ANALYSIS OF IMPACTS OF EO-SAPR ON OE-JOB SATISFACTION BASED ON DEOCS LONGITUDINAL DATA USING DECISION TREE

Wei Wan, Ph.D.
Senior Faculty Research Fellow, DEOMI
Claflin University
Orangeburg, SC 29115

Dr. Daniel P. McDonald, Director
Hope Research Center
Defense Equal Opportunity Management Institute

July 2022

Defense Equal Opportunity Management Institute
366 Tuskegee Drive, Patrick Space Force Base, FL 32925

DEOMI PRESS

SFR Tracking No. 22-01

Contents

Abstract 2

Introduction 6

The Defense Equal Opportunity Climate Survey (DEOCS) 11

Principal Component Analysis (PCA) 7

Decision Tree Analysis 8

Impacts of EO-SAPR on Job Satisfaction 12

Data Reduction by Principal Component Analysis (PCA) 13

Data 14

Validation of Assumptions of PCA 14

Outcomes of PCA 17

Decision Tree Analysis 24

Set up of Decision Tree for EO-SAPR Variable Over Time 24

Outcome of Decision Tree for EO-SAPR Variables Over Time 27

Analysis of Decision Rules from CHAID 27

Analysis of Decision Rules from CRT 30

Analysis of Outcome of Decision Tree by Gender 33

Further Study on OE Variables by Decision Tree 36

Analysis of the Relationship between EO-SAPR Elements and “Group Coh” 36

Analysis of the Relationship between EO-SAPR Elements and “Engagement” 38

Analysis of the Relationship between EO-SAPR Elements and “Inclusion” 39

Analysis of the Relationship between EO-SAPR Elements and “TrustinLeader” 40

Identify the Key Question in “Sexual Harassment Retaliation Climate” 42

Study Decision Tree without PCA....48

Summary....52

References....55

List of Tables....58

List of Figures....60

Appendices....61

- Appendix A – SPSS Syntax for PCA....61
- Appendix B – SPSS Python Syntax for Decision Tree....62
- Appendix C – SPSS Python for Decision Tree for Separate Gender....71
- Appendix D – SPSS Python Syntax for Decision Tree for Other OE Variables....85
- Appendix E – SPSS Syntax for Studying Six SAPR Variables....88
- Appendix F – SPSS Syntax for Decision Tree without PCA....91

Abstract

The purpose of this paper is to explore the factors to affect U.S. military personnel's job satisfaction using decision tree and conduct longitudinal study of these factors over a period of three years. To improve applicability of decision trees a method of data reduction is used to identify the input variables in the decision tree.

The methodology. Dataset-Defense Equal Opportunity Climate Survey (DEOCS) 4.1 is a longitudinal survey data of 2,592,036 cases, which are from 34 cross-sectional survey from October 2017 to July 2020. The DEOCS is a confidential, command-requested organization development survey focusing on issues of organizational effectiveness, equal opportunity, and sexual assault response and prevention. The DEOCS program is managed and administered by DEOMI. Decision Tree method is used to identify the key Equal Opportunity (EO) and Sexual Assault Prevention and Response (SAPR) items, which affect job satisfaction, and other Organizational Effectiveness (OE) items. Each cross-sectional data is used to build a decision tree. EO-SAPR variables are selected as the inputs for the decision trees and the categorized job satisfaction variable are the output of the decision trees. Two splitting methods are used to build the decision trees including the Chi-squared Automatic Interaction Detector (CHAID) and Classification and Regression Tree (CRT). To improve applicability of the decision tree models, principal component analysis technique is used to select the most principal-components-correlated variables as input variables, then to reduce the complexity of decision trees.

Findings. From the results of the principal component analysis, a simplified decision tree is built using seven input variables, which include three demographical variables and four EO-SAPR variables. From the decision rules, which are used to identify "at-risk" (low-level job satisfaction) personnel by EO-SAPR elements for each period of time, the dominating element is

"SexHarRetaliation" - Sexual Harassment Retaliation Climate. Comparison study of the decision rules over time also help identify the trend of change of the EO and SAPR elements, which impact organizational climate.

Originality. EO-SAPR climates affect military personnel's job satisfaction level. This study applies decision tree method to produce decision rules to identify "at-risk" persons, and then assess and analyze how EO-SAPR climates impact job satisfaction. Furthermore, the decision rule itself and the longitudinal study of changes of decision rules can be used in reverse way. They help leaders of the organization to identify the key factors to improve organizational climate and test the effectiveness of strategies implemented before. In the end, the applicability of decision rule from decision trees practically depends on the number of input variables. This study used a statistical technique to decrease the number of input variables and thereby improve usefulness of outputs of the decision trees.

Analysis of Impacts on Equal Opportunity-Sexual Assault Prevention and Response Organizational Effectiveness-Job Satisfaction Based on Defense Equal Opportunity Climate Survey Longitudinal Data using Decision Tree

Introduction

The Defense Organizational Climate Survey (DEOCS)

Both organizational climate and culture are aggregated individual-level long-lasting perceptions of the working environment and culture of the business they work for. They are differentiated by scale, temporality, and specificity. In a military context, due to the hierarchical nature of command structures in the military, leaders are believed to have an outsized role in shaping the climate of any given command (Doty &Gellineau, 2008). Thus, command climate is a substantiation of organizational climate in a military context, which can be defined as the climate for or in a command.

The DEOCS is a congressionally mandated unit-level climate survey that provides commanders with unit-specific information on critical personnel topics so that they can take immediate steps to improve their command climate (Office of People Analytics, 2021). The DEOCS is a commander's management tool that allows them to proactively assess critical organizational climate dimensions that can have an impact on effectiveness within the organization. DEOCS provides a diagnosis of potential organizational issues that can be addressed. Respondents answer questions that affect a unit's readiness and formal and informal policies, practices, and procedures that occur or are likely to occur within the organization (Defense Equal Opportunity Management Institute, n.d.). The questionnaire has four focus areas:

PART I – Demographics

PART II - Organizational Effectiveness (OE)

PART III - Equal Opportunity/Equal Employment Opportunity (EO/EEO)/Fair Treatment

PART IV - Sexual Assault Prevention and Response (SAPR).

Principal Components Analysis (PCA)

PCA is a variable-reduction technique. It is used to reduce a larger set of variables into a smaller set of ‘artificial’ variables, called 'principal components', which account for most of the variance in the original variables (Gorsuch, 1974; Leech, Barrett, & Morgan, 2015)). The typical applications of PCA are as follows (Comrey, & Lee, 1992):

(1) Removing highly correlated variables in a questionnaire.

(2) Testing if the variables which have been chosen can sufficiently represent the construct (EO or SAPR climate in this research) and creating a new questionnaire.

(3) Testing if some questions in the existing questionnaire are superfluous, and then shortening to include fewer items or creating new and more interesting questions, which could improve response rates.

In order to use PCA technique to conduct analysis and produce valid results, following five assumptions need to be satisfied.

(1) The variables should be measured at the continuous level, and the ordinal variables with a wide range of scale, which is like 7-point scale are also frequently used in PCA.

(2) There exists a linear relationship between all variables since PCA technique is based on Pearson correlation coefficients. This assumption can be tested by a matrix scatterplot, but in practice this assumption could be relaxed, or fixed by nonlinear transformation.

(3) Sampling adequacy is required for PCA to produce a reliable result, which implies the sample size should be large enough. The requirement of sample is usually given in terms

of the minimum sample size or the minimum ratio of sample size to the number of variables (MacCallum, Widaman, Zhang, & Hong, 1999; Lingard, & Rowlinson, 2006). The classical guidance is from (Comrey, & Lee, 1992), which determines the adequacy of sample size as follows: 100 = poor, 200 = fair, 300 = good, 500 = very good, 1,000 or more = excellent. The methods to detect sampling adequacy includes the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy for the overall data set and the KMO measure for each individual variable. Statistical Package for Social Sciences (SPSS) statistics provides KMO measure.

(4) Suitability of dataset for data reduction. The variables in the dataset should have adequate correlations in order to be reduced to a smaller number of components. SPSS Statistics provides Bartlett's test of sphericity to test this suitability.

(5) No significant outliers.

Decision Tree

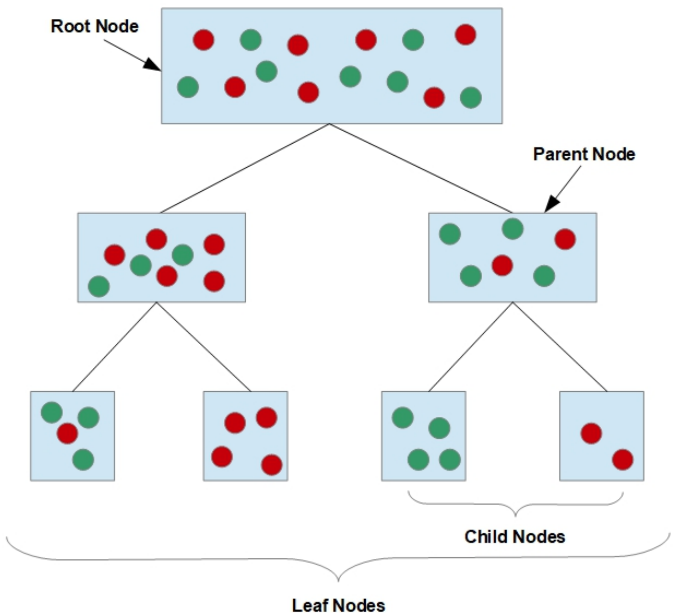
Decision Tree is one of non-parametric supervised machine learning algorithms. Since 1960's when it was first introduced (Smith, & Koning, 2017), It has been widely used for classification and regression problems. It produces a classification rule in terms of a set of input variables (splitter) to splits the population or sample into several homogeneous subsets, where homogeneity is measured in terms of the target variable. Decision tree has been widely used in different research fields, which include (Smith & Koning, 2017; Maimon & Rokach, 2015):

- Variable selection.
- Assessing the relative importance of variables.
- Handling of missing values
- Prediction

- Data manipulation.

The key terminologies used in decision tree method is as follows:

Fig. 1 Sample Decision Tree



Root Node represents the entire population or sample, and is the node that starts the tree graph, where we evaluate and select the variable that best splits the population. Leaf Nodes are nodes which do not have any child node. Thus, they are the final nodes of the tree, where the predictions of a category are made. A node which is divided into sub-nodes is called as Parent Node, and the sub-nodes are called as Child Nodes.

The key idea of a decision tree algorithm is it starts from the Root Node to split nodes into sub-nodes, and the splitting process will continue until the nodes become homogenous/pure in terms of the target variable. In each splitting, the algorithm decides which input variable to split. The algorithm identifies the input variable that is the most related to the outcomes. The closest relationship will produce the purest Child nodes. A broad range of criteria and methods have been developed to evaluate the relationship between the input and target variable. Depending on whether the target variable being continuous or categorical, these criteria can be divided into two classes.

For continuous target variable Reduction in Variance, the classical splitting method is Reduction in Variance. Variance is used to measure the homogeneity of a node. At each split, the variance of each child node is calculated, then the variance of this split is calculated as the weighted average variance of all child nodes. Then, this parent node is split with the input variable which gives the lowest variance. Split will continue until completely homogeneous nodes are achieved.

For categorical target variable, Information Gain is a method to split the node. Entropy is calculated by following formula:

$$Entropy = -\sum_{i=1}^{n} p_i \log_2 p_i$$

which is used to evaluate the purity of a node. If the value of entropy is lower, then the purity of the node is higher, which implies the entropy of a homogeneous node is zero. The Information Gain is defined by

$$information\ Gain = 1 - Entropy$$

which implies purer nodes have higher value of information gain. Thus, Information Gain method is implemented as follows: for each split, the entropy of each child node is calculated, and the entropy of each split is calculated as the weighted average entropy of child nodes. Then, the split with the highest information gain is selected. This process continues until homogeneous nodes are achieved.

For categorical target variable, Gini Impurity method is similar to Information Gain method. Gini is computed by following formula:

$$Gini = \sum_{i=1}^{n} p_i^2$$

and Gini Impurity is computed by

$$Gini\ Impurity = 1 - Gini = 1 - \sum_{i=1}^{n} p_i^2$$

Thus, lower Gini Impurity implies higher homogeneity of the node. The split with lowest Gini Impurity will be selected.

The next important method for categorical target values is based on Chi-square. It can make two or more than two splits. At each split, this method calculates the Chi-Square value of each child node by taking the sum of Chi-Square values for each class in a node and select the input variable with highest Chi-Square value to split; or calculates a chi-square test of independence for each input on the target variable, which determines the extent to which the target variable depends on the input variable. A lower p values indicates the higher dependence. Thus, the method chooses the input variable that generates the lowest p value to split.

The IBM SPSS Decision Tree package includes four growing methods as follows (IBM Corp, 2021).

Table 1

IBM SPSS Growing Methods

	Input variables	**Dependent variable**	**Measure used to select input variable**	**Split at each node**	**Pruning**
CHAID	Categorical/ Continuous	Categorical	Chi-square	Multiple	Pre-pruning using

					Chi-square test for independence
Exhaustive CHAID	Categorical/ Continuous	Categorical	Chi-square	Multiple	Pre-pruning using Chi-square test for independence
CRT	Categorical/ Continuous	Categorical/ Continuous	Gini index; Twoing criteria	Binary; Split on linear combinations	Pre-pruning using a single-pass algorithm
QUEST	Categorical/ Continuous	Categorical	Chi-square for categorical variables; J-way ANOVA for continuous or ordinal variables	Binary; Split on linear combinations	Post-pruning

Impacts of Equal Opportunity (EO)-Sexual Assault Prevention and Response (SAPR) on Job Satisfaction

In DEOCS, Part 2-4 areas express how command climate is affected and evaluated by individual perception of each area. In fact, there are enormous research about how an organization affect its personnel, and vice versa. One part of foundational research in organizational climate is characterized by exploring which elements are causes or effects, which are objective attributes of an organization or subjective perceptions of individuals. Although there are a lot of discussion about whether organizational climate is just a measure of job satisfaction (Johannesson, 1973), job satisfaction is vital for improved organizational performances (Obiekwe & Obibhunun, 2019), and then the aggregated measure of job satisfaction is an important indication of organizational effectiveness. In the domain of human resource management and organizational behavior, job satisfaction is defined as a "pleasurable or positive emotional state resulting from the appraisal of one's job or job experiences" (Schneider & Snyder, 1975). Job satisfaction is a key element of work motivation, which is a fundamental determinant of one's behavior in an organization (Ćulibrk, Delić, Mitrović, & Ćulibrk, 2018).

