**Towards Unveiling Relationship between Clan Culture and Job Satisfaction  
—Preliminary Analysis**

**Minjeong Seo**

**Department of Organizational Leadership Policy and Development**

**PSY** **8712: Data Science in Psychology**

**Dr. Richard N. Landers**

**May 04, 2024**

**Towards Unveiling Relationship between Clan Culture and Job Satisfaction  
—Preliminary Analysis**

**Project Overview**

Introducing Millennials and Generation Z (hereafter referred to as the MZ Generation) into Korean organizations has posed a significant challenge due to a high turnover rate. This issue is underscored by findings from the '2019 Graduates Occupational Mobility Survey' conducted by the Korea Employment Information Service. In 2019, the turnover rate for young worker was notably high at 17.4%, which stands out even during a period marked by heightened employment difficulties (Korea Employment Information Service, 2020). In response, organizations are implementing strategies such as enhancing job satisfaction, which is intricately linked to positive attitudes and improved performance (Song et al., 2022; Bang, 2019). Moreover, these strategies are closely tied to organizational culture.

Despite the acknowledged importance of job satisfaction and its positive impact on reducing turnover rates, the MZ generation exhibits lower levels of job satisfaction. While previous research has focused on enhancing job satisfaction and decreasing turnover rates, there exists a gap in comprehensive analyses that integrate these dimensions within the context of clan culture. To address this gap, this study presents an approach by investigating the relationships among organizational culture and job satisfaction. Thus, the project aims to finding the existence of these relationships, confirming control variables, and identifying the most suitable analytical models as a first step. Accordingly, this project explores an Ordinary Least Squares (OLS) model alongside various machine learning models to determine the actual relationships and find the most appropriate model. Additionally, a Shiny application has been developed to aid in comprehending the outcomes of this study.

**Hypotheses and Research Questions**

To achieve the project purpose, the study poses the following research questions:

*Research Question 1: Does a significant relationship exist between IV and DV?*

*Research Question 2: Which control variables can be included to ensure a more precise study?*

*Research Question 3: Which model is the most suitable for conducting the research?*

**Method**

**Open Science Materials**

To enhance accessibility, promote reproducibility and collaboration, and facilitate the sharing and publishing of my analysis, I developed a web-based binder and uploaded all project materials to GitHub. The web-based binder and GitHub provide a straightforward way for researchers to access and use the materials associated with this study. The binder is available at <https://mybinder.org/v2/gh/min9509/psy8712-final.git/HEAD>. For detailed information about the study materials and their usage, including the readme file. To access the detailed information in the readme file about the study materials and their usage, please follow these steps: Visit the GitHub repository dedicated to this study by clicking on <https://github.com/min9509/psy8712-final>. Once on the repository's main page, you will find a list of files and folders. Locate the "README.md" file, typically displayed prominently at the top of the list or within the repository's main folder. Click on the "README.md" file to open it and access comprehensive information about the study materials. This includes detailed instructions on how to use them effectively and any additional resources or references relevant to your understanding and usage of the materials.

**Data**

To address the research questions, I utilized secondary data from the 'Human Capital Corporate Panel' (HCCP), which has been developed and made publicly available by the South Korean government. The use of this data was authorized under approval number 389003 from Statistics Korea. Specifically, for this study, I employed the most recent data from HCCP Wave 2 conducted in 2022.

**Measures**

Organizational culture, especially clan culture, were established as independent variables to achieve this objective, with job satisfaction as the dependent variable. Gender, marital status, and job status (permeant or temporary) were designated as control variables. Independent and dependent variables were assessed using a 5-point Likert scale. Other control variables are categorical variables. The detailed questions for each variable are outlined in the following table.

Table 1. The explanation of Variables

|  |  |  |
| --- | --- | --- |
| **Category** | **Question** | **Code** |
| Clan Culture | A family-like organizational atmosphere has been formed. | W21Q25D |
| My company/team value unity | W21Q25E |
| My company/team places greater importance on teamwork | W21Q25F |
| Job Satisfaction | I am satisfied with the content of the work I am currently doing. | W21Q26A |
| I am satisfied with the salary I am currently receiving. | W21Q26B |
| I am satisfied with the interpersonal relationships at my current workplace. | W21Q26C |
| Gender | What is your gender? | W21DQ01 |
| Marital status | Are you currently married, single, divorced, separated, or widowed? | W21DQ03 |
| Job status | Are your job currently permanent job or temporary job? | W21Q28 |

**Procedure**

This project involved conducting descriptive statistics, correlation, and regression analyses to validate relationships and probabilities. Additionally, a Shiny app was developed to enhance comprehension, and machine learning techniques were applied to identify optimal models.

