With the provided dataset, I first produce a table that identifies and justifies the data type (nominal, ordinal, interval, or ratio) for each variable; the table will present relevant summary statistics (number of observations, mean, median, mode, minimum, maximum and standard deviation) as per the data type for all variables presented.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data Field** | **Data Type** | **Justification** | **Summary Statistics** | |
| EmpID | Nominal | EmpID is a unique identifier and label for each employee. It does not possess any numerical meaning or ranking. Hence, it is considered Nominal. | Number of observations | 1480 |
| Mean | - |
| Median | - |
| Mode | NA |
| Minimum | - |
| Maximum | - |
| Standard Deviation | - |
| Age | Ratio | Age is a continuous numerical variable possessing a true zero which allows meaningful comparisons. Hence, it is considered as Ratio. | Number of observations | 1480 |
| Mean | 36.92 |
| Median | 36 |
| Mode | 35 |
| Minimum | 18 |
| Maximum | 60 |
| Standard Deviation | 9.13 |
| Gender | Nominal | Gender consists of categories such as Female and Male. There is no inherent order or ranking between these two categories. Hence, it is considered Nominal. | Number of observations | 1466 |
| Mean | - |
| Median | - |
| Mode | Male |
| Minimum | - |
| Maximum | - |
| Standard Deviation | - |
| Department | Nominal | Department is a categorical variable, and it does not possess meaningful numerical order amongst the departments Hence, it is considered Nominal. | Number of observations | 1480 |
| Mean | - |
| Median | - |
| Mode | Research & Development |
| Minimum | - |
| Maximum | - |
| Standard Deviation | - |
| Distance From Home | Ratio | Distance from home is a continuous numerical variable that possess a true zero, allowing meaningful ratio comparisons. Hence, it is considered as Ratio. | Number of observations | 1480 |
| Mean | 9.22 |
| Median | 7 |
| Mode | 2 |
| Minimum | 1 |
| Maximum | 29 |
| Standard Deviation | 8.13 |
| Monthly Income | Ratio | Monthly income is a continuous numerical variable that possess a true zero, allowing meaningful ratio comparisons. Hence, it is considered as Ratio. | Number of observations | 1480 |
| Mean | 6571.90 |
| Median | 4933 |
| Mode | 6142 |
| Minimum | 1009 |
| Maximum | 70000 |
| Standard Deviation | 5216.21 |
| Total Working Years | Ratio | The total working years can represent the number of years an employee has worked, stemming from zero. Thus, it possesses a true zero allowing meaningful ratio comparisons. Hence, it is considered as Ratio. | Number of observations | 1480 |
| Mean | 11.26 |
| Median | 10 |
| Mode | 10 |
| Minimum | -8 |
| Maximum | 40 |
| Standard Deviation | 7.81 |
| Training Times Last Year | Ratio | Training sessions attended are numeric and possess a true zero value which allows meaningful comparisons. Hence, it is considered as Ratio. | Number of observations | 1480 |
| Mean | 2.80 |
| Median | 3 |
| Mode | 2 |
| Minimum | 0 |
| Maximum | 6 |
| Standard Deviation | 1.29 |
| Environment Satisfaction | Ordinal | Environment satisfaction is measured on a scale ranging from 1 to 4, with each value representing a distinct level of satisfaction towards workplace environment. The values are sorted meaningfully with each number bearing a significant difference in satisfaction level. Hence, it is considered as Ordinal. | Number of observations | 1480 |
| Mean | 2.72 |
| Median | 3 |
| Mode | 3 |
| Minimum | 1 |
| Maximum | 4 |
| Standard Deviation | 1.09 |
| Relationship Satisfaction | Ordinal | Relationship satisfaction is measured on a scale ranging from 1 to 4, with each value representing a distinct level of satisfaction in their working relationship with peers. The values are sorted meaningfully with each number bearing a significant difference in level of relationship satisfaction. Hence, it is considered as Ordinal. | Number of observations | 1480 |
| Mean | 2.71 |
| Median | 3 |
| Mode | 3 |
| Minimum | 1 |
| Maximum | 4 |
| Standard Deviation | 1.08 |
| Job Involvement | Ordinal | Job involvement is measured on a scale ranging from 1 to 4, with each value representing a distinct level of satisfaction in their jobs. The values are sorted meaningfully with each number bearing a significant difference in satisfaction level. Hence, it is considered as Ordinal. | Number of observations | 1480 |
| Mean | 2.73 |
| Median | 3 |
| Mode | 3 |
| Minimum | 1 |
| Maximum | 4 |
| Standard Deviation | 0.71 |
| Work Life Balance | Ordinal | Work life balance is measured on a scale ranging from 1 to 4, with each value representing a distinct level of satisfaction that employees have towards their work life balance. The values are sorted meaningfully with each number bearing a significant difference in satisfaction level. Hence, it is considered as Ordinal. | Number of observations | 1480 |
| Mean | 2.76 |
| Median | 3 |
| Mode | 3 |
| Minimum | 1 |
| Maximum | 4 |
| Standard Deviation | 0.71 |
| Job Satisfaction | Ordinal | Job satisfaction is measured on a scale ranging from 1 to 4, with each value representing a distinct level of employee satisfaction towards their job. The values are sorted meaningfully with each number bearing a significant difference in satisfaction level. Hence, it is considered as Ordinal. | Number of observations | 1480 |
| Mean | 2.73 |
| Median | 3 |
| Mode | 4 |
| Minimum | 1 |
| Maximum | 4 |
| Standard Deviation | 1.10 |
| Attrition | Nominal | Attrition consists of categories such as Yes and No. There is no inherent order or ranking between these two categories. Hence, it is considered Nominal. | Number of observations | 1480 |
| Mean | - |
| Median | - |
| Mode | No |
| Minimum | - |
| Maximum | - |
| Standard Deviation | - |

