Evaluating the Impact of Employee Performance Metrics Across Multiple Variables Using MANOVA

Shinjon Ghosh

Department of Mathematics, Illinois State University

Abstract:

Employee training and development programs are vital for organizations aiming to improve performance, productivity, and engagement across departments. This study aims to provide a comprehensive view of how several variables influences various performance metrics (Productivity, Satisfaction Rate, Programs Completed, Feedback Score) allowing for targeted improvements. The hypothesis testing involves Multivariate Analysis of Variance (MANOVA) – whether employee performance matrices with demographic and organizational variables are significant or not. By examining a range of features, the aim to uncover patterns and relationships that influence gender, job positions and departments. The results highlight the employee performance metrics are significant and there is influential relationship between demographic (Gender) and organizational variables (Job Position, Department).

Keywords: Multivariate Analysis of Variance, employee performance metrics, outlier test, Boxplot, normality checking, Box-Cox transformation.

Introduction:

The impact of training and development programs play vital role in company's improvement. Three types of job training (on-the-job training, off-the-job training, and job rotation) have a positive relationship with organizational performance. Among the variables, job rotation showed the strongest correlation with organizational performance (0.829), followed by on-the-job training (0.771) and off-the-job training (0.703) [3]. Soft skills training is essential for organizations to enhance employee performance and achieve long-term success. The authors illustrate that prioritizing soft skills training should be an integral part of any organization's employee development policy. By developing a workforce proficient in both technical and interpersonal skills, organizations can foster a positive and productive work environment, leading to increased profitability, improved customer relations, and higher employee retention [1]. The employee development programs are increases day by day and includes several diversity tasks on the training program. The MANOVA provides insights into factors that related to employee performance metrics.

Background & Literature Review:

The importance of training and development programs to the specific needs of both the organization and its employees and evaluating the effectiveness of these programs to measure their impact on organizational performance. A high percentage of respondents agreed that training and development initiatives can boost job performance (82%), improve motivation and engagement (85%), influence retention rates (88%), keep employees updated with industry trends (90%), and enhance productivity (87%) [4]. The training programs at Brakes India significantly improve employee development, engagement, and organizational success. The study highlights key trends, including a predominantly male demographic, varying perceptions across age groups, and positive correlations between training participation and improved attendance, commitment to organizational goals, and employee satisfaction. The authors emphasize the need to adapt training programs continuously to meet evolving employee needs and align with organizational goals [8]. The proposed matrix variate p-value framework can be extended to various other multivariate hypothesis testing scenarios beyond MANOVA, including testing the equality of covariance matrices. The authors conclude by emphasizing the potential of this new matrix p-value approach to offer a more robust and accurate method for analyzing multivariate data across various statistical applications [5]. The Box–Cox transformation is a valuable technique in linear regression, particularly when the assumptions of the regression model are not met. It involves applying a power transformation to the response variable, aiming to stabilize variances and improve the fit to the normality assumption. They emphasize that by focusing on efficient computation of summary statistics, like the Box–Cox information array, statisticians can develop more effective and scalable methods for handling massive datasets [6].

Objectives:

- Set a hypothesis test on MANOVA based on independent variables (Gender, Position, Department) and employee performance metrics.
- Identify key factors influencing employee performance metrics (Programs Completed, Productivity, Satisfaction Rate, Feedback Score).

Methods & Materials:

Data Source & Overview

The Employee Performance Metrics dataset collects from Kaggle (secondary data resources) which provides demographics, organizational and performance of employees. The dataset contains 200 observations and 11 features. Among 11 variables position (Analyst, Intern, Junior Developer, Senior Developer, Manager, Team lead), department (Finance, HR, IT, Marketing, Sales), gender (Male & Female) are categorical features and age, salary, productivity, projects completed, satisfaction rate, feedback score are numerical variables. Moreover, the dataset has employee join date and name columns. Each row corresponds to a specific user, identifiable by their unique name. This dataset allows us to explore and analyze employee performance metrics based on demographic and organizational variables. It is a useful resource for hypothesis testing on MANOVA and find out impact on performance matrices.

Data Preparation

The aim of our project is to set a hypothesis on MANOVA and find out effect of demographic and organizational variables with employee performance metrics. The categorical variables are factorized by labels where gender (Male = 1 & Female = 2), position (Analyst = 1, Manager = 2, Team lead = 3, Junior Developer = 4, Senior Developer = 5, & Intern = 6), department (Marketing = 1, IT = 2, Sales = 3, HR = 4, & Finance = 5). Later, I define gender, position and department as independent variable. On the other hand, I set performance metrics variables are dependent variables.

