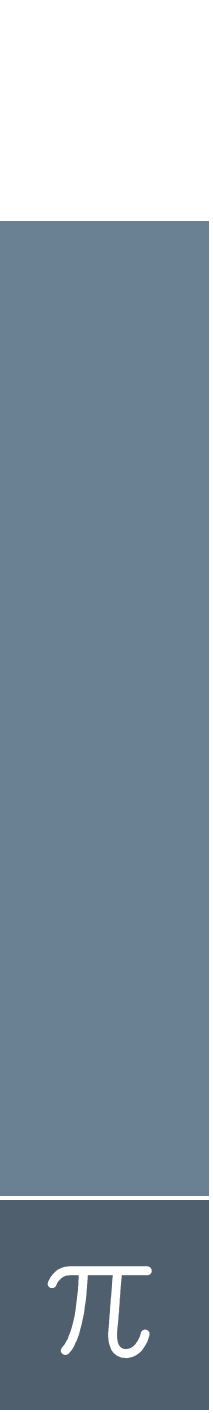
Project 2: Prediction Models

Raag Patel
DS6306
Winter 21-22

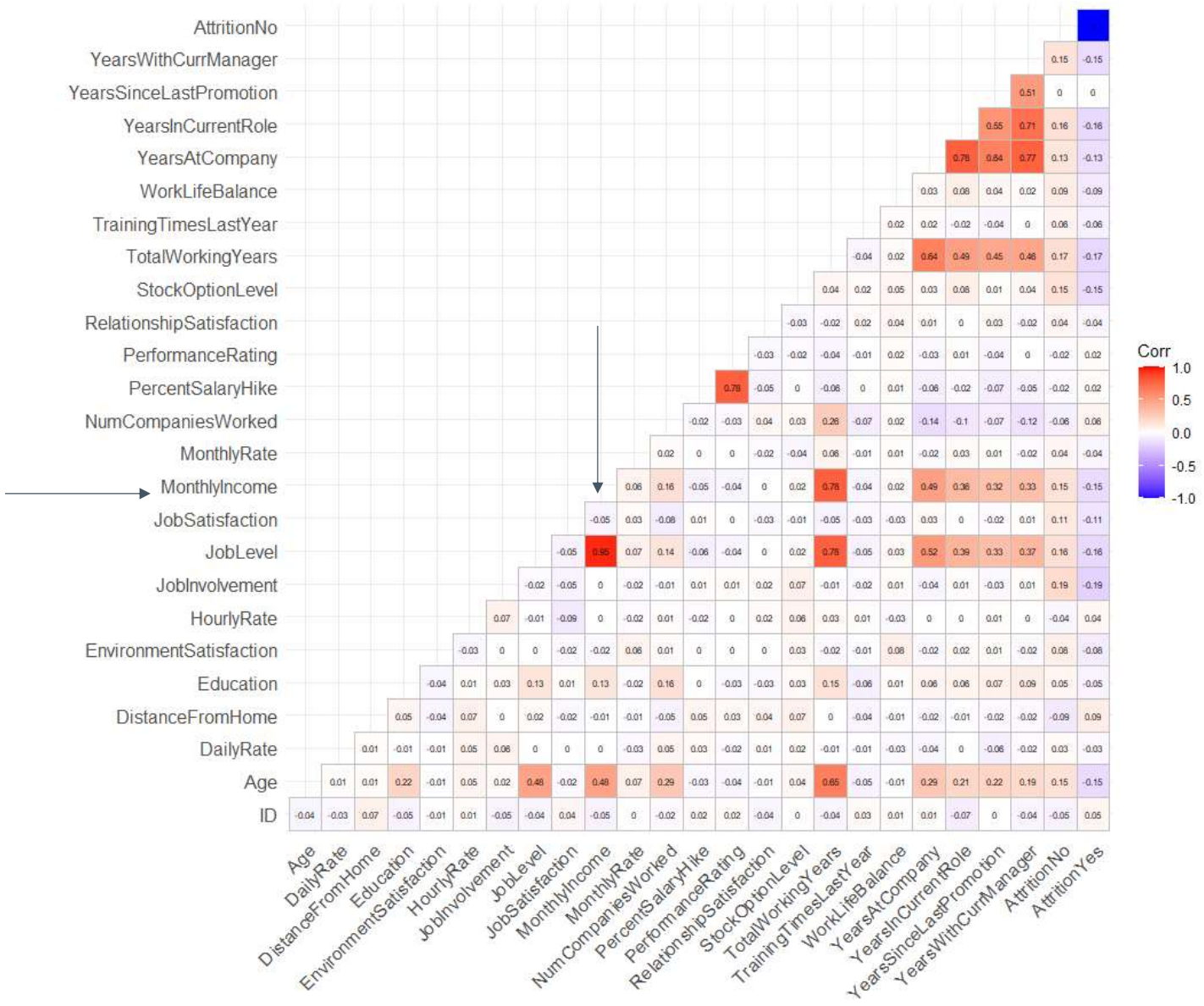


Goal

- Build Predictive Models
- Predicting Attrition/Salary
- High Accuracy

A Look Inside

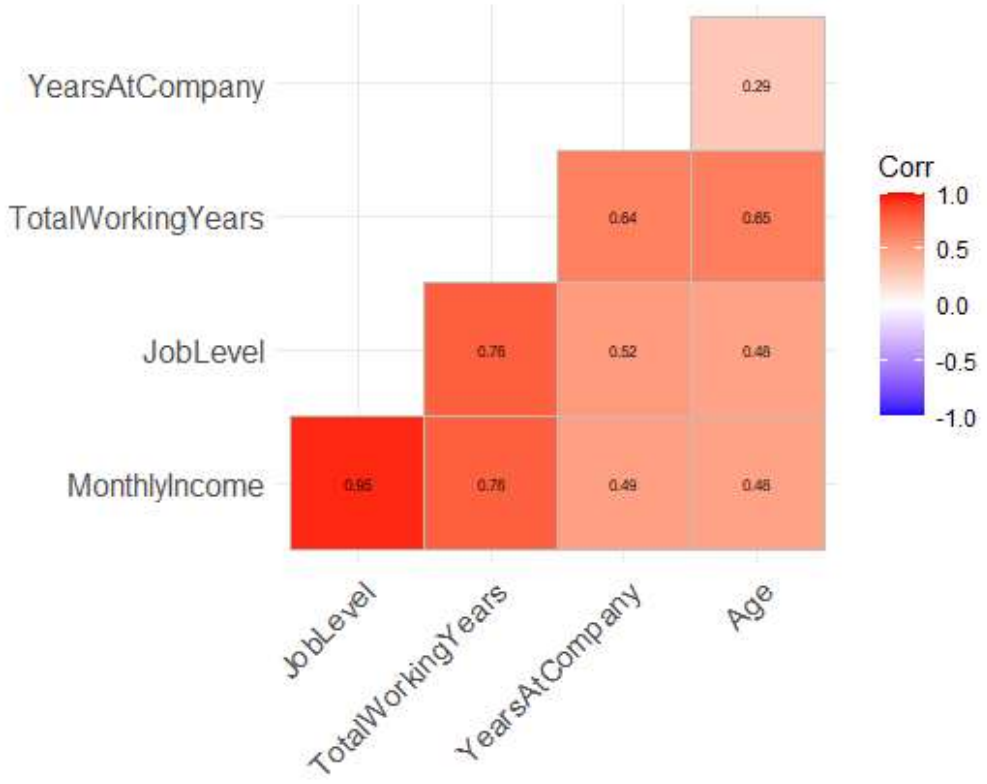
- Categorical, Continuous, Factors, Numeric
- Correlation
- Cleaning Data


AttritionNo
YearsWithCurrManager
YearsSinceLastPromotion
YearsInCurrentRole
YearsAtCompany
WorkLifeBalance
TrainingTimesLastYear
TotalWorkingYears
StockOptionLevel
RelationshipSatisfaction
PerformanceRating
PercentSalaryHike
NumCompaniesWorked
MonthlyRate
MonthlyIncome
JobSatisfaction
JobLevel
JobInvolvement
HourlyRate
EnvironmentSatisfaction
Education
DistanceFromHome
DailyRate
Age
ID
Age
DailyRate
DistanceFromHome
Education
EnvironmentSatisfaction
HourlyRate
JobInvolvement
JobLevel
JobSatisfaction
MonthlyIncome
MonthlyRate
NumCompaniesWorked
PercentSalaryHike
PerformanceRating
RelationshipSatisfaction
StockOptionLevel
TotalWorkingYears
TrainingTimesLastYear
WorkLifeBalance
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrManager
AttritionNo
AttritionYes
Corr
1.0
0.5
0.0
-0.5
-1.0

