SAS Project: Analyzing Employee Attrition and Compensation Patterns

Wethanie Law

Hsiu-Ping Lin

California State University, Los Angeles

CIS 5250 Visual Analytics

Dr. Shilpa Balan

October 31, 2023

Table of Contents

SAS Project: Analyzing Employee Attrition and Compensation Patterns

A. Dataset Source/URL

https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

B. Introduction

Employee attrition and workplace satisfaction are two pivotal aspects in modern organizations that have a profound impact on workforce dynamics and overall performance. The motivation behind this project is driven by the increasing recognition of these factors as significant determinants of business success. As highlighted in recent research by Li et al. (2021), employee turnover is a critical cost for businesses and an essential human capital metric. Despite its relevance, many firms do not publicly disclose this measure. To address this gap, our project delves into the IBM HR Analytics Employee Attrition & Performance dataset to explore the relationships between employee attrition, workplace satisfaction, and various other employee-related variables. Through comprehensive analyses and visualizations, we aim to shed light on the dynamics at play, motivated by the need to understand the implications of attrition and satisfaction on organizational performance.

The business purpose of this project is multifaceted. First and foremost, it seeks to offer actionable insights to organizations, enabling them to formulate evidence-based strategies for addressing employee attrition and enhancing workplace satisfaction. In today's competitive job market, understanding the factors that influence attrition and satisfaction is crucial for retaining top talent and maintaining a productive workforce. The project strives to uncover the specific drivers of attrition and the facets of workplace satisfaction that significantly impact employee well-being and retention, as proposed by De Winne et al. (2018). Additionally, this analysis aims to assist organizations in optimizing their workforce planning, talent management, and human resources policies.

The findings of this project can be employed for various HR practices, including talent acquisition, performance management, compensation strategies, and employee engagement initiatives. By better understanding the relationships between attrition, satisfaction, and other employee variables, organizations can create a more inclusive and equitable work environment, ultimately enhancing their competitiveness and performance in the marketplace.

C. Data Description

Screenshot showing five rows of all columns of the entire dataset.

Enlarged screenshots showing the complete header name.

	A	B	C	D	E	F	G	H	I	J	K
1	Age	Gender	Attrition	BusinessTravel	Department		DistanceFromHome	Education	EducationField	EnvironmentSatisfaction	JobInvolvement	JobLevel
2	41	Female	Yes	Travel_Rarely	Sales		1	Some College	Life Sciences	Medium	High	Junior
3	49	Male	No	Travel_Frequently	Research & Development		9.19	Below College	Life Sciences	High	Medium	Junior
4	37	Male	Yes	Travel_Rarely	Research & Development		2	Some College	Other	Very High	Medium	EntryLevel
5	33	Female	No	Travel_Frequently	Research & Development		3	Master	Life Sciences	Very High	High	EntryLevel

	L	M	N	O	P	Q	R	S	T
1	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	WorkLifeBalance
2	Sales Executive	Very High	Single	5993	Yes	11	Excellent	Low	Bad
3	Research Scientist	Medium	Married	5130	No	23	Outstanding	Very High	Better
4	Laboratory Technician	High	Single	2090	Yes	15	Excellent	Medium	Better
5	Research Scientist	High	Married	2909	Yes	11	Excellent	High	Better

	U	V	W	X	Y	Z	AA
1	NumCompaniesWorked	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
2	8	8	0	6	4	0	5
3	1	10	3	10	7	1	7
4	6	7	3	0	0	0	0
5	1	8	3	8	7	3	0

	Field Name	Description	Example Value
1	Age	Represents the age of the employee	41
2	Gender	Specifies the gender of the employees (Male, Female)	Female
3	Attrition	Indicates whether the employee has left the company (Yes or No)	Yes
4	BusinessTravel	Specifies the frequency of business travel for the employee (Non Travel, Travel Frequently, Travel Rarely)	Travel_Rarely
5	Department	Denotes the department in which the employee works(HR, R&D, Sales)	Sales
6	DistanceFromHome	Represents the distance between an employee's home and workplace	1