Result & Discussion:

I find out outliers and extreme observations from employee performance metrics variables using summary statistics. From the SAS output, I get 0,1, 24 & 25 values are extreme observation in Projects Completed variables and 0 & greater than 95 values are outliers in Productivity variables. Moreover, I observe in Satisfaction Rate variables those values are less than 2 and larger than 96, these are extreme observations. Furthermore, in Feedback Score variables values which are smaller than 1.2 and larger than 4.7, these are high leverage points. After finding extreme observations, I remove those values from the dataset because those values deviate the results.

The SAS System

The UNIVARIATE Procedure
Variable: Projects_Completed (Projects_Completed)

Extreme Observations

Lowest			Highest		
Value	Name	Obs	Value	Name	Obs
0	Susan Foley	93	24	Mary Barber	100
0	Allison Smith	77	24	Patrick Green	108
1	Mrs. Heidi Villa	147	24	Timothy Johnson	143
1	John Elliott	132	25	Daniel Mata	79
1	Nicholas Johnson	123	25	Michael Oneal	164

The SAS System

The UNIVARIATE Procedure
Variable: Satisfaction_Rate (Satisfaction_Rate)

Extreme Observations

Lowest			Highest		
Value	Name	Obs	Value	Name	Obs
0	Joseph Friedman	191	97	Donald Davis	13
0	Manuel Sanchez	188	98	Ryan Avila	161
0	Susan Foley	93	99	Kim Larson	32
0	Jerry Mccoy	88	100	Emily Palmer	33
1	Daniel Mata	79	100	Drew Rojas	194

The SAS System

The UNIVARIATE Procedure
Variable: Productivity (Productivity)

Extreme Observations

Lowest			Highest		
Value	Name	Obs	Value	Name	Obs
0	Tyrone Underwood	186	96	Michael Hernandez	199
0	Susan Hill	157	97	Shannon Gray	174
0	Joshua White	152	97	Autumn Taylor	177
0	Steven Anderson	34	98	Patrick Manning	134
0	Heather Mueller	27	98	Ryan Avila	161

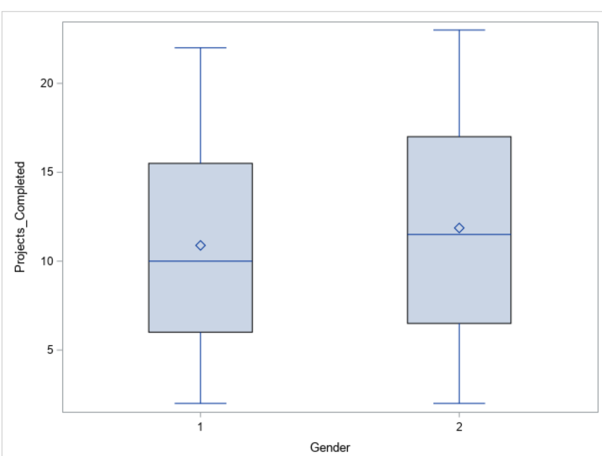
The SAS System

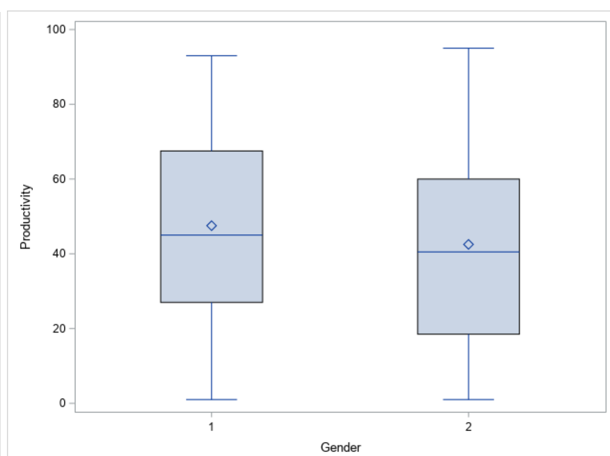
The UNIVARIATE Procedure
Variable: Feedback_Score (Feedback_Score)

