

IBM HR EMPLOYEE ATTRITION MANAGEMENT ATTRITION PREDICTION

3X DATA MINING GROUP MEMBERS

IGOR PEDEVANI, MOHAMMED TOPIWALLA ,PATRICIA LONDONO, MILICENT OTCHERE, SANCHITA KUMARI, VALERIO TROTTA

PROBLEM STATEMENT

- OUR CLIENT IS IBM A LEADING FIRM AND IN THE IT SECTOR. IT IS RECENTLY FACING A STEEP INCREASE IN ITS EMPLOYEE ATTRITION . EMPLOYEE ATTRITION HAS GONE UP FROM 14% TO 25% IN THE LAST 1 YEAR . WE ARE ASKED TO PREPARE A STRATEGY TO IMMEDIATELY TACKLE THIS ISSUE SUCH THAT THE FIRM'S BUSINESS IS NOT HAMPERED AND ALSO TO PROPOSE AN EFFICIENT EMPLOYEE SATISFACTION PROGRAM FOR THE LONG RUN. CURRENTLY, NO SUCH PROGRAM IS IN PLACE . FURTHER SALARY HIKES ARE NOT AN OPTION.
- THE ATTRITION PROBLEM IS NOT ONLY UNIQUE TO IBM BUT TO OTHER IT COMPANIES SUCH AS INFOSYS, INDIA'S SECOND LARGEST IT SERVICES COMPANY, THAT IS ALSO BATTLING HIGH ATTRITION, WITH A PEAK ATTRITION OF 20.4 % IN THE OCTOBER-DECEMBER QUARTER OF FY15.

HOW CAN WE REDUCE IBM COMPANY'S ATTRITION RATE BY PREDICTING IF A CANDIDATE WILL EXIT IN INDIA WITHIN THE YEAR?

- **SPECIFIC** :- TO INDIAN GEOGRAPHY IN IBM
- **MEASURABLE**:- TO REDUCE ATTRITION RATE(BY AT LEAST 5%)
- **ACTION ORIENTED**:- REDUCE EMPLOYEE ATTRITION & SUGGEST EMPLOYEE ENGAGEMENT & SATISFACTION PROGRAMS
- **RELEVANT**:- DIRECT IMPACT ON COMPANY'S TOP AND BOTTOM LINE
- **TIME BOUND** :- 12 MONTHS

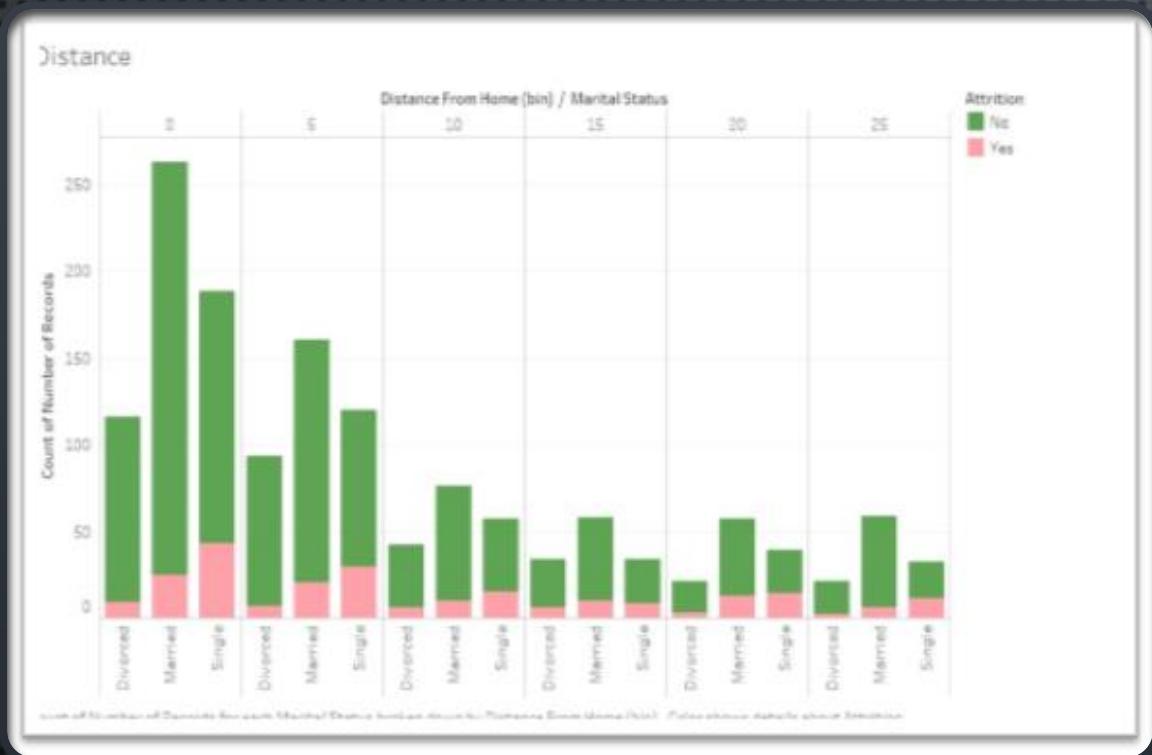
* The Smart framework is from Fractal analytics



