**Employee Performance Analysis**

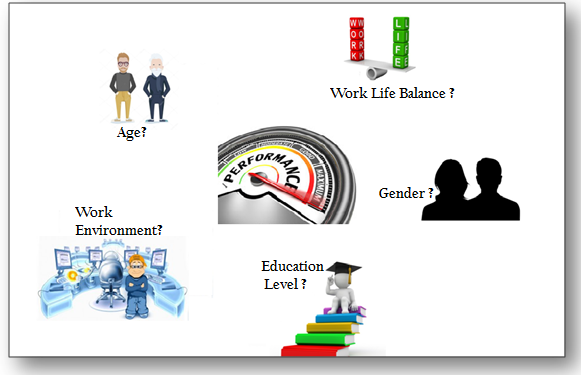
**Submitted By:**

Vidyashree Venkatesh

vidyashree31390@gmail.com

Project Summary

Every Company small or big ,the performance of its employees plays a key role on the company's growth. Here in this project "Employee Performance Analysis" , concern of Mr.Brain ,CEO INX Future on the decrease in their employees performance which is affecting their reputation in market is being addressed and analysis being done to identify the causing factors and to develop a Machine Learning Algorithm helping in future recruitment.



Data Source from the Third Party :



Analysis and Summary of the Data:

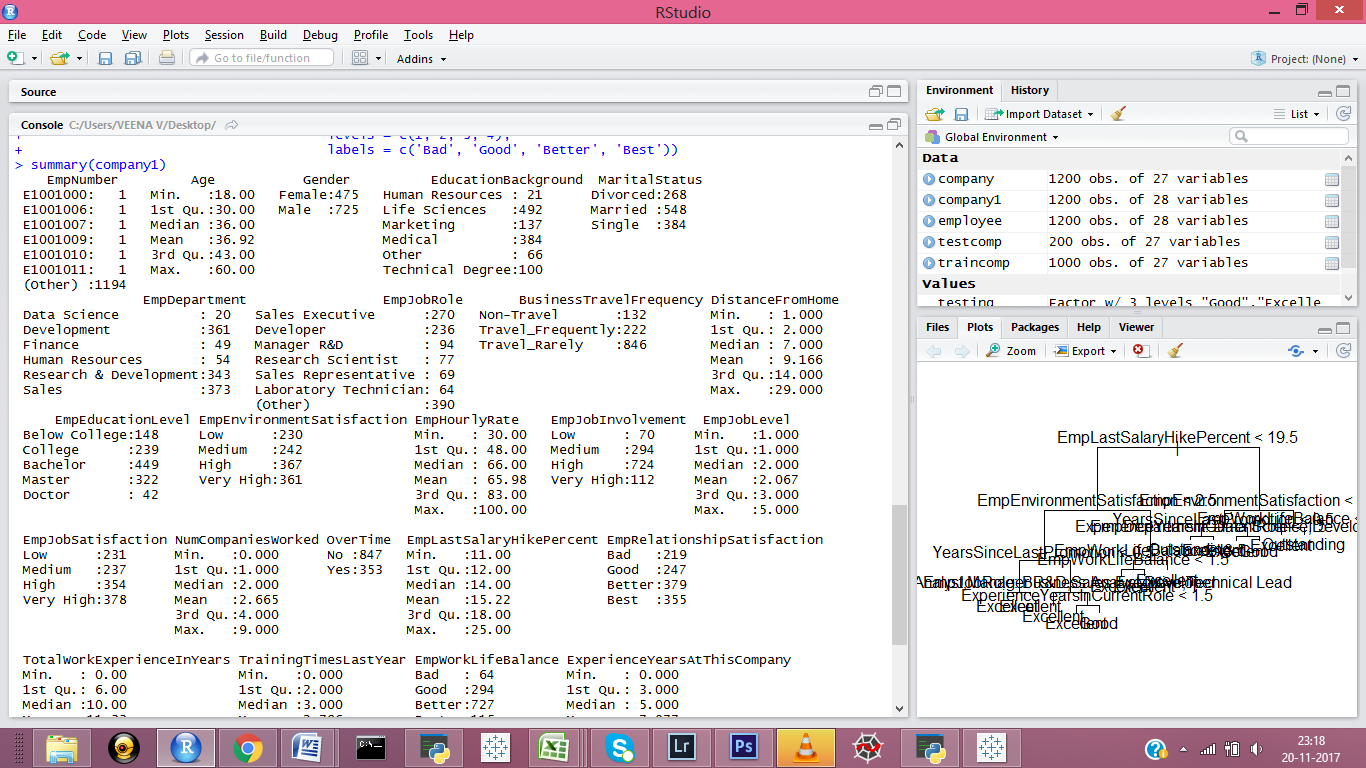
Data seems to be more or less normalised, it's not skewed hence Normalisation is not required. Few insights on data is as below.