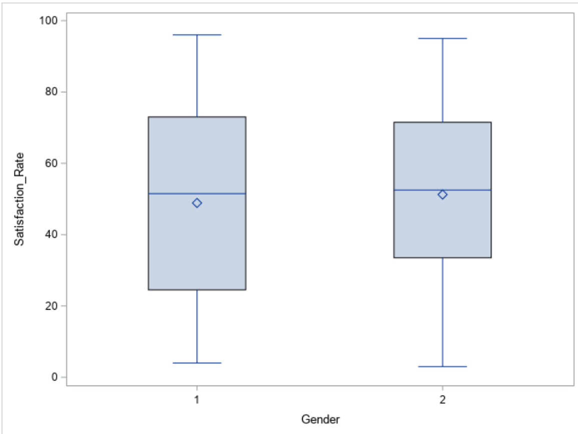
Extreme Observations

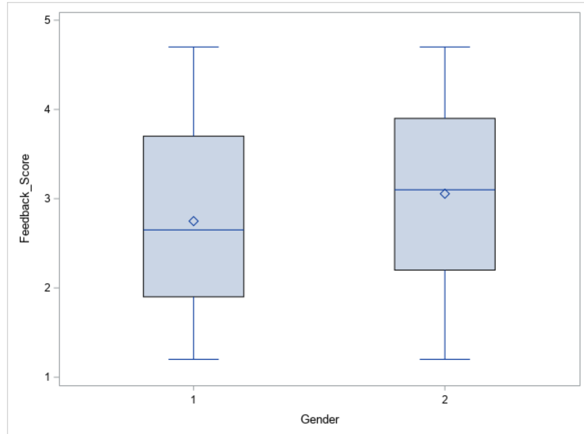
Lowest			Highest		
Value	Name	Obs	Value	Name	Obs
1.0	Autumn Taylor	177	4.8	Alyssa Montoya	136
1.1	Brian Delgado	187	4.9	Michelle Perez	31
1.1	Brett Butler	150	4.9	Amy Carr	97
1.1	Patricia Collins	148	4.9	Matthew Moran	156
1.1	Michael Moore	118	4.9	Michael Oneal	164

After removing extreme observations from the dataset, I find the total observation size is 156. Then I draw boxplot to check out more outliers in the dataset. I use performance metrics variables against gender. From the box plot, we can say there is not any outliers in performance metrics variables.