HOW CAN WE REDUCE IBM COMPANY'S ATTRITION RATE BY PREDICTING IF A CANDIDATE WILL EXIT IN INDIA WITHIN THE YEAR

1)Background	4)Constraints
<ul style="list-style-type: none">• Best performing IT company in terms of topline and bottom line.• Employee Attrition has increased from 14% to 25% in the last 1 year, much higher than the industry average of 16%• Committed to curtailing attrition as it is not sustainable in the long run	<ul style="list-style-type: none">• Salary hike cannot be considered
2)Desired Outcome	5)Stakeholders
<ul style="list-style-type: none">• Reduce attrition rate by 5% in the next 18 months• Saving recruitment cost and improving employee satisfaction rate• develop a holistic employee satisfaction program	<ul style="list-style-type: none">• CEO/ HR Head/ BU Heads• Attrition cell
3)Scope	6)Resources
<ul style="list-style-type: none">• In-house attrition analysis tool and Early Warning System to identify individuals who are likely to leave and prioritize action items for immediate intervention	<ul style="list-style-type: none">• Interviews with HR head, Attrition cell, recruitment team, BU Heads• Insights based on industry best practices and secondary research• Review of exit interviews and HR attrition data to observe trends across departments, gender, experience level , etc.

EXPLORATORY DATA ANALYSIS



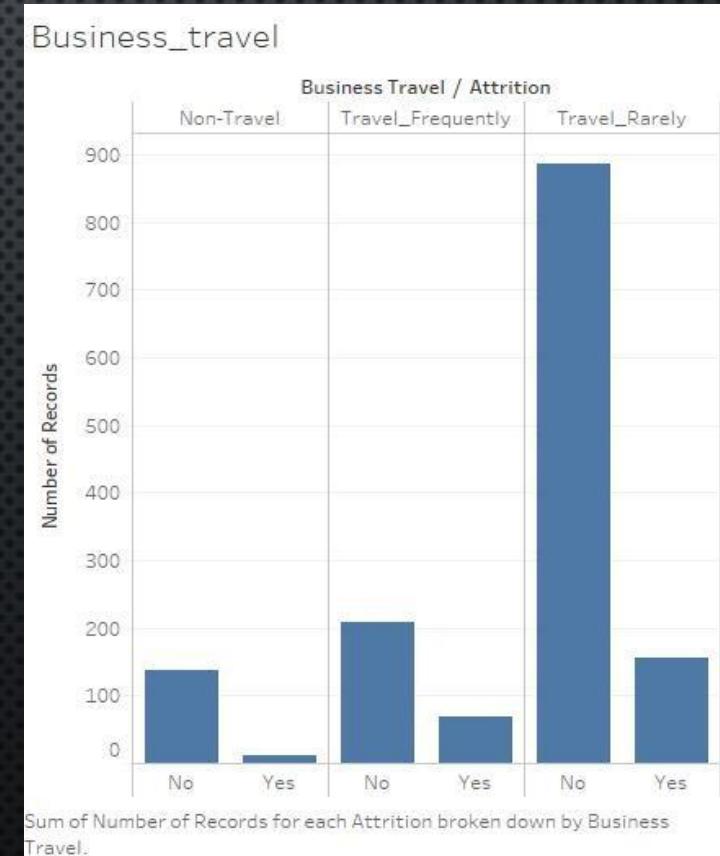
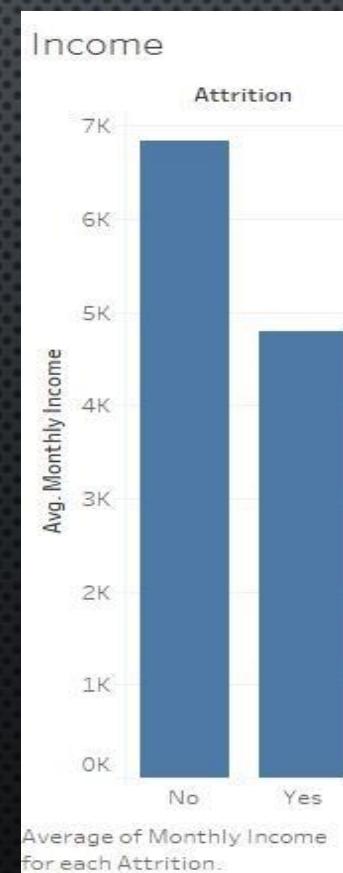
OUR AIM IS TO ANALYSE THE IBM ATTRITION DATA AND FIND WHAT TRENDS OR PATTERNS ITS SHOWING TOWARDS ATTRITION. TO UNDERSTAND THE DATA BETTER, WE DEEP DIVE INTO THE DATA AND TRY TO FIND WHAT COULD BE THE POSSIBLE REASON FOR THE ATTRITION (THE ANALYSIS HAS BEEN DONE STEP WISE TAKING EACH VARIABLE).

FROM OUR GRAPHS WE REALIZED SEVERAL INTERESTING PATTERNS, LIKE THAT PEOPLE OF AGE BETWEEN 25-35 AND PEOPLE WHO DIDN'T GET CHANCE FOR BUSINESS TRAVEL HAVE HIGH ATTRITION RATE, OR THAT EMPLOYEES WHO ARE MARRIED AND LIVING AT A DISTANT PLACE FROM COMPANY HAVE HIGH ATTRITION RATE, TOO.