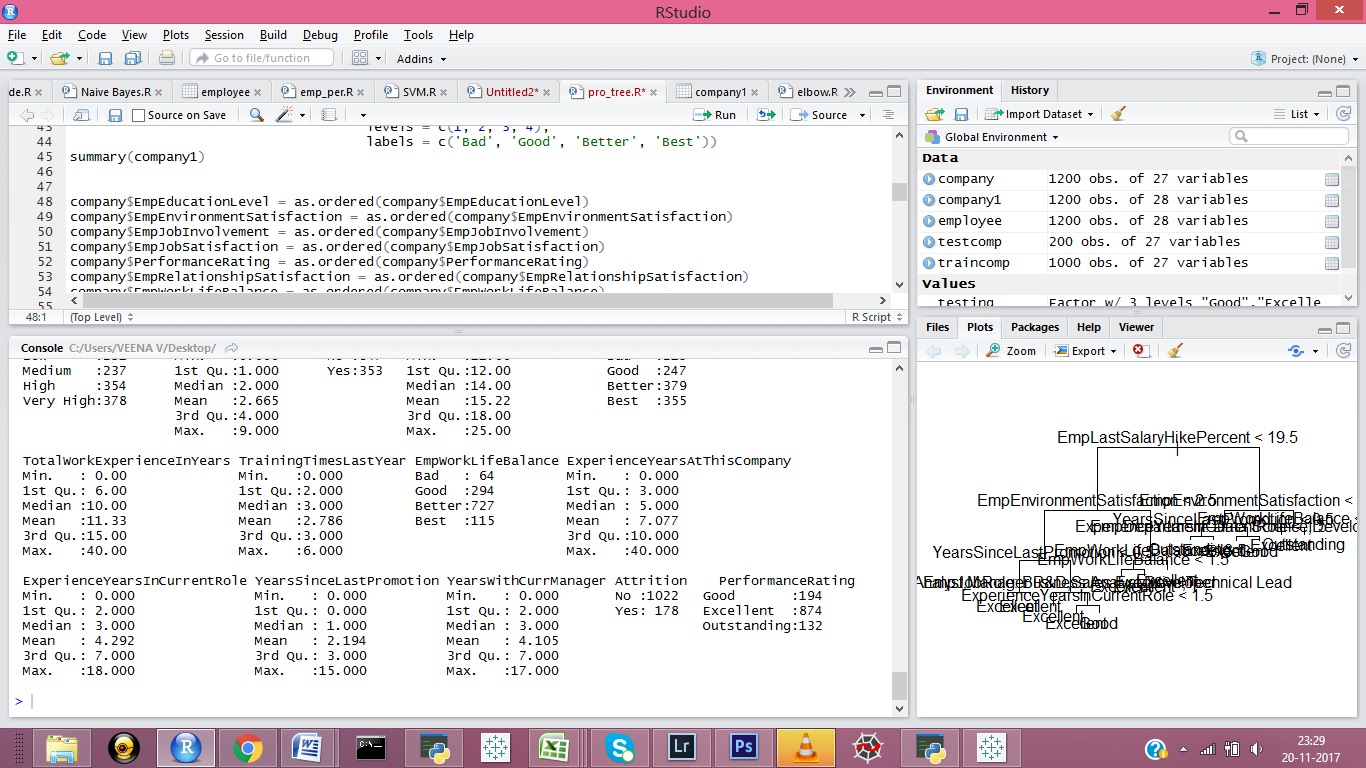
Number of Rows -1200

Features -28 , The data contains many categorical mainly ordinal .

Primary Feature - PerformanceRating [1. 'Low',2. 'Good' ,3 'Excellent', 4.'Outstanding']

Summary of the Data :





Analysis and Approach:

Process flow:

Predictive model is built using multiple Machine Algorithms

Process the data /cleansing data

Extract features from provided raw data.