Attrition Model

- 81.91% Accuracy
- 83.46% Sensitivity
- 74.00% Specificity

```
Confusion Matrix and Statistics


predictor_attr  No Yes
           No  212  13
           Yes  42  37

               Accuracy : 0.8191
                 95% CI : (0.7711, 0.8607)
    No Information Rate : 0.8355
    P-Value [Acc > NIR] : 0.8039676

                  Kappa : 0.4661

 Mcnemar's Test P-Value : 0.0001597

            Sensitivity : 0.8346
            Specificity : 0.7400
         Pos Pred Value : 0.9422
         Neg Pred Value : 0.4684
             Prevalence : 0.8355
         Detection Rate : 0.6974
   Detection Prevalence : 0.7401
      Balanced Accuracy : 0.7873

       'Positive' Class : No
```


YearsAtCompany
TotalWorkingYears
JobLevel
MonthlyIncome
0.29
0.64
0.65
0.76
0.52
0.48
0.95
0.78
0.49
0.46
JobLevel
TotalWorkingYears
YearsAtCompany
Age
Corr
1.0
0.5
0.0
-0.5
-1.0

Salary Model

- Linear Regression Model
- MRSE > 3000

```
# Final, best fit.
salary_fit = lm(MonthlyIncome ~ JobLevel + TotalWorkingYears + YearsAtCompany, data = CS_Salary_Model)
summary(salary_fit)
```

```
##
## Call:
## lm(formula = MonthlyIncome ~ JobLevel + TotalWorkingYears + YearsAtCompany,
##     data = CS_Salary_Model)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5625.1  -888.5    42.3   725.5  3968.3
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -1764.37     100.04 -17.637  < 2e-16 ***
## JobLevel           3724.98      68.94  54.035  < 2e-16 ***
## TotalWorkingYears    70.76      11.07   6.394 2.64e-10 ***
## YearsAtCompany      -32.04      10.11  -3.170  0.00158 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1383 on 866 degrees of freedom
## Multiple R-squared:  0.9099, Adjusted R-squared:  0.9096
## F-statistic:  2915 on 3 and 866 DF,  p-value: < 2.2e-16
```

Models Applied

```
##     ID MonthlyIncome
## 1  871      5956.677
## 2  872      2218.318
## 3  873     14225.179
## 4  874      2154.243
## 5  875      2231.696
## 6  876      6998.072
## 7  877      6027.441
## 8  878      2038.063
## 9  879      2347.875
## 10 880     15027.525
## 11 881     11167.705
## 12 882      2327.808
## 13 883      5917.951
## 14 884      5911.262
## 15 885      6393.232
## 16 886      5872.535
## 17 887      6239.734
## 18 888      6180.940
## 19 889      6168.970
## 20 890      2225.007
```

```
summary(Case2Predictions_Salary)
```

```
##        ID         MonthlyIncome
##  Min.   : 871.0   Min.   : 1961
##  1st Qu.: 945.8   1st Qu.: 2287
##  Median :1020.5   Median : 6063
##  Mean   :1020.5   Mean   : 6236
##  3rd Qu.:1095.2   3rd Qu.: 6979
##  Max.   :1170.0   Max.   :18698
```

```
##      ID Attrition
## 1  1171        No
## 2  1172        No
## 3  1173       Yes
## 4  1174        No
## 5  1175        No
## 6  1176        No
## 7  1177        No
## 8  1178        No
## 9  1179        No
## 10 1180       Yes
## 11 1181        No
## 12 1182       Yes
## 13 1183        No
## 14 1184        No
## 15 1185        No
## 16 1186        No
## 17 1187       Yes
## 18 1188        No
## 19 1189        No
## 20 1190        No
```

```
summary(Case2Predictions_Attrition)
```

```
##        ID       Attrition
##  Min.   :1171   No :193
##  1st Qu.:1246   Yes:107
##  Median :1320
##  Mean   :1320
##  3rd Qu.:1395
##  Max.   :1470
```

Questions?
Comments?
Concerns?

Email and/or Comment!