7/16/2023

Employee Attrition in Marvelous Construction

Project Final Report

NAME : KOBINATH AMIRTHALINGAM
INDEX NO : 200308F

Problem Overview

The CEO of Marvelous Construction has hired me as a data scientist to analyze the data within the company and understand the situation regarding employee attrition. The company has been experiencing a high number of employee resignations, which is concerning. My task is to analyze the provided dataset, which includes employee details, attendance, leaves, and salary information extracted from the company's ERP system.

My goal as a data scientist is to analyze this data to gain insights into the factors contributing to employee attrition within Marvelous Construction. This analysis will help the CEO and the Human Resources department understand the situation better and make informed decisions to address the issue of employee resignations.

Dataset description

Dataset has 997non-null entries and 19 data variables.

Data variable	Data Type	Dtype
Employee_No	Numerical discrete	int64
Employee_Code	Numerical discrete	int64
Name	Categorical nominal	object
Title	Categorical nominal	object
Gender	Categorical nominal	object
Religion_ID	Numerical discrete	int64
Marital_Status	Categorical nominal	object
Designation_ID	Numerical discrete	int64
Date_Joined	Categorical ordinal	object
Date_Resigned	Categorical ordinal	object
Status	Categorical nominal	object
Inactive_Date	Categorical ordinal	object
Reporting_emp_1	Numerical discrete	object
Reporting_emp_2	Numerical discrete	Object
Employment_Category	Categorical nominal	Object
Employment_Type	Categorical nominal	Object
Religion	Categorical nominal	object
Designation	Categorical nominal	Object
Year_of_Birth	Numerical continuous	int64

```
RangeIndex: 997 entries, 0 to 996
Data columns (total 19 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Employee_No         997 non-null    int64
 1   Employee_Code       997 non-null    int64
 2   Name                997 non-null    object
 3   Title               997 non-null    object
 4   Gender              997 non-null    object
 5   Religion_ID         997 non-null    int64
 6   Marital_Status      997 non-null    object
 7   Designation_ID      997 non-null    int64
 8   Date_Joined         997 non-null    object
 9   Date_Resigned       997 non-null    object
 10  Status              997 non-null    object
 11  Inactive_Date       997 non-null    object
 12  Reporting_emp_1     997 non-null    object
 13  Reporting_emp_2     997 non-null    object
 14  Employment_Category 997 non-null    object
 15  Employment_Type     997 non-null    object
 16  Religion            997 non-null    object
 17  Designation         997 non-null    object
 18  Year_of_Birth       997 non-null    int64
dtypes: int64(5), object(14)
memory usage: 148.1+ KB
```

Data pre-processing

Data Cleaning

- When I checked for null values, I found that the column named "Reporting_emp_2" has null values for all rows. Therefore, I dropped that column.

- After discovering incorrect values in the "Year_of_Birth" column, I proceeded to convert non-numeric values to NaN. Then, I calculated the mode of the column and replaced the NaN values with the mode value to ensure consistency and accuracy in the data.

- I converted the data type of the "Year_of_Birth" column to integer values and the "Date_Joined" column to datetime objects.

- I fill missing "Marital_Status" values within each group using the mode of the respective group, and then check for any remaining null values.

- Finally, I perform various checks and modifications on the "Date_Resigned" and "Inactive_Date" columns based on specific conditions to handle missing or incorrect values appropriately. After completing the necessary operations, I drop the "Inactive_Date" column because the data in "Date_Resigned" and "Inactive_Date" are the same.

Insights from data analysis

I have found some useful insights from the data analysis.

➤ The table indicates a higher number of inactive employees among those who are married compared to those who are single. This could suggest that married employees may be more likely to leave the company. The company should analyze the reasons behind this trend and consider implementing strategies to improve employee retention for married individuals.


Contingency Table: Marital_Status vs Status
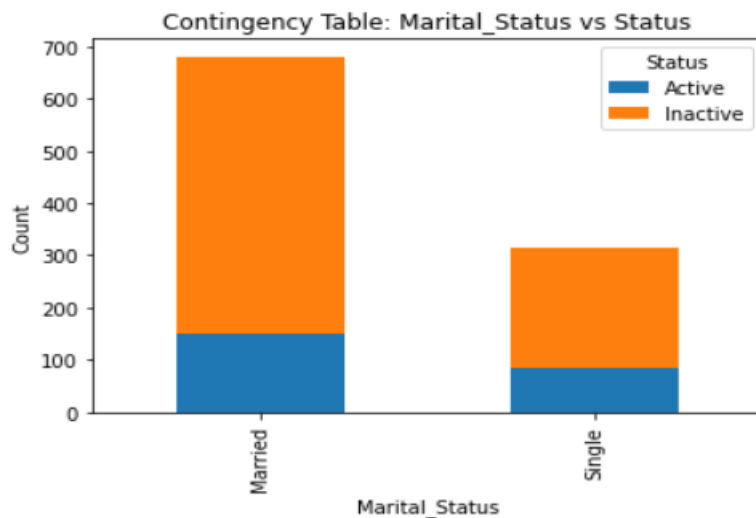
Status	Active	Inactive
Marital_Status		
Married	150	531
Single	83	233

➤ The graph shows a rising trend in employee inactivity for individuals born between 1990 and 1995. This could imply that employees from this particular age group are more likely to leave or become inactive within the company.