	Field Name	Description	Example Value
7	Education	Represents the employee's level of education (1:Below College, 2: College, 3: Bachelor, 4: Master, 5: Doctor)	Below College
8	EducationField	Specifies the field of education or specialization of the employee (HR, Life Sciences, Marketing, Medical Sciences, Others, Technical)	Life Sciences
9	EnvironmentSatisfaction	Represents the level of satisfaction with the work environment (1: Low, 2: Medium, 3: High, 4: Very High)	Medium
10	JobInvolvement	Represents the level of job involvement or engagement (1: Low, 2: Medium, 3: High, 4: Very High)	High
11	JobLevel	The job level (1: EntryLevel, 2:Junior 3:Intermediate, 4:MidLevel, 5:Senior)	Junior
12	JobRole	Specifies the role or position of the employee within the organization (HC Rep, HR, Lab Technician, Manager, Managing Director, Research Director, Research Scientist, Sales Executive, Sales Representative)	Sales Executive
13	JobSatisfaction	Represents the level of satisfaction with the job (1: Low, 2: Medium, 3: High 4: Very High)	Very High
14	MaritalStatus	Indicates the marital status of the employee (Divorced, Married, Single)	Single
15	MonthlyIncome	Represents the monthly salary income of the employee	5993
16	OverTime	Specifies whether the employee gets overtime pay (Yes) or not (No)	Yes
17	PercentSalaryHike	Indicates the percentage salary increase for the previous year	11
18	PerformanceRating	Denotes the employee's performance rating (1: Low, 2: Good, 3: Excellent, 4: Outstanding)	Excellent
19	RelationshipSatisfaction	Represents the level of satisfaction with work-related relationships (1: Low, 2: Medium, 3: High, 4: Very High)	Low
20	WorkLifeBalance	Describes how is work-life balance (1: Bad, 2: Good, 3: Better, 4: Best)	Bad

	Field Name	Description	Example Value
21	NumCompaniesWorked	Indicates the number of companies the employee has worked for	8
22	TotalWorkingYears	Represents the total number of years the employee has worked	8
23	TrainingTimesLastYear	Indicates the number of hours the employee spent on training in the previous year	0
24	YearsAtCompany	Indicates the total number of years the employee has been with the company	6
25	YearsAtCurrentRole	Represents the number of years the employee has been in their current role	4
26	YearsSinceLastPromotion	Denotes the number of years since the employee's last promotion	0
27	YearsWithCurrentManager	Represents the number of years the employee has spent with their current manager	5

D. Data Cleaning

1. Categorical Value Mapping with vlookup

Pre-cleaning

	F	G	H	I	J
1	Education	EducationField	EnvironmentSatisfaction	JobInvolvement	JobLevel
2	2	Life Sciences	2	3	2
3	1	Life Sciences	3	2	2
4	2	Other	4	2	1
5	4	Life Sciences	4	3	1
6	1	Medical	1	3	1
7	2	Life Sciences	4	3	1
8	3	Medical	3	4	1

Post-cleaning

	F	G	H	I	J
1	Education	EducationField	EnvironmentSatisfaction	JobInvolvement	JobLevel
2	College	Life Sciences	Medium	High	Junior
3	Below College	Life Sciences	High	Medium	Junior
4	College	Other	Very High	Medium	EntryLevel
5	Master	Life Sciences	Very High	High	EntryLevel
6	Below College	Medical	Low	High	EntryLevel
7	College	Life Sciences	Very High	High	EntryLevel
8	Bachelor	Medical	High	Very High	EntryLevel

Explanation

In the first data cleaning category, called "Categorical Value Mapping," we used Excel's VLOOKUP function to transform numerical categorical values into more easily interpretable text categories. For instance, consider the "Education" column, which initially featured numeric categories (1, 2, 3, 4, or 5). To interpret the meaning of each number, users would have had to refer to a separate reference table or legend, which could be cumbersome. By converting these numerical codes into descriptive text categories (e.g., changing "1" to "Below College"), we eliminated the necessity for external references. This optimization was aimed at enhancing the dataset's readability and interpretability. As a result, users can promptly comprehend the meaning of categorical values without the need for additional cross-referencing. This enhancement simplifies the subsequent analysis in SAS by offering clear and immediately comprehensible categorical data, thus facilitating more efficient and user-friendly data analysis.

2. Missing Value Imputation

Pre-cleaning

	E	F
1	Department	DistanceFromHome
51	Research & Development	
141	Human Resources	
181	Research & Development	
252	Research & Development	
631	Human Resources	
831	Sales	
920	Sales	
933	Research & Development	
1335	Sales	
1340	Sales	
1375	Research & Development	

	E	F
1	Department	DistanceFromHome
2	Sales	1
3	Research & Development	8
4	Research & Development	2
5	Research & Development	3
6	Research & Development	2
7	Research & Development	2
8	Research & Development	3
9	Research & Development	24
10	Research & Development	23
11	Research & Development	27

original | highlight | keep | final_data

Average: 9.194653873   Count: 1460   Sum: 13415

Post-cleaning

	E	F
1	Department	DistanceFromHome
51	Research & Development	9.19
141	Human Resources	9.19
181	Research & Development	9.19
252	Research & Development	9.19
631	Human Resources	9.19
831	Sales	9.19
920	Sales	9.19
933	Research & Development	9.19
1335	Sales	9.19
1340	Sales	9.19
1375	Research & Development	9.19

Explanation

We focused on "Missing Value Imputation" in the third data cleaning category. After identifying missing values in the "DistanceFromHome" column, we decided to fill these blank cells with the mean of this column. Since this column features numerical values, and the data appears to be missing at random, replacing the blanks with the column's mean ensured that our analysis in SAS could proceed without hindrances. By highlighting the entire "DistanceFromHome" column, we can easily find the mean value, which, in this case, is 9.19. We replaced the blank cells with this calculated mean. Handling missing data is crucial because unaddressed missing values can lead to several issues in statistical analysis. These issues include a reduction in sample size, potential bias, incomplete or inaccurate results, and invalid inferences. Properly managing missing data is essential to ensure our analysis results are reliable, representative, and accurate.

