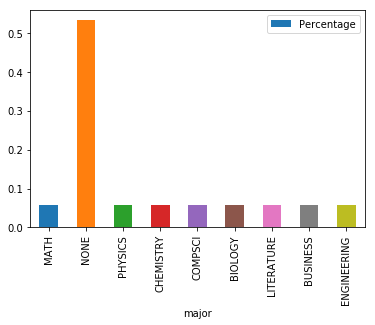
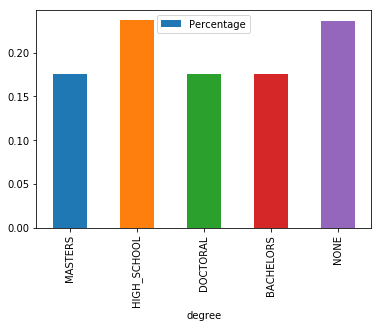
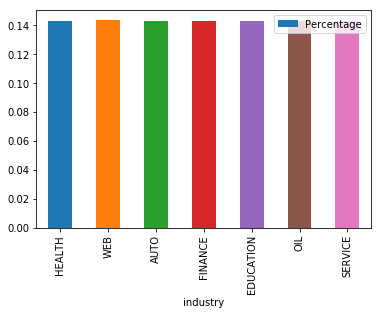
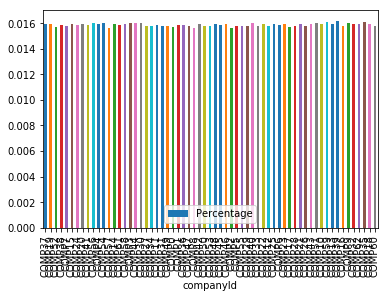
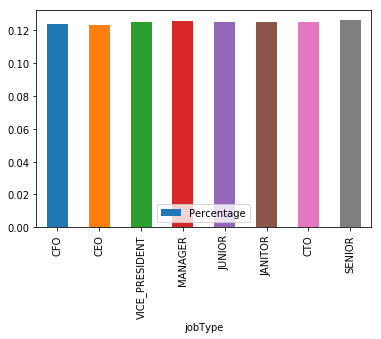
# 

# 2.0 APPENDIX ( PLOTS AND FIGURES)

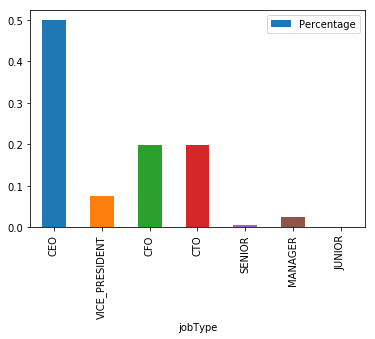
1.0 Distribution of columns without high paying Jobs.





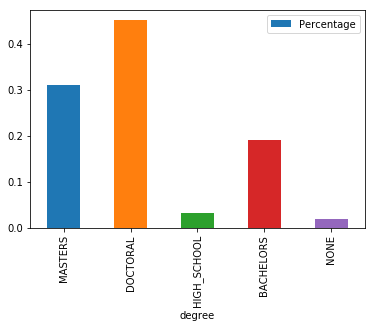
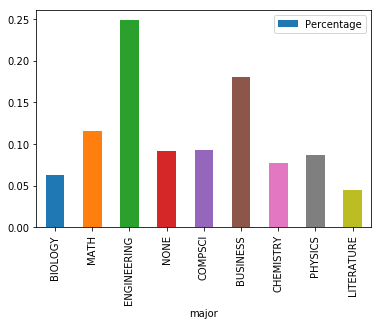


