HR Analytics

#Simran Agichani

>Data Exploration and Transformation for all three DV's:

Perfscoreid	Empsatisfaction	Absences
		Without log

```> table(df$perfscoreid)```
 	
``` 1   2   3   4```
```13 18 243 37```
	```> table(df$empsatisfaction)```
	
	``` 1 2 3 4 5```
	``` 2 9 108 94 98```
		With log transform

We can see that most employees are having average performance.	We can see that employees are generally satisfied, with very few having 1 or 2 as satisfaction index.	

Feature Engineering:

>The dateofhire, dateoftermination, dob, lastperformancereviewdate are all in excel's serial date format which need to be converted to date format for analysis.

>Used dateofhire and dateoftermination to extract number of days of employment as empdays.

>Used sysdate for calculating no. of days of employment for people with NA's in dateoftermination, assuming that they are still working.

>date transformations are done to answer the analysis question for finding probabilities of survival after 5 years.

>The performance and Empsatisfaction are both censored variables as they are ratings/index. Therefore, I will be using tobit model for modelling these two dependent variables.

>Histogram for Absences is a non-normal distribution and count type data, therefore I will be using glm poisson model.

>Since the data is multilevel as it has department which has employees and managers which have employees working under tham. Therefore, converted the variables of department and managername as fixed effects.

>After checking correlations between numeric variables, it was found that for particular dv's there was no issue with the chosen variables.
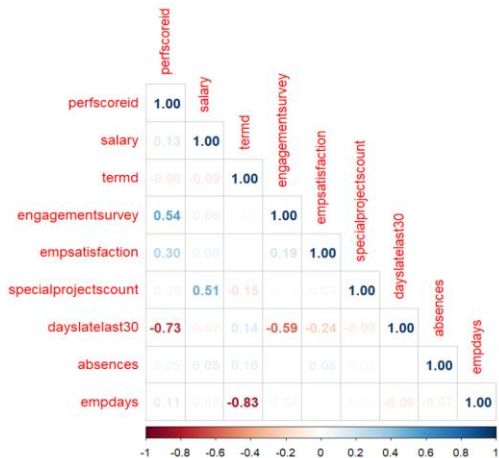
Q4: What is the probability that a typical employee will continue working in each department of this company after 5 years? (2 points)

For getting probabilities of an employee continuing the job, survival model can be used which could help get the survival probabilities when number of days are taken as 1825 which is basically 5 years as asked. The variables considered for this model would be:

Predictor for Probabilities	Effect	Rationale
Department	+/-	We want to examine the effect of Departments on termination outcomes.
Empdays	-	As people stay longer, they might tend to switch the job.
Termd	NA	To run the probability model, we need a status predictor to tell if a person was terminated or not.

Here the event is termination from job, and we are using Kaplan-Meier analysis to get probabilities to model time to event. We will be looking for values for each department that correspond to 1825 empdays mas asked for 5 years.

Probabilities by Departments:

There is 77% probability that an employee in Admin Office continues the job after 5 years. Similarly, the probabilities for other departments are:
Executive Office: 100% (might be that people are extremely satisfied in this dept.)
IT/IS: 80%
Production: 72%
Sales: 83%
Software Engineering: 72%

Q1: 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? (6 points)

Predictor Table for finding top three predictors for performance, satisfaction, absences.
Tobit models for Performance and Satisfaction as they are censored. Glm fixed effects for Absences as it is non normal count data.

Predictor	Performance	Satisfaction	Absences
Employee Name,EmpID	No, Unique identifier	No, Unique identifier	No, Unique identifier
MarriedID	No, doesn't explain better than status.	No, doesn't explain better than status.	No, doesn't explain better than status.
MaritalStatusID	No, Taken marital description because of better understanding.	No, Taken marital description because of better understanding.	No, Taken marital description because of better understanding.
EmpStatusID	No, Taken emp status because of better understanding.	No, Taken emp status because of better understanding.	No, Taken emp status because of better understanding.
DeptID	No, Taken department for clear understanding.	No, Taken department for clear understanding.	No, Taken department for clear understanding.
PerfScoreID	DV	Yes+, Good performance might improve satisfaction.	Yes+, Good performance might reduce absences.
FromDiversityJobFairID	No, it is covered in recruitment process.	No, it is covered in recruitment process.	No, it is covered in recruitment process.
Salary	Yes+, Might entice people to work more and hence has an effect.	Yes+, More pay leads to more financial stability hence would affect satisfaction.	Yes? More pay might mean more work, therefore less absence.
Termd	No, Status variable for termination, has been captured under employment status.	No, Status variable for termination, has been captured under employment status.	No, Status variable for termination, has been captured under employment status.
PositionID/Position	No, might be correlated with department.	No, might be correlated with department.	No, might be correlated with department.
State	No, Work is done irrespective of the location, nowadays remote option is also available.	No, Work is done irrespective of the location, nowadays remote option is also available.	Yes, Location of work if not remote might impact the feasibility and thereby absence.
Zip	No, Work is done irrespective of the location, nowadays remote option is also available.	No, Work is done irrespective of the location, nowadays remote option is also available.	Captured under state.
DOB	No, demographic factor.	No, demographic factor.	No, demographic factor.