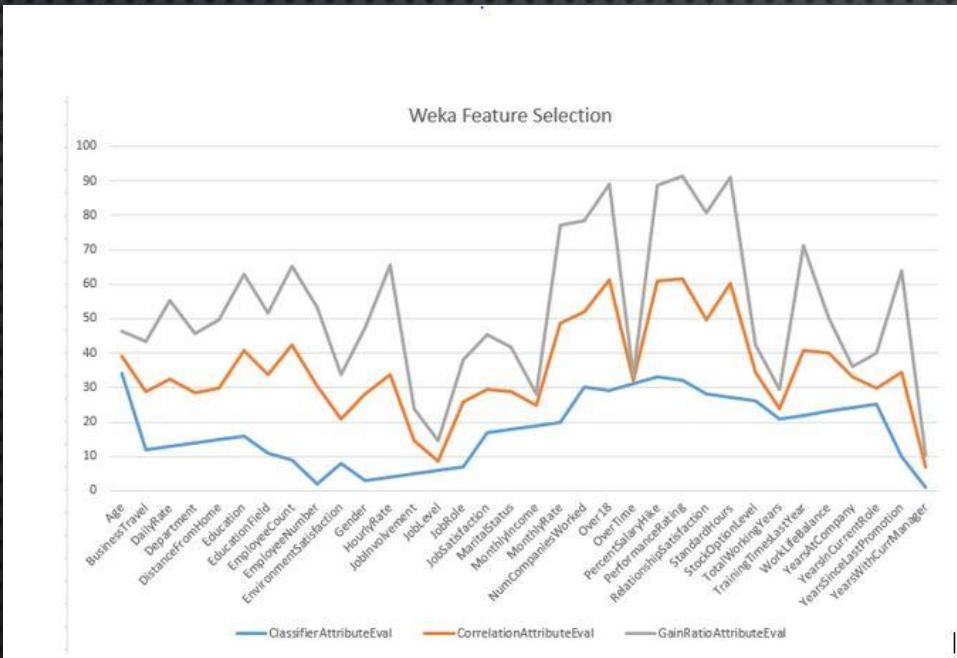
FURTHERMORE, WE HAVE BEEN ABLE TO UNDERSTAND THE PRINCIPAL CATEGORIES WITH THE HIGHEST ATTRITION RATES (I.E.: SALES DEPARTMENT, HR & TECHNICAL DEGREE IN EDUCATION, SINGLES, MALES, LOW INCOME, ETC...)

EXPLORATORY DATA ANALYSIS

- FROM THE GRAPH WE OBSERVE THAT AVERAGE MONTHLY INCOME AFFECT ATTRITION RATE. ALSO PEOPLE WHO DO NOT GET THE CHANCE TO TRAVEL FOR BUSINESS HAVE HIGHER ATTRITION RATES.



FEATURE SELECTION



SEVERAL TECHNIQUES WERE USED IN WEKA FOR FEATURE SELECTION

- CLASSIFIER ATTRIBUTE EVALUATION, CORRELATION ATTRIBUTE EVALUATION AND GAIN RATIO ATTRIBUTE EVALUATION. THE CHART BELOW SUMMARIZES THE RESULTS THAT THESE MODELS FOUND

FEATURE SELECTION

CfsSubsetEval		
GreedyStepwise		
10 fold cross-validation (stratified), seed: 1		
number of features (%)		attribute
10	100 %)	1 Age
10	100 %)	2 BusinessTravel
10	100 %)	22 OverTime
10	100 %)	27 StockOptionLevel
10	100 %)	34 YearsWithCurrManager
9	90 %)	13 JobInvolvement
9	90 %)	31 YearsAtCompany
8	80 %)	18 MonthlyIncome
7	70 %)	10 EnvironmentSatisfaction
7	70 %)	14 JobLevel
7	70 %)	28 TotalWorkingYears
6	60 %)	16 JobSatisfaction
6	60 %)	30 WorkLifeBalance
2	20 %)	15 JobRole
0	0 %)	3 DailyRate
0	0 %)	4 Department
0	0 %)	5 DistanceFromHome

- A CORRELATION BASED FEATURE SELECTION SUBSET EVALUATION METHOD WAS ALSO RUN .
- USING THIS METHOD OF WEKA WE WERE ABLE TO IDENTIFY 13 IMPORTANT VARIABLES USEFUL FOR CLASSIFICATION
- USING THE BACKWARD ELIMINATION METHOD, WE WERE ABLE TO IDENTIFY THE FOLLOWING 9 VARIABLES THAT WERE NOT USEFUL IN PREDICTING ATTRITION IN OUR CLASSIFICATION MODEL:

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	Education	4	29	0.7877	0.9401
2	PerformanceRating	1	28	0.0779	0.7801
3	Department	2	27	0.5313	0.7667
4	PercentSalaryHike	1	26	0.2384	0.6253
5	MonthlyRate	1	25	0.5345	0.4647
6	HourlyRate	1	24	0.6436	0.4224
7	MaritalStatus	2	23	2.2497	0.3247
8	MonthlyIncome	1	22	1.8706	0.1714
9	DailyRate	1	21	3.7932	0.0515

DATA CLEANING AND BALANCING

- WELL OUR DATA IS SERIOUS IMBALANCED, SO THE FIRST CHALLENGE WE FACED WAS BALANCING IT (83/16)%
- SECOND WE HAD A LOT OF FACTOR VARIABLES WHICH WE NEEDED TO CONVERT TO DUMMIES BEFORE BEGINNING
- EXCEPT THIS OUR DATA WAS SQUEAKY CLEAN