Many factors affecting job satisfaction usually includes working conditions; opportunity for advancement; workload and stress Level; respect from Co-Workers; relationship with supervisors and financial rewards. A causal model relating military respondents' attitudes toward equal opportunity (EO)-related fairness to job satisfaction, organization commitment, and perceptions of work group efficacy (McIntyre, Bartle, Landis, & Dansby, 2002) shows that work group EO fairness influences their organizational commitment, job satisfaction, and perceived work group efficacy. And the structural equation model (McIntyre, Bartle, Landis, & Dansby, 2002) establishes causal linkages between perceived work group efficacy, job satisfaction, and organizational commitment.

Substantial research has identified workplace sexual harassment as one of the most damaging and ubiquitous barriers to career success and job satisfaction and withdrawal (Willness, Steel, &Lee, 2007; Fitzgerald, Shullman, Bailey, Richards, Swecker, Gold, & Weitman, 1988). The study (Hutagalung & Ishak, 2012) indicates that sexual harassment can be a predictor of job satisfaction and work stress. Its result shows a significant negative relationship between sexual harassment experience with job satisfaction and significant positive relationship between sexual harassment experience with work stress. However, study by Dr. Brenda L. Moore (Moore, 2010) shows that sexual harassment has a strong significant negative effect on perceived unit effectiveness and percent retention, and a significant effect on unit cohesion, but no significant effect on job satisfaction or unit commitment in U.S. Air Force units.

Another research done by Rebecca S Merkin and Muhammad Kamal Shah (Merkin & Shah, 2014) compares and contrasts how differences in perceptions of sexual harassment impact productive work environments for employees in Pakistan and the U.S. and shows how sexual harassment impact on job satisfaction. Significant results indicated that employees who were sexually harassed reported (a) a decrease in job satisfaction (b) greater turnover intentions and (c) a higher rate of absenteeism.

In DEOCS, Job satisfaction is one of elements in Part 2 - OE. This study aims to explore and assess the relationship between job satisfaction and the elements in EO-SAPR, and then to identify and locate the EO-SAPR elements which affect job satisfaction in U.S. military context, and to exhibit how these impacts change over time.

Data Reduction by Principal Component Analysis (PCA)

Data

The dataset of this study is DEOCS 4.1 data (Clean_Oct17-July20.sav). This data set contains 2,592,036 cases, 294 variables. It is a longitudinal data set, containing DEOCS survey at 34 points in time from Oct 17 to Jul 20 with time interval being MONTH.

Validation of Assumptions of PCA

This study aims to identify EO-SAPR variables which impact job satisfaction, to assess the degree of impact, and to study how this impact changes over time. Decision Trees, as a non-parametric supervised learning method for classification and regression is used to assess the relationship between EO-SAPR variables and job satisfaction. The EO-SAPR variables are listed in Table 2, 3, respectively. They are intended to be input variables of decision tree method, but in order to reduce the complexity of tree, and then improve the applicability of decision rules, PCA method is used to deduct the number of items/variables in EO-SAPR. PCA is implemented by IBM SPSS. The syntax for PCA is Appendix A.

Table 2

Equal Opportunities (EO) variables

Index	Name	Label	Measure
236	Hazing	Hazing Scale Score - Mean Percentage of Hazing behaviors selected	scale
237	Bullying	Bullying Scale Score - Mean Percentage of Bullying behaviors selected	scale
238	Discrimination	Discrimination Scale Score	scale

Table 3

Sexual Assault Prevention and Response (SAPR) Variables

Index	Name	Label	Measure
239	SexHar	Sexual Harassment	scale
240	SAPRPrevent	Sexual Assault Prevention Climate Scale Score	scale
241	SAPRRepKnowledge	Sexual Assault Reporting Knowledge Scale Score	scale
242	SAPRResponse	Sexual Assault Response Climate Scale Score	scale
243	SAPRRetaliation	Sexual Assault Retaliation Climate Scale Score	scale
244	SexHarRetaliation	Sexual Harassment Retaliation Climate Scale Score	scale

From the outcome of PCA results, we are able to check to make sure that the assumptions of using PCA method are satisfied and identify the major components in this set of variables.

Assumptions is PCA is validated as follows:

(1) Continuity of measurement is satisfied. All variables in Table 2 and 3 are mean of the ordinal variables with 7-point scale.

(2) There exists a linear relationship between all variables in Table 2 and 3. This assumption is tested by Correlation Matrix (Table 4).

(3) Assumption of sampling adequacy is satisfied. The adequacy of sample is at the level of excellent. The Kaiser-Meyer-Olkin (KMO) Measure in Table 5 indicates sufficient items for each factor.

(4) Assumption of suitability of dataset for data reduction in terms of having adequate correlations is satisfied. Bartlett's Test of Sphericity in Table 5 indicates that the correlation matrix is significantly different from an identity matrix.

(5) Assumption of no significant outliers is satisfied.

Table 4

Correlation Matrix

Table 4. Correlation Matrix

		Hazing Scale Score - Mean Percentage of Hazing behaviors selected	Bullying Scale Score - Mean Percentage of Bullying behaviors selected	Discrimination Scale Score	Sexual Harassment	Sexual Assault Prevention Climate Scale Score	Sexual Assault Reporting Knowledge Scale Score	Sexual Assault Response Climate Scale Score	Sexual Assault Retaliation Climate Scale Score	Sexual Harassment Retaliation Climate Scale Score
Correlation	Hazing Scale Score - Mean Percentage of Hazing behaviors selected	1.000	.375	.018	-.122	-.109	-.037	-.128	-.118	-.125
	Bullying Scale Score - Mean Percentage of Bullying behaviors selected	.375	1.000	.040	-.247	-.226	-.048	-.264	-.247	-.270
	Discrimination Scale Score	.018	.040	1.000	-.065	-.083	-.006	-.083	-.081	-.081
	Sexual Harassment	-.122	-.247	-.065	1.000	.526	.178	.582	.552	.582
	Sexual Assault Prevention Climate Scale Score	-.109	-.226	-.083	.526	1.000	.170	.560	.481	.510
	Sexual Assault Reporting Knowledge Scale Score	-.037	-.048	-.006	.178	.170	1.000	.178	.239	.215
	Sexual Assault Response Climate Scale Score	-.128	-.264	-.083	.582	.560	.178	1.000	.590	.632
	Sexual Assault Retaliation Climate Scale Score	-.118	-.247	-.081	.552	.481	.239	.590	1.000	.861
	Sexual Harassment Retaliation Climate Scale Score	-.125	-.270	-.081	.582	.510	.215	.632	.861	1.000

Table 5

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.819
Bartlett's Test of Sphericity	Approx. Chi-Square	7860198.013
	df	36
	Sig.	.000

Outcome of PCA

Since PCA method is based on analysis of correlation, basic correlation analysis is the first step of analysis. If the correlation of two variables is more than 0.9, then these two variables are just needed to linearly combine to make a new variable, and these two variables form one component. If the correlation of two variable is less than 0.3, which depends on the real problem, then each of these two variables form one component. From the correlation matrix in Table 4, the correlation of following two variables with other variables are less than 0.3, so they naturally form two major components to explain the variance in EO and SAPR, and they will be chosen as independent inputs of decision tree.

- Discrimination Scale Score
- Sexual Assault Reporting Knowledge Scale Score

The table of Communalities (Table 6) provides the proportion of each variable's variance that can be explained by the principal components. As what is indicated in correlation matrix, following two variables composed of two independent components.

- Discrimination Scale Score: 0.999
- Sexual Assault Reporting Knowledge Scale Score: 0.990

From Table 7. Total Variance Explained, and Figure 2 Scree Plot, four principal components whose eigenvalues are 0.9 or greater are extracted, and account for about 75% of the total variance.

Table 8. Component Matrix contains component loadings, which are the correlations between the variable and the component. We identify the “representing” variable for each component as the variable which has the largest correlation with this component. Thus, it is obvious that the representing variable for Component 3 and Component 4 follows. Furthermore, from Table 9. Component Score Coefficient Matrix, the score of variable “Discrimination Scale Score” for component 3 is far greater than other variables; the core of variable “Sexual Assault Reporting Knowledge Scale Score” for component 4 is far greater than other variables. It reinforces the conclusion from choosing the largest correlation in Table 8, and correlation analysis.

- Discrimination Scale Score: Component 3
- Sexual Assault Reporting Knowledge Scale Score: Component 4

Another two representing variables for Components 1 and 2 are also identified as the variables which have the largest correlation in Table 8, and the largest component score in Table 9.

- Sexual Harassment Retaliation Climate Scale Score: Component 1.
- Hazing Scale Score - Mean Percentage of Hazing behaviors selected: Component 2.

Table 6

Communalities

	Initial	Extraction
Hazing Scale Score - Mean Percentage of Hazing behaviors selected	1.000	.746
Bullying Scale Score - Mean Percentage of Bullying behaviors selected	1.000	.660
Discrimination Scale Score	1.000	.999
Sexual Harassment	1.000	.620
Sexual Assault Prevention Climate Scale Score	1.000	.545
Sexual Assault Reporting Knowledge Scale Score	1.000	.990
Sexual Assault Response Climate Scale Score	1.000	.676
Sexual Assault Retaliation Climate Scale Score	1.000	.744
Sexual Harassment Retaliation Climate Scale Score	1.000	.784

Extraction Method: Principal Component Analysis.

Table 7

Total Variance Explained

Table 7. Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.623	40.257	40.257	3.623	40.257	40.257	3.370	37.440	37.440
2	1.230	13.670	53.927	1.230	13.670	53.927	1.387	15.409	52.849
3	.997	11.080	65.007	.997	11.080	65.007	1.006	11.181	64.030
4	.913	10.139	75.147	.913	10.139	75.147	1.001	11.117	75.147
5	.640	7.106	82.252						
6	.604	6.713	88.966						
7	.459	5.100	94.066						
8	.398	4.426	98.492						
9	.136	1.508	100.000						

Extraction Method: Principal Component Analysis.

Figure 2

Scree Plot

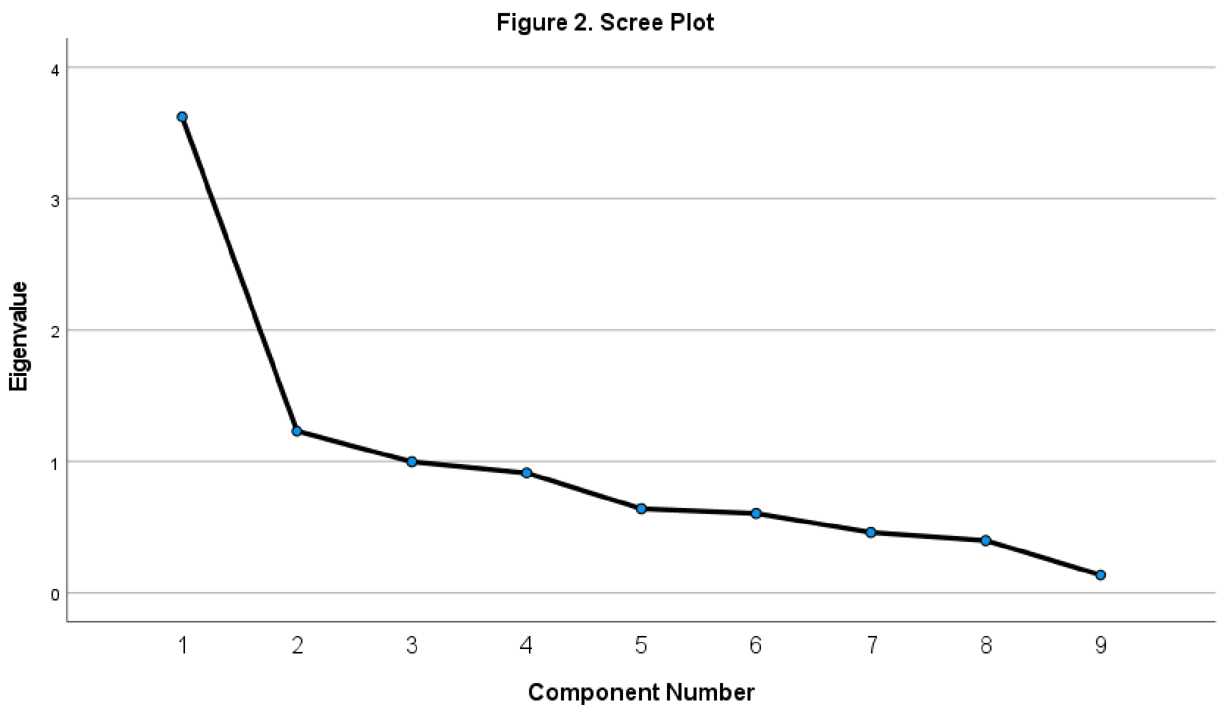


Table 8

Component Matrix

	Component 1	2	3	4
Sexual Harassment Retaliation Climate Scale Score	.872	.131		
Sexual Assault Retaliation Climate Scale Score	.848	.148		
Sexual Assault Response Climate Scale Score	.809			.122
Sexual Harassment	.774			.119
Sexual Assault Prevention Climate Scale Score	.725			.104
Hazing Scale Score - Mean Percentage of Hazing behaviors selected	-.252	.810		.152
Bullying Scale Score - Mean Percentage of Bullying behaviors selected	-.430	.689		
Discrimination Scale Score	-.130		.942	.309
Sexual Assault Reporting Knowledge Scale Score	.315	.197	.322	-.865