r

Testing & Evaluation is done to identify Optimal Model

Identify the top factors affecting Performance

On analysing the data the following decisions are made on the approach

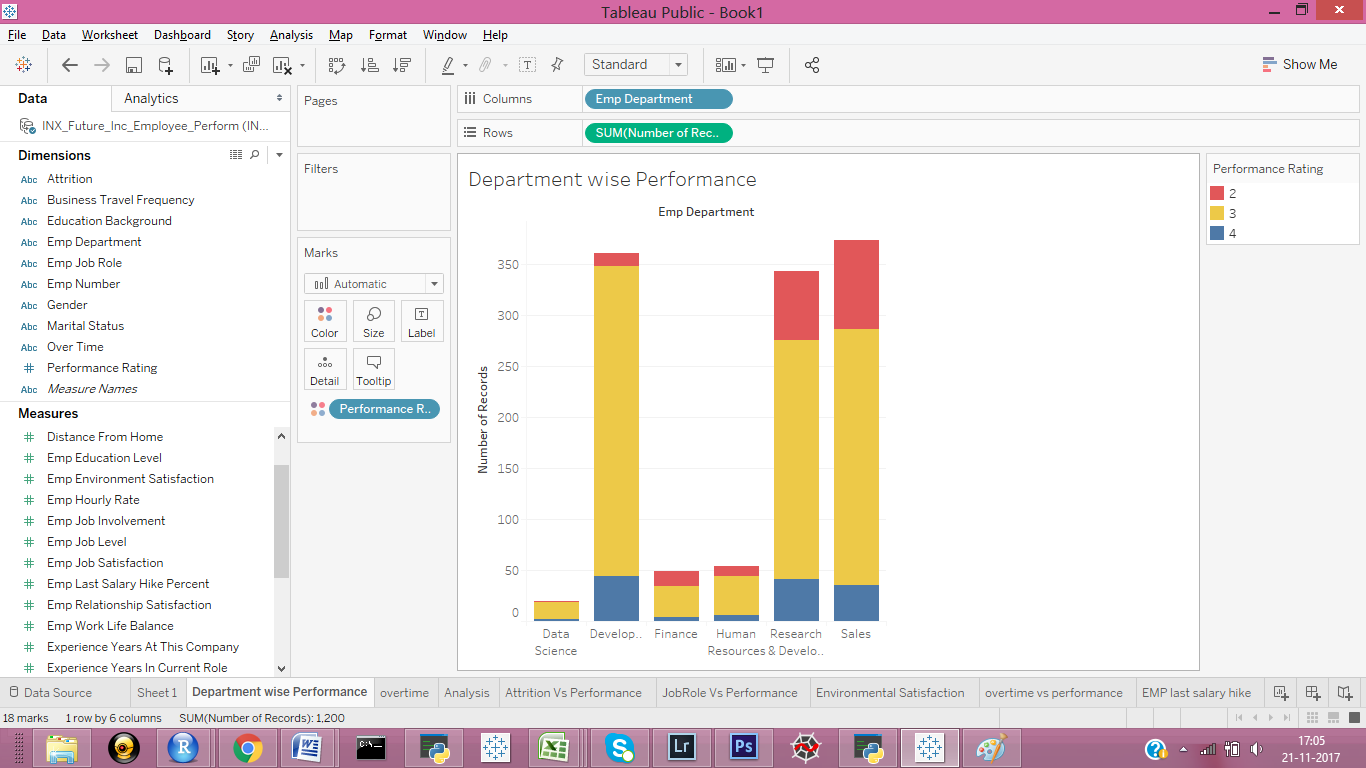
* It's a Supervised Learning
* The Primary output feature is a Categorical output.
* Classification Algorithm to be used .
* Algorithms such as Decision tree, Neural net, Support Vector , Radom Forest can be used.
* Tableau provides better visualisations for identifying the relation between features.

Methodology & Tools

|  |  |
| --- | --- |
| Machine Learning Algorithms | Tools |
| * Decision Tree * Random Forest * SVM | * Jupyter Notebook * R Studio * Tableau |