CLASSIFICATION & INTERPRETATION

IN ORDER TO CREATE AN EARLY WARNING SYSTEM WE CREATED A MACHINE LEARNING MODEL THAT WOULD PERFORM CLASSIFICATION TO CHECK IF AN EMPLOYEE COULD QUIT THE COMPANY

Model	Data	ACC	ROC	Kappa
Logistic Regression	All data	88.04%	0.67	0.42
	Smote 1	81.52%	0.82	0.63
	Smote 2	79.25	0.79	0.58
	Fs via Backward elimination	86.41%	0.64	0.34
	Fs via CFS subset evaluation	85.32%	0.52	0.05
Decision Tree	All Data	84.23%	0.61	0.25
	Smote 1	85.57%	0.86	0.71
	Smote 2	92.22%	0.92	0.84
	Fs via Backward elimination	83.96%	0.64	0.31
	Fs via CFS subset evaluation	84.23%	0.61	0.25
Random Forest	All data	87.72%	0.59	0.26
ANN	All data	85.86%	0.64	0.33
SVM	All data	83.9%	0.58	0.19
XGBOOST	All data	88.04%	0.66	0.4
	Smote 1	91.41%	0.92	0.82
	Smote 2	87.35%	0.87	0.74

CLASSIFICATION & INTERPRETATION

FROM THE PREVIOUS RESULTS ITS CLEAR THAT DECISION TREE STOLE THE SHOW!

HOWEVER LETS THINK PRACTICALLY

- IT IS OFTEN REQUIRED TO EXPLAIN THE BUSINESS WHY WE THINK A PERSON COULD LEAVE, IN THIS CASE WE NEED A MODEL WHOSE OUTPUT WE CAN EXPLAIN. IN OUR CASE A DECISION TREE OR LOGISTIC REGRESSION
- SOMETIMES HR WOULD JUST LIKE TO RUN OUR MODEL ON RANDOM DATA SETS , SO ITS NOT ALWAYS POSSIBLE TO BALANCE OUR DATASETS USING TECHNIQUES LIKE SMOTE
- OUR MODEL SHOULD JUST BE ABLE TO PREDICT BETTER THAN RANDOM BUT IMAGINE THE COST OF ENTERTAINING AN EMPLOYEE WHO WAS NOT GOING TO LEAVE BUT OUR SYSTEM TAGGED HIM – THIS IS A FUTURE IMPROVEMENT FOR OUR MODEL
- XGBOOST MODEL CREATED A NICE ENSEMBLE OF TREES FOR US, WHOSE ACCURACY COULD INCREASE MORE THAN THE DECISION TREE IF WE GET MORE DATA

CLUSTERING & INTERPRETATION

WE SERVED OURSELVES OF CLUSTER TECHNIQUES BUT FOR CLASSIFICATION PURPOSES. WE USED THE K-MEANS METHOD BECAUSE IT ALLOWS US TO PREDETERMINE THE NUMBER OF OUTPUT CLUSTERS. SINCE WE ARE STUDYING ATTRITION, THAT IN THE DATASET IS A DISCRETE BINARY CLASS, WE TUNED THE ALGORITHM TO GET JUST TWO CLUSTERS.

WE THEN USE THE TRAIN-TEST SET SPLITTING AND APPLIED THE CLUSTERING ON THE TRAINING SET. AFTER THAT WE EVALUATED BOTH THE VALIDITY OF THE CLUSTERING OUTPUT AND THE CLASSIFYING POWER OF THE SAME, THROUGH A SET OF MEASURES (DISTORTION, SILHOUETTE, PRECISION AND RECALL, ROC AREA UNDER THE CURVE)

GOING THROUGH, WE REITERATE THE SAME EXACT PROCESS BUT WITH TWO OTHER DIFFERENT DATASETS, OBTAINED WITH DIFFERENT FEATURE SELECTION TECHNIQUES. THE FIRST ONE WAS OBTAINED WITH A BACKWARD ELIMINATION BASED ON A LOGISTIC REGRESSION MODEL AND RUN ON SAS, THE SECOND ONE WAS OBTAINED WITH A CORRELATION-BASED FEATURE SELECTION (CFS) RUN ON WEKA.

- AFTER THAT WE COLLECTED THE RESULT FROM EVERY ATTEMPT AND COMPARED THEM, AND AS A RESULT THE MOST EFFICIENT WITH RESPECT TO THE CLASSIFICATION GOAL WAS THE BACKWARD ELIMINATION THAT HAD A 0.78 PRECISION.

ASSOCIATION RULE MINING

- ATTRITION=NO JOBROLE=SALES EXECUTIVE 269 ==> DEPARTMENT=SALES 269 <CONF:(1)> LIFT:(3.3) LEV:(0.13) [187] CONV:(187.39)
- ATTRITION=NO JOBROLE=RESEARCH SCIENTIST 245 ==> DEPARTMENT=RESEARCH & DEVELOPMENT 245 <CONF:(1)> LIFT:(1.53) LEV:(0.06) [84] CONV:(84.83)
- ATTRITION=NO JOBROLE=RESEARCH SCIENTIST MONTHLYINCOME=POOR 244 ==> DEPARTMENT=RESEARCH & DEVELOPMENT 244 <CONF:(1)> LIFT:(1.53) LEV:(0.06) [84] CONV:(84.49)

CONCLUSIONS

WE SUCCESSFULLY CREATED AN EARLY WARNING SYSTEM WHICH IMMEDIATELY TELLS THE HUMAN RESOURCES DEPARTMENT IF AN EMPLOYEE IS PRUNE TO LEAVE OR NOT.