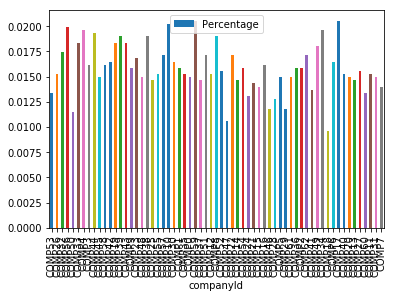
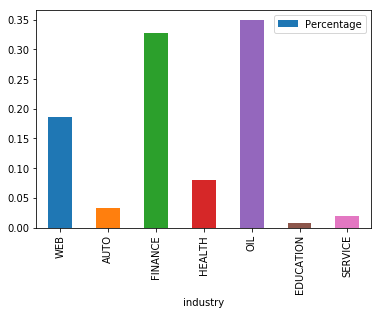
1.1 Distribution of columns without outliers



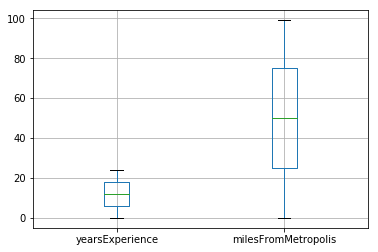
we see a huge difference in the jobType distribution for higher salaries. This difference prompts up to have two models. One for

CEO,CFO,CTO,VICE\_PRESIDENT and another model for rest of the jobs.

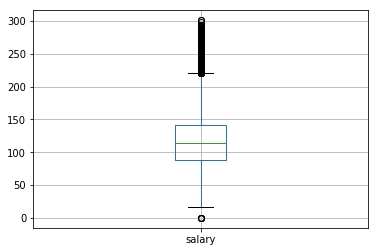
 

1.2 Box plot of ‘years of experience’ and ‘miles from metropolis’

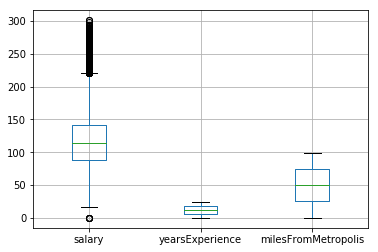


We don’t see outliers in these quantitative columns.

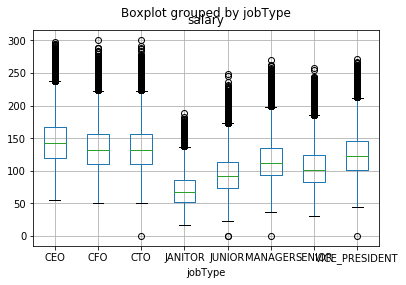
Salary Boxplots,



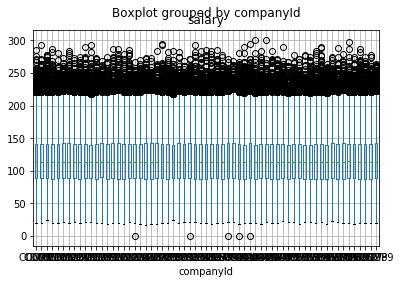
There are many outliers seen in the box plot of salary.



Box plot of salary by jobType



We do see outliers here, for each type of jobType. Since each job type are having outliers, it seems like there are may be some companies who normally pays higher than majority of the companies. So removing these outliers doesn’t make sense from practical point of view.

