**HR Analytics Report**

The dependent variables areemployee performance, satisfaction, absences and Salary.

Employee performance and Satisfaction are categorical variables.

Absences is a count variable.

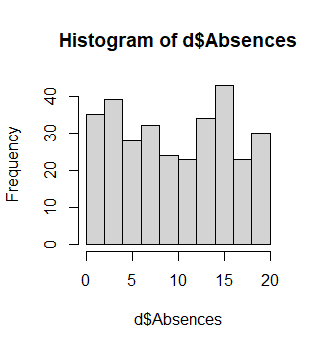
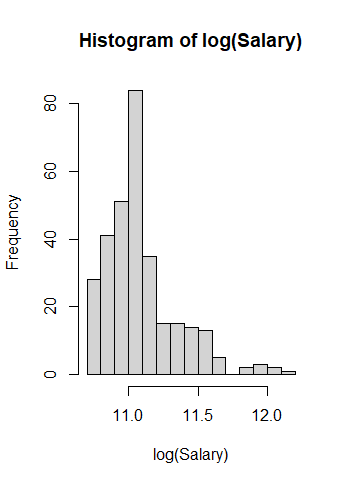
Let’s look at the distribution of these dependent variables in question.

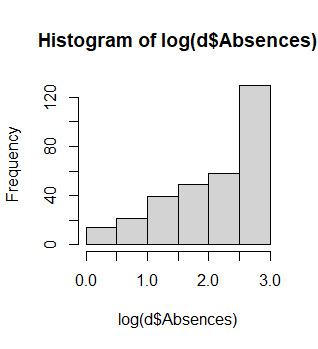
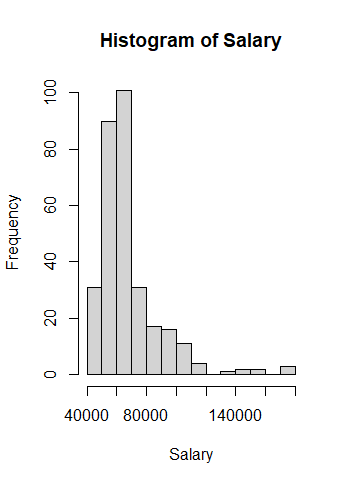
Table(d$PerformanceScore) table(EmpSatisfaction)

|  |  |  |  |
| --- | --- | --- | --- |
| Exceeds | Fully Meets | Needs Improvement | PIP |
| 37 | 243 | 18 | 13 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| 2 | 9 | 108 | 94 | 98 |

Logistic regression will not be a good fit because we have more than 3 levels in the dependent variables. Multinomial logit and Ordered logit (Because they have an order to the levels) would be a better choice of modelling these DV’s. Also, another option would be to run Tobit models because the DV’s are bounded between a range.

More than 50 % of the employees have been rated as “Fully meets” and more than 60 % of the employees have a satisfactory level of 3 or above.



The distribution of Absences does not seem to be normal, and OLS will not be a good choice for modelling it. The log transformation has only made it left-skewed and Poisson models will be over dispersed. We can try running quasi-Poisson or negative binomial models on the absences variable. The salary looks approximately normal after a log transformation and hence we can try glm models with link=log on it.

Descriptive Analysis

There are 206 employees that are currently working and 104 are terminated.

The mean and median rating of engagement survey is 4 which looks good for the company.

The mean and median absences over the last 30 days is 10 days which is quite high.

The sales department seems to have the highest average number of absences over the last 30 days.

Salaries of employees range from 45k to 250k. But the average salary is around 62 to 70K. This gap is due to the high salary of President and CEO which are 250k and 220k respectively.

Only 1 employee has been hired from a Web application source, and 2 employees have been hired from “Other” sources.

240 out of 311 employees do not have a special project.

Correlation Analysis

A picture containing diagram

Description automatically generated

High correlation between DeptID and SpecialProjectsCount. And, between Performance Score ID and DaysLateLast30. Hence, we will not include them both as predictors in the model.