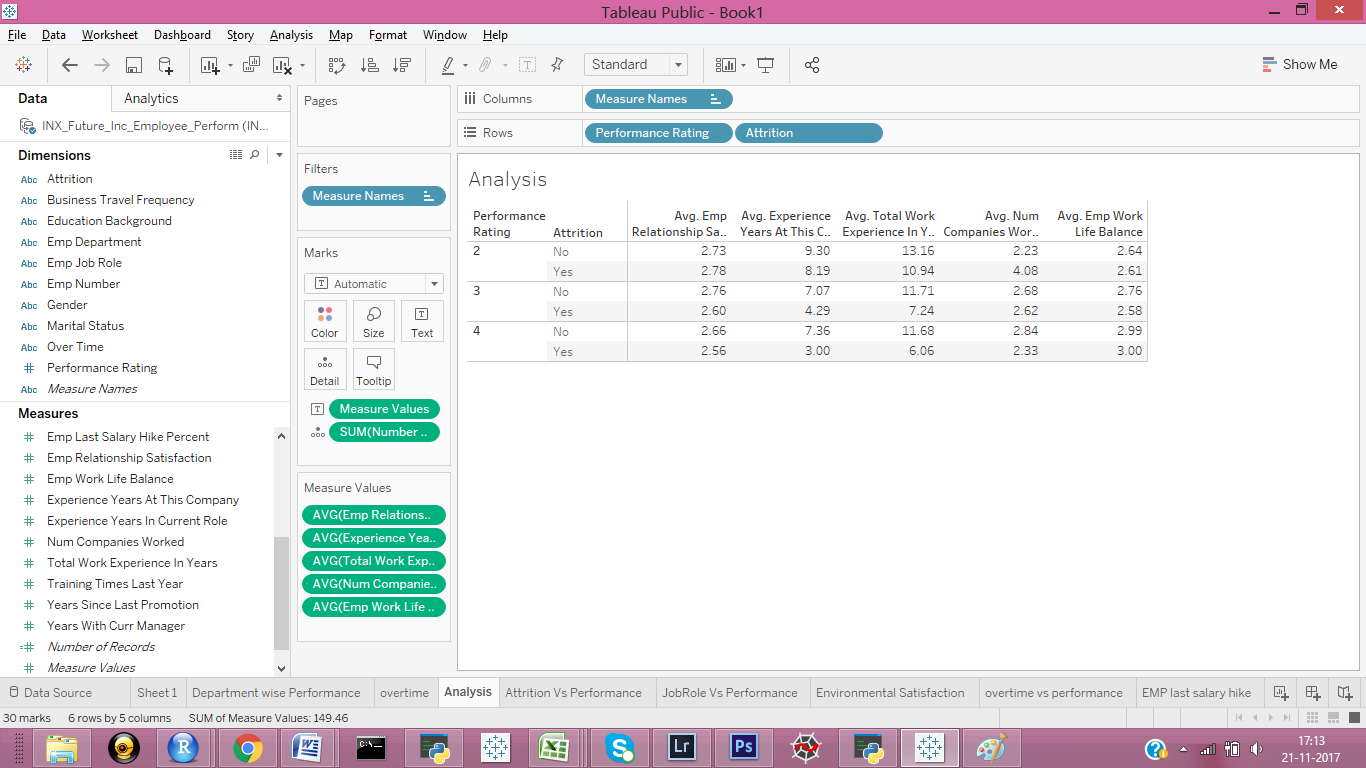
Evaluation & Visualisations:

* Department wise Performance:



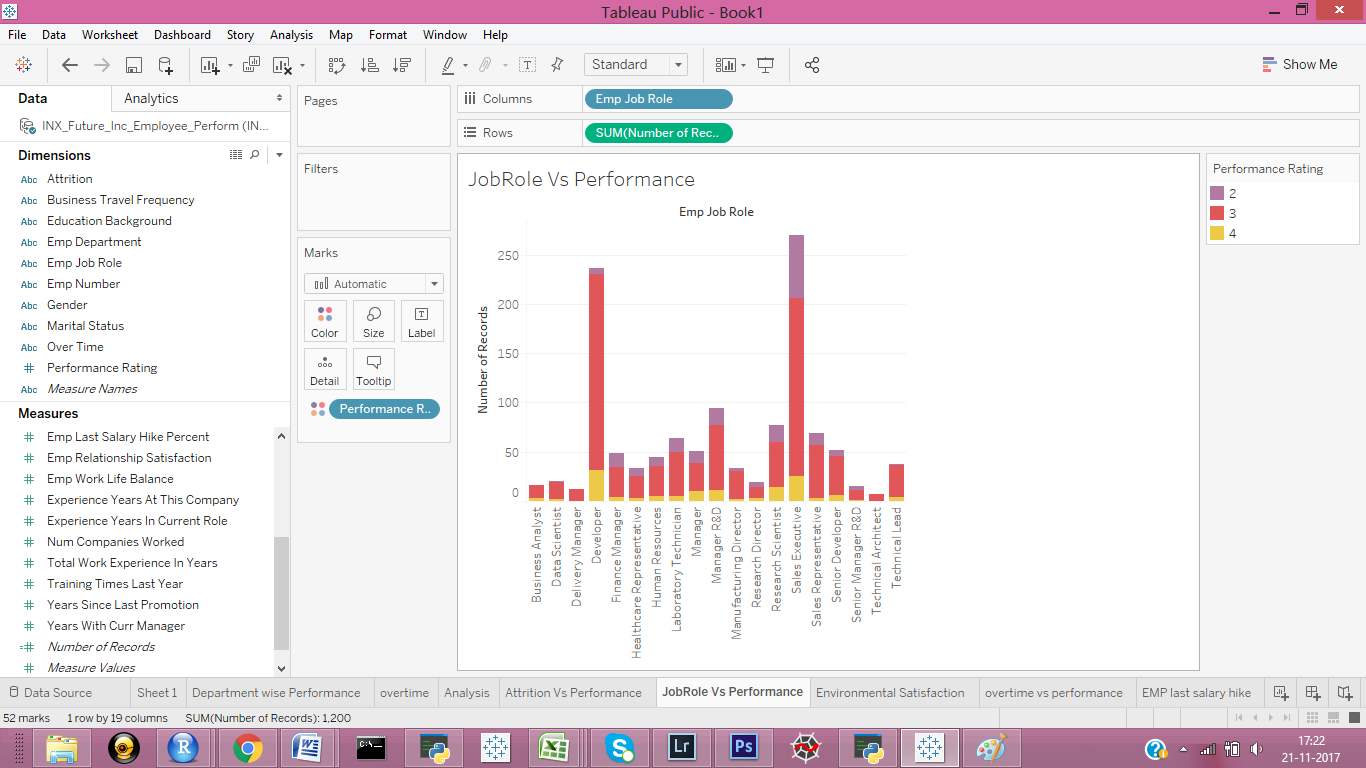
Sales Department made up the largest part of the data. However Developers have better performance ratio as they have the least low performing employees.