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

Table 9

Component Score Coefficient Matrix

	Component			
	1	2	3	4
Hazing Scale Score - Mean Percentage of Hazing behaviors selected	.122	.671	-.066	-.003
Bullying Scale Score - Mean Percentage of Bullying behaviors selected	.018	.568	.067	-.005
Discrimination Scale Score	.034	-.006	-.009	1.003
Sexual Harassment	.246	.026	-.069	.034
Sexual Assault Prevention Climate Scale Score	.231	.032	-.070	-.013
Sexual Assault Reporting Knowledge Scale Score	-.086	-.001	1.014	-.008
Sexual Assault Response Climate Scale Score	.256	.024	-.079	.009
Sexual Assault Retaliation Climate Scale Score	.257	.057	.038	.019
Sexual Harassment Retaliation Climate Scale Score	.270	.048	-.008	.024

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Decision Tree Analysis

Set up of Decision Tree for EO-SAPR Variable Over Time

This study aims to study the relationship between EO-SAPR items and Job Satisfaction. Decision tree method is used to identify and assess the impact of EO-SAPR items on Job Satisfaction over time. Thus, the target dependent variables in the dataset are as follows (Table 10):

Table 10

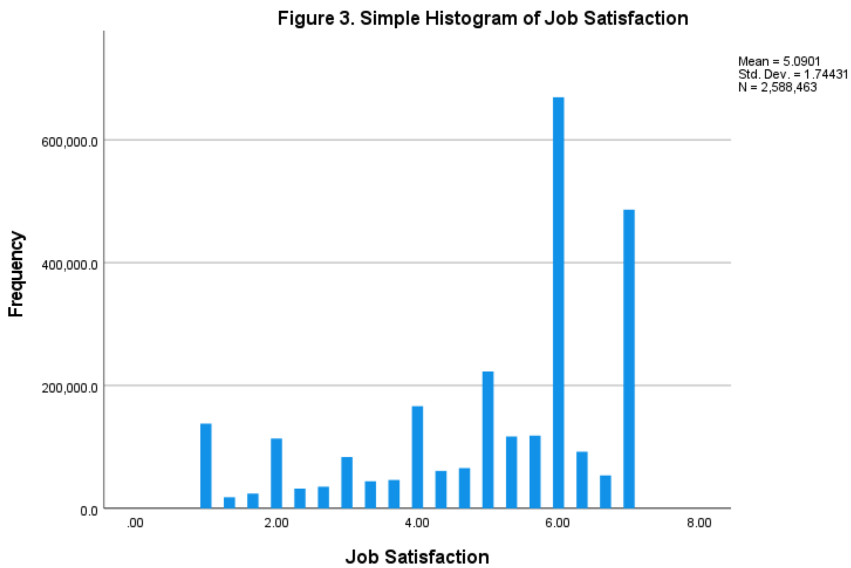
Target OE Variable

Index	Name	Label	Values
79	Oe24	JobSat1: I like my current job.	1 = Strongly Disagree 2 = Disagree 3 = Slightly Disagree 4 = Neither Agree nor Disagree 5 = Slightly Agree 6 = Agree 7 = Strongly Agree
80	Oe25	JobSat2: I feel satisfied with my current job.	1 = Strongly Disagree 2 = Disagree 3 = Slightly Disagree 4 = Neither Agree nor Disagree 5 = Slightly Agree 6 = Agree 7 = Strongly Agree
81	Oe26	JobSat3: I am happy with my current job.	1 = Strongly Disagree 2 = Disagree 3 = Slightly Disagree 4 = Neither Agree nor Disagree 5 = Slightly Agree 6 = Agree 7 = Strongly Agree
232	JobSat		JobSat = average of above three variables

The target dependent variable is “JobSat”, which is numerical variable. The distribution of “JobSat” is in Figure 3.

Figure 3

Simple Histogram of Job Satisfaction



To simplify analysis by decision tree, a new categorical variable for job satisfaction is created as follows (Table 11), which is used as dependent variable of decision tree.

Table 11

Depend Variables of Decision Tree

Index	**Name**	**Label**	**Values**
295	Cate_JobSat	Categorization of variable “JobSat”. 1=at-risk 2=neutral 3=happy 4=missing value	"1" if JobSat < 3 "2" if JobSat >= 3 & JobSat <= 5 "3" if JobSat >5 "4" if JobSat missing value.

The following demographic variables (i.e., Table 12) are independent variable of decision tree.

Table 12

Independent variables - Demographics variables

Index	Name	Label	Values
221	Race	Combined race variable.	0 = Race Missing 1 = American Indian or Alaskan Native 2 = Asian 3 = Black 4 = Native Hawaiian or Other Pacific Islander 5 = White 6 = Declined to Respond 7 = Selected Multiple Races
225	MilCiv	Military versus Civilian	1 = Military 2 = Civilian 3 = Other
224	Rank	Rank variable used for DEOCS CMD Report Body & DEOCS Roll Up Reports (combines Jr. Enl & NCOs and WOs & JrOffs).	1 = Junior Enlisted & NCO (E1-E6) 2 = Senior Enlisted (E7-E9) 3 = Junior Officer (WO & O1-O3) 4 = Senior Officer (O4-O6) 5 = Junior Federal Civilian (GS 1-12) 6 = Senior Federal Civilian (GS13 - SES) 7 = Wage Grade 8 = NAF 9 = Other

Following variables (Table 13, 14) are independent variable of decision tree, which are outcome of PCA.

Table 13

Independent variables - Equal Opportunities (EO) variables

Index	Name	Label	Values
236	Hazing	Hazing Scale Score - Mean Percentage of Hazing behaviors selected	
238	Discrimination	Discrimination Scale Score	

Table 14

Independent variables - Sexual Assault Prevention and Response (SAPR) variables

Index	Name	Label	Values
241	SAPRRepKnowledge	Sexual Assault Reporting Knowledge Scale Score	
244	SexHarRetaliation	Sexual Harassment Retaliation Climate Scale Score	

Table 15

Time variables

Index	Name	Label	Values
182	MMYR	Month and Year of DEOCS Collection	1 = Oct 17 ... 34 = Jul20 (34 values)

Outcome of Decision Tree for EO-SAPR Variables Over Time

Decision trees are built up for each value of time variable - "MMYR" by CHAID and CRT. In SPSS, a Python program is imbedded to iterate on the time variable to generate a decision tree. Syntax is in Appendix B.

The outcomes of decision tree identify the how the input variables are related to the dependent variable over time. The target value of dependent variable is "1= at-risk", that is "JobSat" < 3. The decision tree provides the decision rule to classify a person as "at-risk".

Analysis of Decision Rules from CHAID

From Table 16, the first important variable to classify at-risk personnel is "SexHarRetaliation" among all 34 decision rules. Here the order of importance is calculated by the extent the variables are related to target variable. For CHAID, the larger Chi-Square statistic value implies closer relationship. Furthermore, among the decision rules to classify a person as "at-risk", in terms of the first important variable, the 91.3%of 34 decision rules is: "SexHarRetaliation $\leq$ 4"; the 4.3% of them is "SexHarRetaliation $\in (4,5.667)$"; the 4.3% of them is "SexHarRetaliation = missing".

Table 16

First Important Variable Impacting Job Satisfaction from Decision Tree Rule (CHAID)

Variable	Percentage of occurrences over 34 Cross-sectional times	Corresponding Split Values	Percentage of occurrences over 34 Cross-sectional times
SexHarRetaliation	**100%**	≤ 4	**91.3%**
		missing	**4.3%**
		$(4, 5.667)$	**4.3%**

The second important variable is exhibited in Table 17. This table is interpreted as follows. The second important variable being "MilCiv" takes 50% of decision rules which classify at-risk person; and when "MilCiv" is the second important variable, 95.5% of these decision rules is "MilCiv is equal to Military"; the 4.5% of them is "MilCiv is equal to Civilian; Other." The other rows of this table are interpreted similarly.

Table 17

Second Important Variable Impacting Job Satisfaction from Decision Tree Rule (CHAID)

Variable	Percentage of occurrences over 34 Cross-sectional times	Corresponding Split Values	Percentage of occurrences over 34 Cross-sectional times
MilCiv	**50%**	Military	**95.5%**
		Civilian; Other	4.5%
Rank	**29.5%**	E1-E6	**71.4%**
		WO&O1-O3	28.6%
		O4-O6	12.5%
		GS13-SES	4.2%
		missing	12.5%

Hazing	**13.6%**	> 25 < 25	**83.3%** 16.7%
Race	**6.8%**	Decline to Respond	**66.7%**
		White	33.3%

The third important variable is exhibited in Table 18. This table is interpreted as follows. The third important variable being “Hazing” takes 40.5% of decision rules which classify at-risk person; and when “Hazing” is the third important variable, 86.7% of these decision rules is “Hazing > 25”; the 13.3% of them is “Hazing < 25”. The other rows of this table are interpreted similarly.

Table 18

Third Important Variable Impacting Job Satisfaction from Decision Tree Rule (CHAID)

Variable	**Percentage of occurrences over 34 Cross-sectional times**	**Corresponding Split Values**	**Percentage of occurrences over 34 Cross-sectional times**
Hazing	**40.5%**	> 25	**86.7%**
		< 25	13.3%
Race	35.1%	Black	7.7%
		American Indian or Alaskan Native	7.7%
		Declined to Respond	**92.3%**
		Selected Multiple Races	46.2%
		White	23.1%
Discrimination	10.8%	(3.917,5.143) (4,5.25)	**50%** 25%
		> 5.286; missing	25%
SexPRRepKnowledge	8.1%	>0.667	**66.7%**
		(0.5, 0.667)	33.3%
Rank	2.7%	O-4–O-6	100%
MilCiv	2.7%	military	100%

Analysis of Decision Rules from CRT

From Table 19, the first important variable to classify at-risk personnel is still "SexHarRetaliation" among all 34 decision rules. Here the order of importance is calculated by Gini value for CRT. Among the decision rules to classify a person as "at-risk", in terms of the first important variable, the 56.1% of 34 decision rules is: "SexHarRetaliation < 5.917"; the 12.2% of them is "SexHarRetaliation <4.917", etc.

Table 19

First Important Variable Impacting Job Satisfaction from Decision Tree Rule (CRT)

Variable	**Percentage of occurrences over 34 Cross-sectional times**	**Corresponding Split Values**	**Percentage of occurrences over 34 Cross-sectional times**
SexHarRetaliation	**83.3%**	<4.417	5.0%
		<4.25	10.0%
		<4.208	15.0%
		<3.917	25.0%
		<3.75	5.0%
		<3.583	10.0%
		<3.417	15.0%
		<3.25	5.0%
		<3.083	10.0%
Rank	8.3%	E-1–E-6	100%
		E-7–E-9	50%
		WO&O1-O3	50%
		GS-13–SES	50%
Hazing		>37.5	100%
MilCiv	4.2%	Military	100%

The second important variable is exhibited in Table 20. This table is interpreted as follows. The second important variable being "Rank" takes 62.5% of decision rules which classify at-risk person; and when "Rank" is the second important variable, 100% of these decision rules contains

“Rank is equal to E1-E6”; 60.0% of them contains “Rank is equal to WO&O-1–O-3.” The other rows of this table are interpreted similarly.

Table 20

Second Important Variable Impacting Job Satisfaction from Decision Tree Rule (CRT)

Variable	**Percentage of occurrences over 34 Cross-sectional times**	**Corresponding Split Values**	**Percentage of occurrences over 34 Cross-sectional times**
Rank	**62.5%**	E1-E6	**100.0%**
		WO&O1-O3	60.0%
		E7-E9	25.0%
		O4-O6	5.0%
SexHarRetaliation	21.9%	<4.583	14.3%
		<3.917	57.1%
		<3.583	14.3%
		>2.583	14.3%
Hazing		>37.5	50.0%
		<37.5	50.0%
MilCiv	3.1%	Military	100%

The third important variable is exhibited in Table 21. This table is interpreted as follows. The third important variable being “SexHarRetaliation” takes 83.3% of decision rules which classify at-risk person; and when “SexHarRetaliation” is the third important variable, 25% of these decision rules is “SexHarRetaliation<3.917”, etc. When “Rank” is the third important variable, 100.0% of them contains “Rank is equal to E1-E6”, and 50% of them contains “Rank is equal to E7-E9”. The other rows of this table are interpreted similarly.