3. Data Splitting

Pre-cleaning

	A	B	C	
1	Age & Gender	Attrition	BusinessTravel	
2	41, Female	Yes	Travel_Rarely	
3	49, Male	No	Travel_Frequently	
4	37, Male	Yes	Travel_Rarely	
5	33, Female	No	Travel_Frequently	
6	27, Male	No	Travel_Rarely	
7	32, Male	No	Travel_Frequently	
8	59, Female	No	Travel_Rarely	

Post-cleaning

	A	B	C	D
1	Age	Gender	Attrition	BusinessTravel
2	41	Female	Yes	Travel_Rarely
3	49	Male	No	Travel_Frequently
4	37	Male	Yes	Travel_Rarely
5	33	Female	No	Travel_Frequently
6	27	Male	No	Travel_Rarely
7	32	Male	No	Travel_Frequently
8	59	Female	No	Travel_Rarely

Explanation

The third and final category for data cleaning that we selected is "data splitting." In this category, we need to separate the combined data encountered within the original dataset. The amalgamation of "Age" and "Gender" information into a singular column hindered our ability to conduct comprehensive analysis using SAS. In order to address this issue, we employed the text manipulation functions available in Excel. We utilized the formula =LEFT(A1, FIND(",", A1)-1) to extract the variable "Age" from the data in column A, resulting in the creation of a separate column specifically for this variable. Furthermore, we used the formula =RIGHT(A1, LEN(A1)-FIND(",", A1) to isolate "Gender" and establish a standalone column for the variable "Gender". The utilization of the separation of variables technique enables us to conduct accurate analysis on the numerical variable "Age" and the categorical variable "Gender" with SAS. This partitioning guarantees a higher level of precision, concentration, and sophistication in the analysis of the data.

E. Analysis & Visualizations

1. Question of Analysis One

What is the average monthly income and average total years worked for each age year?

Average Monthly Income and Average Total Years Worked by Age

Explanation

The line chart provides a clear visual of how both average monthly income and average total years worked change with age. As age increases, there is a consistent upward trend in average monthly income, which aligns with the idea that individuals typically earn more as they gain experience and age. This observation echoes the concept presented in York's article (2019), emphasizing the importance of considering the life cycle of income in discussions about income inequality.

The chart also illustrates an inverted-U shape pattern for income, peaking as individuals near retirement age before gently declining as they retire. Even among those aged 65 and over, above-average incomes persist, reflecting the influence of the baby boomer generation reaching peak earning years. The connection between increasing average monthly income and the gradual rise in average total years worked with age underscores the significance of education and experience in shaping income growth over one's lifetime. This chart offers valuable insights into income dynamics and emphasizes

the need for a more comprehensive understanding of income inequality, considering the
age-related shifts and the evolving financial well-being of individuals.