* Average Analysis of different factors:



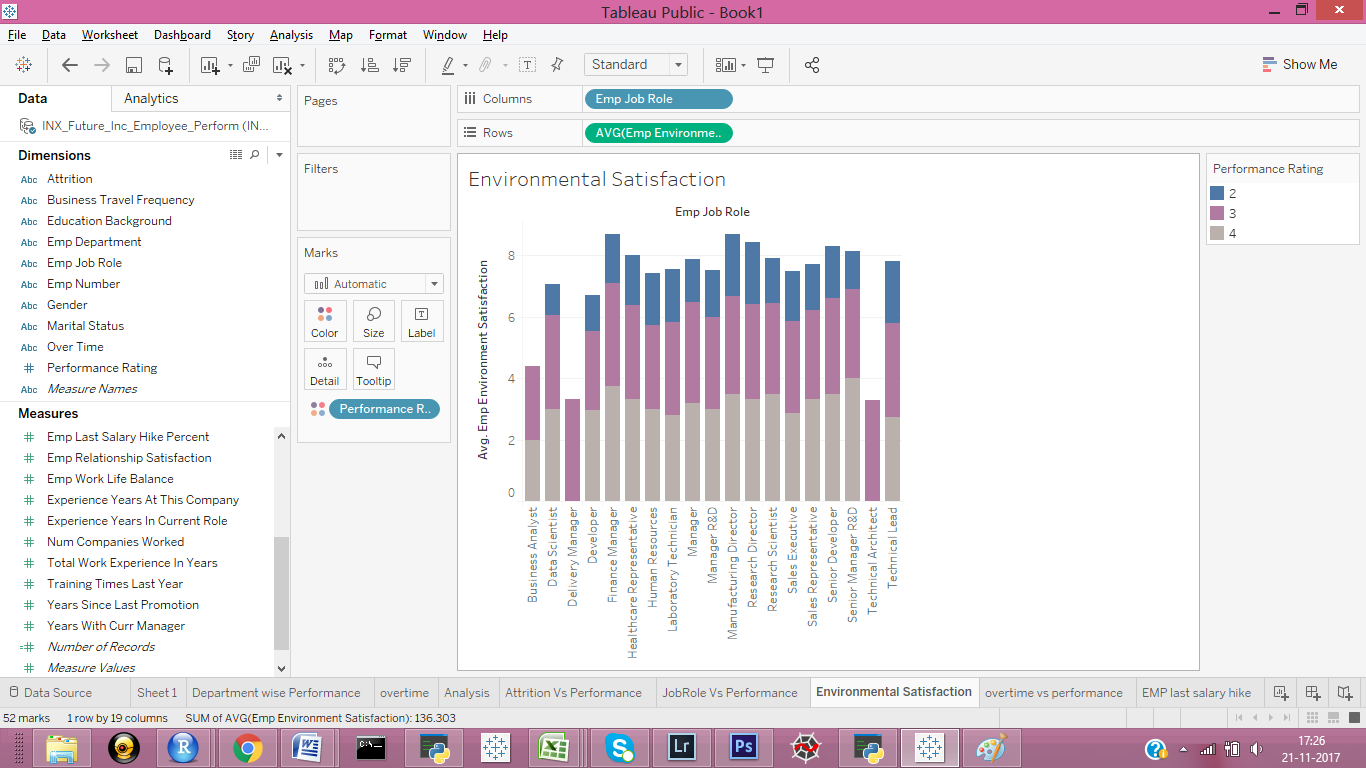
It can be derived that on an average higher the employee work life balance higher is the performance. Low performing employees seem to have more number of experience years.