Sex	Yes? To analyze variation based on gender.	Yes? To analyze variation based on gender.	Yes? To analyze variation based on gender.
MaritalDesc	Yes? Might affect performance as someone's personal life also defines their mental wellbeing.	Yes? Might affect satisfaction as someone's personal life also defines their mental wellbeing.	Yes? Might affect absences as someone's personal life might involve responsibilities.
CitizenDesc	No, Employment opportunities tend to not be biased and provide inclusivity.	Yes, Employment opportunities for eligible/noneligible citizens might affect their satisfaction.	No, Employment opportunities tend to not be biased and provide inclusivity.
HispanicLatino/RaceDesc	No, Demographic factor doesn't tell the knowledge of a person, hence not considered.	No, Demographic factor doesn't tell the mental state of a person, hence not considered.	No, Demographic factor doesn't tell the knowledge of a person, hence not considered.
DateofHire/Date of Termination	No, not an aspect that could define performance as it is a time variable. Used for getting empdays.	No, not an aspect that could define satisfaction as it is a time variable. Used for getting empdays.	No, not an aspect that could define absences as it is a time variable. Used for getting empdays.
TermReason	No, it can be the effect and not the cause that could determine performance.	No, it can be the effect and not the cause that could determine satisfaction.	No, It can be the effect and not the cause that could determine absences.
EmploymentStatus	Yes? Might affect the mental state of a person impacting performance.	Yes? Might affect the mental state of a person impacting satisfaction.	No, absence makes no sense if taken for terminated people, reverse causality.
Department	Yes? Used for analysis to answer the questions on dept level.	Yes? Used for analysis to answer the questions on dept level.	Yes? Some departments might have more work load, might lead to more absence.
ManagerName	Yes? Used for analysis to answer the questions.	Yes? Used for analysis to answer the questions.	No, Absence might depend on type of work but not manager.
ManagerID	No, not as descriptive as ManagerName, hence not taken.	No, not as descriptive as ManagerName, hence not taken.	No, not as descriptive as ManagerName, hence not taken.
RecruitmentSource	Yes? Taken for analysis of given questions.	Yes? Taken for analysis of given questions.	No, the source from where a person is hired does not define how much absent he/she will be.
PerformanceScore	No, taken perfscoreid.	No, taken perfscoreid	No, taken perfscoreid
EngagementSurvey	Yes+, taken as it might tell the interest of a	Yes+, taken as it might tell the interest of a	Yes+, taken as it might tell the interest of a

	person which could determine performance.	person which could determine satisfaction.	person which could impact absences.
EmpSatisfaction	Yes+, more satisfied person might perform better.	DV	Yes+, more satisfied person might be less unavailable.
SpecialProjectsCount	Yes+, It may define new opportunities hence can affect the performance.	Yes+, It may define new opportunities hence can affect the satisfaction.	Yes+, It may define new opportunities hence can affect the absences.
LastPerformanceReviewDate	No, cannot be related to how a person performs. No cause effect.	No, cannot be related to how much a person satisfied. No cause effect.	No, cannot be related to how much a person is absent. No cause effect.
DaysLateLast30	Yes, more late a person arrives, lower might be the performance.	No, absence captures more.	No, can fully define absence.
Absences	Taken dayslatelast30, to capture the effect of unavailability.	Yes, higher the unavailability, lower might be the satisfaction.	DV

Stargazer:

Dependent variable:

```
                                -------------------------------------------------------------

                                perfscoreid      empsatisfaction       absences

                                   Tobit             Tobit             Poisson

                                    (1)               (2)                (3)

--------------------------------------------------------------------------------------------

perfscoreid                                       0.392*** (0.098)    0.064* (0.038)

salary                      0.00000 (0.00000)   0.00000 (0.00000)  0.00001*** (0.00000)

sexM                                              -0.066 (0.101)

maritaldescMarried                                -0.178 (0.174)      0.042 (0.067)

maritaldescSeparated                              -0.150 (0.295)     -0.302*** (0.117)

maritaldescSingle                                 -0.055 (0.176)      -0.100 (0.067)

maritaldescWidowed                                 0.215 (0.358)      -0.055 (0.126)

citizendescNon-Citizen                             0.329 (0.497)

citizendescUS Citizen                             -0.209 (0.252)

employmentstatusTerminated for Cause   0.207** (0.105)      0.082 (0.229)
```