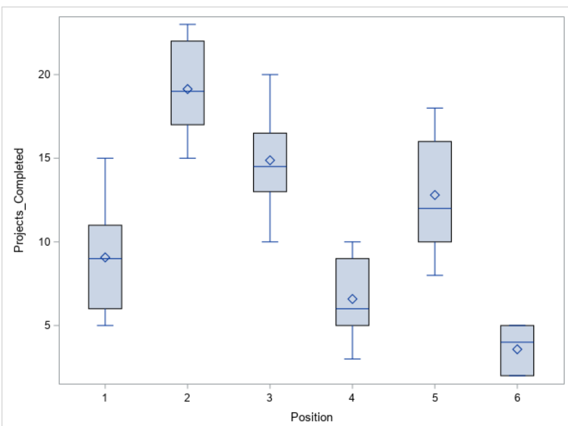


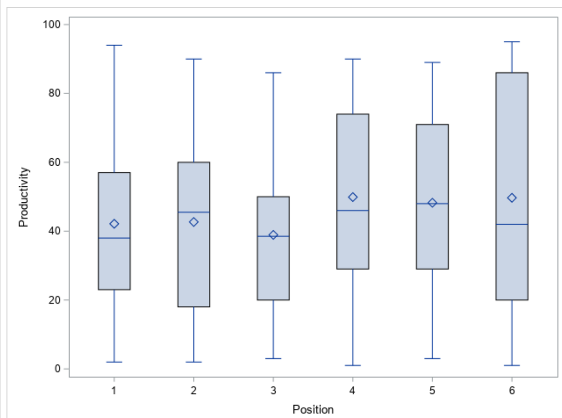


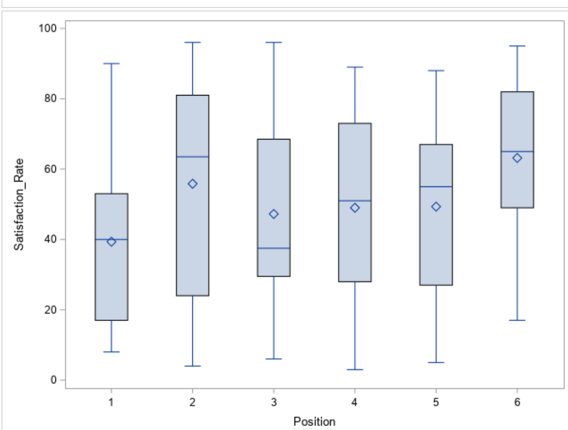


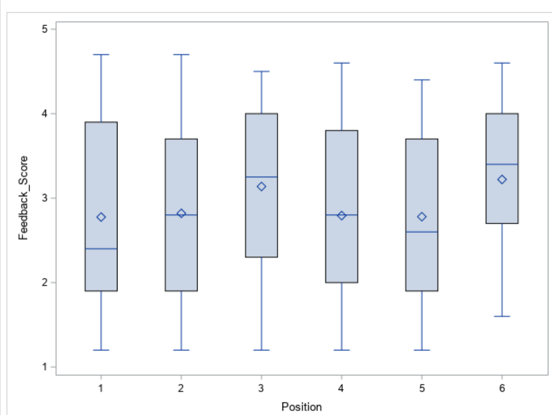


Later, I checked outliers of performance metrics variables against position variables by box plot. From the output of box plot, I do not find any extreme observations.

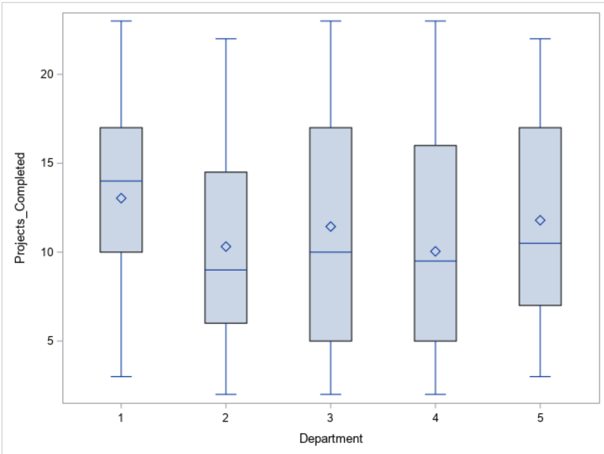


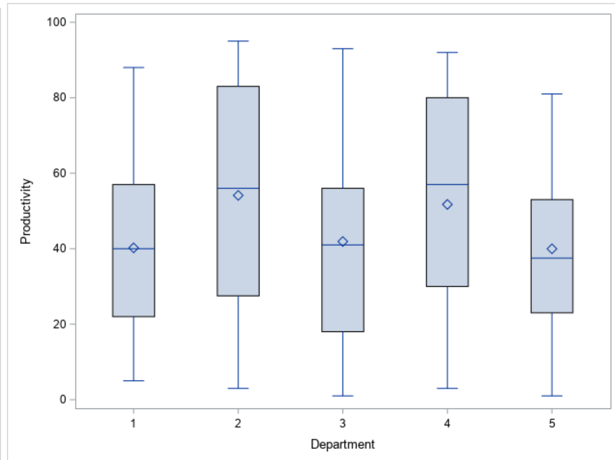


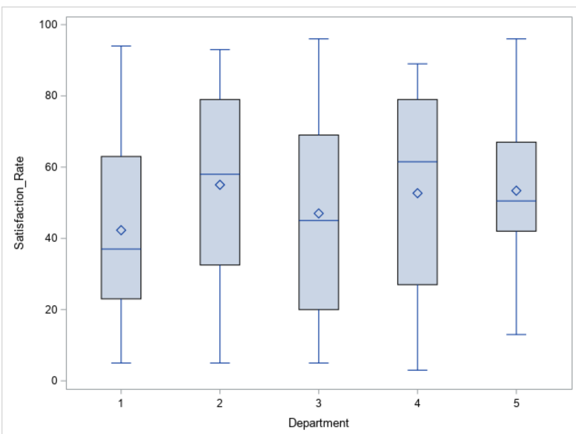


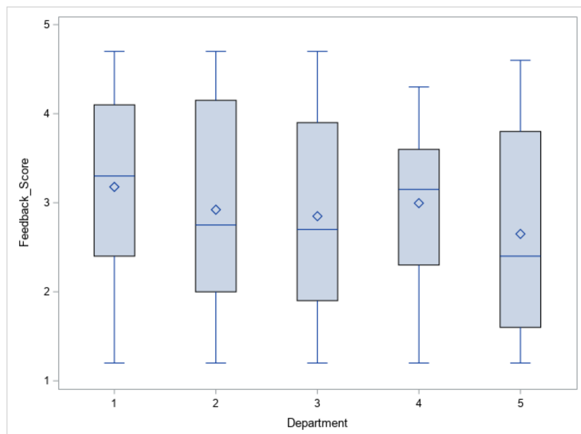


Moreover, I draw box plot of performance metrics variables against independent variables department. From the box plot analysis, I do not find any outliers.