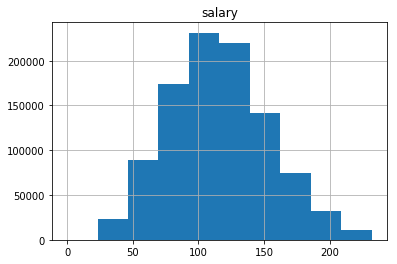


1.3 Distribution of Salary

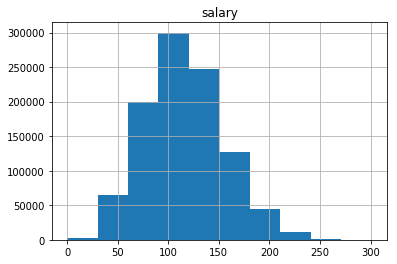
Salary mean = 116.061818

Salary standard deviation = 38.71791702215943

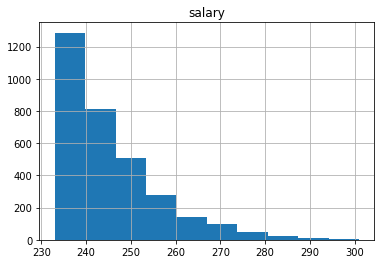
Without outliers,



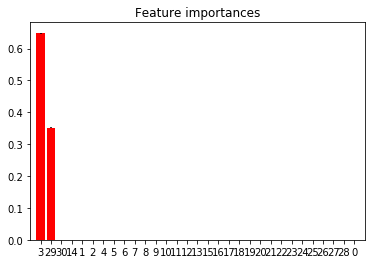
Including outliers:



Just the outliers:



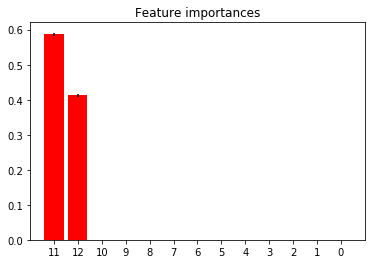
1.4 Variable Importance after including the salary outliers (Plus minus two Sigmas from mean)



Important features:

‘yearsOfExperience’ and ’jobType\_JANITOR’.

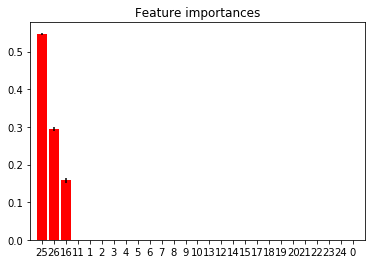
1.5 Variable importance just for Janitor Jobs



Important features:

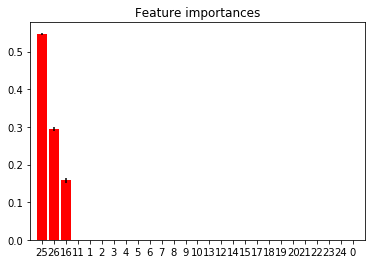
‘yearsOfExperience’ and milesFromMetropolis’.

1.6 Variable importance for jobs CEO,CFO,VICE\_PRESIDENTS,CTO.



Important features ‘yearsExperience’, ’ milesFromMetropolis’, ‘major\_CHEMISTRY’

1.7 Variable importance for jobs other than CEO,CFO,VICE\_PRESIDENTS, CTO.



Important features ‘yearsExperience’, ’ milesFromMetropolis’, ‘major\_NONE’

1.8 MODELS AND MSE GRAPHS

Random Forest Model for CEO,CTO,VICE\_PRESIDENT, CFO. Plot of MES Vs No Of trees,

No\_Of\_tree MSE

0 25 852.944674

1 50 853.399370

2 75 852.792093

3 100 852.852200

4 150 852.813222

5 175 852.980848

6 200 853.098130

depth\_List MSE

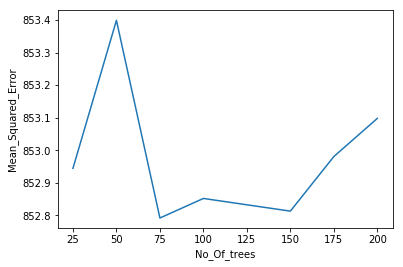
0 2 852.852200

1 4 726.855936

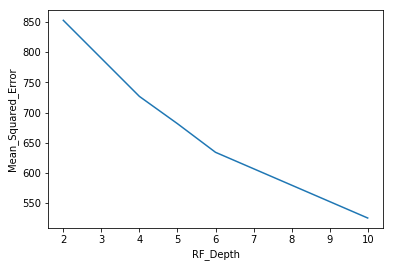
2 5 681.689620

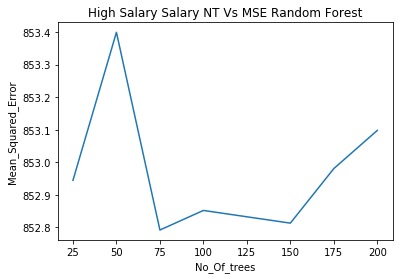
3 6 634.289191

4 10 525.460155



Random Forest Model for High Salary CEO, CTO, VICE\_PRESIDENT, CFO. Plot of RF Depth Vs No Of trees,