Inspect the dataset for quality issues and prepare the data using appropriate treatment methods. The data quality issues identified, and their respective applied treatment will be displayed with the aid of screenshots.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.1a: Conditional Formatting*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.1b: Duplicate EmpID*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.1c: After Treatment*

I utilised conditional formatting, enabled highlight cell rules containing duplicates in the EmpID column and filter to sort by colour. The 10 duplicated EmpID were removed.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.2a: Blank entries in Gender*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.2b: After treatment*

Using filter, select “(Blanks)” to filter and identify blanks in Gender column. I identified 14 records with blanks (Figure 1.2a). I manually replace blanks with the mode, “Male” for these records.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.3a: Inconsistency in Department names*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.3 b: Research & Development (After treatment)*

A screenshot of a computer

AI-generated content may be incorrect.

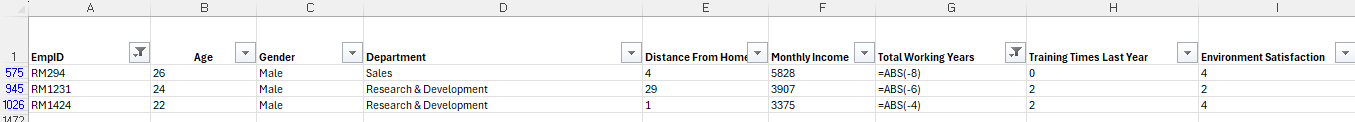
*Figure 1.3c: Human Resources (After treatment)*

Using filter; Department column has inconsistency in naming such as 2 entries being abbreviated as “HR” and “R&D” instead of “Human Resources” and “Research & Development” respectively (Figure 1.3a). I used the find and replace function to replace all abbreviated entries of with their respective full department name of “Human Resources” and “Research & Development” (Figure 1.3b, 1.3c)

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.4a: Inconsistent Negative values in Total Woking Years Column*



*Figure 1.4b: ABS Formula application*

A screenshot of a spreadsheet

AI-generated content may be incorrect.

*Figure 1.4c: After treatment*

Using filter; 3 records contain negative values which instead should be positive (Figure 1.4a). ABS Formula is applied to the three columns to display positive numbers (Figure 1.4b, 1.4c).

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.5a: Inconsistent decimal place values in Training Times Last Year*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.5b: Format cells*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.5c: After treatment*

Using filter; 2 entries in Training Time Last Year Column (Figure 1.5a) consist of decimal points instead of whole numbers (Figure 1.5b). Using format cells function for the whole column, columns were formatted to only display numbers in 0 decimal places (Figure 1.5c).