WE ACHIEVED THIS EARLY WARNING SYSTEM BASED ON SEVERAL DATA MINING TECHNIQUES IN ORDER TO BE VERY ACCURATE ON SUPERVISED CLASSIFICATION MODELLING

ISSUE TREE

ADDITIONAL SLIDES EXPLAINING BUSINESS PROBLEM CONVERSION TO DATA SCIENCE PROBLEM

how can we reduce
ABC it company's
attrition rate from 25% to
14% in India within the
next 18 months

voluntary attrition

involuntary attrition i.e.
employee has been
asked to leave the
organization

A1

employee satisfied with
his / her job

A2

employee un-satisfied
with his / her job

in scope

out of scope/constraint

A1

employee satisfied with
his / her job

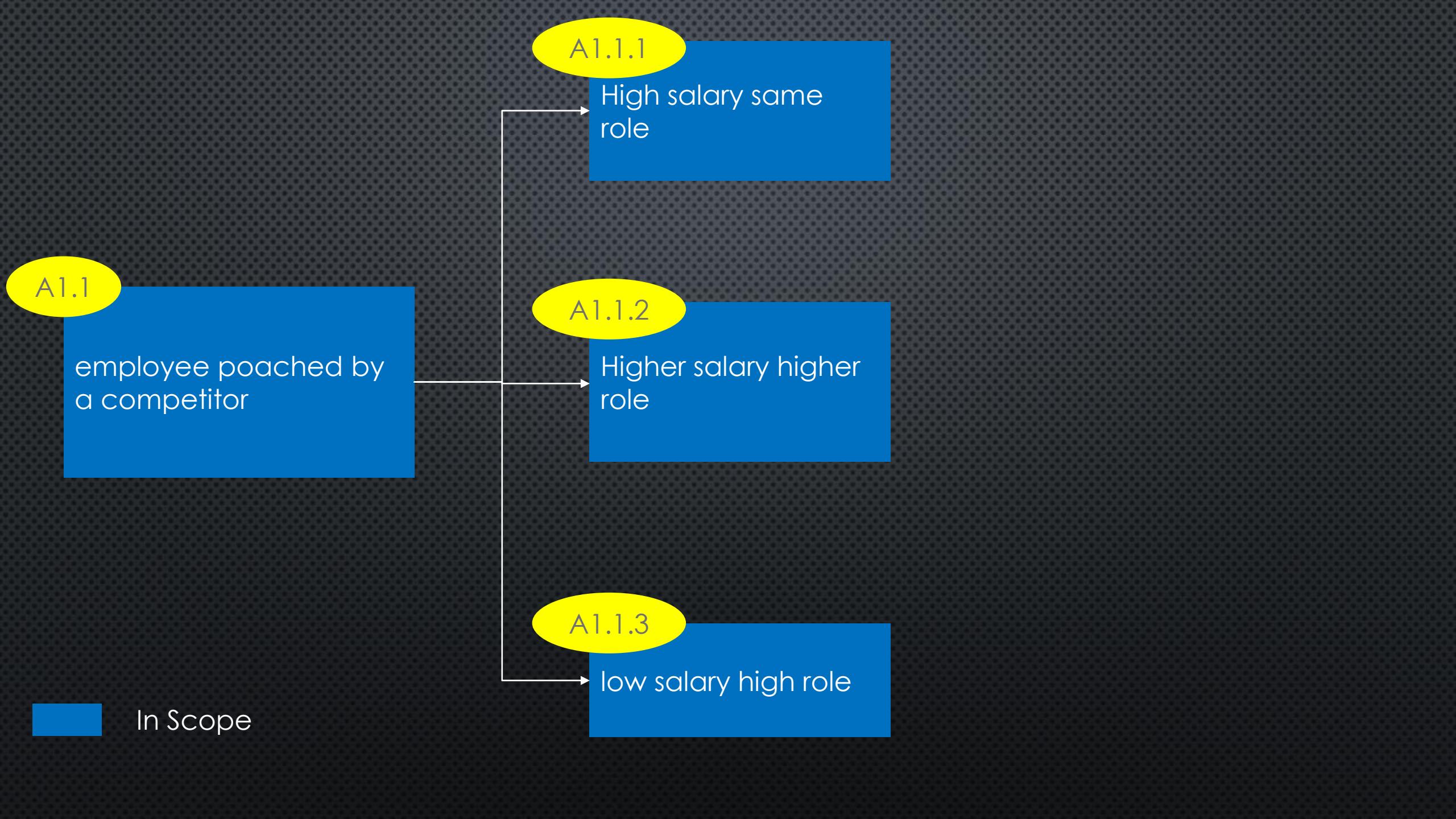
