To: Mr. I. Luv Reg

From: Jacob Igel

Date: March 26th, 2021

Subject: Analysis Memo of Employee Satisfaction

I am writing this memo to analyze Employee Satisfaction for Mr. I Luv Regs’ Company. After analyzing the data, I believe I can point out the strengths and weaknesses of this particular dataset. In this memo, I will touch on several regressions and models that will all be shown in the appendix pages at the conclusion of this memo.

**General Dissatisfaction**

When looking over the Departments, Recruitment Source, Position, and Races within the data set, some conclusions came from that. It appears that the Executive Office is the most dissatisfied when compared to the Administrative Offices (Step 1 a). Based on the high p-value that accounts for the whole data and the low R square or the strength of the relationship between recruitment and satisfaction, we can see that Recruitment source probably would not play a factor into Employee Satisfaction (Step 1 b). It appears that being in the Administrative Position causes more employee dissatisfaction (Step 1 c). In relation to American Indian or Alaskan Native employees, Hispanic Employees seem to be the most dissatisfied (Step 1 d).

**What Model to Use**

When trying two separate models, I came to the conclusion that using a Higher-Order Model was better than using a First-Order Model. With the First-Order Model, we are just comparing individual values grouped together compared to employee satisfaction while with the Higher-Order Model, we are combing individual values and then also some values compared with each other inside the same model compared to employee satisfaction. This does lead to more error but the variance of this graph or the relationship of values compared to employee satisfaction is greater (pt. Higher-order Regression Model Part 2C-g) than the First-Order Model having less variance/guessing (Analysis Step 2 1d). This can give a better guess of what to use to measure employee satisfaction going forward.

Pre-processing Steps (6 points) —- PUT IN THE APPENDIX

The data set contains variables and observations that need to be dropped before creating an effective multiple linear regression model.

Steps Needed: 1) (3 pts) You need to perform the analysis only on employees who are active. In other words, the observations for which the variable EmploymentStatus is equal to Active. HR <- read.csv('HR\_DAE.csv') HR = HR[which(HR$EmploymentStatus == "Active"),]

HR = read.csv('HR\_DAE.csv')

HR = HR[which(HR$EmploymentStatus == "Active"),]

2) (3 pts) Drop all unnecessary variables. In this part, you should drop

• Status Variables: EmploymentStatus, DateofTermination, and TermReason since we only keep “Active” employees.

• All date variables: DOB, DateofHire(we will learn later how to effectively include them) • ID variables: EmpID

• Names: Employee\_Name, ManagerName

• Zip Code and State: State, Zip (we will learn later how to effective include them)

TIP: You can use either subsetting, e.g., data = data[,-colnumber] to delete the unneccessary variables or data$variablename = NULL. To show you completed successfully these steps paste the code below and the str() of your dataset.

HR$EmploymentStatus = NULL

HR$DateofTermination = NULL

HR$TermReason = NULL

HR$DOB = NULL

HR$DateofHire = NULL

HR$EmpID = NULL

HR$Employee\_Name = NULL

HR$ManagerName = NULL

HR$State =NULL

HR$Zip = NULL

> str(HR)

'data.frame': 182 obs. of 12 variables:

$ PayRate : num 55 32 32 56 60 ...

$ Position : chr "Area Sales Manager" "Production Technician I" "Production Technician I" "Area Sales Manager" ...

$ Sex : chr "F" "F" "F" "M " ...

$ MaritalDesc : chr "Single" "Single" "Single" "Married" ...

$ CitizenDesc : chr "US Citizen" "Non-Citizen" "US Citizen" "US Citizen" ...

$ RaceDesc : chr "White" "Black or African American" "White" "American Indian or Alaska Native" ...

$ Department : chr "Sales" "Production " "Production " "Sales" ...

$ RecruitmentSource : chr "Other" "Pay Per Click - Google" "Monster.com" "Pay Per Click - Google" ...

$ PerformanceScore : chr "Exceeds" "Fully Meets" "Fully Meets" "Fully Meets" ...

$ EngagementSurvey : num 4.77 3.58 2.79 3.6 3.02 3.02 1.53 5 3.01 4.5 ...

$ SpecialProjectsCount: int 1 6 0 2 0 3 4 4 7 7 ...

$ EmpSatisfaction : num 53.3 53.1 52.9 56.9 52.7 ...

Analysis Step 1 (10 Points) —- PUT IN THE APPENDIX. INCLUDE CONCLUSIONS IN THE MEMO

The IT company does not want you to include the following variables in the model: department, RecruitmentSource, Positionand Race. Instead, the company wants you to describe if you find any indications that there is statistical evidence of dissatisfaction considering the information included in them.