2. **Question of Analysis Two**

   What is the distribution of monthly income for employees by attrition and gender?
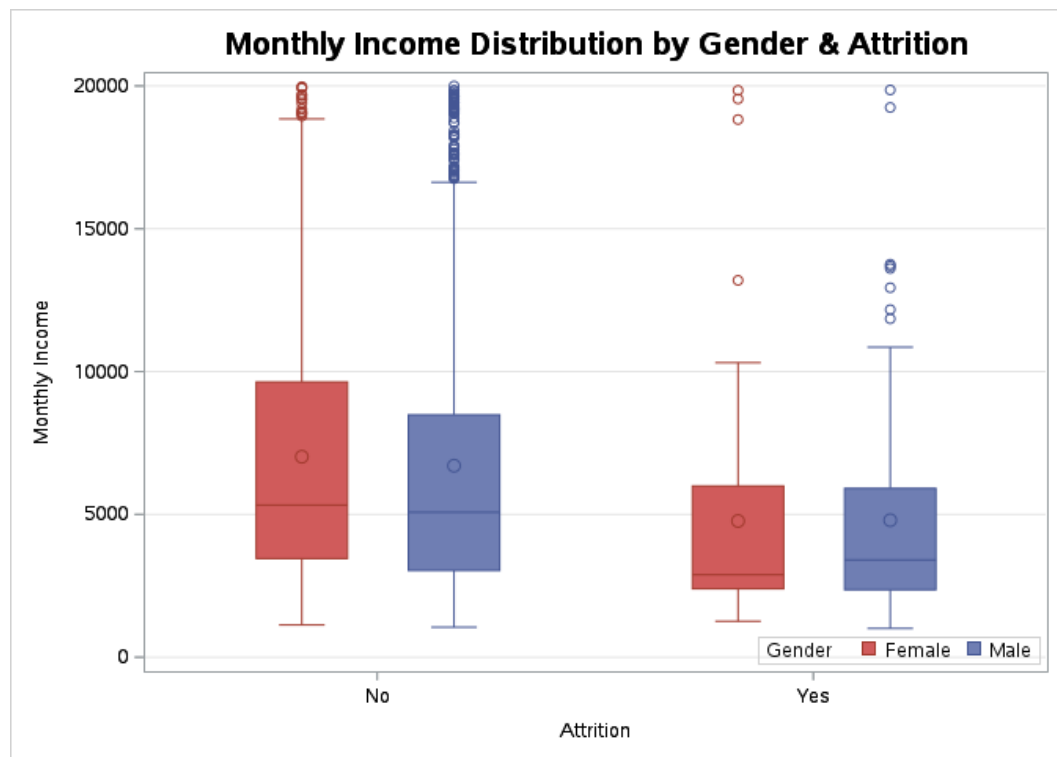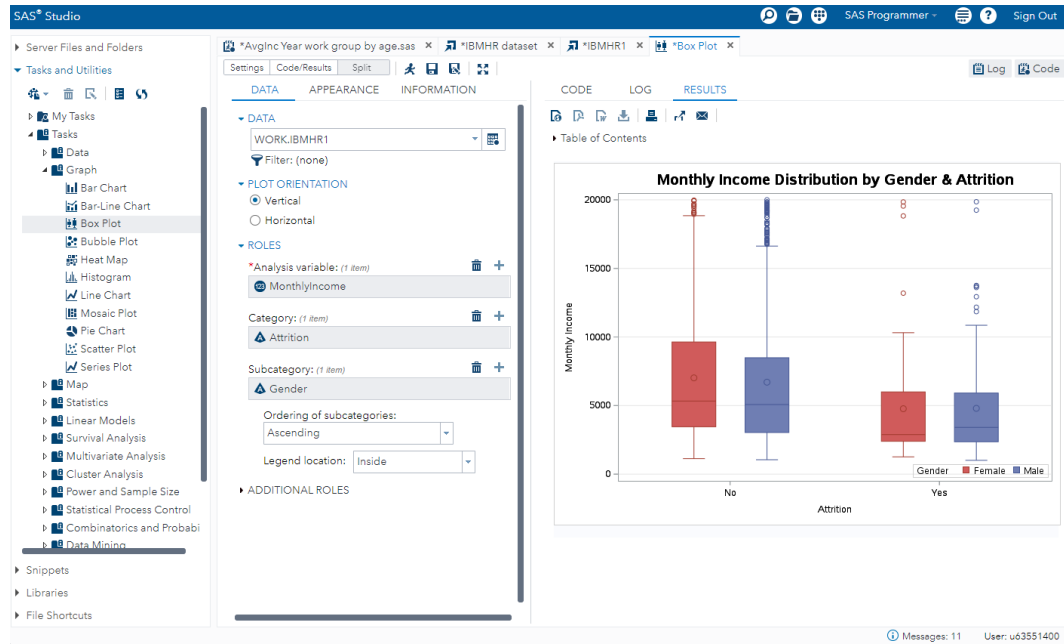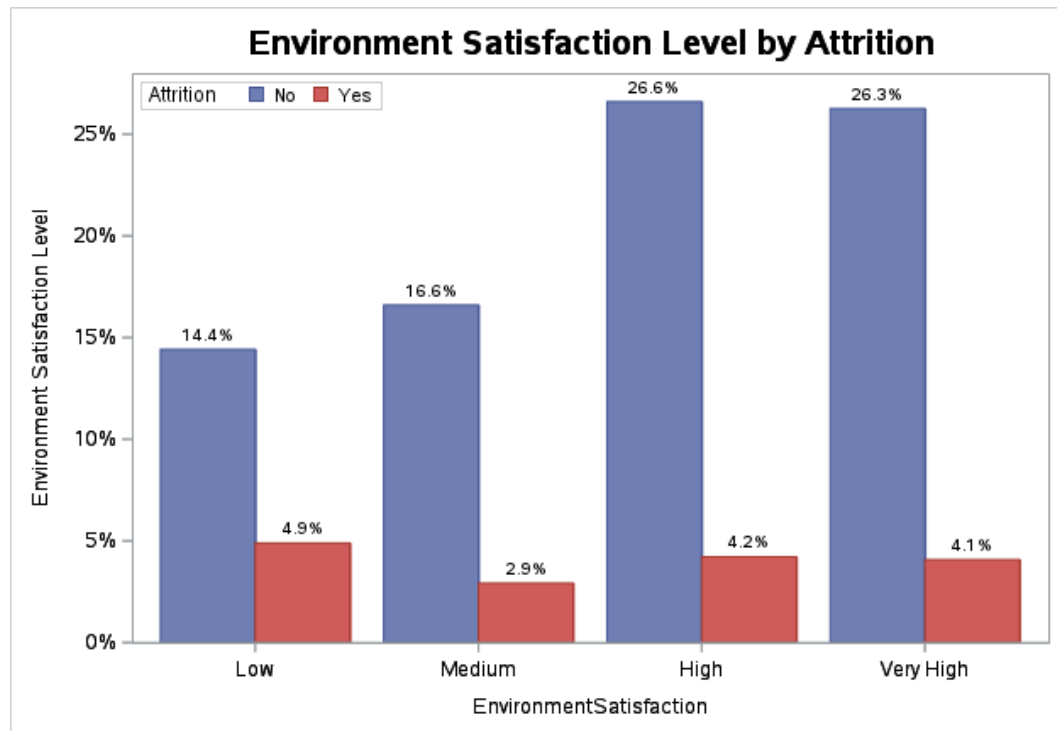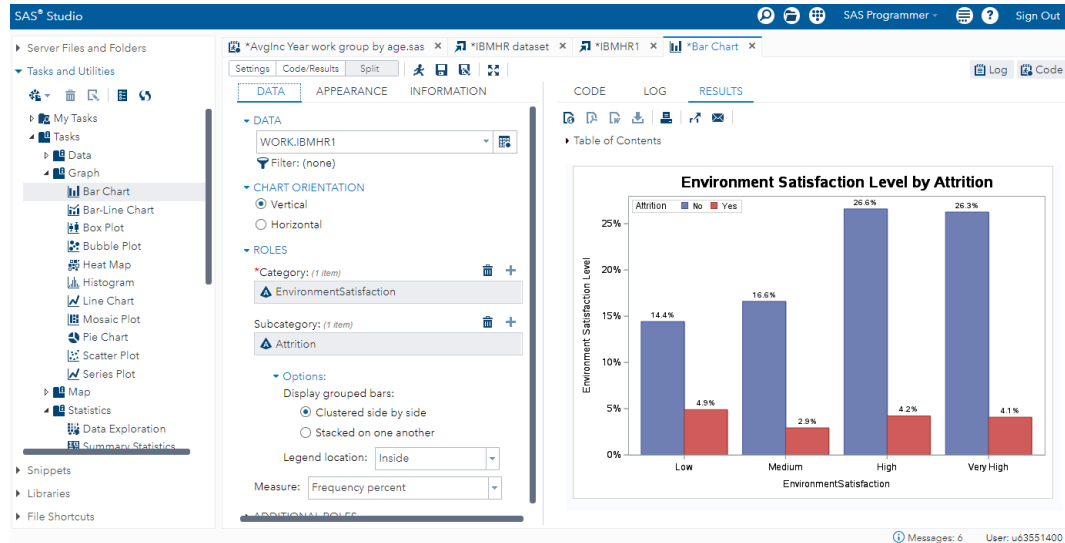
Explanation