* Job Role Impact:



Developer roles have a decent performance.

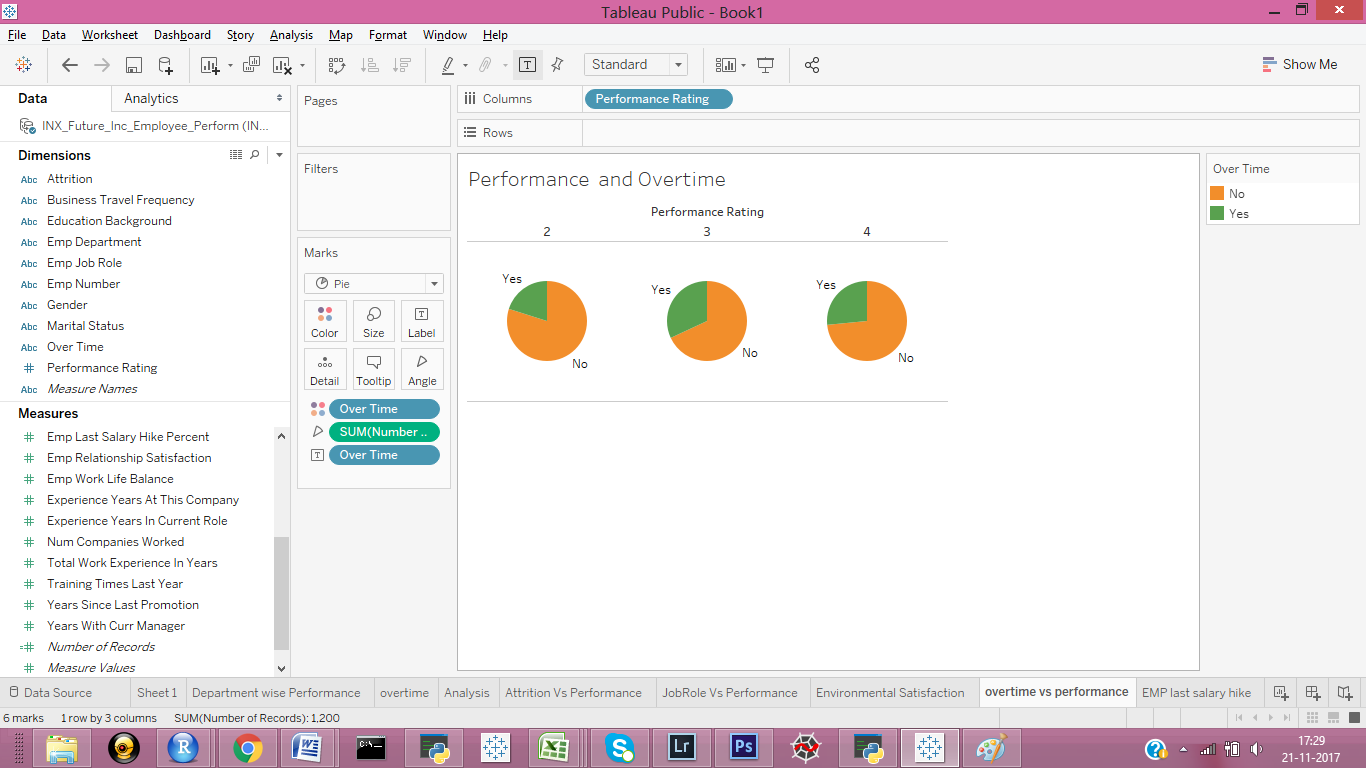
* Employee Work Environment Satisfaction:



Performance of the employee is proportional to the environment satisfaction.