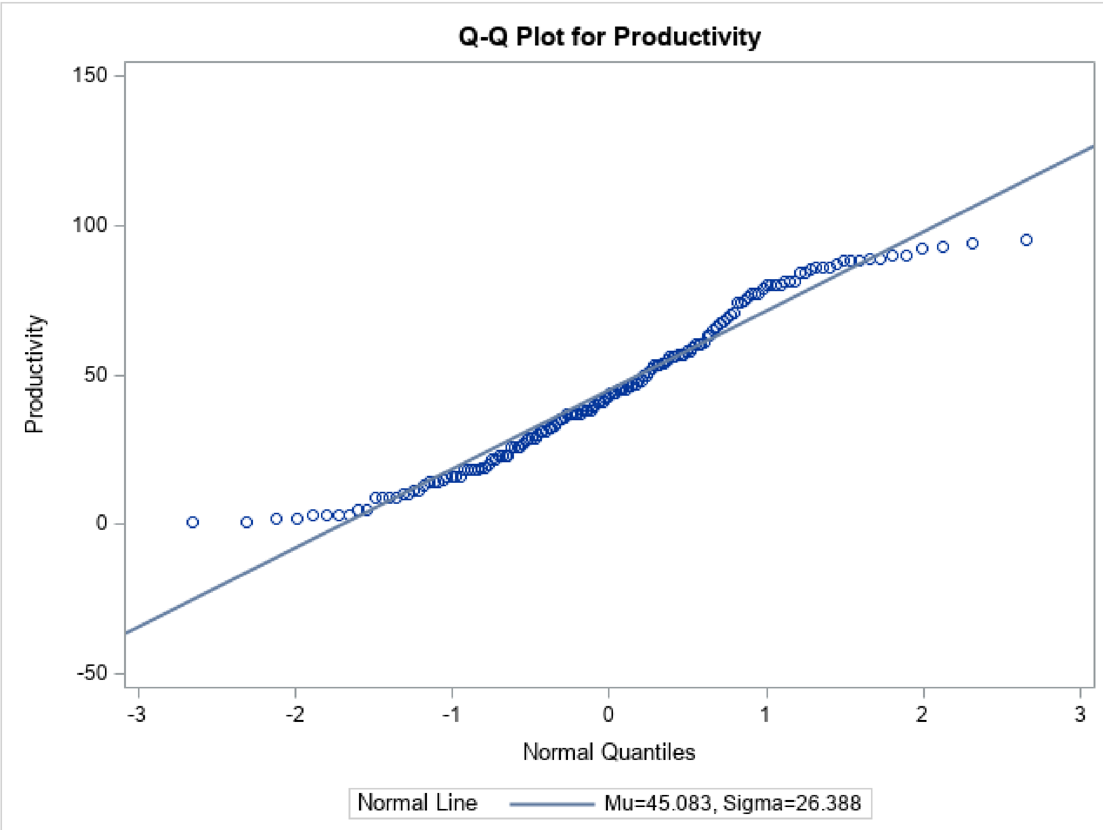




I check normality assumptions of performance metrics variables by Shapiro-wilk test, Kolmogorov test, and Q-Q plot.