Status by Year of Birth
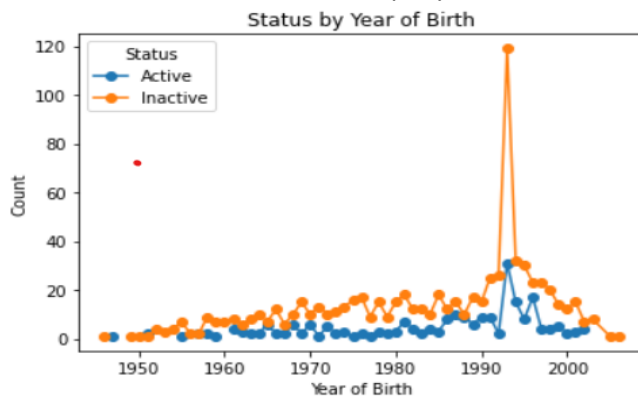
➤ The contingency table shows the count of employees based on their "Status" (Active or Inactive) and "Employment_Type" (Contract Basis or Permanent).
The high number of inactive employees under a contract basis (754) may indicate a potential issue. It's important for the company to investigate the reasons behind this high rate of

inactivity. Possible reasons could include dissatisfaction with the contract terms, lack of job security, or limited career growth opportunities. The company should conduct employee surveys or interviews to understand the underlying causes and address any issues that may be leading to employee disengagement.

Employee Status by Employment Type
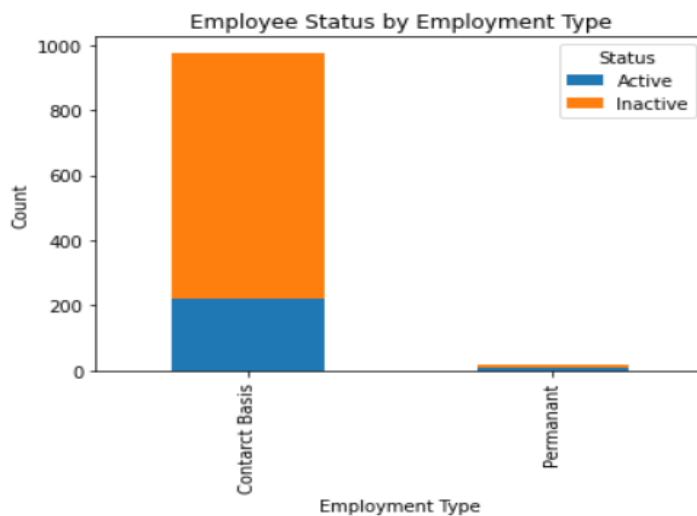
```
Status          Active   Inactive
Employment_Type
Contarct Basis     223        754
Permanant           10         10
```

➢ Based on the bar chart, it is observed that a significant number of employees in the "Labour" employment category are inactive. This finding raises concerns about the engagement and retention of labour employees within the company.

Percentage of Active and Inactive Employees by Employment_Category
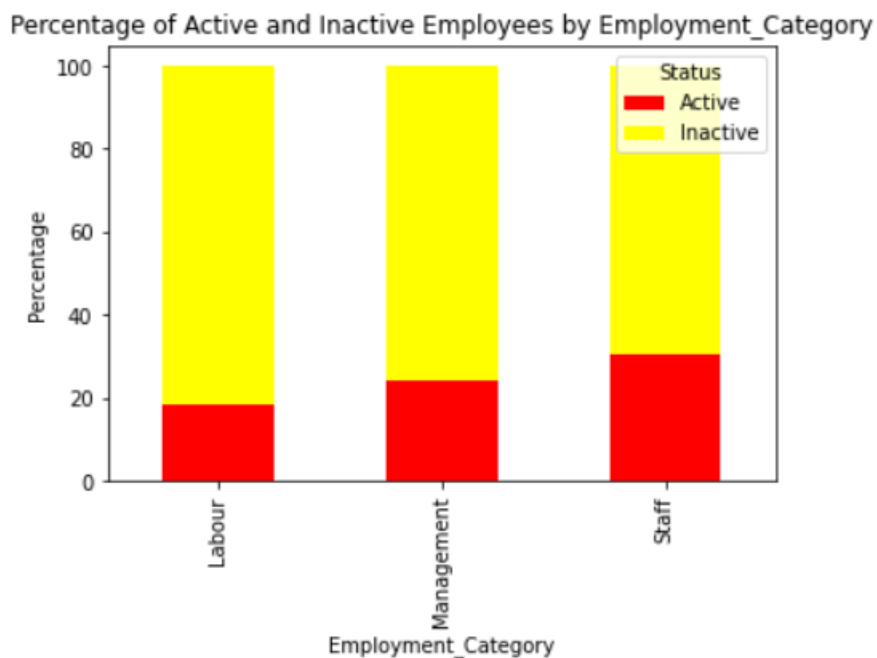
➤ Based on the data provided, it is evident that a larger proportion of male employees are inactive compared to female employees. This gender disparity in the inactive status raises concerns and may indicate potential challenges or issues specific to male employees within the company.

Percentage of Active and Inactive Employees by Employment_Category