisfaction level ( less than 10% or between 30% to 50%) left the company more.

Many employees having very high satisfaction level (70% to 90%) also left the company.

**Evaluation Vs Left company :**



**Link to Tableau public :**

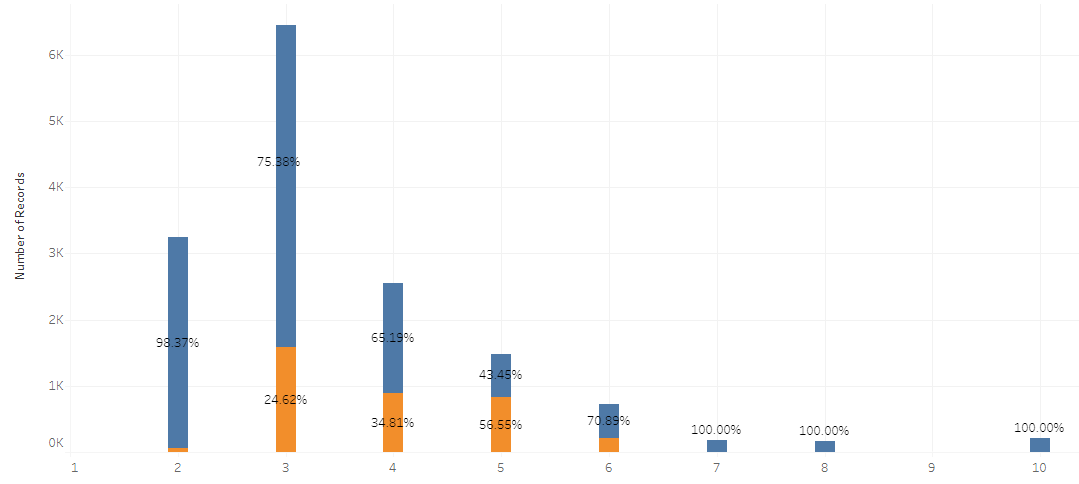
https://public.tableau.com/profile/mehak.narang6693#!/vizhome/EmployeeRetentionAnalysis/EvaluationDistribution

**Data Insight :**

Employee with low evaluation or very high evaluation left the company more.

Employee with average evaluation (46% to 70%) tend to stay in the company.

**Experience in company vs left company :**



**Link to Tableau public :**

https://public.tableau.com/profile/mehak.narang6693#!/vizhome/EmployeeRetentionAnalysis/ExperienceinCompany

**Data Insight :**

Most of the employees leave the company after around 4 to 6 years.