Deliverable: Include outputs of the code and regression(s) you create along with the answers to these questions:

a) Are employees for specific departments more dissatisfied than others?

* reg1 = lm(EmpSatisfaction ~ Department, data = HR)

Text

Description automatically generated with low confidence

* It appears that the Executive Office is the most dissatisfied when compared to the Administrative Offices.

b) Is there any indication that the Recruitment Source (RecruitmentSource) matters in how satisfied employees are?

* reg2 = lm(EmpSatisfaction ~ RecruitmentSource, data = HR)

Text

Description automatically generated

Text

Description automatically generated

* Based on the high p-value that accounts for the whole data and the low R square or the strength of the relationship between recruitment and satisfaction, we can see that Recruitment source probably would not play a factor into Employee Satisfaction.

c) Is there any indication that any particular position (Position) is more likely to be dissatisfied than others?

* reg3 = lm(EmpSatisfaction ~ Position, data = HR)

A picture containing timeline

Description automatically generated

* It appears that being in the Administrative Position causes more employee dissatisfaction.

d) Are employees of specific races more dissatisfied than others?

* reg4 = lm(EmpSatisfaction ~ RaceDesc, data = HR)

Text

Description automatically generated

* In relation to American Indian or Alaskan Native employees, Hispanic Employees seem to be the most dissatisfied.

Analysis Step 2 (20 Points) —- PUT IN THE APPENDIX. INCLUDE CONCLUSIONS IN THE MEMO

First-order Regression Model (10 points)

1. Fit the multiple regression model, regressing Y (EmpSatisfaction) on X1 (PayRate), X2 (Sex), X3 (CitizenDesc), X4 (PerformanceScore), X5 (EngagementSurvey) and X6(SpecialProjectsCount).

* reg5 = lm(EmpSatisfaction ~ PayRate + Sex + CitizenDesc + PerformanceScore + EngagementSurvey + SpecialProjectsCount, data = HR)

1. Conduct a test of overall model significance. State H0 and Ha, the F-statistic, degrees of freedom, p-value, and conclusion. (2 pts) Text

   Description automatically generated with medium confidence
   1. H0: Bk = 0

Ha: Bk ≠ 0

* 1. F-Statistic = 3.846
  2. DF = 173
  3. As of right now, without dropping any values, it appears that we would reject the H0 in favor of the HA due to the low overall p-value.

1. Conduct tests of partial significance for each slope coefficient (X1, X2, X3, X4, X5, X6). For each, state H0 and Ha, the test statistic, degrees of freedom, p-value, and conclusion. (2 pts)
   1. X1: Pay Rate
      1. H0: B1 = 0

Ha: B1 ≠ 0

* + 1. Test Statistic = 2.553
    2. DF = 172
    3. P-value = 0.0099471
    4. Conclusion: With the low p-value and no other levels interfering, we can see that pay rate is related to employee satisfaction when compared to the base level of SexF.
  1. X2: Sex
     1. H0: B2 = 0

Ha: B2 ≠ 0

* + 1. Test Statistic M = -0.766
    2. Test Statistic F = 8.508
    3. DF = 172
    4. P-value = 0.23710400.2371040
    5. Conclusion: Since the overall p-value and the individual SexM level are both over our alpha of .1, this does not contribute to employee satisfaction when compared to the base level of SexF.
  1. X3: Citizen Desc
     1. H0: B3 = 0

Ha: B3 ≠ 0

* + 1. Test Statistic (Non-Citizen) = 1.216
    2. Test Statistic (Citizen)= -0.828
    3. DF = 172
    4. P-value = 0.2304827
    5. Conclusion: Since the overall p-value and the CitizenDescNon-Citizen & CitizenDescUS Citizen levels are both over our alpha of .1, this does not contribute to employee satisfaction when compared to the base level of SexF.
  1. X4: Performance Score
     1. H0: B4 = 0

Ha: B4 ≠ 0

* + 1. Test Statistic = PerfomanceScoreFully = -0.304
    2. Test Statistic = PerfomanceScoreNeeds Improvement = -1.627
    3. Test Statistic = PerfomanceScorePIP = -4.197
    4. DF = 172
    5. P-value = 0.0001191
    6. Conclusion: Since one of the levels of performance score is below the alpha = .1, that means that performance score is related to employee satisfaction when compared to the base level of SexF.
  1. X5: Engagement Survey
     1. H0: B5 = 0

Ha: B5 ≠ 0

* + 1. Test Statistic = -0.478
    2. DF = 172
    3. P-value = 0.6468278
    4. Conclusion: Since the overall p-value of the engagement survey is above the alpha level of .1, we can conclude that it is not significant to employee satisfaction when compared to the base level of SexF.
  1. X6: Special Projects Count
     1. H0: B6 = 0

Ha: B6 ≠ 0

* + 1. Test Statistic = -1.125
    2. DF = 172
    3. P-value = 0.2622276
    4. Conclusion: Since the overall p-value of the Special Projects Count is above the alpha level of .1, we can conclude that it is not significant to employee satisfaction when compared to the base level of SexF.