The box plot chart above categorizing monthly income data by gender (female and male) and attrition (yes and no), offers intriguing insights into potential relationships between compensation and turnover intention, as explored in the article "The Influence of Organizational Commitment, Compensation, and Employee Engagement on Turnover Intention" by Purnama et al. (2022). The chart reveals clear patterns in monthly income distributions within the organization. Notably, the medians for both male and female employees who have not experienced attrition (Attrition = No) appear relatively higher, implying that employees who remain in their positions might have more consistent and potentially better incomes. Conversely, the lower medians for both male and female employees who have experienced attrition (Attrition = Yes) suggest that those who leave the company may be dissatisfied with their compensation. This is in alignment with the article's findings, which indicate that compensation indeed has a significant effect on turnover intention. The box plot underscores that a higher median income might contribute to reduced turnover intention by enhancing employee satisfaction and commitment, which is consistent with the study's conclusions.

Moreover, the presence of outliers in the box plot, particularly among employees who did not experience attrition, is worth noting. These outliers represent individuals with unusually high incomes, which can indicate that the organization values certain roles more than others or rewards exceptional performance. These findings align with the broader discussion in the article about how different forms of compensation, including both direct and indirect components, can influence employees' intentions to stay or leave. In summary, the box plot provides a visual confirmation of the link between compensation and turnover intention explored in the article, highlighting the importance

of fair and competitive compensation packages in retaining talented employees and preventing attrition in organizations like PT Sriwijaya Air Group.

3. **Question of Analysis Three**

What is the percentage of each environment satisfaction level group by attrition?