Table 21

Third Important Variable Impacting Job Satisfaction from Decision Tree Rule (CRT)

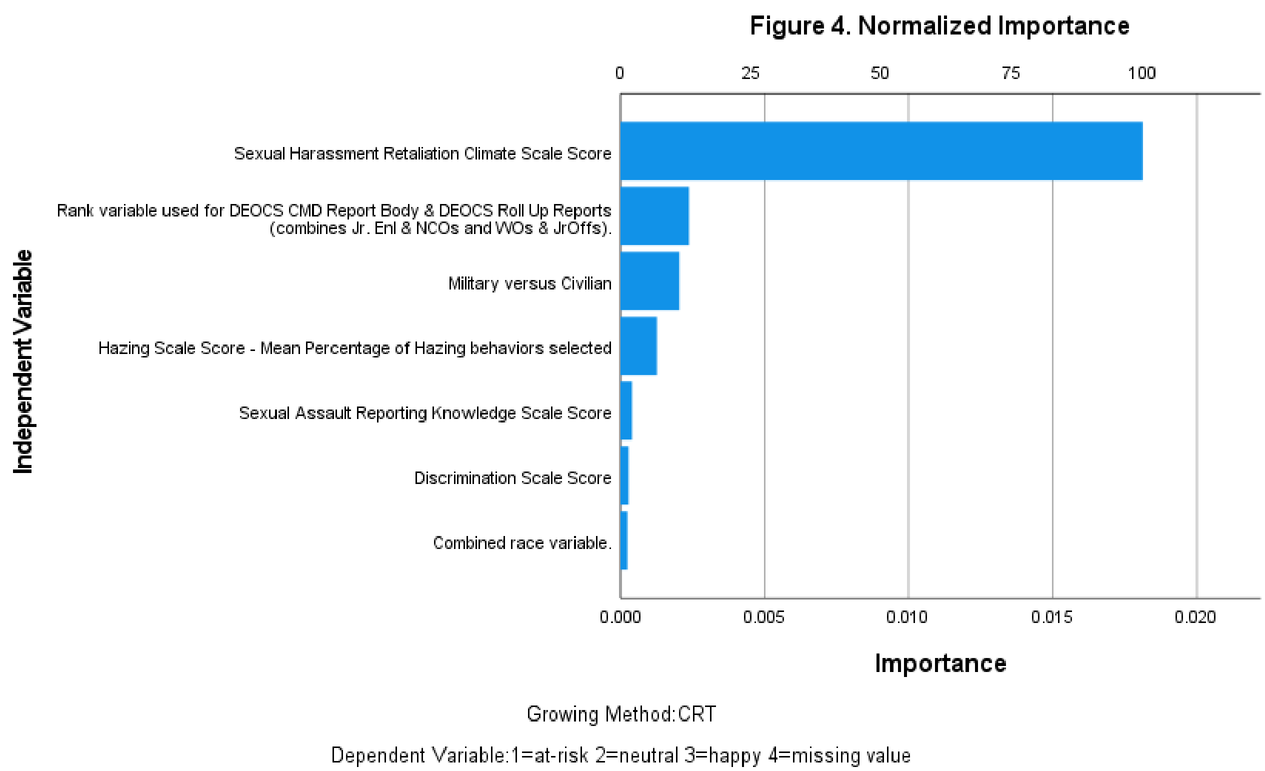
Variable	**Percentage of occurrences over 34 Cross-sectional times**	**Corresponding Split Values**	**Percentage of occurrences over 34 Cross-sectional times**
SexHarRetaliation	**83.3%**	<4.417	5.0%

		<4.25	10.0%
		<4.208	15.0%
		<3.917	25.0%
		<3.75	5.0%
		<3.583	10.0%
		<3.417	15.0%
		<3.25	5.0%
		<3.083	10.0%
Rank	8.3%	E1-E6	100%
		E7-E9	50%
		WO&O1-O3	50%
		GS13-SES	50%
Hazing		>37.5	100%
MilCiv	4.2%	Military	100%

Furthermore, CRT method also provides the quantitative importance of independent variables to the target variable over time. Following Fig. 4 is for October 2017. It is noticed that the importance of the variable - “SexHarRetaliation” dominates other variables. It is true for all the time.

Figure 4

Normalized Importance



Analysis of Outcome of Decision Tree by Gender

In 4.2 we have analyzed how EO-SAPR variables impact Job Satisfaction. In this section, we analyze how this impact differs between male and female. Thus, the decision tree method is applied to the split dataset files by gender. The syntax for decision tree by gender is Appendix C.

Figure 5 exhibits the first important variable to classify at-risk personnel over time. It is obvious that the variable - “SexHarRetaliation” are the first important variable for both male and female for all the time. Figure 6 exhibits the second important variable “converges” to the variable of “Hazing” for male, and to the variable of “Rank” for female. And lower rank of female tends to be classified as at-risk. Figure 7 exhibits the third important variable “converges” to the variable of "Rank" for male, and to the variable of "MilCiv" for female. And military female tends to be classified as at-risk.

Figure 5

Comparison of First Important Variables Impacting JobSat – Male vs Female

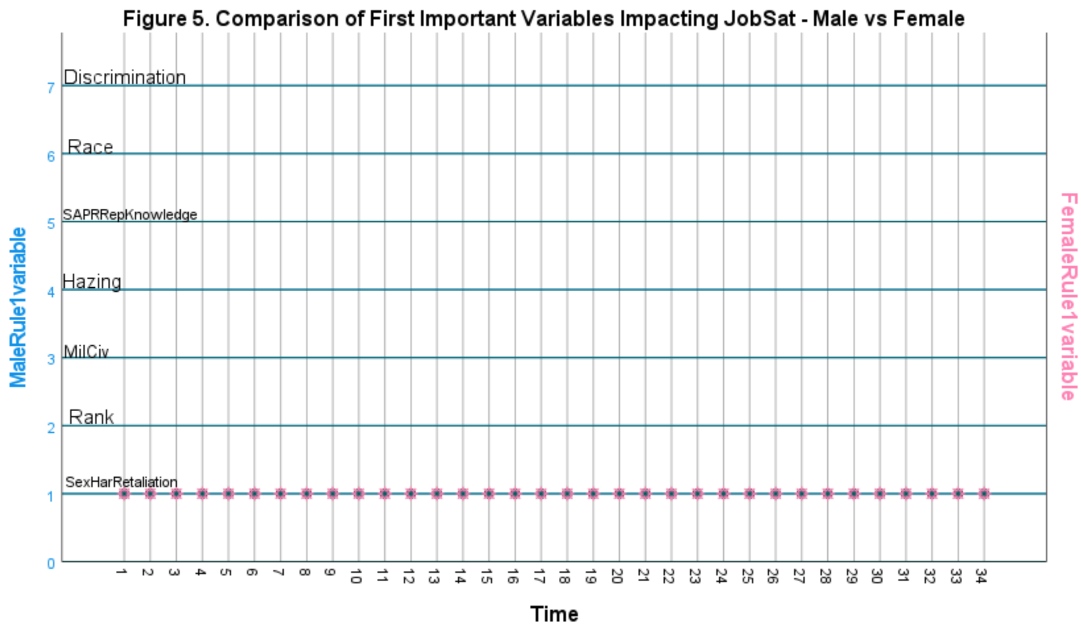


Figure 6

Comparison of Second Important/Variable Impacting JobSat – Male vs Female

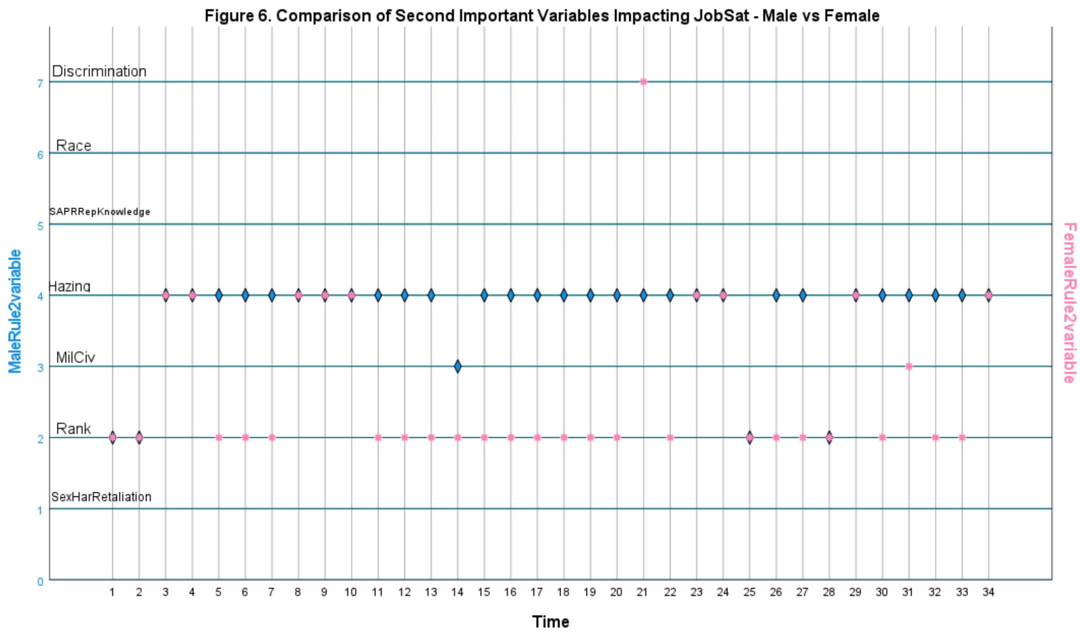
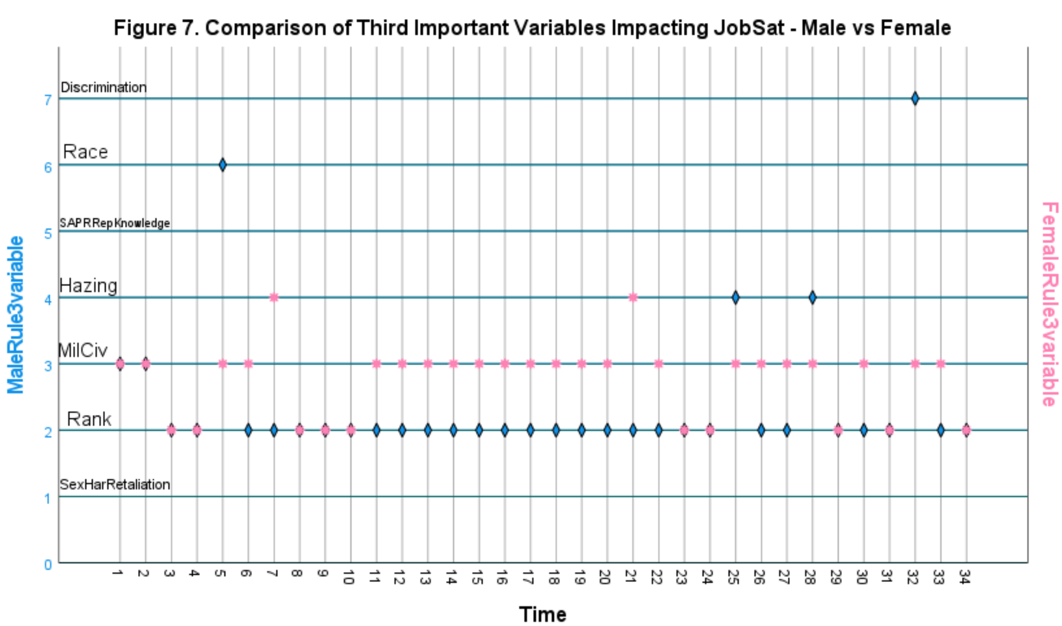


Figure 7

Comparison of Third Important Variables Impacting JobSat -Male vs Female



Further Study on OE Variables by Decision Tree

In previous study, we have identified the key variables in EO-SAPR, and analyzed how they are related to the OE variable - “JobSat” over time. In this section, we extend the previous research into other OE variables - “TrustinLeader, Inclusion, GroupCoh, Engagement” by decision tree (Appendix D). The objective is to identify how the EO-SAPR variables are related to or affect OE variables.

Analysis of the Relationship between EO-SAPR Elements and “GroupCoh”

The questions for OE variables – “GroupCoh” is as follows. Its value is computed as the mean value of scores from these three questions. Its distribution is in Figure 8.

Group Cohesion: *A dynamic process that is reflected in the tendency for a group to stick together within your immediate workplace and remain united in the pursuit of its objectives and/or for the satisfaction of participants’ interpersonal needs.*

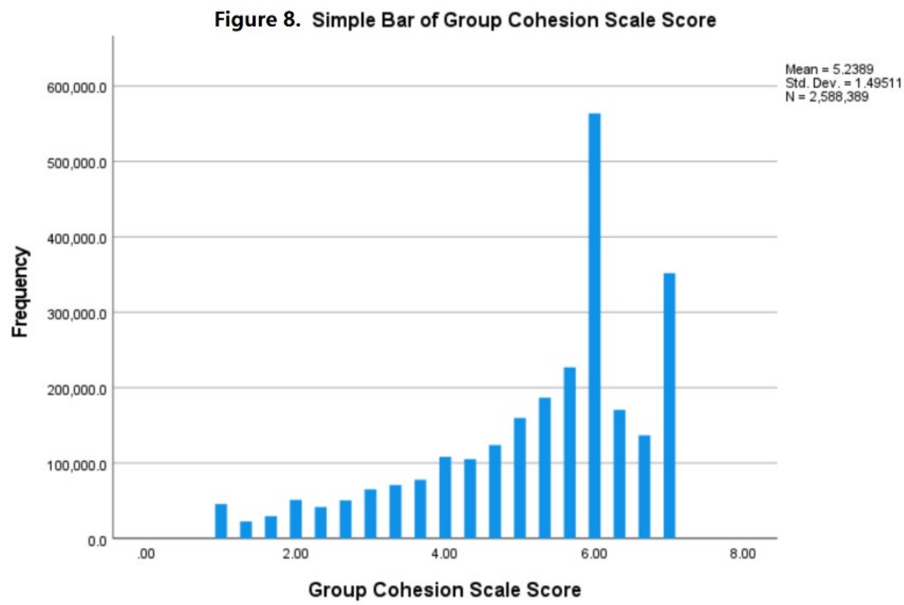
20. My workgroup is united in trying to reach its goals for performance.

21. We all take responsibility for the performance of the workgroup.

22. If members of our workgroup have problems in the workplace, everyone wants to help them so we can get back on task.

Figure 8

Simple Bar of Group Cohesion Scale Score



Decision tree – CRT method is used to identify the relation between "Hazing, Discrimination, SAPRRepKnowledge, SexHarRetaliation, MilCiv, Rank, Race" and "GroupCoh". Importance of these EO-SAPR variable on "GroupCoh" is exhibited in Table 22. This table is interpreted as follows. The first row shows that the order of importance exhibited in a decision rule is: "SexHarRetaliation" is much more important than "Discrimination", which is more important than "Hazing", and this order has appeared 32 out of 34 cross-sectional datasets. The second row has similar interpretation.