No\_Of\_tree MSE

0 25 852.944674

1 50 853.399370

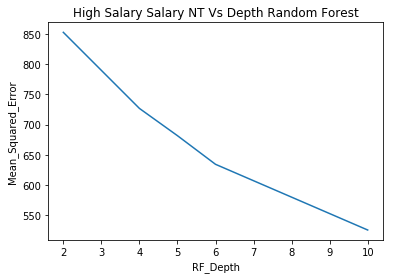
2 75 852.792093

3 100 852.852200

4 150 852.813222

5 175 852.980848

6 200 853.098130



depth\_List MSE

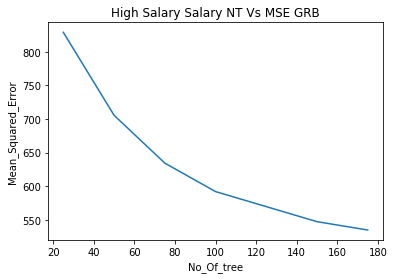
0 2 852.852200

1 4 726.855936

2 5 681.689620

3 6 634.289191

4 10 525.460155



No\_Of\_tree MSE

0 25 828.886444

1 50 705.631787

2 75 634.522471

3 100 592.194487

4 150 547.476024

5 175 535.065467

Gradient Boost model for other jobs than CXOs ( OK Salary)

No\_Of\_tree MSE

0 25 642.485300

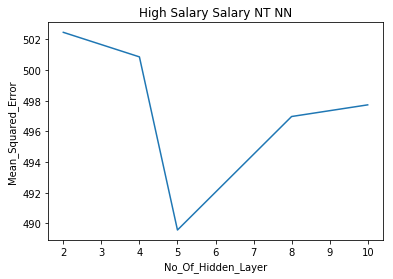
1 50 489.733292

2 75 409.577186

3 100 364.218324

4 150 316.728191

5 175 303.464853



No\_Of\_Hidden\_Layer MSE

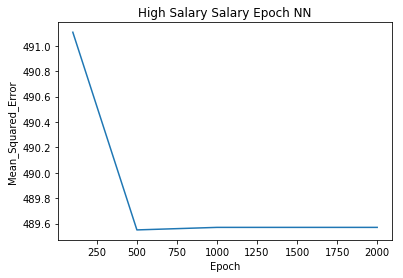
0 2 502.465899

1 4 500.865074

2 5 489.570640

3 8 496.974921

4 10 497.734747



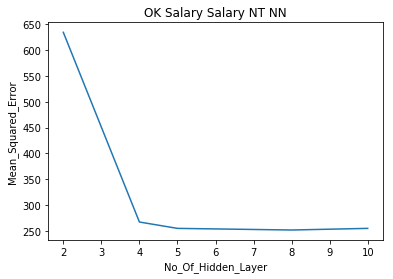
No\_Of\_Epoches MSE

0 100 491.105719

1 500 489.550675

2 1000 489.570640

3 2000 489.570640



No\_Of\_Hidden\_Layer MSE

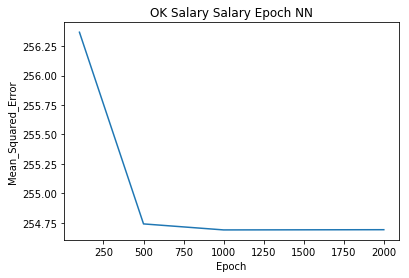
0 2 634.578750

1 4 267.015169

2 5 254.688369

3 8 251.526180

4 10 254.699023



No\_Of\_Epoches MSE

0 100 256.367671

1 500 254.739429

2 1000 254.688369

3 2000 254.690219