Explanation

The presented bar chart provides valuable insights into the relationship between employees' environment satisfaction levels and their attrition status. It segregates employees into two groups: those who have not experienced attrition (No) and those who have experienced attrition (Yes). The chart delineates four levels of environment satisfaction: Low, Medium, High, and Very High. It is evident from the chart that a notable percentage of employees who have not experienced attrition report "High" and "Very High" satisfaction levels, constituting 27% and 26%, respectively. In contrast, a mere 5% of this group expresses "Low" satisfaction. This data suggests that employees who choose to stay with the organization tend to exhibit higher levels of satisfaction with their work environment. Conversely, employees who have experienced attrition display a less favorable distribution of satisfaction levels, with only 4% in both the "High" and "Very High" categories, while 17% and 3% report "Medium" and "Low" satisfaction levels, respectively. This stark contrast highlights that employees who decide to leave the organization often exhibit lower levels of environment satisfaction.

The insights derived from this bar chart resonate with the findings of the earlier research article by Kurniawaty et al. (2019), which delved into the factors impacting employee turnover intention, including work environment and job satisfaction. The chart underscores that higher levels of environment satisfaction are linked with lower attrition rates, reinforcing the premise that an improved work environment may hold the potential to reduce employee turnover. This alignment underscores the importance for organizations, such as Mandiri Bank, to prioritize the enhancement of their work environment as a strategy to retain valuable employees and more effectively address challenges associated with turnover (Kurniawaty et al., 2019).

F. Statistical Summary

1. Analysis Variable: Age

Analysis Variable : Age				
Mean	**Std Dev**	**Minimum**	**Maximum**	**N**
36.9238095	9.1353735	18.0000000	60.0000000	1470

Statistical Measure	Explanation
Mean (Average) 36.9238095	The average age of employees in the dataset is approximately 36.92. It indicates the central tendency of ages, and it can serve as a reference point to understand the typical age of employees in the organization. The mean age suggests that the majority of employees fall around the late 30s, which might indicate a relatively mature workforce.
Std Dev (Standard Deviation) 9.1353735	The standard deviation for age is approximately 9.14. It measures the dispersion or spread of the age values. A higher standard deviation indicates more variability in ages, while a lower value suggests ages are more tightly clustered around the mean. A standard deviation of 9.14 suggests that the ages of employees vary moderately. Some employees may be significantly younger or older than the mean age.
Minimum 18.0000000	The minimum age in the dataset is 18 years. This is the youngest age among all the employees in the dataset. The presence of employees as young as 18 suggests that the organization employs individuals who may be just starting their careers. This diversity in age groups can be beneficial for knowledge transfer and fresh perspectives within the workforce.
Maximum 60.0000000	The maximum age in the dataset is 60 years. This is the oldest age among all the employees in the dataset. The existence of employees up to the age of 60 indicates that the organization values experienced and seasoned professionals. It's important to consider the needs and expectations of senior employees, including potential retirement planning and career progression.
N (Sample Size) 1470	The sample size for the "Age" variable is 1470, meaning there are 1470 data points for age in the dataset. With a this amount of sample size, the age statistics are likely to be representative of the entire employee population.

2. Analysis Variable: Monthly Income

Analysis Variable : MonthlyIncome				
Mean	Std Dev	Minimum	Maximum	N
6502.93	4707.96	1009.00	19999.00	1470

Statistical Measure	Explanation
Mean (Average) 6502.93	The mean monthly income of employees in the dataset is approximately $6,502.93. This number represents the typical or average income for employees in the organization. It suggests that the average employee's monthly income is around $6,502.93.The mean income serves as a reference point for understanding the central income level in the organization. It can be used for benchmarking and setting salary structures.
Std Dev (Standard Deviation) 4707.96	The standard deviation for monthly income is approximately $4,707.96. It measures the spread or variability in income levels. A higher standard deviation implies more income variation among employees. A standard deviation of $4,707.96 indicates significant income diversity within the organization. This variation could be due to differences in roles, experience, or other factors.
Minimum 1009.00	The minimum monthly income in the dataset is $1,009.00. This number is the lowest reported income among all employees.The existence of employees with a minimum income of $1,009.00 suggests that there are individuals in the organization with relatively lower incomes.
Maximum 19999.00	The maximum monthly income in the dataset is $19,999.00. This is the highest reported income among all employees. The presence of employees with a maximum income of $19,999.00 indicates that the organization employs individuals with higher incomes, possibly in senior or specialized roles.
N (Sample Size) 1470	The sample size for the "MonthlyIncome" variable is 1470, meaning there are 1470 data points for monthly income in the dataset. With a sample size of 1470, the income statistics are based on a substantial dataset, providing a reliable basis for compensation decisions and salary-related policies.