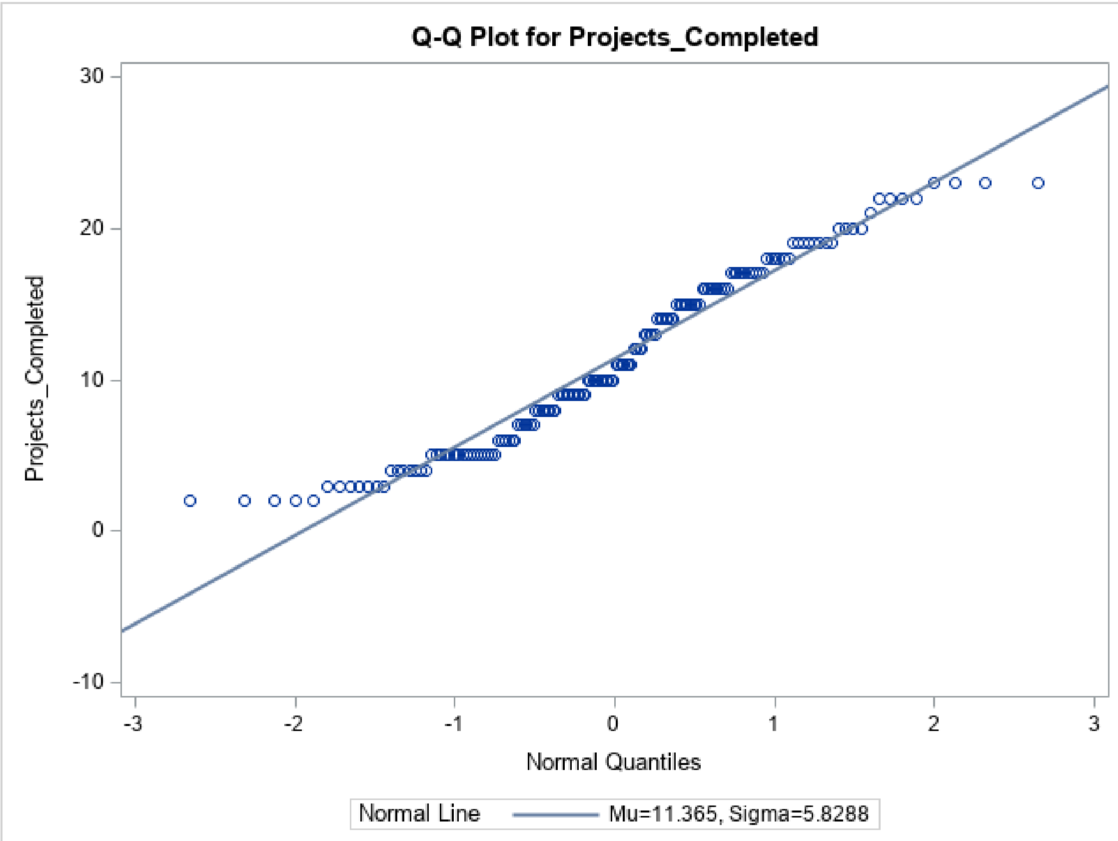
Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	W	0.95915	Pr < W	0.0001
Kolmogorov-Smirnov	D	0.07496	Pr > D	0.0310
Cramer-von Mises	W-Sq	0.184245	Pr > W-Sq	0.0085
Anderson-Darling	A-Sq	1.484241	Pr > A-Sq	<0.0050



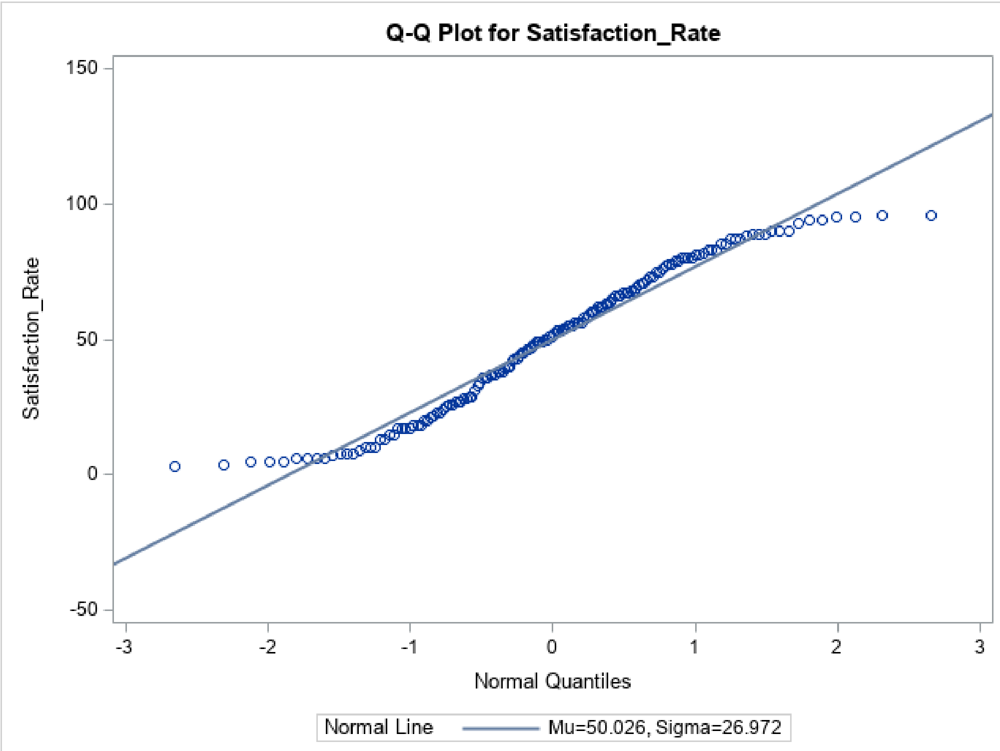
From the output of Shapiro-wilk test, Kolmogorov test, and Q-Q plot, I observe the performance metrics variables productivity is not normal because the p-value is less than 0.05 and points are not perfectly align on the line in Q-Q plot.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.95427	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.093364	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.334698	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.100585	Pr > A-Sq	<0.0050