A1.1

employee poached by
a competitor

A1.2

employee not poached

In Scope





In Scope



Out of Scope/Constraint

Terminal
node

A1.1.2

Higher salary higher role

Offer similar or higher salary

Offer higher role

A1.1.3

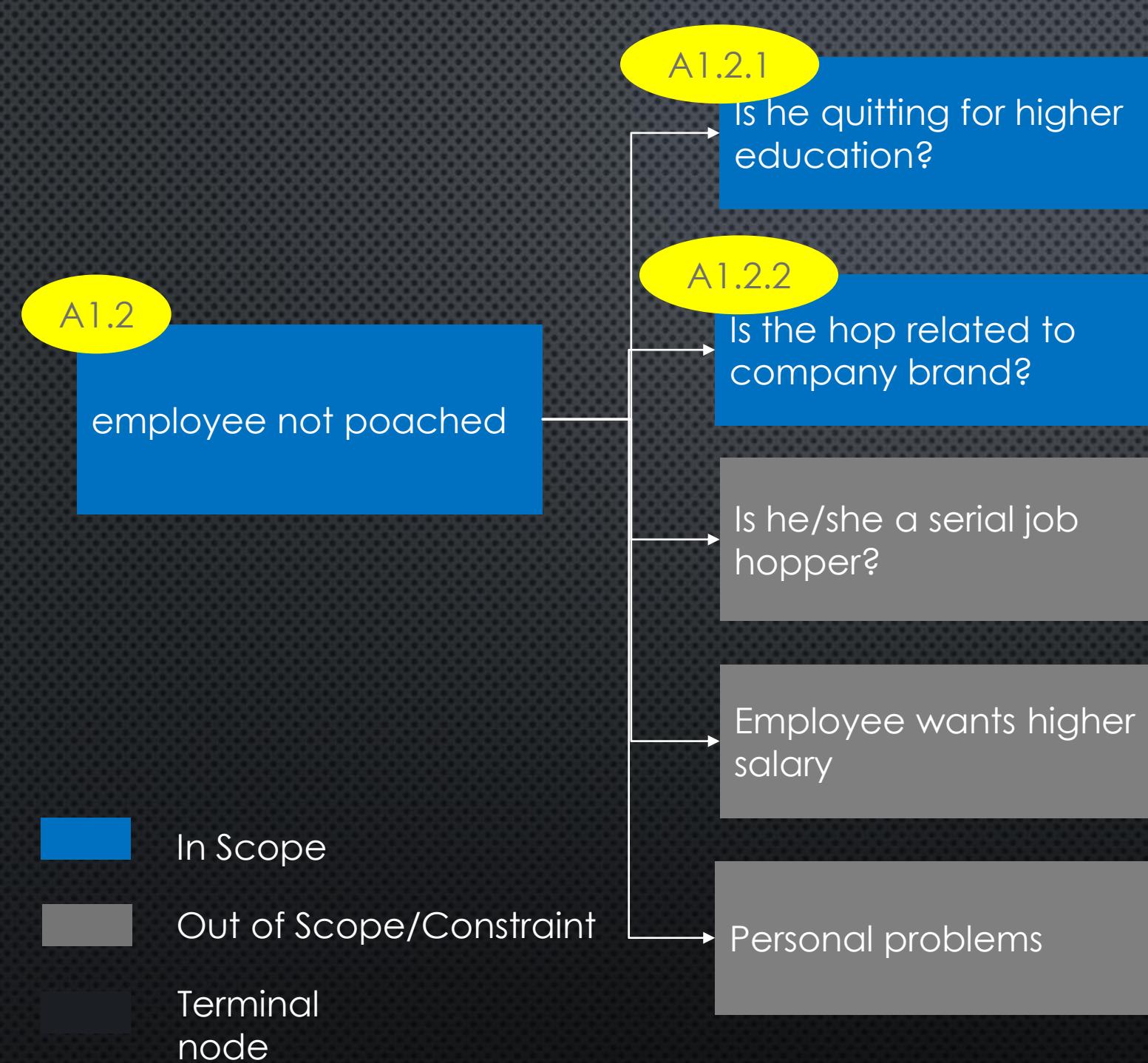
low salary high role

Offer higher role

In Scope

Out of Scope/Constraint

Terminal node



In Scope

Out of Scope/Constraint

Terminal node

A1.2.1

Is he quitting for higher education?

Relevant to organization

Not relevant to organization

Can organization finance and give break?

In Scope
Terminal node

A1.2.2

Is the hop related to company brand?