Table 22

Order of Importance of EO-SAPR Variables to "GroupCoh"

SexHarRetaliation >>	Discrimination >	Hazing	32/34
SexHarRetaliation>>	Discrimination>	Rank	2/34

Analysis of the Relationship between EO-SAPR Elements and "Engagement"

The questions for OE variables – "Engagement" is as follows. Its value is computed as the mean value of scores from these three questions. Its distribution is in Figure 9.

Engagement: *Engagement refers to a persistent positive and fulfilling state of mind characterized by mental resilience, dedication, and immersion in the work role.*

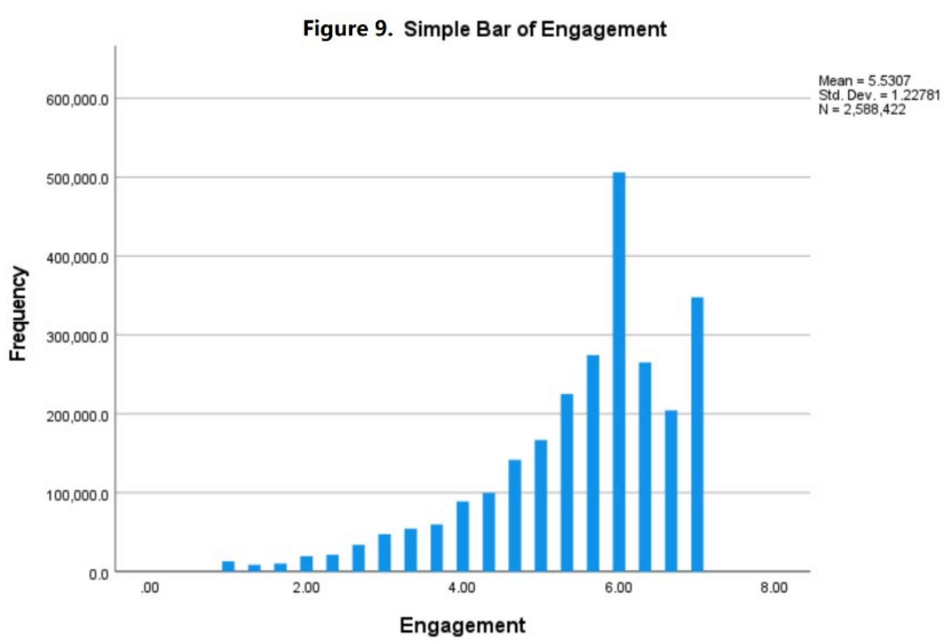
39. At my workplace, I am mentally resilient *(mentally resilient is defined as: able to press on and adapt to psychologically challenging work situations while still maintaining your sense of confidence*).

40. I am enthusiastic about my work.

41. Time flies when I am working.

Figure 9

Simple Bar of Engagement



Similarly, Decision tree – CRT method is used to identify the relation between “Hazing, Discrimination, SAPRRepKnowledge, SexHarRetaliation, MilCiv, Rank, Race” and “Engagement”. Importance of these EO-SAPR variable on “Engagement” is exhibited in Table 23.

Table 23

Order of Importance of EO-SAPR Variables to “Engagement”

SexHarRetaliation >>	Rank >	Discrimination	29/34
SexHarRetaliation>>	Discrimination>	Rank	4/34

Analysis of the Relationship between EO-SAPR Elements and “Inclusion”

The questions for OE variables - “Inclusion” is as follows. Its value is computed as the mean value of scores from these six questions. Its distribution is in Figure 10.

Inclusion at work: *Involves the ways in which organizations, groups, leaders, and military members/employees allow everyone (diverse in identities, cultures, and ways of thinking and acting) to participate, contribute, have a voice, and feel that they are connected and belong, all without losing individual uniqueness or having to give up valuable identities or aspects of themselves.*

42. Coworkers are treated as valued members of the team without losing their unique identities.

43. I feel excluded by my workgroup because I am different.

44. Within my workgroup, I am encouraged to offer ideas on how to improve operations.

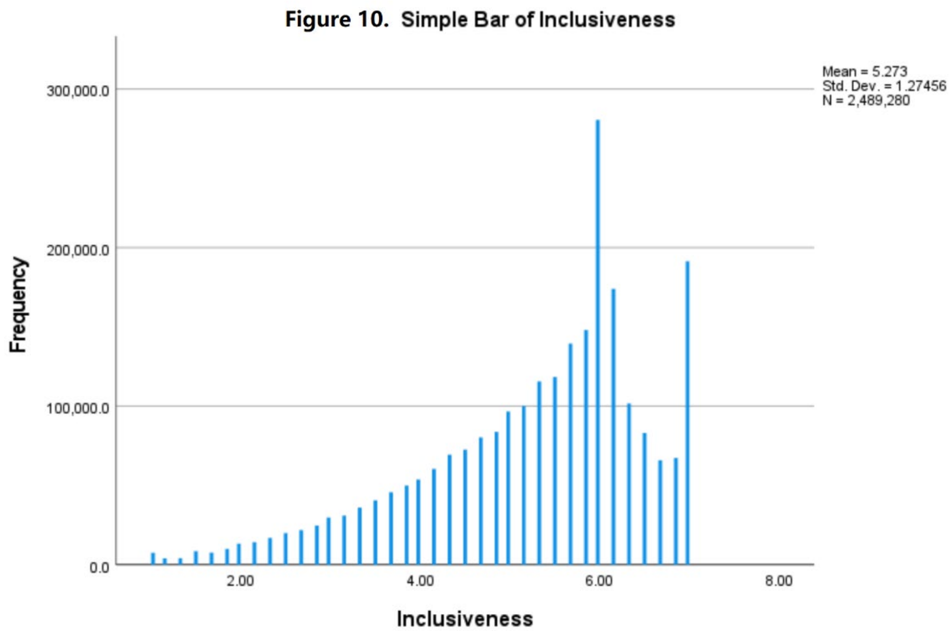
45. Military members/employees in my workgroup are empowered to make work-related decisions on their own.

46. Outcomes (e.g., training opportunities, awards, and recognition) are fairly distributed among military members/employees of my workgroup.

47. The decision-making processes that impact my workgroup are fair.

Figure 10

Simple Bar of Inclusiveness



Similarly, Decision tree – CRT method is used to identify the relation between “Hazing, Discrimination, SAPRRepKnowledge, SexHarRetaliation, MilCiv, Rank, Race” and “Inclusion”. Importance of these EO-SAPR variable on “Inclusion” is exhibited in Table 24.

Table 24

Order of Importance of EO-SAPR Variables to “Inclusion”

SexHarRetaliation >>	Discrimination >	Hazing	30/34
SexHarRetaliation>>	Discrimination>	Rank	4/34

Analysis the relationship between EO-SAPR elements and “TrustinLeader”

The questions for OE variables - “TrustinLeader” is as follows. Its value is computed as the mean value of scores from these five questions. Its distribution is in Figure 11.

Trust in Leadership: *The expectation that a leader will act in your organization's best interest that he or she will follow through with actions which affect the outcomes of others, and that he or she will act in a fair and equitable manner.*

23. I can rely on my immediate supervisor to act in my organization's best interest.

24. My immediate supervisor follows through with commitments he or she makes.

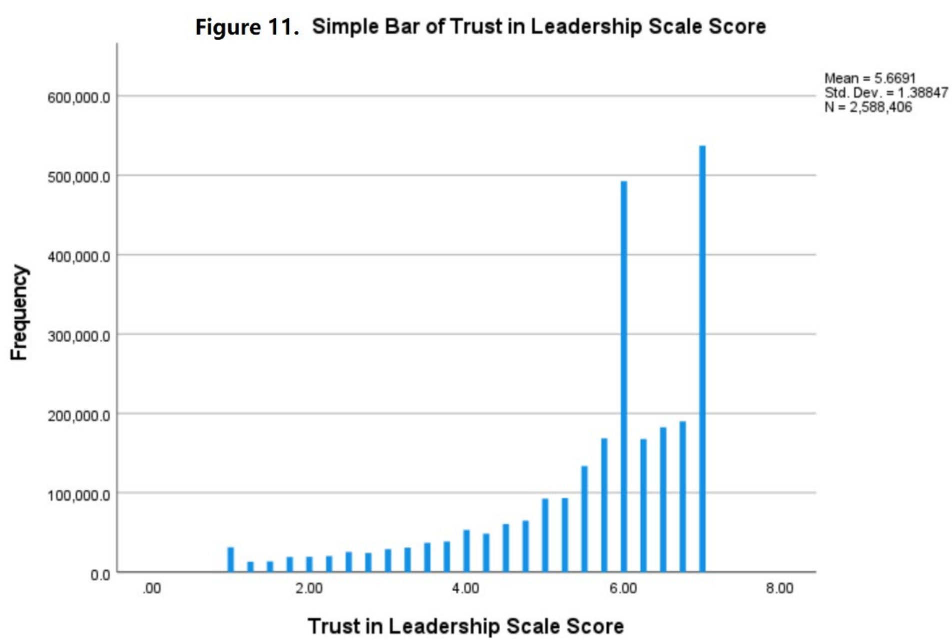
25. Please select "Disagree" for this item.

26. I feel comfortable sharing my work difficulties with my immediate supervisor.

27. My immediate supervisor treats me fairly.

Figure 11

Simple Bar of Trust in Leadership Scale Score



Similarly, Decision tree – CRT method is used to identify the relation between "Hazing, Discrimination, SAPRRepKnowledge, SexHarRetaliation, MilCiv, Rank, Race" and

“TrustinLeader”. Importance of these EO-SAPR variable on “Inclusion” is exhibited in Table 25.

Table 25

Order of Importance of EO-SAPR Variables to “TrustinLeader”

SexHarRetaliation >>	Discrimination >	Hazing	29/34
SexHarRetaliation>>	Hazing >	Discrimination	5/34

In this part of investigation, we use Decision Tree-CRT method to assess the relation between a set of EO-SAPR variables - “Hazing, Discrimination, SAPRRepKnowledge, SexHarRetaliation, MilCiv, Rank, Race” with each OE variable - “TrustinLeader, Inclusion, GroupCoh, Engagement” over time. We found that the SAPR variable – “SexHarRetaliation” has the most important and dominating impact on OE variables. And other EO variables - “Discrimination, Hazing” have similar impact, and demographics variable – “Rank” ranks the third. Furthermore, we have used PCA methods to reduce the number of input variables of decision tree. From above analysis of decision tree, “SexHarRetaliation” as a representing variable of SAPR variables, so we guess SAPR is more closely related to the low-score-OE personnel, that is SAPR has more influence on OE. In order to improve OE climate, we need to improve SAPR climate. We will verify this guess with decision tree with full set of EO-SAPR variables.

Identifying the Key Questions in “Sexual Harassment Retaliation Climate”

In previous study, we have studied and identified the key variables in EO-SAPR variables which impact the OE variables - “JobSat, TrustinLeader, Inclusion, GroupCoh, Engagement”, and the finding is that the first important and dominating variable among EO-SAPR is “SexHarRetaliation”. The questions in “SexHarRetaliation” are as follows. The following study is to identify which questions from Sexual Harassment Retaliation Climate make the variable - "SexHarRetaliation" to be the dominating variable, which impacts EO-SAPR, and be the key question(s), which could represent the most of variance among the variables in Sexual Harassment Retaliation Climate.

Sexual Harassment Retaliation Climate: *Military member's/employee's perception of whether retaliation would occur if a sexual harassment complaint was made in their unit/organization.*

66. In my workgroup, military members/employees who file a sexual harassment complaint would be:

a. Excluded from the social interactions or conversations.

b. Subjected to insulting or disrespectful remarks or jokes.

c. Blamed for causing problems.

d. Denied career opportunities (e.g., denied training, awards, or promotions).

e. Disciplined or given other corrective action.

f. Discouraged from moving forward with the complaint.

By Principal Components Analysis (PCA), from Correlation Matrix (Table 26) SARetaliat4 and SARetaliat5 are strongly correlated ($r > 0.99$), but weakly correlated to SARetaliat1, SARetaliat2, SARetaliat3 ($r \approx 0.03 < 0.1$). Thus, variables of SARetaliat4 and SARetaliat5 seem to be measuring the same thing, and they form one component, or would be combined to a new in some way.

Table 26

Correlation Matrix

Table 26. Correlation Matrix

		SARetaliat1: Excluded from social interactions or conversations. (In my workgroup, reporters of sexual assault would be:)	SARetaliat2: Subjected to insulting or disrespectful remarks or jokes. (In my workgroup, reporters of sexual assault would be:)	SARetaliat3: Blamed for causing problems. (In my workgroup, reporters of sexual assault would be:)	SARetaliat4: Denied career opportunities. (In my workgroup, reporters of sexual assault would be:)	SARetaliat5: Disciplined or given other corrective action. (In my workgroup, reporters of sexual assault would be:)	SARetaliat6: Discouraged from moving forward with the report. (In my workgroup, reporters of sexual assault would be:)
Correlation	SARetaliat1: Excluded from social interactions or conversations. (In my workgroup, reporters of sexual assault would be:)	1.000	.675	.645	.030	.031	.595
	SARetaliat2: Subjected to insulting or disrespectful remarks or jokes. (In my workgroup, reporters of sexual assault would be:)	.675	1.000	.838	.034	.034	.760
	SARetaliat3: Blamed for causing problems. (In my workgroup, reporters of sexual assault would be:)	.645	.838	1.000	.037	.034	.801
	SARetaliat4: Denied career opportunities. (In my workgroup, reporters of sexual assault would be:)	.030	.034	.037	1.000	.999	.088
	SARetaliat5: Disciplined or given other corrective action. (In my workgroup, reporters of sexual assault would be:)	.031	.034	.034	.999	1.000	.087
	SARetaliat6: Discouraged from moving forward with the report. (In my workgroup, reporters of sexual assault would be:)	.595	.760	.801	.088	.087	1.000

PCA extract two components. From "Component Matrix" in Table 27, it contains component loadings, which are the correlations between the variable and the component. Thus, by comparing the loadings for these six variables, the first component mostly explains the variance from following four variables:

SARetaliat1: Excluded from social interactions or conversations. (In my workgroup, reporters of sexual assault would be:)

SARetaliat2: Subjected to insulting or disrespectful remarks or jokes. (In my workgroup, reporters of sexual assault would be:)

SARetaliat3: Blamed for causing problems. (In my workgroup, reporters of sexual assault would be:)

SARetaliat6: Discouraged from moving forward with the report. (In my workgroup, reporters of sexual assault would be:)

And the variable "SARetaliat3" has the largest loading in this component, and be regarded as "representing" variable for this component.