employmentstatusVoluntarily Terminated	-0.052 (0.053)	-0.023 (0.125)	
departmentExecutive Office	-0.488 (0.701)	-0.248 (1.614)	0.133 (0.702)
departmentIT/IS	0.357 (0.517)	-0.060 (1.264)	-1.017** (0.442)
departmentProduction	0.368 (0.382)	-0.854 (0.864)	-0.685*** (0.241)
departmentSales	-0.226 (0.541)	-0.447 (1.220)	-0.649* (0.349)
departmentSoftware Engineering	0.230 (0.477)	-0.833 (1.107)	-1.473*** (0.445)
managernameAmy Dunn	-0.017 (0.349)	0.309 (0.850)	-0.585 (0.418)
managernameBoard of Directors	0.082 (0.715)	-1.393 (1.659)	-2.746*** (0.732)
managernameBrandon R. LeBlanc	0.136 (0.516)	-1.175 (1.190)	-1.517*** (0.475)
managernameBrannon Miller	-0.042 (0.348)	-0.119 (0.843)	-0.637 (0.416)
managernameBrian Champaigne	-0.283 (0.400)	-1.104 (0.910)	-0.229 (0.447)
managernameDavid Stanley	-0.142 (0.349)	0.378 (0.849)	-0.480 (0.417)
managernameDebra Houlihan	0.234 (0.566)	0.402 (1.316)	-0.741 (0.529)
managernameElijiah Gray	-0.039 (0.347)	0.349 (0.846)	-0.646 (0.417)
managernameEric Dougall	-0.008 (0.454)	-1.583 (1.043)	-0.288 (0.497)
managernameJanet King	-0.059 (0.324)	-0.093 (0.795)	-0.841** (0.401)
managernameJennifer Zamora	-0.111 (0.334)	-0.739 (0.768)	-0.786* (0.411)
managernameJohn Smith	0.343 (0.536)	0.216 (1.257)	-0.296 (0.500)
managernameKelley Spirea	-0.127 (0.349)	0.212 (0.846)	-0.378 (0.417)
managernameKetsia Liebig	-0.102 (0.348)	0.423 (0.845)	-0.609 (0.418)
managernameKissy Sullivan	-0.094 (0.349)	0.323 (0.848)	-0.451 (0.417)
managernameLynn Daneault	0.621 (0.538)	0.237 (1.263)	-0.615 (0.503)
managernameMichael Albert	-0.159 (0.348)	0.456 (0.844)	-0.406 (0.417)
managernamePeter Monroe	-0.362 (0.415)	-0.958 (0.949)	-0.252 (0.459)
managernameSimon Roup	-0.270 (0.390)	-1.255 (0.887)	-0.535 (0.445)
managernameWebster Butler	-0.109 (0.352)	0.413 (0.849)	-0.335 (0.417)
recruitmentsourceDiversity Job Fair	-0.009 (0.108)	0.215 (0.250)	
recruitmentsourceEmployee Referral	-0.031 (0.109)	0.234 (0.249)	
recruitmentsourceGoogle Search	-0.169* (0.096)	0.468** (0.215)	

recruitmentsourceIndeed   -0.144 (0.091)   0.349* (0.206)

recruitmentsourceLinkedIn   -0.088 (0.091)   0.283 (0.204)

recruitmentsourceOn-line Web application   -0.272 (0.377)   1.204 (0.860)

recruitmentsourceOther   -0.186 (0.378)   0.026 (0.905)

recruitmentsourceWebsite   -0.282** (0.136)   -0.229 (0.312)

engagementsurvey   0.115*** (0.034)   0.092 (0.073)   -0.031 (0.028)

empsatisfaction   0.078*** (0.025)

dayslatelast30   -0.297*** (0.021)

specialprojectscount   0.108 (0.098)   -0.004 (0.039)

absences   0.006 (0.008)

Constant   1.992*** (0.535)   2.808** (1.296)   3.078*** (0.493)

Observations   311   311   311

Log Likelihood   -122.825   -373.403   -1,137.723

Akaike Inf. Crit.   2,343.445

Wald Test   517.289*** (df = 39) 85.571*** (df = 47)

==
===========

When not considering the sign (positive/negative) the top predictors according for performance, satisfaction, absence by controlling for other variables are:

Performance	Satisfaction	Absences
Department, Manager, RecruitmentSource	Department, Manager, Citizendesc	Manager, Department, Maritaldesc

Q2: 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? (2 points)

Q3: 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? (1 point)

	Managers	Departments	Recruitment Source
Performance	As compared to Alex Sweetwater, Lowest: Peter Monroe, Brian Champaigne, Simon Roup Highest: Lynn Daneault, John Smith, Debra Houlihan	As compared to admin office: Lowest: Executive office, sales Highest: Production, IT/IS	Top2: Career builder, diversity job fair, Bottom 2: Online web application, website
Satisfaction	As compared to Alex Sweetwater, employees with: Amy Dunn, Debra Houlihan, Ketsia Liebig are more satisfied and employees with: Eric Dougall,Board of Directors, Simon Roup are least satisfied.	Order of satisfaction by dept: Admin Office, IT/IS, Executive OFIce	Top 2: Online web application, google search Bottom2: website, career builder

Q5: 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? (3 points)

Some interactions between department and sex/racedec/performacescore/satisfaction might work, but unsure how to interpret them. Ran the model though.