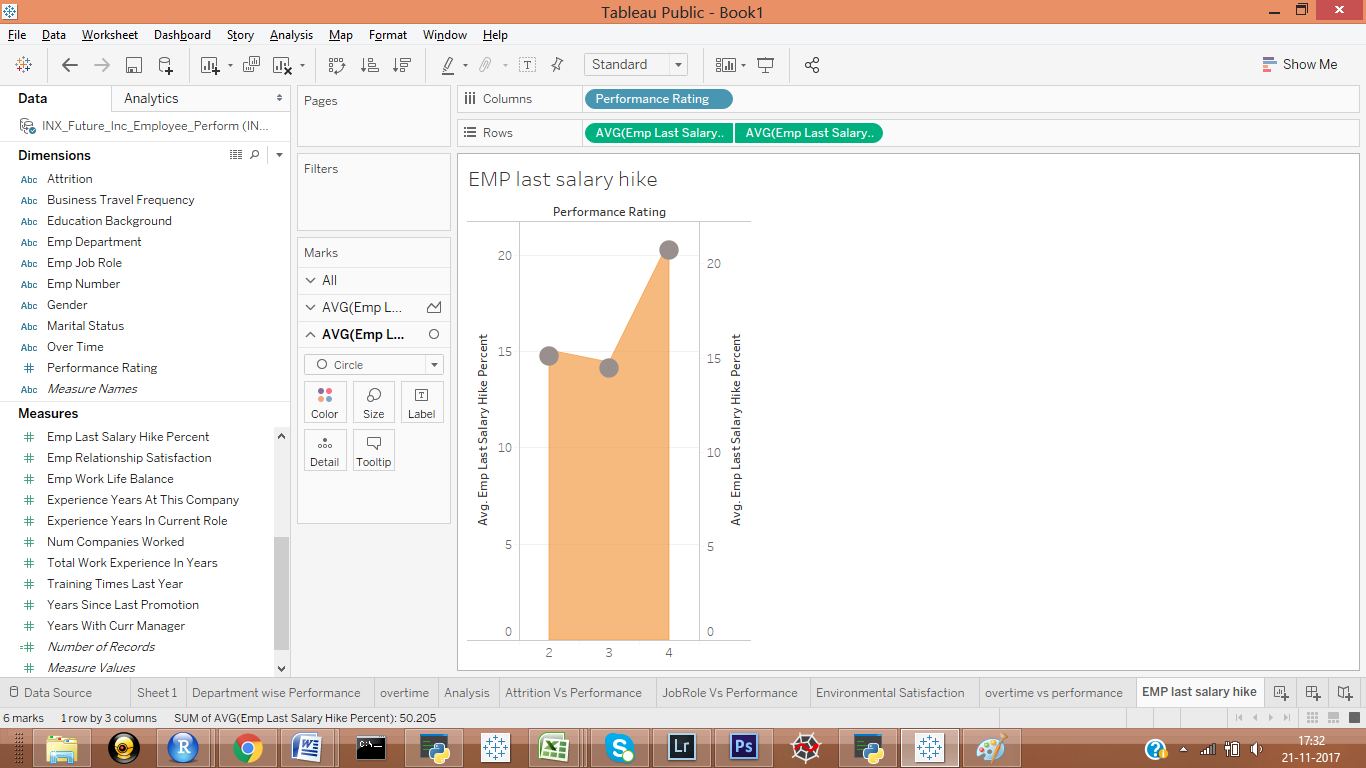
* Impact from Overtime:

Overtime doesn't have much impact on performance.



* Employee Last salary hike:

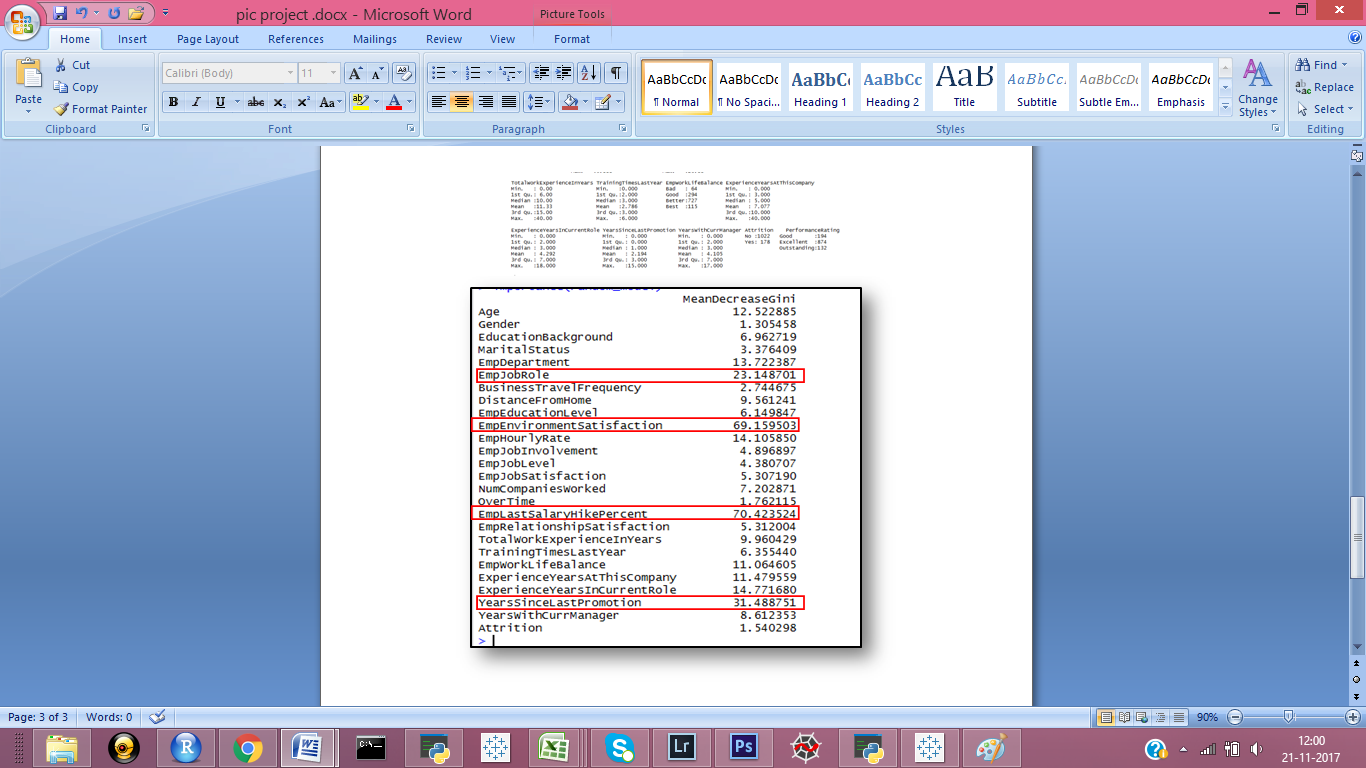
High performers have a better hike previously.