Variable Selection- Predictor table for Performance Score, Employee Satisfaction and Absences

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Effect on Performance Score | Effect on Employee Satisfaction | Effect on Absences | Rationale |
| Salary | + | +/- | ? | Higher pay may lead to higher motivation and job satisfaction, resulting in better performance. Its impact on absences is unknown. |
| Department | +/- | +/- | ? | Departmental culture, resources, and management can either positively or negatively impact performance and satisfaction depending on various factors specific to a department. Its effect on absences is unknown. We also need to it answer Q.2 |
| Manager | +/- | +/- | ? | Managers might have a difference in styles on how they lead their teams, and it could impact differently on different employee’s performance and job satisfaction. It is also needed to answer Q2 |
| Recruitment Source | ? | ? | ? | Recruitment sources is needed to answer Q3. |
| Race | ? | ? | ? | Required to answer Q5. |
| Sex | ? | ? | ? | Required to answer Q5. |
| Employee Performance | N/A | + | N/A | This DV can be a predictor for Employee satisfaction as high performing employees can be more satisfied with their job than others. |
| Employee Satisfaction | + | N/A | N/A | This DV can be a predictor for employee performance as employees who are satisfied with their job, would be performing well. |
| Absences | - | N/A | N/A | Absences is included to predict Employee performance as high absences may lead to lower performance. |

• Employee Name, EmpID, MaritalStatusID, EmpStatusID, FromDiversityJobFairID, Termd, PositionID, State, Zip, DOB, Sex, MaritalDesc, CitizenDesc, HispanicLatino, RaceDesc, DateofHire, DateofTermination, TermReason, EmploymentStatus, ManagerName, and ManagerID (Correlates with Manager Name) were excluded as they do not provide any direct insight into employee performance and are not expected to have a significant impact on performance.

• PerformanceScore and EmpSatisfaction were excluded as they are outcome variables that we are trying to predict and including them in the model would lead to biased results. But they were included as predictors in other models.

•SpecialProjectsCount and DaysLast30 were excluded due to high correlation as mentioned in the above correlation plot.

Predictor table for Salary-

|  |  |  |
| --- | --- | --- |
| Variable | Estimated effect | Rationale |
| Department | +/- | Employees in certain departments might earn more. |
| Race | ? | To answer Q5 |
| Sex | ? | To answer Q5 |
| Performance Score | +/- | Higher performance may lead to higher salary and vice versa. |

Models-

1. Multinomial Models

per\_ml <- multinom(PerformanceScore ~ Salary + EmpSatisfaction + Department \* ManagerName

+ Absences + RecruitmentSource + RaceDesc + Sex, data=d, trace=FALSE)

sat\_ml = multinom(EmpSatisfaction ~ Salary + PerformanceScore + Department \* ManagerName +

RecruitmentSource + RaceDesc + Sex , data=d,subset = Termd=="0", trace=FALSE)

1. Negative Binomial Model for Absences

library(MASS)

ab\_nb = glm.nb(Absences ~ Department + PerformanceScore + Salary + ManagerName + RecruitmentSource + RaceDesc + Sex , data = d,link = log)

1. Ordinal Models for PerformanceScore and EmpSatisfaction

per\_ol <- polr(PerformanceScore ~ Salary + EmpSatisfaction + Department \* ManagerName

+ Absences + RecruitmentSource + RaceDesc + Sex, data=d, Hess=TRUE)

1. Tobit models on PerformanceScore and EmpSatisfaction

library(AER)

tobi\_per = tobit(PerfScoreID ~ EmpSatisfaction + Department + ManagerName + RecruitmentSource, left=1, right=4, data=d)

summary(tobi\_per)

tobi\_sat = tobit(EmpSatisfaction ~ PerformanceScore + Department + ManagerName + RecruitmentSource, left=1, right=5, data=d)

summary(tobi\_sat)

