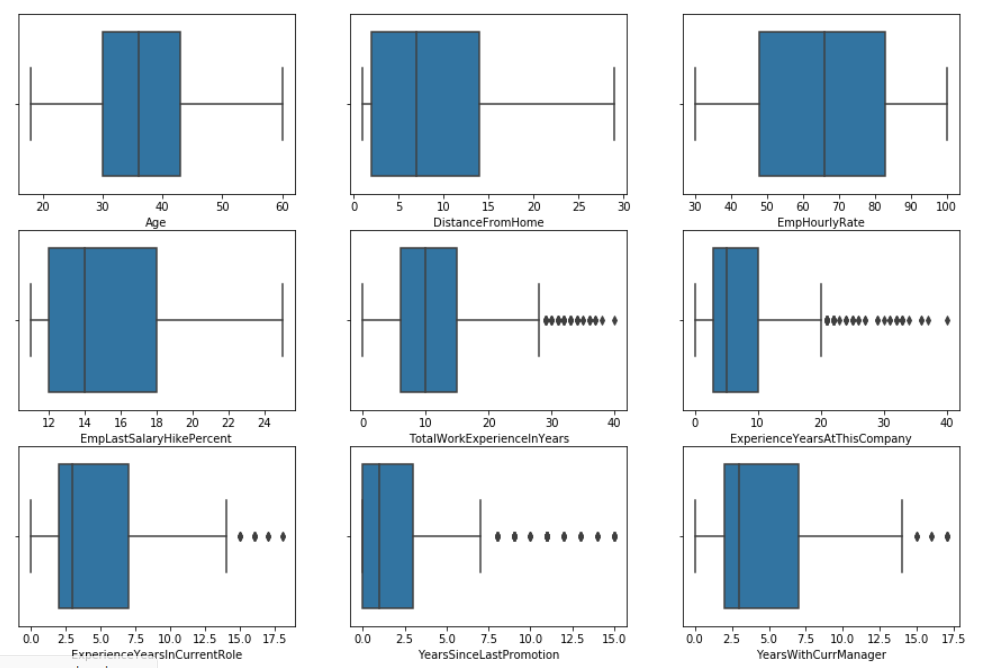
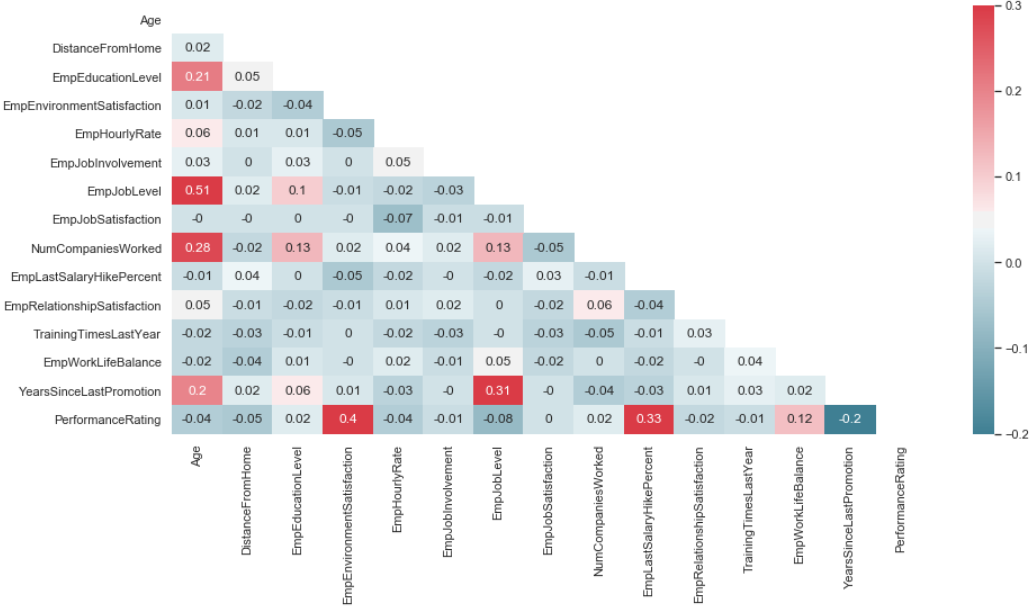
**Feature Engineering:**

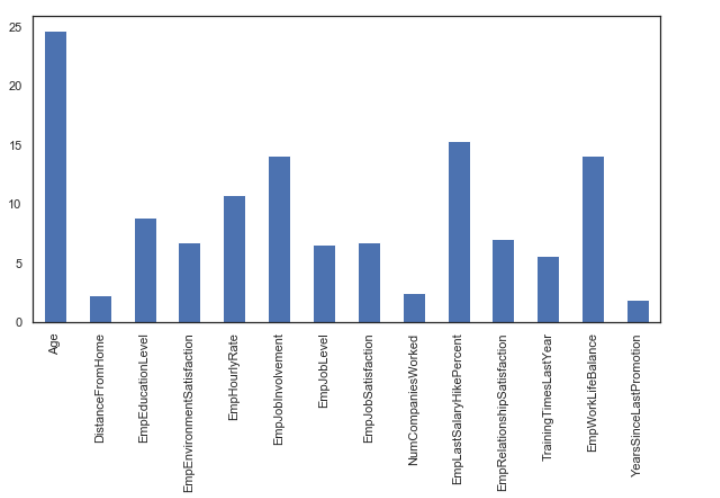
---------------------------------------------------------------------------------------------------------------------------

* Totally there are 27 features :
  + 26: predictor features
  + One target vaiable having three classes
* The continous variables having outliers and it is handled with IQR



* Collinearity :
  + Correlation Matrix : Binary correlation : Dropping highly correlated features
    - Col TotalWorkExperienceInYears and EmpJobLevel are highly correlated
    - Col ExperienceYearsInCurrentRole and ExperienceYearsAtThisCompany are highly correlated
    - Col YearsWithCurrManager exibits strong correlation with ExperienceYearsInCurrentRole and ExperienceYearsAtThisCompany
  + VIF(Variable Inflation Factor) : To handle Multicollinearity :
    - Below Columns exhibit Multicollinearity i.e VIF >10

'Age','EmpJobInvolvement','EmpLastSalaryHikePercent','EmpWorkLifeBalance' .

* + - But dropping all above columns negatively affecting the accuracy hence dropping only 'Age' column which is having highest VIF i.e 25
* Label Encoding the categorical features:
  + - ['Gender' ,'EducationBackground' ,'MaritalStatus','EmpDepartment','EmpJobRole' ,

'BusinessTravelFrequency','OverTime' ,'Attrition' ]

* The significant features for better accuracy are derived from **SelectKBest** Technique.