* Summary of top impacting factors:

Top factors impacting performance are identified by **using Random forest** .

* EmpJobRole
* EmpEnvironmentSatisfaction
* EmpLastSalaryHike
* YrsSinceLastPromotion



* Model & Code:

Decision Tree Algorithm Built in Jupyter NoteBook.

Code:



R Studio was used to build 3 algorithms(Tree,RandomForest,SVM) and choose best one on comparing results.

Code:



Confusion Matrix and Statistics: Decision Tree

Reference

Prediction Good Excellent Outstanding

Good 58 6 1

Excellent 6 253 9

Outstanding 0 2 25

Overall Statistics

Accuracy : 0.9333

95% CI : (0.9024, 0.9568)

No Information Rate : 0.725

P-Value [Acc > NIR] : <2e-16

Kappa : 0.8416

Mcnemar's Test P-Value : 0.1414

Confusion Matrix and Statistics:RandomForest

Reference

Prediction Good Excellent Outstanding

Good 58 7 1

Excellent 6 253 10

Outstanding 0 1 24

Overall Statistics

Accuracy : 0.9306

95% CI : (0.8992, 0.9546)

No Information Rate : 0.725

P-Value [Acc > NIR] : < 2e-16

Kappa : 0.8342

Mcnemar's Test P-Value : 0.03773

SVM Performance :

performance\_predicitions Good Excellent Outstanding

Good 45 11 0

Excellent 19 247 12

Outstanding 0 3 23

> agreement <- performance\_predicitions == testcomp$PerformanceRating

> table(agreement)

agreement

FALSE TRUE

45 315

> prop.table(table(agreement))

agreement

FALSE TRUE

0.125 0.875

Supporting Screen Shots for the performance accuracy:



Models Accuracy Percentage:

|  |  |
| --- | --- |
| Machine Learning Algorithm | Accuracy Percentage |
| Decision Trees | 93.33 |
| Random Forest  Support Vector Machine | 93.06  87.5 |

Summary:

An analysis is made on the current employee data and the underlying causes affecting performance are identified. And a reliable performance predicting model is built to help in hiring future employees. Below are the findings noted from the study on the data.

Insights and Recommendation:

Insights

* Developer roles have better performance ratio compared to other departments.
* The top 3 factors that impacted the performance as predicted by RandomForest:

1. EmpEnvironmentSatisfaction
2. EmpLastSalaryHike
3. YrsSinceLastPromotion

* A trained model was built using different algorithms Decision tree and Random Forest have similar accuracy of ~93%. This can be used in future recruitment.
* Employees with less work environment satisfaction tend to show low performance.
* Employees who had higher hike and recent promotions are most likely to perform well.
* Employee JobRole, Worklifebalance , HourlyRate do contribute significantly on performance.
* The longer the employee stays in their current role, their performance degrades.

Recommendations

* It's advisable to initiate a investigation on root cause of employee dissatisfaction of work environment .
* Promote work life balance within the company and affected departments.
* Involve employees in training programs and encourage taking up different Job roles and responsibilities , which broadens theirs skillets.

This intern helps employees career growth and performance factor to the company.