Output-

|  |  |  |
| --- | --- | --- |
|  | | |
|  | *Dependent variable:* | |
|  |  | |
|  | PerfScoreID | EmpSatisfaction |
|  | (1) | (2) |
|  | | |
| Salary | 0.00001\* (0.00000) |  |
| EmpSatisfaction | 0.234\*\*\* (0.043) |  |
| PerformanceScore.L |  | 1.387\*\*\* (0.269) |
| PerformanceScore.Q |  | -0.450\* (0.253) |
| PerformanceScore.C |  | 0.178 (0.220) |
| DepartmentAdmin Offices | -0.070 (0.916) | 0.231 (1.629) |
| DepartmentIT/IS | -0.541 (0.740) | 2.472\* (1.276) |
| DepartmentProduction | 0.296 (0.668) | -0.642 (1.191) |
| DepartmentSales | -0.564 (0.973) | -0.281 (1.615) |
| ManagerNameAmy Dunn | -0.217 (0.723) | 0.0002 (1.277) |
| ManagerNameBoard of Directors | 0.251 (1.311) | -1.642 (2.331) |
| ManagerNameBrandon R. LeBlanc | 0.238 (0.993) | -0.952 (1.752) |
| ManagerNameBrannon Miller | -0.211 (0.726) | -0.350 (1.283) |
| ManagerNameBrian Champaigne | 0.445 (0.802) | -2.973\*\* (1.409) |
| ManagerNameDavid Stanley | -0.176 (0.726) | 0.099 (1.283) |
| ManagerNameDebra Houlihan | 0.215 (1.104) | 0.330 (1.815) |
| ManagerNameElijiah Gray | -0.150 (0.722) | 0.124 (1.279) |
| ManagerNameEric Dougall | 1.052 (0.882) | -3.708\*\* (1.457) |
| ManagerNameJanet King | -0.090 (0.680) | -0.432 (1.216) |
| ManagerNameJennifer Zamora | 0.300 (0.682) | -1.942 (1.213) |
| ManagerNameJohn Smith | 0.337 (1.056) | 0.306 (1.702) |
| ManagerNameKelley Spirea | -0.055 (0.726) | -0.040 (1.285) |
| ManagerNameKetsia Liebig | -0.137 (0.726) | 0.240 (1.282) |
| ManagerNameKissy Sullivan | -0.195 (0.726) | 0.153 (1.284) |
| ManagerNameLynn Daneault | 0.636 (1.059) | 0.475 (1.704) |
| ManagerNameMichael Albert | -0.316 (0.724) | 0.283 (1.279) |
| ManagerNamePeter Monroe | 0.323 (0.826) | -2.868\*\* (1.391) |
| ManagerNameSimon Roup | 0.544 (0.789) | -3.221\*\* (1.379) |
| ManagerNameWebster Butler | -0.157 (0.731) | 0.231 (1.291) |
| RecruitmentSourceCareerBuilder | 0.132 (0.662) | -7.044 (728.813) |
| RecruitmentSourceDiversity Job Fair | 0.119 (0.665) | -6.813 (728.813) |
| RecruitmentSourceEmployee Referral | 0.246 (0.664) | -6.544 (728.813) |
| RecruitmentSourceGoogle Search | -0.003 (0.655) | -6.304 (728.813) |
| RecruitmentSourceIndeed | 0.037 (0.654) | -6.492 (728.813) |
| RecruitmentSourceLinkedIn | 0.106 (0.653) | -6.532 (728.813) |
| RecruitmentSourceOther | -0.013 (0.921) | -6.835 (728.814) |
| RecruitmentSourceWebsite | 0.040 (0.682) | -7.312 (728.813) |
| Constant | 1.538\* (0.787) | 10.830 (728.813) |
|  | | |
| Observations | 309 | 309 |
| Log Likelihood | -313.255 | -413.778 |
| Wald Test | 49.577\*\* (df = 34) | 61.252\*\*\* (df = 35) |
|  | | |
| *Note:* | \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 | |
|  |  | |

Rationale:

The Multinomial models were giving very high coefficients( Inf) and I tried with ordinal regression but the models did not converge. Lastly, the tobit model ran without hassle as the DV’s were bound in nature. We will interpret them to answer the questions. Also, it passes all the LINE assumptions and multicollinearity assumptions.

1. What are the top three predictors that contribute the most (positively or negatively) to employee performance, satisfaction, and absences, after controlling for other variables, and by how much?