The second component contains following two variables:

SARetaliat4: Denied career opportunities. (In my workgroup, reporters of sexual assault would be:)

SARetaliat5: Disciplined or given other corrective action. (In my workgroup, reporters of sexual assault would be:)

Furthermore, these two variables have pretty high correlation, and contain almost same loading in second component, so either of them can be the "representing" variable for second component.

Table 27

Component Matrix[a]

Table 27. Component Matrix[a]

	Component	
	1	2
SARetaliat1: Excluded from social interactions or conversations. (In my workgroup, reporters of sexual assault would be:)	.805	-.089
SARetaliat2: Subjected to insulting or disrespectful remarks or jokes. (In my workgroup, reporters of sexual assault would be:)	.919	-.099
SARetaliat3: Blamed for causing problems. (In my workgroup, reporters of sexual assault would be:)	.923	-.097
SARetaliat4: Denied career opportunities. (In my workgroup, reporters of sexual assault would be:)	.142	.990
SARetaliat5: Disciplined or given other corrective action. (In my workgroup, reporters of sexual assault would be:)	.141	.990
SARetaliat6: Discouraged from moving forward with the report. (In my workgroup, reporters of sexual assault would be:)	.891	-.032

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Two decision tree methods - CRT and CHAID have been used to identify "SexHarRetaliation" to be the most important variable impacting EO-SAPR. The variable "SexHarRetaliation" is the mean of variables "SARetaliat1, SARetaliat2, SARetaliat3, SARetaliat4, SARetaliat5, SARetaliat6". CHAID method uses Chi-square statistic to measure dependence of each input variable on the target variable and choose the variable with the highest dependence to split. To filter out the most important variables, which make "SexHarRetaliation" to be outstanding variable, we calculate the Chi-Square Test statistic X^2 between each variable of "SARetaliat1, SARetaliat2, SARetaliat3, SARetaliat4, SARetaliat5, SARetaliat6" with job satisfaction, which is exhibited in Table 28. The larger Chi-Square statistic value implies the higher dependence on "Job Satisfaction". The top two variables with the largest value are: Sapr56c, Sapr56d, which belong to Component 1, and Component 2, respectively. Thus, we would like to state that these two variables are the two key variables in Sexual Harassment Retaliation Climate to make the variable "SexHarRetaliation" to be related to "JobSat" much more than other EO-SAPR elements. SPSS Syntax for studying six SAPR variables is Appendix F.

Table 28

Chi-Square Statistic Value between "Sexual Harassment Retaliation Climate" and "Job Satisfaction"

No.	Variable	Label	chi square statistic with Job Satisfaction	Components
145	Sapr56a	SARetaliat1: Excluded from social interactions or conversations. (In my workgroup, reporters of sexual assault would be:)	135575.677432	Component 2
146	Sapr56b	SARetaliat2: Subjected to insulting or disrespectful remarks or jokes. (In my workgroup, reporters of sexual assault would be:)	157078.642389	Component 2
147	Sapr56c	SARetaliat3: Blamed for	169246.796225	Component2

		causing problems. (In my workgroup, reporters of sexual assault would be:)		
148	Sapr56d	SARetaliat4: Denied career opportunities. (In my workgroup, reporters of sexual assault would be:)	168156.411519	Component 1
149	Sapr56e	SARetaliat5: Disciplined or given other corrective action. (In my workgroup, reporters of sexual assault would be:)	137322.529292	Component 1
150	Sapr56f	SARetaliat6: Discouraged from moving forward with the report. (In my workgroup, reporters of sexual assault would be:)	162442.874573	Component 2

Study Decision Tree without PCA

The goal of this investigation is to assess and identify the relation between a set of EO-SAPR variables - "Hazing, Discrimination, SAPRRepKnowledge, SexHarRetaliation, MilCiv, Rank, Race" with each OE variable. In previous study, to reduce the complexity of decision tree, to improve applicability of decision rules, and to save computing time, PCA method has been used to reduce the number of input variables of decision tree. The representing EO variables are "Hazing Discrimination", and the representing SAPR variables are "SAPRRepKnowledge, SexHarRetaliation". The decision rules from decision tree over time show that the variable - "SexHarRetaliation" is the most important and dominating variable, which impact OE variables. Thus, we guess SAPR climate has stronger connection with OE climate than EO climate, which is a lower SAPR score implies a lower OE score. To prove or disprove this guess, decision tree method is applied with full set of EO-SAPR variables in Table 2 and Table 3 as input of decision tree.

Decision tree - CRT method (Appendix G) is used to construct decision trees for each cross-sectional dataset. The input variables are "Hazing, Bullying, Discrimination, SexHar, SAPRPrevent, SAPRRepKnowledge, SAPRResponse, SAPRRetaliation, SexHarRetaliation", and output variable is "JobSat".

The order of importance of EO-SAPR variables is exhibited in Table 29. This table is interpreted as follows. The first column is the order of importance: "1" stands for the most important, and "9" stands for the least important. In the row of "1", the variable - "SAPRPrevent" (Sexual Assault Prevention Climate Scale Score) ranks first 21 out of 34 cross sectional datasets, that is 61.8% of all times; the variable - "SAPRResponse" (Sexual Assault Response Climate Scale Score) ranks first 9 out 34 cross sectional datasets, that is 26.5% of all times; the variable - "SexHar" (Sexual Harassment) ranks first 4 out 34 cross sectional datasets, which is 11.8% of all times. Other rows are interpreted similarly.

Table 29

Order of Important Variable Impacting Job Satisfaction from Decision Tree Rule

Order of importance	Variables	Frequency	Percent

1	Sexual Assault Prevention Climate Scale Score	21	61.8
	Sexual Assault Response Climate Scale Score	9	26.5
	Sexual Harassment	4	11.8
2	Sexual Assault Prevention Climate Scale Score	5	14.7
	Sexual Assault Response Climate Scale Score	18	52.9
	Sexual Harassment	11	32.4
3	Sexual Assault Prevention Climate Scale Score	7	20.6
	Sexual Assault Response Climate Scale Score	7	20.6
	Sexual Harassment	19	55.9
	Sexual Harassment Retaliation Climate Scale Score	1	2.9
4	Sexual Assault Prevention Climate Scale Score	1	2.9
	Sexual Harassment Retaliation Climate Scale Score	33	97.1
5	Sexual Assault Retaliation	34	100.0

	Climate Scale Score		
6	Bullying Scale Score - Mean Percentage of Bullying behaviors selected	34	100.0
7	Discrimination Scale Score	4	11.8
	Hazing Scale Score - Mean Percentage of Hazing behaviors selected	30	88.2
8	Discrimination Scale Score	5	14.7
	Hazing Scale Score - Mean Percentage of Hazing behaviors selected	4	11.8
	Sexual Assault Reporting Knowledge Scale Score	25	73.5
9	Discrimination Scale Score	25	73.5
	Sexual Assault Reporting Knowledge Scale Score	9	26.5

Above table can be summarized as follows. In terms of impacts on job satisfaction, the EO-SAPR variables are sorted as:

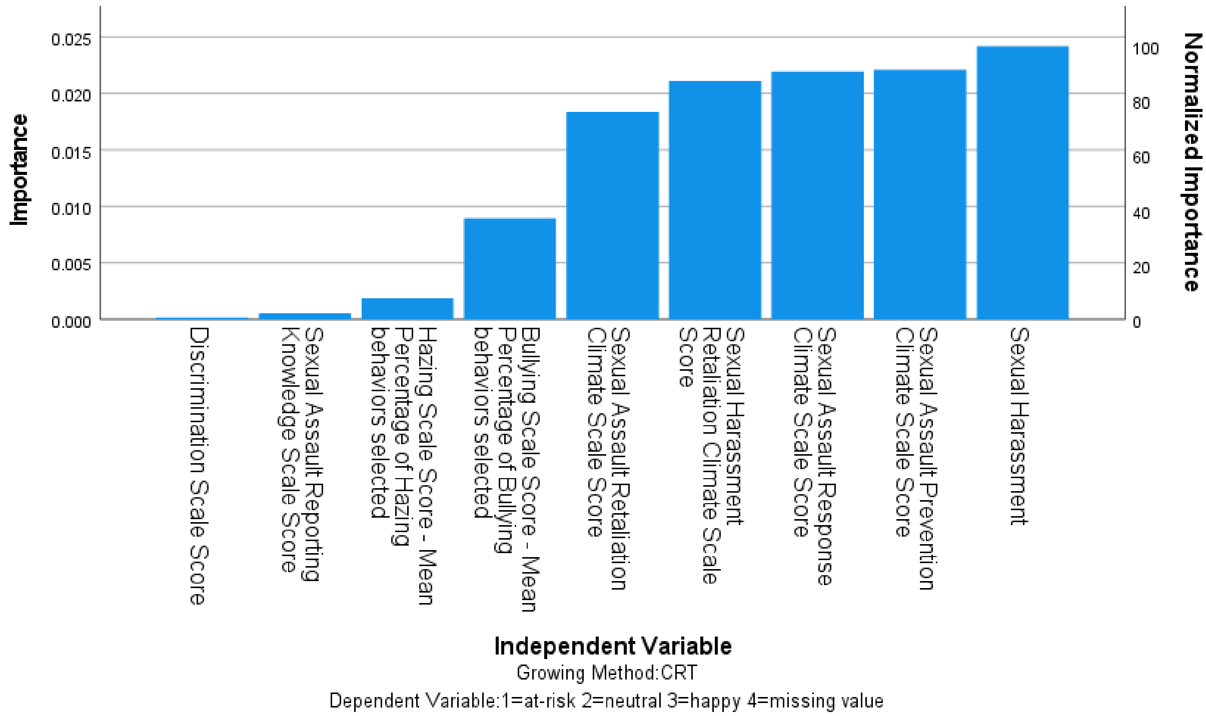
Sexual Assault Prevention Climate Scale Score≅ Sexual Assault Response Climate Scale Score≅ Sexual Harassment≅ Sexual Harassment Retaliation Climate Scale Score≅ Sexual Assault Retaliation Climate Scale Score≥

Bullying Scale Score - Mean Percentage of Bullying behaviors selected≅Hazing Scale Score - Mean Percentage of Hazing behaviors selected≅ Sexual Assault Reporting Knowledge Scale Score, Discrimination Scale Score

Thus, in terms of impacts on job satisfaction, EO-SAPR variables are naturally grouped into two parts: SAPR variables have similar impacts on job satisfaction, and EO variables have similar impacts on job satisfaction. Furthermore, the SAPR variables have obvious larger impacts than EO variables. Following Figure 12, which is from one cross-sectional dataset shows this type of dichotomous separation.

Figure 12

Independent Variable



Summary

Literatures have shown that EO and SAPR climates have impacts on Organizational Effectiveness (OE). The investigation of this paper is to explore how EO and SAPR climates affect OE, and to assess the relationship between EO-SAPR variables and OE variable, especially job satisfaction of U.S. military personnel.

Decision tree method has been adopted to assess and analyze the relationship. To reduce the complexity of decision tree and improve computing efficiency, PCA method has been adopted to extract the components from EO-SAPR variables. The variables which have the largest loading in each component, “SAPRRepKnowledge, SexHarRetaliation, Hazing, Discrimination”, and demographic variables “Race, MilCiv, Rank” are selected as inputs of decision tree, and categorized “JobSat” variable as dependent variable.

Python programs have been developed in SPSS to facilitate the process of generating decision tree for each cross-sectional dataset. From the outcome of decision tree, it is observed that the impacts on job satisfaction of the variable - “SexHarRetaliation” dominates other variables for all the time. Thus, to improve job satisfaction, U.S. military needs to improve Sexual Harassment Retaliation Climate.

Decision tree analysis is then extended to study this relationship for different genders. It is found that “SexHarRetaliation” is the most impacting variable to both male and female, but the second important variable is “Rank” for female, but “Hazing” for male.

Decision tree analysis is then extended to other OE variables. It is found that “SexHarRetaliation” is generally the most impacting variable to impact other OE variables.

In order to compare the accuracy of decision tree analysis with selected variables, decision tree analysis is conducted with full set of EO-SAPR variables as input. The important finding is that SAPR climate impact OE much more than OE climate.

In the end, two growing methods have been adopted to generate decision tree - CRT and CHAID. In terms of accuracy of decision tree, Classification table (Table 30, 31) shows the number of cases classified correctly and incorrectly for each category of the dependent variable – “JobSat”. It is observed that decision rules are accurately predict “satisfied” category (>90%) but are not good at predicting “neutral” and “at-risk” categories (<10%). Furthermore, Risk table provides a measure of the tree's predictive accuracy. The “risk estimate” is the proportion of cases incorrectly classified after adjustment for prior probabilities and misclassification costs. The “risk estimate” ranges from 0.35 to 0.4 for both CRT and CHAID. Thus, further analysis about why decision tree methods are much more accurate to classify “satisfied” is motivated.

Table 30

Classification

	Predicted				
Observed	at-risk	neutral	satisfied	missing value	Percent Correct
at-risk	356	1386	5712	0	4.8%
neutral	329	1815	13963	0	11.3%
satisfied	260	1662	37300	0	95.1%
missing value	0	6	52	0	0.0%

Overall Percentage	1.5%	7.7%	90.7%	0.0%	62.8%

Growing Method: CHAID

Dependent Variable: 1=at-risk 2=neutral 3=happy 4=missing value

Table 31

Classification

	Predicted				
Observed	at-risk	neutral	happy	missing value	Percent Correct
at-risk	349	1816	4484	0	5.2%
neutral	291	3933	15680	0	19.8%
happy	260	2925	33045	0	91.2%
missing value	1	8	49	0	0.0%
Overall Percentage	1.4%	13.8%	84.8%	0.0%	59.4%

Growing Method: CRT

Dependent Variable: 1=at-risk 2=neutral 3=happy 4=missing value

References

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Psychology Press, Taylor & Francis Group.