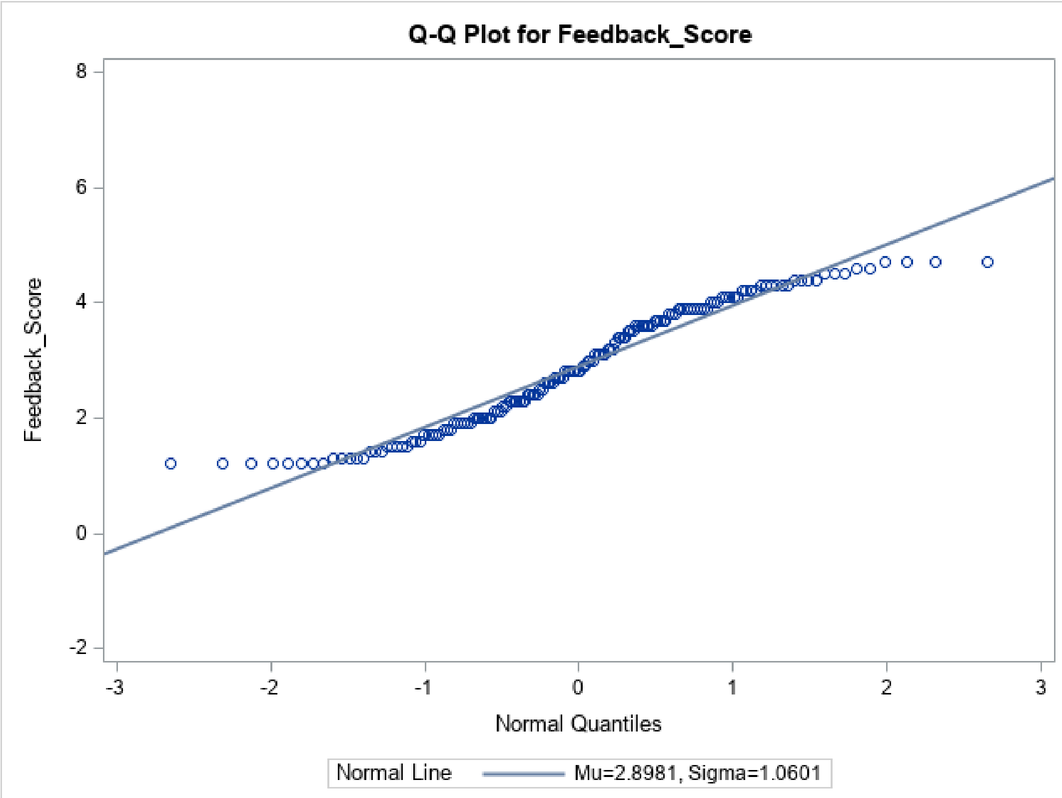
From the output of Shapiro-wilk test, Kolmogorov test, and Q-Q plot, I find out the performance metrics variables projects completed is not normal because the p-value is less than 0.05 and points are not perfectly align on the line in Q-Q plot.

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.956675	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.070626	Pr > D	0.0558
Cramer-von Mises	W-Sq	0.192139	Pr > W-Sq	0.0068
Anderson-Darling	A-Sq	1.526684	Pr > A-Sq	<0.0050



From the output of Shapiro-wilk test, Kolmogorov test and q-q plot, we can say that the satisfaction rate variable is not normal. because the p-value is less than 0.05 and points are curving on the line in Q-Q plot.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.94436	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.105031	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.368774	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.420685	Pr > A-Sq	<0.0050



From the output of Shapiro-wilk test, Kolmogorov test and q-q plot, we can say that the feedback score variable is not normal the p-value is less than 0.05 and points are curving on the line in Q-Q plot.