A screenshot of a spreadsheet

AI-generated content may be incorrect.

*Figure 1.6a: Income sorted from high to low*

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 1.6b: Box plot and summary stats*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 1.6c: After Treatment*

Using filter, I sorted income from highest to lowest (Figure 1.6a). In another column, I applied logarithmic transformation on each income value. The values were displayed via box plot, the second highest income value was observed solely displayed as an outlier while the highest income value was absent (Figure 1.6b). I assumed that two highest income entries ,$70000 and $69000 were outliers and removed them (Figure 1.6c).

**With the treated dataset, four charts will be developed and visualize using Tableau to explain the key patterns found in the data.**

A pie chart with text

AI-generated content may be incorrect.

*Figure 2.1: Organization total attrition rate by department*

The pie chart displays the employee’s attrition percentage by department, insights observed are that Research & Development has the highest attrition rate, 56.12%, followed by Sales with 38.82% and Human resources with the least attrition rate of 5.06%.

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 2.2: Attrition rate across all departments*

The stacked bar chart showcases both the attrition rate and number of employees that were retained and have departed from the organization. Insights derived is Sales has the highest attrition rate of 21%, translating to 1 in 5 employees in this department left the company. It is higher than Research & Development,14% and Human Resource with the lowest attrition rate of 19%. This suggests that driving factors could be slow monthly income progression and poor work relationships with peers within departments.

A graph of different colored bars

AI-generated content may be incorrect.

*Figure 2.3a: Average relationship satisfaction score across departments*

The horizontal bar chart here displays the average relationships satisfaction score of departed and retained employees across each department. It is observed that retained employees in Research & Development and Sales had a higher average relationship satisfaction score of 2.74 and 2.71 respectively as compared to those who departed which is 2.52 and 2.64 respectively. This suggests that relationship satisfaction plays a significant role towards employee attrition rate in these two departments.

A graph with different colored rectangles

AI-generated content may be incorrect.

*Figure 2.3b: Average relationship satisfaction scores across departments (Attrition: Yes) after Parameter)*

The Parameter enables the Reference line which indicates the average relationship satisfaction for (Attrition: No) employees, to view comparison between their average scores.

A graph with blue and orange lines

AI-generated content may be incorrect.

*Figure 2.4: Relationship between Total Working Years and Monthly Income for employees*

The scatter plot showcases that employees with more working years generally earn a higher monthly income. The reference line average monthly income of past and present employees is $6253 across all departments.

A graph with orange and green lines

AI-generated content may be incorrect.

*Figure 2.5: Relationship between Total Working Years and Monthly Income for employees who left the company (Attrition: Yes) after parameter*

Parameters are adopted to analyse attrition trend by department with reference line indicating respective department expected average monthly income. The reference line indicates that employees in Research & Development earning below the average monthly income of $4108 would depart from the company, suggesting that income is a key attrition factor.

**With the four charts developed, they will be integrated into a business performance dashboard. The dashboard will have advanced dashboard navigation functions such as filtering, highlighting and navigating to an external website link to assist users**

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 3.2: Department Attrition Analysis: Relationship Satisfaction Dashboard, Income, and Working Experience after highlighting action*

The highlight action brings the viewer’s attention to department relevant metrics of the dashboard.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 3.3: Department Attrition Analysis: Relationship Satisfaction, Income, and Working Experience Dashboard after filter action*

The filter action enables viewers to view respective charts insights by selecting department from the pie chart which reflects attrition rate across the various departments.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 3.5: Navigating to an external HTTP Link using Benchmark Salaries and cultures on Glassdoor*

*A person looking at something

AI-generated content may be incorrect.*

*Figure 3.5: Salary and Company Benchmark Search External Link* (Glassdoor, n.d.)

A screenshot of a chart

AI-generated content may be incorrect.

*Figure 3.8: A Company Reviews by category* (Glassdoor, n.d.)

A job-recruitment platform listing jobs and companies, offering salaries based on industries, work experience and company culture metrics. Aiding in benchmark standards towards improving workplace dynamics and reduce organizational attrition with revised and informed strategies. (Figure 3.7, 3.8)