Ćulibrk, J., Delić, M., Mitrović, S., & Ćulibrk, D. (2018). Job satisfaction, organizational commitment and job involvement: The mediating role of job involvement. *Frontiers in Psychology, 9*, 132.

Doty, J., & Gellineau, J. (2008, July). Command climate. *Army Magazine, 58*(7), 22–23.

Defense Equal Opportunity Management Institute. (n.d.). *DEOMI organizational climate survey* [Online]. https://www.deocs.net/DocDownloads/DEOCS%20Brochure.pdf

Fitzgerald, L. F., Shullman, S. L., Bailey, N., Richards, M., Swecker, J., Gold, Y., & Weitzman, L. (1988). The incidence and dimensions of sexual harassment in academia and the workplace. *Journal of Vocational Behavior,32*(2)), 152–175.

Gorsuch, R. L. (1974). *Factor analysis*. Sanders

Hutagalung, F., & Ishak, Z. (2012). Sexual harassment: A predictor to job satisfaction and work stress among women employees. *Procedia-Social Behavioral Sciences, 65*, 723–730.

IBM Corp. (2021). *IBM SPSS decision trees 28*. IBM Corp. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Decision_Trees.pdfhttps://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Decision_Trees.pdf

Johannesson, R. E. (1973). Some problems in the measurement of organizational climate. *Organizational Behavior and Human Performance, 10*(1), 118–144.

MacCallum, R. C., Widaman, K. F, Zhang, S. & Hong. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.4.84

Maimon, O. Z., & Rokach, L. (2015). *Data mining with decision trees: Theory and applications, 2nd ed*. World Scientific.

Merkin, R. S., & Shah, M. K. (2014). The impact of sexual harassment on job satisfaction, turnover intentions, and absenteeism: Findings from Pakistan compared to the United States. *SpringerPlus*, 3(1), 1–13.

McIntyre, R. M.., Bartle, S. A., Landis, D., & Dansby, M. R. (2002, January). The effects of equal opportunity fairness attitudes on job satisfaction, organizational commitment, and perceived work group efficacy. *Military Psychology, 14*(4), 299–319.

Moore, B. L. (2010*). Effects of sexual harassment on job satisfaction, retention, cohesion, commitment and unit effectiveness: The case of the Air Force* (Technical Report No. 08-10). Defense Equal Opportunity Management Institute. https://apps.dtic.mil/sti/pdfs/AD1044614.pdf

Obiekwe, O., & Obibhunun, L. (2019). Impact of employee job satisfaction on organizational performance. *Academic Journal of Current Research, 6*(12), 6–12.

Office of People Analytics. (2021, October). *Defense organizational climate survey (DEOCS) redesign: Phase 1 overview report* [OPA Report No. 2021-158). Defense Human Resource Activity. https://cg-872b0f53-3ed8-4ee5-8078-10a830db319c.s3-us-gov-west-1.amazonaws.com/opa.mil-production1/uploads/refinery/opa/asset/upload/2274/DEOCSPhase1_Report.pdf?X-Amz-Expires=600&X-Amz-Date=20221208T194116Z&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAR7FXZINYAB3WC5I4%2F20221208%2Fus-gov-west-1%2Fs3%2Faws4_request&X-Amz-SignedHeaders=host&X-Amz-Signature=32d023c062cf2abded478df3e16eb0c0ec03c4bbb03c2c993750f7413fbb7166

Leech, N. L., Barrett, K.C., & Morgan, G.A. (2015). *IBM SPSS for intermediate statistics: Use and interpretation, fifth edition*. Routledge.

Lingard, H. C., & Rowlinson, S. (2006). *Sample size in factor analysis: Why size matters*. https://safe.menlosecurity.com/doc/docview/viewer/docN6A7CD8CDA5494df93362c6a3d0917bfaa01b8f3c81620bcc5e2a03899869a4c2ad4311a44c43

Schneider, B., & Snyder, R. A. (1975). Some relationships between job satisfaction and organization climate. *Journal of Applied Psychology, 60*(3), 318.

Smith, C., & Koning, M. (2017). *Decision trees and random forests: A visual introduction for beginners*. Blue Windmill Media.

Willness, C. R., Steel, P., & Lee, K. (2007). A meta-analysis of the antecedents and consequences of workplace sexual harassment. *Personnel Psychology, 60*(1), 127–162.

List of Tables

Table 1. IBM SPSS Growing Methods	11
Table 2. Equal Opportunities (EO) variables	15
Table 3. Sexual Assault Prevention and Response (SAPR) variables	15
Table 4. Correlation Matrix	16
Table 5. KMO and Bartlett's Test	17
Table 6. Communalities	19
Table 7. Total Variance Explained	20
Table 8. Component Matrix	22
Table 9. Component Score Coefficient Matrix	23
Table 10. Target OE Variable	24
Table 11. Depend Variables of Decision Tree	25
Table 12. Independent variables - Demographics variables	26
Table 13. Independent variables - Equal Opportunities (EO) variables	26
Table 14. Independent variables - Sexual Assault Prevention and Response(SAPR) variables	27
Table 15. Time variables	27
Table 16. First Important Variable Impacting Job Satisfaction from Decision Tree Rule (CHAID)	28
Table 17. Second Important Variable Impacting Job Satisfaction from Decision Tree Rule (CHAID)	28

Table 18. Third Important Variable Impacting Job Satisfaction from Decision Tree Rule (CHAID)	29
Table 19. First Important Variable Impacting Job Satisfaction from Decision Tree Rule (CRT)	30
Table 20. Second Important Variable Impacting Job Satisfaction from Decision Tree Rule (CRT)	31
Table 21. Third Important Variable Impacting Job Satisfaction from Decision Tree Rule (CRT)	31
Table 22. Order of Importance of EO-SAPR Variables to "GroupCoh"	37
Table 23. Order of Importance of EO-SAPR Variables to "Engagement"	39
Table 24. Order of Importance of EO-SAPR Variables to "Inclusion"	40
Table 25. Order of Importance of EO-SAPR Variables to "TrustinLeader"	42
Table 26. Correlation Matrix	44
Table 27. Component Matrix	46
Table 28. Chi-Square Statistic Value between "Sexual Harassment Retaliation Climate" and "Job Satisfaction"	47
Table 29. Order of Important Variable Impacting Job Satisfaction from Decision Tree Rule	49
Table 30. Classification	53
Table 31. Classification	54

List of Figures

Figure 1. Sample Decision Tree	9
Figure 2. Scree Plot	21
Figure 3. Simple Histogram of Job Satisfaction	25
Figure 4. Normalized Importance	33
Figure 5. Comparison of First Important Variables Impacting JobSat - Male vs Female	34
Figure 6. Comparison of Second Important Variables Impacting JobSat - Male vs Female	35
Figure 7. Comparison of Third Important Variables Impacting JobSat - Male vs Female	35
Figure 8. Simple Bar of Group Cohesion Scale Score	37
Figure 9. Simple Bar of Engagement	38
Figure 10. Simple Bar of Inclusiveness	40
Figure 11. Simple Bar of Trust in Leadership Scale Score	41
Figure 12. Normalized Importance	51

Appendices

Appendix A - SPSS Syntax for PCA

```
FACTOR
  /VARIABLES Hazing Bullying Discrimination SexHar SAPRPrevent SAPRRepKnowledge
SAPRResponse
    SAPRRetaliation SexHarRetaliation
  /MISSING LISTWISE
  /ANALYSIS Hazing Bullying Discrimination SexHar SAPRPrevent SAPRRepKnowledge
SAPRResponse
    SAPRRetaliation SexHarRetaliation
  /PRINT UNIVARIATE INITIAL CORRELATION KMO REPR EXTRACTION ROTATION
FSCORE
  /FORMAT SORT BLANK(.10)
  /PLOT EIGEN ROTATION
  /CRITERIA MINEIGEN(0.8) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /METHOD=CORRELATION.
```

Appendix B - SPSS Python Syntax for Decision Tree

```
*----------Open dataset and Create Categorized Job Satisfaction variable-------*.

BEGIN PROGRAM python3.

import spss,spssaux


spss.Submit(r"""

GET FILE="F:/Data_05312022/DoD SAFE-3/Clean_Oct17-July20.sav".

DATASET NAME alldata.


STRING  Cate_JobSat (A8).

IF  (JobSat < 3) Cate_JobSat="1".

IF ((JobSat >= 3) & (JobSat <= 5)) Cate_JobSat="2".

IF (JobSat >5) Cate_JobSat="3".

IF (SYSMIS(JobSat) = 1) Cate_JobSat="4".

VARIABLE LABELS  Cate_JobSat '1=at-risk 2=neutral 3=happy 4=missing value'.

VALUE LABELS Cate_JobSat '1' 'at-risk' '2' 'neutral' '3' 'satisfied' '4' 'missing value'.


EXECUTE.


DATASET ACTIVATE alldata.

SAVE OUTFILE='F:/Data_05312022/DoD SAFE-3/Clean_Oct17-July20_newvar2.sav'.

new file.

DATASET CLOSE alldata.
```

```
""")
END Program.
*----------split into 34 separate files for each time period-------*.
SPSSINC SPLIT DATASET SPLITVAR=MMYR
/OUTPUT DIRECTORY= "F:\SplitFiles2" DELETECONTENTS=NO
/OPTIONS NAMES=VALUES.
*----------count number of unique values in "MMYR"-------*.
BEGIN PROGRAM python3.
spss.Submit(r"""
GET FILE="F:/Data_05312022/DoD SAFE-3/Clean_Oct17-July20_newvar.sav".
DATASET NAME alldatawithnewvar.
""")


spss.StartDataStep()
datasetObj1 = spss.Dataset(name="alldatawithnewvar")
templist = []
target_index = 0
for i in range(spss.GetVariableCount()):
    if spss.GetVariableName(i) == 'MMYR':
        print(i, spss.GetVariableName(i))
        target_index = i


for i in range(len(datasetObj1.cases)):
```

```
    templist.append(datasetObj1.cases[i][target_index])


new_set = set(templist)
print("No of unique items in the list are:", len(new_set))
spss.EndDataStep()
END Program.
*---------Run SPSS Decision Tree for each time period by CRT-------*.
BEGIN PROGRAM python3.
spss.Submit(r"""
GET FILE="F:/SplitFiles2/1.sav".
DATASET NAME alldata_1.


TREE Cate_JobSat [n] BY Race [n] Rank [n] MilCiv [n] SexHarRetaliation [s] Hazing [s]
   Discrimination [s] SAPRRepKnowledge [s]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
  /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
  /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK TREETABLE
  /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
  /PLOT IMPORTANCE GAIN INDEX RESPONSE INCREMENT=10
  /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
  /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
```

```
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
  /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
  /COSTS EQUAL
  /PRIORS FROMDATA ADJUST=NO
  /MISSING NOMINALMISSING=MISSING.
""")


for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))


    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFiles2/{}.sav".
            DATASET NAME alldata_{}.


            TREE Cate_JobSat [n] BY Race [n] Rank [n] MilCiv [n] SexHarRetaliation [s] Hazing
[s]
                Discrimination [s] SAPRRepKnowledge [s]
               /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
```

```
        /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
        /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK
TREETABLE
        /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
        /PLOT IMPORTANCE GAIN INDEX RESPONSE INCREMENT=10
        /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
        /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
        /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
        /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
        /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
        /COSTS EQUAL
        /PRIORS FROMDATA ADJUST=NO
        /MISSING NOMINALMISSING=MISSING.

       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       """.format(j,j,k,k))
  else:
     spss.Submit(r"""
       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
```

```
        new file.
        """.format(k,k))
End Program.
*----------End-Run decision Tree for each time period-------*.
*----------Run decision Tree for each time period by CHAID-------*.
BEGIN PROGRAM python3.
spss.Submit(r"""
GET FILE="F:/SplitFiles2/1.sav".
DATASET NAME alldata_1.

TREE Cate_JobSat [n] BY Rank [n] MilCiv [n] Discrimination [s] Hazing [s]
SAPRRepKnowledge [s]
   SexHarRetaliation [s] Race [n]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
  /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
  /PRINT MODELSUMMARY CLASSIFICATION RISK TREETABLE
  /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
  /PLOT GAIN INDEX RESPONSE INCREMENT=10
  /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
  /METHOD TYPE=CHAID
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
```

```
  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
  /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO
CHISQUARE=PEARSON CONVERGE=0.001
    MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10
  /COSTS EQUAL
  /MISSING NOMINALMISSING=MISSING.
""")

for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))

    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFiles2/{}.sav".
            DATASET NAME alldata_{}.

            TREE Cate_JobSat [n] BY Rank [n] MilCiv [n] Discrimination [s] Hazing [s]
SAPRRepKnowledge [s]
                SexHarRetaliation [s] Race [n]
              /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
```

```
            /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
            /PRINT MODELSUMMARY CLASSIFICATION RISK TREETABLE
            /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
            /PLOT GAIN INDEX RESPONSE INCREMENT=10
            /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
            /METHOD TYPE=CHAID
            /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
            /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
            /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO
CHISQUARE=PEARSON CONVERGE=0.001
              MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10
            /COSTS EQUAL
            /MISSING NOMINALMISSING=MISSING.


          DATASET ACTIVATE alldata_{}.
          DATASET CLOSE alldata_{}.
          """.format(j,j,k,k))
    else:
        spss.Submit(r"""
          DATASET ACTIVATE alldata_{}.
          DATASET CLOSE alldata_{}.
```

```
    new file.
    """.format(k,k))
End Program.
*----------End-Run decision Tree for each time period-------*.
```