G. Statistical Tests

One-way Frequency Analysis

Analysis Variable: Age

Age	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
18	8	0.54	8	0.54	40	57	3.88	1005	68.37
19	9	0.61	17	1.16	41	40	2.72	1045	71.09
20	11	0.75	28	1.90	42	46	3.13	1091	74.22
21	13	0.88	41	2.79	43	32	2.18	1123	76.39
22	16	1.09	57	3.88	44	33	2.24	1156	78.64
23	14	0.95	71	4.83	45	41	2.79	1197	81.43
24	26	1.77	97	6.60	46	33	2.24	1230	83.67
25	26	1.77	123	8.37	47	24	1.63	1254	85.31
26	39	2.65	162	11.02	48	19	1.29	1273	86.60
27	48	3.27	210	14.29	49	24	1.63	1297	88.23
28	48	3.27	258	17.55	50	30	2.04	1327	90.27
29	68	4.63	326	22.18	51	19	1.29	1346	91.56
30	60	4.08	386	26.26	52	18	1.22	1364	92.79
31	69	4.69	455	30.95	53	19	1.29	1383	94.08
32	61	4.15	516	35.10	54	18	1.22	1401	95.31
33	58	3.95	574	39.05	55	22	1.50	1423	96.80
34	77	5.24	651	44.29	56	14	0.95	1437	97.76
35	78	5.31	729	49.59	57	4	0.27	1441	98.03
36	69	4.69	798	54.29	58	14	0.95	1455	98.98
37	50	3.40	848	57.69	59	10	0.68	1465	99.66
38	58	3.95	906	61.63	60	5	0.34	1470	100.00
39	42	2.86	948	64.49					

Explanation

The one-way frequency table for the "Age" variable offers valuable insights into the age distribution of employees within the organization. It reveals that the most common age groups among employees are concentrated in their late 20s to early 40s, with a peak between the ages of 29 to 34. This observation suggests that a significant portion of the workforce falls within this age range. The distribution appears balanced across different age groups, indicating inclusivity in terms of age representation. Moreover, the cumulative frequency demonstrates a gradual increase in the number of

employees as age advances, with around 99.66% of employees aged 60 or below. These findings align with the summary statistics of the "Age" variable, including the mean age of approximately 36.92 and the standard deviation of around 9.14, confirming that the dataset exhibits diversity in age, with a central tendency in the late 20s to early 40s.

Overall, the frequency table offers a comprehensive overview of the age demographics within the organization and complements the statistical summary, analysis variable, "Age". These insights are invaluable for HR planning, talent management, and the development of age-inclusive workplace policies. The organization can use this information to tailor HR strategies, benefits, and career development programs to accommodate the diverse needs and expectations of employees at different stages of their careers, thus fostering an inclusive and equitable work environment.

Correlation Analysis

Analysis Variable: Monthly Income and Total Working Years

1 With Variables:	TotalWorkingYears
1 Variables:	MonthlyIncome

Pearson Correlation Coefficients, N = 1470	
	MonthlyIncome
TotalWorkingYears	0.77289