Then I use log transformation, square root transformation and box-cox transformation on employee performance metrics variables to meet normality assumptions. For log and square root transformation, I find the results is similar as without transformation output. But in Box – Cox transformation I get different output from performance metrics variables.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.944307	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.111373	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.361262	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.401498	Pr > A-Sq	<0.0050

After Box – Cox transformation, I observe the projects completed variables is not normal according to Kolmogorov test because p-value is less than 0.05.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.96894	Pr < W	0.0014
Kolmogorov-Smirnov	D	0.069448	Pr > D	0.0655
Cramer-von Mises	W-Sq	0.102389	Pr > W-Sq	0.1049
Anderson-Darling	A-Sq	0.949665	Pr > A-Sq	0.0175

After the Box – Cox transformation, I find the productivity variables is normal according to Kolmogorov test because p-value is 0.06 which is larger than 0.05.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.953109	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.068054	Pr > D	0.0770
Cramer-von Mises	W-Sq	0.250495	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.80751	Pr > A-Sq	<0.0050

I notice the satisfaction rate variable is normal according to Kolmogorov test because p-value is 0.07 which is greater than 0.05.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.959672	Pr < W	0.0002
Kolmogorov-Smirnov	D	0.10616	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.288312	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.853401	Pr > A-Sq	<0.0050

After Box – Cox transformation, I observe the feedback score variable is not normal according to Kolmogorov test because p-value is less than 0.05.

Now, I get productivity and satisfaction rate variables follow normality assumption. Then I set null hypothesis and alternative hypothesis for analyze relationship between gender, position, department with employee performance metrics using MANOVA.

Null Hypothesis:

$$H_0 : \mu_i = \mu_j = \mu_k = 0 ; \quad i \epsilon \; Gender\ levels, j \epsilon \; Position\ levels, k \epsilon \; Department\ levels$$

It means there is no significant effect of gender, position & department with employee performance metrics variables.

Alternative Hypothesis:

$$H_a : \mu_{ijk} \neq 0$$

It illustrates that at least one independent variable (Gender, Position, Department) has significant effect on the employee metrics variables (productivity, satisfaction rate).

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall Gender*Positi*Depart Effect
H = Type III SSCP Matrix for Gender*Positi*Depart
E = Error SSCP Matrix

S=2 M=26.5 N=48

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.32662787	1.31	112	196	0.0493
Pillai's Trace	0.85638113	1.32	112	198	0.0437
Hotelling-Lawley Trace	1.50128991	1.30	112	183.9	0.0573
Roy's Greatest Root	0.80694488	1.43	56	99	0.0617

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

From the MANOVA results, we can say that there is significant effect of gender, department, and position with employee performance matrices (productivity, satisfaction Rate).

Limitations:

The Employee Performance Metrics datasets have only 200 observation and the dependent variables are not normal. Moreover, after removing extreme values the analyzable observation was 156.

Conclusion:

This project aims to leverage MANOVA test to detect impact factor on employee performance metrics and provide actionable insights. Specifically, constructing MANOVA testing with the help of SAS software to find out significant factor on employee performance metrics. According to the MANOVA output, I found gender, position, department variables have significant effect on employee performance metrics variables (productivity, Satisfaction Rate). In addition, after the Box-Cox transformation, the employee performance metrices variables follow normality assumption. Overall, these results offer a guideline of company to take effective decision for improving employee performance metrics.

Reference:

[1] Srinath, R., & Bawatharini, A. (2023, June). The impact of soft skills training on employee performance and its implications on organisational success [Conference paper]

[2] https://www.kaggle.com/datasets/adityaab1407/employee-productivity-and-satisfaction-hr-data.

[3] Said, N. S. M., Halim, N. W. A., Manaf, S. M. A., & Adenan, N. D. (2022). The impact of training and development on organisational performance. *Jurnal Intelek, 17*(2), Article 18177. https://doi.org/10.24191/ji.v17i2.18177

[4] Upadhyay, R. (2023). The relationship between employee training and development and organisational performance: an Empirical study. *PSYCHOLOGY AND EDUCATION (2018) 55(1): 74-82 ISSN: 1553-6939*, *55*(1). https://doi.org/10.48047/pne.2018.55.1.04

[5] Díaz-García, J. A., & Caro-Lopera, F. J. (2024). Matrix variate p-value in MANOVA. *arXiv*. https://doi.org/10.48550/arXiv.2409.17309

[6] Tonglin Zhang & Baijian Yang (2017) Box–Cox Transformation in Big Data,

Technometrics, 59:2, 189-201, DOI: 10.1080/00401706.2016.1156025

[7] Anderson, T. W. 1982. An introduction to multivariate statistical Analysis. 2nd edition,

John Wiley & Sons, New York.

[8] Shalini, P., & Suvalakshmi, B. (2024). A study on investigating the impact of training programs on employee performance and organizational success. *International Scientific Journal of Engineering and Management, 3*(5). https://doi.org/10.55041/ISJEM01779