Top 3 predictors for all the DVs-

|  |  |  |
| --- | --- | --- |
| Employee Performance Score | Employee Satisfaction | Absences |
| |  |  | | --- | --- | | ManagerNameEric Dougall | (0.455) | | |  | | --- | | DepartmentIT/IS |   (2.472) | |  |  | | --- | --- | | RecruitmentSourceOther | 2.634\*\* | |
| |  |  | | --- | --- | | RecruitmentSourceEmployee Referral | (0.295) | | |  |  |  | | --- | --- | --- | | PerformanceScore.L |  | 1.387\*\*\* | | |  |  | | --- | --- | | RecruitmentSourceDiversity Job Fair | 2.601\*\* | |
| |  |  | | --- | --- | | EmpSatisfaction | (0.237\*\*\*) | | |  | | --- | | ManagerNameLynn Daneault |   (0.475) | |  |  | | --- | --- | | RecruitmentSourceEmployee Referral | 2.410\*\* | |
| |  |  | | --- | --- | | ManagerNameDebra Houlihan | (-0.538) | | RecruitmentSourceWebsite (-7.31) | |  |  | | --- | --- | | ManagerBoard of Directors | -3.213\*\* | |
| |  |  | | --- | --- | | ManagerNameMichael Albert | (-0.538 |   ) | RecruitmentSourceCareerBuilder  (-7.04) | |  |  | | --- | --- | | ManagerNameBrandon R. LeBlanc | -1.772\* | |
| |  |  | | --- | --- | | ManagerNameAmy Dunn | (-0.444) | | RecruitmentSourceOther(-6.83) | |  |  | | --- | --- | | RaceDescHispanic | -1.457 | |

2. Is there any relationship between the manager and/or department an employee works for and their performance and satisfaction scores? If so, which managers and/or departments have the highest AND lowest employee performance and satisfaction scores, and what is the extent of this gap?

There seems to be no relation between the manager and the department an employee works for based on their performance and satisfactory scores. But there seems to be differences among them.

Production department has the highest performance score compared to the Software Engineering dept by 29.6 % while the lowest performance score is for Sales department by 56 % lower than the Engineering dept with the difference among the highest and lowest depts being 26.4%.

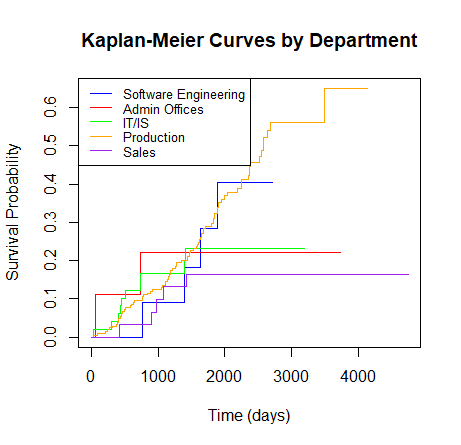
Eric Dougall from the IT dept has the highest performance score of 105 % compared to the Alex Sweetwater. The lowest performance score is for the manager Michael Albert which is 31 % less than Alex Sweetwater. The difference between these two is 74%.

3. Which recruitment source results in the highest AND lowest performing and most AND least satisfied employees and what is the gap in performance and satisfaction between the two recruitment sources?

Employee referral has the highest performing score compared to the Web-application source by 24 %. RecruitmentSourceWebsite has the lowest satisafaction score of 73 % less compared to the Web application recruitment source.

Recruitment source other has the lowest performing score compared to the web-application source by 1.3 % less than it.

4. What is the probability that a typical employee will continue working in each department of this company after 5 years?



From the output of km model, we can see that the estimated probabilities of survival at 1825 days(5 years) are:

the probability that a typical employee will continue working in each department of this company after 5 years is:

Software Engineering: 0.284

Admin Offices: 0.222

Executive Office: 0 (no events occurred)

IT/IS: 0.211

Production: 0.297

Sales: 0.137

5. Is there any department in the company where there is pay inequity on the basis of race or gender? How much is this pay inequity contributing to poor performance and/or low satisfaction in those departments?

This question is being answer by the nb model (output not shown here)

nb = glm.nb(Salary ~ Department \* (RaceDesc + Sex + PerformanceScore +EmpSatisfaction ) , data = d)

There seems to be a pay inequity among certain departments on the basis of race and gender explained below.

Asians in the Sales and IT departments are being paid 18.5 % less relative to Software Engineering departments with Blacks.

Due to this inequity, The sales department is performing 22.5 % lower compared to the Software engineering dept and IT dept is performing almost 50 % lower than the Software engineering dept.

But the employee satisfaction seems to be positive in these two departments by 8 and 10 % respectively higher than the Software engineering dept.

Being a Male in an IT dept pays 18 % more than being a female in Software engineering dept. All the males in all departments seems to be getting high pay compared to the females.

Relative to the females in Software engineering dept, The difference looks like this:

|  |  |
| --- | --- |
| DepartmentAdmin Offices:SexM | 0.088 (0.250) |
| DepartmentIT/IS:SexM | 0.271 (0.185) |
| DepartmentProduction:SexM | 0.188 (0.179) |
| DepartmentSales:SexM | 0.143 (0.191) |