Scatter Plot
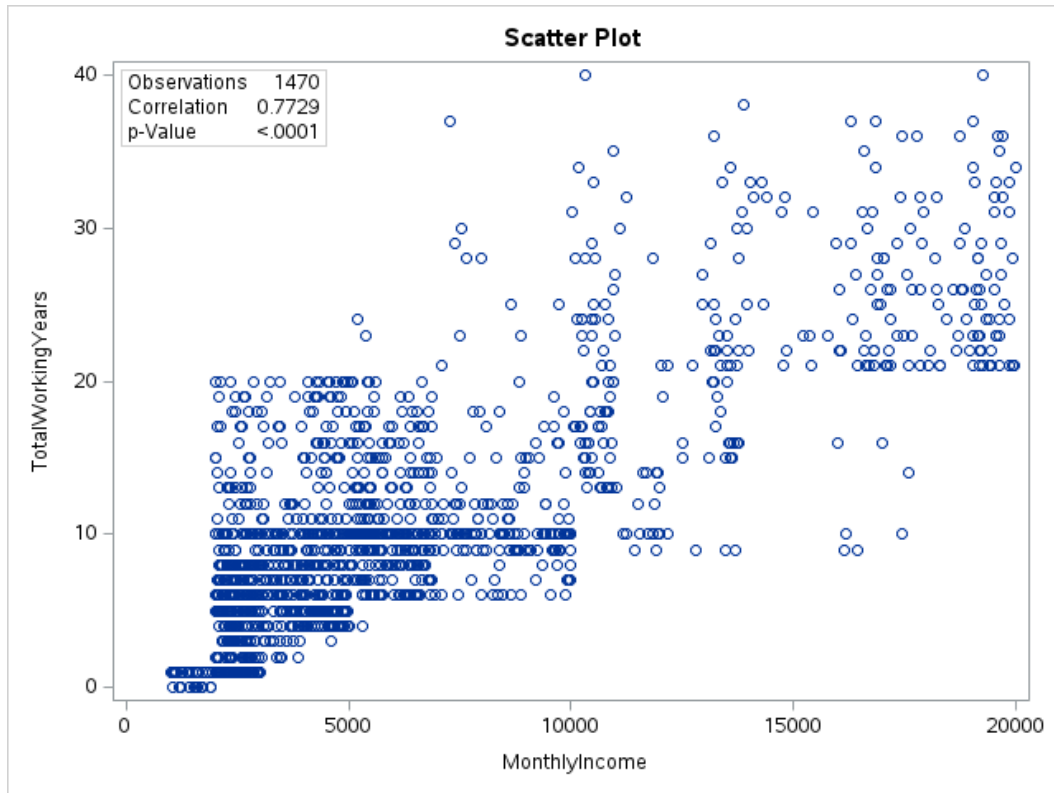
Explanation

The correlation analysis between "TotalWorkingYears" and "MonthlyIncome" reveals a substantial and positive relationship between these two key variables, with a Pearson correlation coefficient (r) of 0.77289. This high correlation coefficient suggests that as employees accumulate more years of working experience, their monthly income tends to increase proportionally. In other words, the longer an employee has been in the workforce, the higher their monthly earnings tend to be. This insight is consistent with the initial line chart created for the first analysis question, which depicted a nearly parallel trend between average total working years and average monthly income. The strong positive correlation not only confirms the visual observation but also quantifies the strength of this relationship, underlining the significance of experience in shaping an employee's income trajectory.

The robust correlation coefficient of 0.77289 signifies the relevance of an employee's tenure in the workforce to their financial compensation, and it aligns with the graphical representation in the line chart, providing a numerical validation of the observed trend. This finding carries practical implications for both employees and organizations, as it emphasizes the importance of recognizing and rewarding experience in the workplace. As employees accumulate more years of service, organizations may need to consider adjusting compensation structures and opportunities for career growth and development. On an individual level, employees can use this insight to make informed decisions about their career paths and potential income trajectories based on their total working years, ultimately contributing to better financial planning and professional development.

T-Test Analysis

Analysis Variable: Monthly Income and Attrition

Attrition	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
No		1233	6832.7	4818.2	137.2	1051.0	19999.0
Yes		237	4787.1	3640.2	236.5	1009.0	19859.0
Diff (1-2)	Pooled		2045.6	4649.0	329.7		
Diff (1-2)	Satterthwaite		2045.6		273.4		

Attrition	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		6832.7	6563.5	7101.9	4818.2	4635.3	5016.3
Yes		4787.1	4321.3	5252.9	3640.2	3339.3	4001.1
Diff (1-2)	Pooled	2045.6	1398.8	2692.4	4649.0	4486.8	4823.5
Diff (1-2)	Satterthwaite	2045.6	1508.2	2583.1			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1468	6.20	<.0001
Satterthwaite	Unequal	412.74	7.48	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	1232	236	1.75	<.0001

Explanation

        We conducted a t-test analysis to investigate whether there exists a significant difference in monthly income between the two groups in the variable attrition, Yes and No. Our analysis was guided by the following null and alternate hypotheses:

        Null Hypothesis (H_0):

There is no difference in monthly income between the two attrition groups (Yes and No).

        Alternative Hypothesis (H_a):

There is a difference in monthly income between the two attrition groups (Yes and No).

        Upon examining the results, we found that the p-value for our analysis is <0.0001 which is much less than the commonly used alpha value of 0.05. As a result, we rejected the null hypothesis, H_0. This rejection indicates that there is a statistically significant difference in monthly income between employees who have experienced attrition (Yes) and those who have not (No). This finding aligns with the trends observed in the box plot chart created for questions of analysis two above.

        The t-test results provide statistical evidence supporting the existence of variations in monthly income between the two attrition groups, underscoring the importance of considering attrition as a factor when analyzing income disparities within the organization.

References

De Winne, S., Marescaux, E., Sels, L., Beveren, I. V., & Vanormelingen, S. (2019). The impact of employee turnover and turnover volatility on labor productivity: a flexible non-linear approach. *The International Journal of Human Resource Management, 30*(21), pp. 3049-3079. doi:10.1080/09585192.2018.1449129

Kurniawaty, K., Ramly, M., & Ramlawati, R. (2019). The effect of work environment, stress, and job satisfaction on employee turnover intention. *Management Science Letters, 9*(6), pp. 877-886. doi:10.5267/j.msl.2019.3.001

Li, Q., Lourie, B., Nekrasove, A., & Shevlin, T. (2022). Employee Turnover and Firm Performance: Large-Sample Archival Evidence. *Management Science, 68*(8), pp. 5557-6354. Retrieved October 28, 2023, from https://doi.org/10.1287/mnsc.2021.4199

Purnama, E., Nugroho, A., Widayati, C. C., Widjaja, P. H., & Marsita, Y. (2022, April 07). The Influence of Organizational Commitment, Compensation and Employee Engagement on Turnover Intention. *Calitatea, 23(190)*, pp. 362–370. Retrieved October 23, 2023, from https://doi.org/10.47750/QAS/23.190.38

York, E. (2019, March 21). *Average Income Tends to Rise with Age*. Retrieved October 22, 2023, from The Tax Foundation Web site: https://taxfoundation.org/data/all/federal/average-income-age/