Negative Brand Image

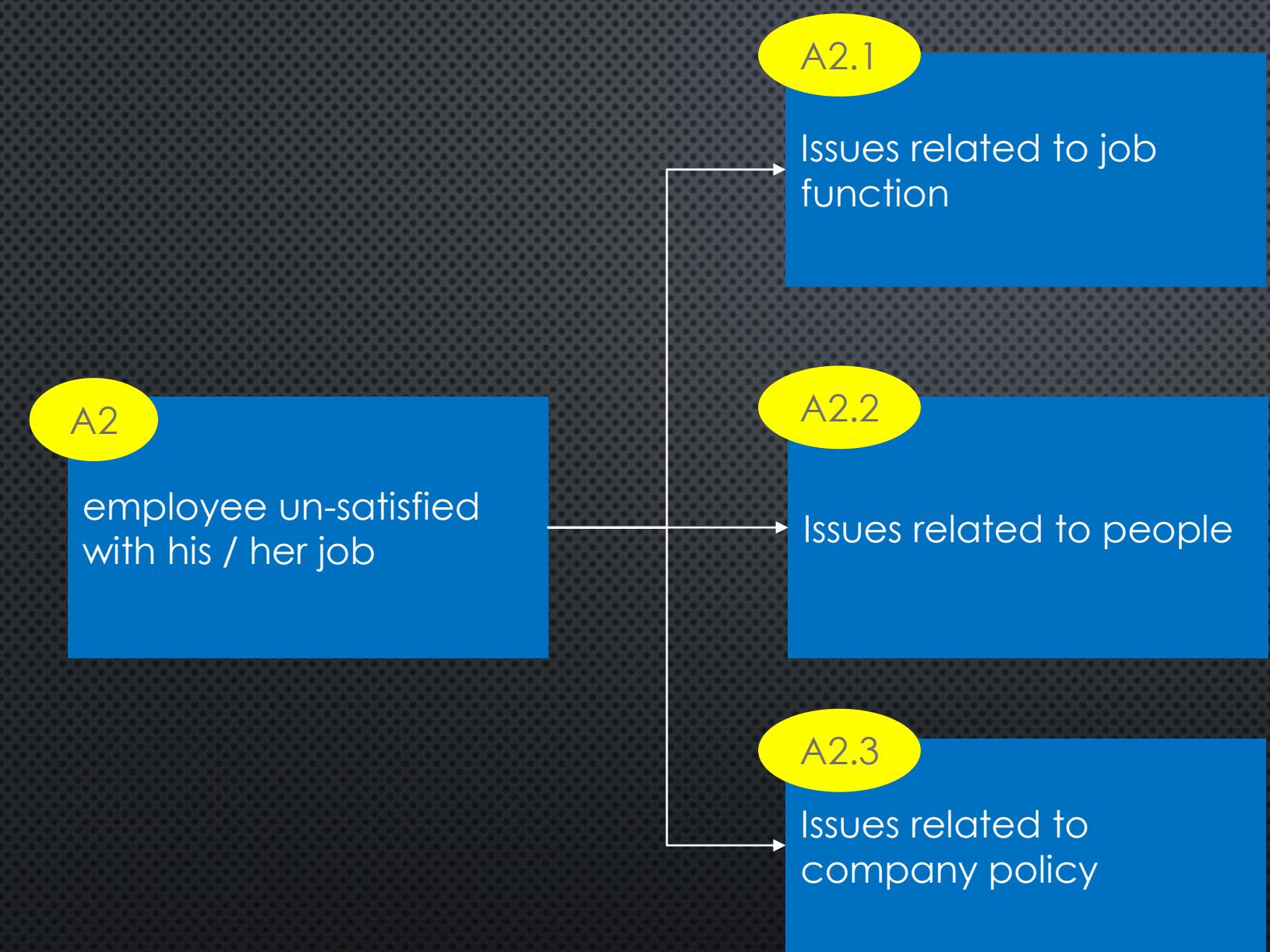
Want for a better brand

Sr. Mgmt to clarify / clear any negative perceptions and communicate business and growth strategy

In Scope

Out of Scope/Constraint

Terminal node



In Scope



Out of Scope/Constraint



A2.1.1

Skill mismatch

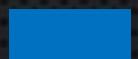
Employee is overqualified

Can Employee be moved to another suitable role?

Employee is Underqualified

Can Employee be trained?

Can Employee be moved to another suitable role?