**Analyses**

**Data Cleaning**

**I imported the original data and performed cleaning operations using the select and mutate functions. Initially, I selected specific questions based on variables and subsequently calculated the average job satisfaction using three questions and the average clan culture using three questions. Additionally, for categorical variables, I converted numerical values to corresponding characters, such as changing 1 or 2 to Male or Female, respectively.**

**Descriptive Statistics and Static Visualizations**

I utilized descriptive statistics techniques to analyze the dataset, extracting key metrics such as mean, mode, minimum, and maximum values for numerical variables (e.g., job satisfaction, clan culture) using the summary function. Additionally, I generated histograms and calculated skewness and kurtosis to assess the normal distribution of the data. Frequency counts for categorical variables were determined using sum functions. Based on the descriptive statistics and histograms, both the independent and dependent variables exhibit normal distributions.

Table 2. Descriptive Statistics for Numerical Variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Median** | **Min** | **Max** |
| Job Satisfaction | 3.447 | 3.333 | 1 | 5 |
| Clan Culture | 3.350 | 3.333 | 1 | 5 |
| \*The higher the number, the higher the job satisfaction and clan culture. | | | | |

Table 3. Descriptive Statistics for Categorical Variables

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | | **N** | **Total** |
| Gender | Male | 6839 | 9516 |
| Female | 2677 |
| Marital Status | Single | 4115 | 9516 |
| Married | 5294 |
| Divorced/Widowed | 107 |
| Permanent Employment | Permanent | 9400 | 9516 |
| Temporary | 116 |

A graph of a graph

Description automatically generated

Fig 1. The Histogram of Job Satisfaction

A graph of a person with histogram

Description automatically generated

Fig 2. The Histogram of Clan Culture

Table 5. Skewness and kurtosis for Numerical Variables

|  |  |  |
| --- | --- | --- |
| **Variable** | **Mean** | **Median** |
| Job Satisfaction | 3.447 | 3.333 |
| Clan Culture | 3.350 | 3.333 |

**Correlation and Regression**

I conducted correlation and regression analyses to validate the relationships and probabilities, utilizing the cor.test and lm functions. The findings revealed significant relationships, with independent variables significantly impacting dependent variables. I also included a scatter plot depicting the relationship between the independent and dependent variables along with the regression line. Significant correlations and regressions were observed, indicating a positive relationship (coefficient = .592, p-value < .00) and a strong likelihood (coefficient = .526, p-value < .00).

**Interactive Visualization**

I developed a Shiny app aimed at enhancing comprehension about control variables for me and readers. This app generated scatter plots with control variables such as Gender, Marital status, and Permanent employment. App users can select exclude or include Gender, Marital Status, Permanent Employment. The shiny application highlighted instances where specific combinations (such as Female-single-permanent) were absent in the control variables, suggesting a need for greater consideration of these control variables.

**Machin Learning**

I employed machine learning techniques to identify the optimal models. Training was conducted on four different models: the OLS model, Elastic Net model, Random Forest model, and XGB model. The respective 10-fold cross-validation R squared values were obtained, along with the final holdout cross-validation R squared value, to determine the most suitable model for the analysis.  According to the tests, the results between models varied slightly, with OLS Regression and Elastic Net models achieving the highest R-squared values for both cross-validated (CV) and holdout sets.

Table 6. Skewness and kurtosis for Numerical Variables

|  |  |  |
| --- | --- | --- |
| **Model** | **CV R square** | **Holdout R square** |
| OLS Model | 1.0000000 | 1.0000000 |
| Elastic Net Model | 0.9999964 | 0.9999974 |
| Random Forest Model | 0.9973938 | 0.9999233 |
| XGB Model | 0.9973192 | 0.9999432 |

**Reflection**

In reflecting on my approach, I plan to adopt specific strategies to enhance the depth and accuracy of my analysis. Firstly, I will continue to leverage the significant, positive, and strong relationship between Clan culture and Job satisfaction, recognizing the model's suitability for advanced studies based on its predictive power. However, I will also address the absence of specific combinations in the control variables, such as Female-single-permanent, by incorporating new variables to enrich the analysis. Additionally, I found the variation in results across different models valuable, particularly noting the high predictive power of the Ordinary Least Squares (OLS) Regression and Elastic Net models. Therefore, for the main research, I intend to incorporate both the OLS and Elastic Net models, capitalizing on their strengths to yield comprehensive and reliable insights. These strategic adjustments refine the analysis and provide a robust foundation for drawing meaningful conclusions.

**Reference**

Bang, M. H. (2019). Moderating Effect Analysis of Interactional Justice on Relation of Distributive Justice and Procedural Justice to Job Satisfaction and Organizational Commitment. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 9(7), 911-923. http://dx.doi.org/10.35873/ajmahs.2019.9.7.082

Song, S. I., Kim, J. H., & Mo, Y. (2022). The Longitudinal study on the Impact of Innovative Organizational Culture on Organizational Commitment. *Journal of the Korea Convergence Society*, 13(4), 383–396. https://doi.org/10.15207/JKCS.2022.13.04.383