1. Dropping any insignificant variable at α = 0.1 (one by one starting by the more complex asnd insignificant variables, and then by significance level), what is the final first order least squares regression fit? (2 pts)
   1. reg5 = lm(EmpSatisfaction ~ PayRate + PerformanceScore, data = HR)Text

      Description automatically generated
2. What percentage of the variation in EmpSatisfaction can be explained by using your first-order final model in c)? (2 pts)
   1. .1443 or 14.43%
3. Explain the effect that each of the predictors in final model c) has on the response. (2 points)
   1. Exceeds = base
   2. Pay Rate
      1. When pay rate increases by one, employee satisfaction increases by .098 when compared to an exceeding performance score while holding everything else constant.
   3. Performance Score
      1. When Performance Score Fully Meets increases by one, employee satisfaction increases by 0.57 when compared to an exceeding performance score while holding everything else constant.
      2. When Performance Score Needs Improvement increases by one, employee satisfaction decreases by 6.06 when compared to an exceeding performance score while holding everything else constant.
      3. When Performance Score PIP increases by one, employee satisfaction decreases by 15.89 when compared to an exceeding performance score while holding everything else constant.

Higher-order Regression Model (10 points)

1. Fit the multiple regression model, regressing Y (EmpSatisfaction) on X1 (PayRate), X2 (Sex), X3 (CitizenDesc), X4 (PerformanceScore), X5 (EngagementSurvey), X6(SpecialProjectsCount), X7 (Quadratic of Special Projects Count: I(SpecialProjectsCountˆ2), X8 (Interaction Between Pay Rate and Special Projects Count: PayRate:SpecialProjectsCount) .

- reg6 = lm(EmpSatisfaction ~ PayRate + Sex + CitizenDesc + PerformanceScore + EngagementSurvey+ SpecialProjectsCount + I(SpecialProjectsCount\*SpecialProjectsCount) + PayRate:SpecialProjectsCount, data = HR)

summary(reg6)

Text

Description automatically generated

a) Start dropping insignificant variables at α = 0.10 (one by one starting by the more complex and insignificant variables, and then by significance level) until you get a model with all significant terms. (4 pts)

b) State your final higher-order model. (2 pts)

1. reg6 = lm(EmpSatisfaction ~ PayRate + CitizenDesc + PerformanceScore + SpecialProjectsCount + I(SpecialProjectsCount\*SpecialProjectsCount) + PayRate:SpecialProjectsCount, data = HR)

c) Explain the effect of each predictor on the response (2 pts) d) What percentage of the variation in satisfaction\_level can be explained by using your higher-order final model in b)? (2 pts)

1. Pay Rate
   1. When pay rate increases by one, employee satisfaction increases by .06 when compared to SexF while holding everything else constant.
2. CitizenDesc
   1. When the CitizensDesc Non-Citizen increases by one, employee satisfaction increases by 16.7 when compared to SexF while holding everything else constant.
   2. When the CitizensDesc Citizen increases by one, employee satisfaction decreases by -2.24 when compared to SexF while holding everything else constant.
3. Performance Score
   1. When Performance Score Fully Meets increases by one, employee satisfaction decreases by -.31 when compared to SexF while holding everything else constant.
   2. When Performance Score Needs Improvement increases by one, employee satisfaction decreases by 6.65 when compared to SexF while holding everything else constant.
   3. When Performance Score PIP increases by one, employee satisfaction decreases by 15.46 when compared to SexF while holding everything else constant.
4. Special Projects Count
   1. When Special Projects Count increase by one, employee satisfaction increases by .61 when compared to SexF while holding everything else constant.
5. Special Projects^2
   1. The variable of the special projects will be dependent on the exact value/number of special projects. As this increase by one, the employee satisfaction will decrease by 0.-56 when compared to SexF while holding everything else constant.
6. Payrate:SpecialProjectsCount
   1. When comparing both Pay Rate and Special Projects together, we can see that they are correlated and as the constant payrate and special projects increase by one, the employee satisfaction increases by .05 when compared to SexF while holding everything else constant.
7. 20% of variation can be explained by this model