Appendix C - SPSS Python Syntax for Decision Tree for Separate Gender

```
BEGIN PROGRAM python3.

import spss,spssaux

spss.Submit(r"""

GET FILE="F:/Data_05312022/DoD SAFE-3/Clean_Oct17-July20_newvar2.sav".

DATASET NAME alldata.""")

END Program.


*---split file into two file based on Gender variable---.

DATASET ACTIVATE DataSet1.

SPSSINC SPLIT DATASET SPLITVAR=Gender

/OUTPUT DIRECTORY= "F:\SplitFile3_Gender" DELETECONTENTS=NO

/OPTIONS NAMES=VALUES.

*----------split above two files into 34 separate files for each time period, separately-------*.

*--male---.

DATASET ACTIVATE DataSet4.

SPSSINC SPLIT DATASET SPLITVAR=MMYR

/OUTPUT DIRECTORY= "F:\SplitFile3_Gender\1_male_34sections"

DELETECONTENTS=NO

/OPTIONS NAMES=VALUES.

*--female---.

DATASET ACTIVATE DataSet5.

SPSSINC SPLIT DATASET SPLITVAR=MMYR
```

```
/OUTPUT DIRECTORY= "F:\SplitFile3_Gender\2_female_34sections"
DELETECONTENTS=NO
/OPTIONS NAMES=VALUES.
*----------Run decision Tree for each time period by CHAID just for male -------*.
BEGIN PROGRAM python3.
spss.Submit(r"""
GET FILE="F:/SplitFile3_Gender/1_male_34sections/1.sav".
DATASET NAME alldata_1.

TREE Cate_JobSat [n] BY Rank [n] MilCiv [n] Discrimination [s] Hazing [s]
SAPRRepKnowledge [s]
    SexHarRetaliation [s] Race [n]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
  /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
  /PRINT MODELSUMMARY CLASSIFICATION RISK TREETABLE
  /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
  /PLOT GAIN INDEX RESPONSE INCREMENT=10
  /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
  /METHOD TYPE=CHAID
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
```

```
  /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO
CHISQUARE=PEARSON CONVERGE=0.001
    MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10
  /COSTS EQUAL
  /MISSING NOMINALMISSING=MISSING.
""")


for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))


    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFile3_Gender/1_male_34sections/{}.sav".
            DATASET NAME alldata_{}.


            TREE Cate_JobSat [n] BY Rank [n] MilCiv [n] Discrimination [s] Hazing [s]
SAPRRepKnowledge [s]
                SexHarRetaliation [s] Race [n]
              /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
              /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
```

```
        /PRINT MODELSUMMARY CLASSIFICATION RISK TREETABLE
        /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
        /PLOT GAIN INDEX RESPONSE INCREMENT=10
        /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
        /METHOD TYPE=CHAID
        /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
        /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
        /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO
CHISQUARE=PEARSON CONVERGE=0.001
         MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10
        /COSTS EQUAL
        /MISSING NOMINALMISSING=MISSING.


       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       """.format(j,j,k,k))
  else:
     spss.Submit(r"""
       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       new file.
```

```
    """.format(k,k))
End Program.
*----------End-Run decision Tree for each time period CHAID-------*.
*----------Run decision Tree for each time period CRT for male-------*.
BEGIN PROGRAM python3.
spss.Submit(r"""
GET FILE="F:/SplitFile3_Gender/1_male_34sections/1.sav".
DATASET NAME alldata_1.

TREE Cate_JobSat [n] BY Race [n] Rank [n] MilCiv [n] SexHarRetaliation [s] Hazing [s]
   Discrimination [s] SAPRRepKnowledge [s]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
  /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
  /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK TREETABLE
  /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
  /PLOT IMPORTANCE GAIN INDEX RESPONSE INCREMENT=10
  /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
  /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
  /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
```

```
  /COSTS EQUAL
  /PRIORS FROMDATA ADJUST=NO
  /MISSING NOMINALMISSING=MISSING.
""")

for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))

    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFile3_Gender/1_male_34sections/{}.sav".
            DATASET NAME alldata_{}.

            TREE Cate_JobSat [n] BY Race [n] Rank [n] MilCiv [n] SexHarRetaliation [s] Hazing
[s]
                Discrimination [s] SAPRRepKnowledge [s]
              /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
              /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
              /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK
TREETABLE
```

```
        /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
        /PLOT IMPORTANCE GAIN INDEX RESPONSE INCREMENT=10
        /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
        /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
        /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
        /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
        /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
        /COSTS EQUAL
        /PRIORS FROMDATA ADJUST=NO
        /MISSING NOMINALMISSING=MISSING.

        DATASET ACTIVATE alldata_{}.
        DATASET CLOSE alldata_{}.
        """.format(j,j,k,k))
  else:
      spss.Submit(r"""
        DATASET ACTIVATE alldata_{}.
        DATASET CLOSE alldata_{}.
        new file.
        """.format(k,k))
End Program.
```

```
*----------End-Run decision Tree for each time period CRT-------*.

*----------Run decision Tree for each time period by CHAID just for female -------*.

BEGIN PROGRAM python3.

spss.Submit(r"""

GET FILE="F:/SplitFile3_Gender/2_female_34sections/1.sav".

DATASET NAME alldata_1.


TREE Cate_JobSat [n] BY Rank [n] MilCiv [n] Discrimination [s] Hazing [s]

SAPRRepKnowledge [s]

   SexHarRetaliation [s] Race [n]

  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES

NODEDEFS=YES SCALE=AUTO

  /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']

  /PRINT MODELSUMMARY CLASSIFICATION RISK TREETABLE

  /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING

CUMULATIVE=NO

  /PLOT GAIN INDEX RESPONSE INCREMENT=10

  /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES

  /METHOD TYPE=CHAID

  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50

  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES

  /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO

CHISQUARE=PEARSON CONVERGE=0.001
```

```
  MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10
 /COSTS EQUAL
 /MISSING NOMINALMISSING=MISSING.
""")


for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))


    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFile3_Gender/2_female_34sections/{}.sav".
            DATASET NAME alldata_{}.


            TREE Cate_JobSat [n] BY Rank [n] MilCiv [n] Discrimination [s] Hazing [s]
SAPRRepKnowledge [s]
              SexHarRetaliation [s] Race [n]
             /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
             /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
             /PRINT MODELSUMMARY CLASSIFICATION RISK TREETABLE
```

```
        /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
        /PLOT GAIN INDEX RESPONSE INCREMENT=10
        /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
        /METHOD TYPE=CHAID
        /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
        /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
        /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO
CHISQUARE=PEARSON CONVERGE=0.001
         MAXITERATIONS=100 ADJUST=BONFERRONI INTERVALS=10
        /COSTS EQUAL
        /MISSING NOMINALMISSING=MISSING.

       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       """.format(j,j,k,k))
  else:
     spss.Submit(r"""
       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       new file.
       """.format(k,k))
```

```
End Program.

*----------End-Run decision Tree for each time period CHAID-------*.

*----------Run decision Tree for each time period CRT for female-------*.

BEGIN PROGRAM python3.

spss.Submit(r"""

GET FILE="F:/SplitFile3_Gender/2_female_34sections/1.sav".

DATASET NAME alldata_1.


TREE Cate_JobSat [n] BY Race [n] Rank [n] MilCiv [n] SexHarRetaliation [s] Hazing [s]
   Discrimination [s] SAPRRepKnowledge [s]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
  /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
  /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK TREETABLE
  /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
  /PLOT IMPORTANCE GAIN INDEX RESPONSE INCREMENT=10
  /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
  /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
  /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
  /COSTS EQUAL
```

```
  /PRIORS FROMDATA ADJUST=NO
  /MISSING NOMINALMISSING=MISSING.
""")

for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))

    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFile3_Gender/2_female_34sections/{}.sav".
            DATASET NAME alldata_{}.

            TREE Cate_JobSat [n] BY Race [n] Rank [n] MilCiv [n] SexHarRetaliation [s] Hazing
[s]
              Discrimination [s] SAPRRepKnowledge [s]
             /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
             /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
             /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK
TREETABLE
```

```
            /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
            /PLOT IMPORTANCE GAIN INDEX RESPONSE INCREMENT=10
            /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
            /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
            /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
            /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
            /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
            /COSTS EQUAL
            /PRIORS FROMDATA ADJUST=NO
            /MISSING NOMINALMISSING=MISSING.

        DATASET ACTIVATE alldata_{}.
        DATASET CLOSE alldata_{}.
        """.format(j,j,k,k))
    else:
        spss.Submit(r"""
        DATASET ACTIVATE alldata_{}.
        DATASET CLOSE alldata_{}.
        new file.
        """.format(k,k))
End Program.
```

----------End-Run decision Tree for each time period CRT-------.

Appendix D - SPSS Python Syntax for Decision Tree for Other OE variables

```
*----------Run decision Tree for each time period by CRT: use variable "Engagement" as
outcome variable-------*.
BEGIN PROGRAM python3.
import spss,spssaux
spss.Submit(r"""
GET FILE="F:/SplitFiles3_Cate_TrustinLeader/1.sav".
DATASET NAME alldata_1.

TREE Engagement [s] BY Hazing [s] Discrimination [s] SAPRRepKnowledge [s]
SexHarRetaliation [s]
    MilCiv [n] Rank [n] Race [n]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
  /PRINT MODELSUMMARY RISK
  /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
  /PLOT IMPORTANCE
  /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
  /CRT MINIMPROVEMENT=0.0001
  /MISSING NOMINALMISSING=MISSING.
```

```
""")

for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))

    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFiles3_Cate_TrustinLeader/{}.sav".
            DATASET NAME alldata_{}.

            TREE Engagement [s] BY Hazing [s] Discrimination [s] SAPRRepKnowledge [s]
SexHarRetaliation [s]
                MilCiv [n] Rank [n] Race [n]
              /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
              /PRINT MODELSUMMARY RISK
              /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
              /PLOT IMPORTANCE
              /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
```

```
        /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
        /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
        /CRT MINIMPROVEMENT=0.0001
        /MISSING NOMINALMISSING=MISSING.


       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       """.format(j,j,k,k))
  else:
     spss.Submit(r"""
       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       new file.
       """.format(k,k))
End Program.
*----------End-Run decision Tree for each time period-------*.
```

Appendix E - SPSS Syntax for Studying Six SAPR variables

```
*---PCA analysis on six variables---.
FACTOR
  /VARIABLES Sapr56a Sapr56b Sapr56c Sapr56d Sapr56e Sapr56f
  /MISSING LISTWISE
  /ANALYSIS Sapr56a Sapr56b Sapr56c Sapr56d Sapr56e Sapr56f
  /PRINT UNIVARIATE INITIAL CORRELATION SIG KMO EXTRACTION ROTATION
FSCORE
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /METHOD=CORRELATION.
*Crosstab analysi for each varaible.

CROSSTABS
  /TABLES=Sapr56a BY Cate_JobSat
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.

CROSSTABS
```

```
  /TABLES=Sapr56b BY Cate_JobSat
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.

CROSSTABS
  /TABLES=Sapr56c BY Cate_JobSat
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.

CROSSTABS
  /TABLES=Sapr56d BY Cate_JobSat
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.

CROSSTABS
  /TABLES=Sapr56e BY Cate_JobSat
  /FORMAT=AVALUE TABLES
```

```
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.

CROSSTABS
  /TABLES=Sapr56f BY Cate_JobSat
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.
```

Appendix F - SPSS Syntax for Decision Tree without PCA

```
*----------Run decision Tree (without PCA) for each time period by CRT -------*.
BEGIN PROGRAM python3.
import spss,spssaux


spss.Submit(r"""
GET FILE="F:/SplitFiles2/1.sav".
DATASET NAME alldata_1.


TREE Cate_JobSat [n] BY Hazing [s] Bullying [s] Discrimination [s] SexHar [s] SAPRPrevent
[s]
    SAPRRepKnowledge [s] SAPRResponse [s] SAPRRetaliation [s] SexHarRetaliation [s]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
  /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
  /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK
  /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
  /PLOT IMPORTANCE
  /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
  /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100 MINCHILDSIZE=50
  /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
```

```
  /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
  /COSTS EQUAL
  /PRIORS FROMDATA ADJUST=NO.
""")

for i in range(34):
    k = i+1
    j = k+1
    print("Above is for {} file".format(k))

    if j < 35:
        spss.Submit(r"""
            GET FILE="F:/SplitFiles2/{}.sav".
            DATASET NAME alldata_{}.

            TREE Cate_JobSat [n] BY Hazing [s] Bullying [s] Discrimination [s] SexHar [s]
SAPRPrevent [s]
                SAPRRepKnowledge [s] SAPRResponse [s] SAPRRetaliation [s] SexHarRetaliation
[s]
             /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES
NODEDEFS=YES SCALE=AUTO
             /DEPCATEGORIES USEVALUES=['1' '2' '3' '4'] TARGET=['1' '3']
             /PRINT MODELSUMMARY IMPORTANCE CLASSIFICATION RISK
```

```
        /GAIN CATEGORYTABLE=YES TYPE=[NODE] SORT=DESCENDING
CUMULATIVE=NO
        /PLOT IMPORTANCE
        /RULES NODES=TERMINAL SYNTAX=GENERIC LABELS=YES
        /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=SE(1)
        /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=100
MINCHILDSIZE=50
        /VALIDATION TYPE=CROSSVALIDATION(10) OUTPUT=BOTHSAMPLES
        /CRT IMPURITY=GINI MINIMPROVEMENT=0.0001
        /COSTS EQUAL
        /PRIORS FROMDATA ADJUST=NO.


       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       """.format(j,j,k,k))
  else:
     spss.Submit(r"""
       DATASET ACTIVATE alldata_{}.
       DATASET CLOSE alldata_{}.
       new file.
       """.format(k,k))
End Program.
*----------End-Run decision Tree (without PCA) for each time period CRT-------*.
```