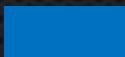


In Scope

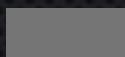


Out of Scope/Constraint

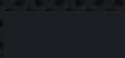
Terminal node



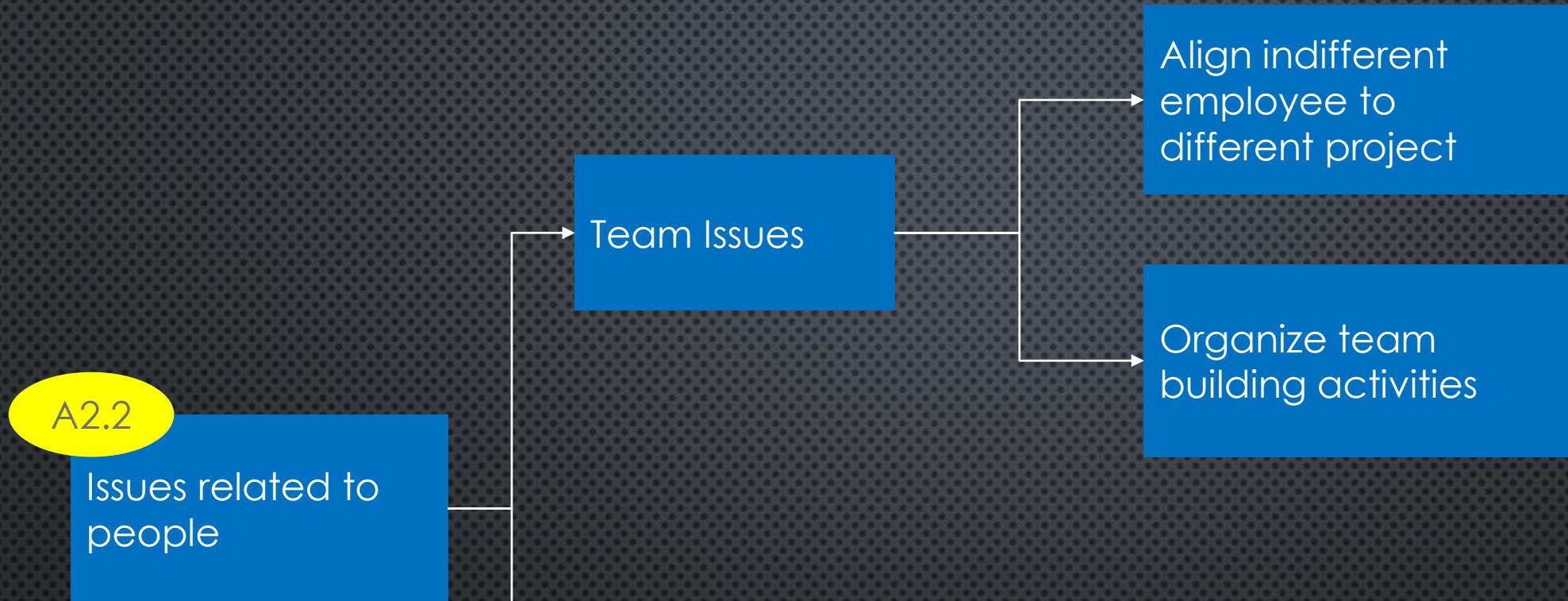
In Scope



Out of Scope/Constraint



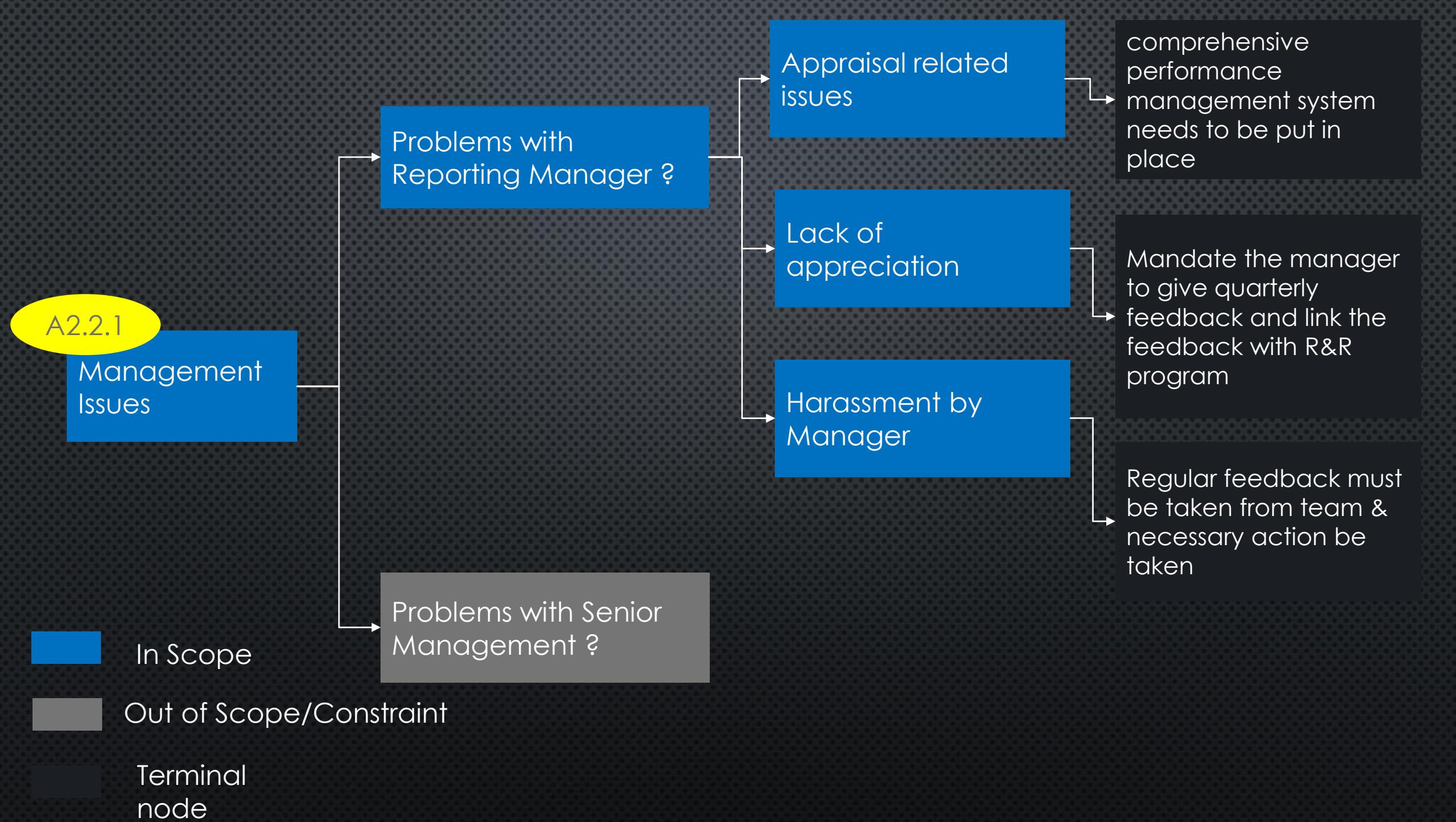
Terminal node



In Scope

Out of Scope/Constraint

Terminal node



A2.3

Issues related to
company policy

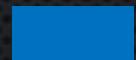
Commuting and
transportation issues

Time and shift issues

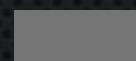
Promotion related

Overseas
Deputation and
travel policies

review policies (internal / external)
based on industry
benchmarks



In Scope



Out of Scope/Constraint

Terminal
node

EARLY WARNING SYSTEM - MODELLING

