

Final Project Report

SIRIWARDANA H B K S

200607V

Problem overview

Analyse and derive valuable insights that would be useful for the Marvelous Construction Company to make strategic decisions to improve retention using the given employee details, attendance, leaves, and salary extracted from the ERP of Marvelous Construction

Dataset Description

1. Attendance: Contains employee attendance, check-in/check-out times, shift details ,work hours .
2. Employees: including personal details, job-related attributes, and employment history.
3. Holidays: Contains holidays.
4. Leaves: leave records.
5. Salary: Contains salary-related data, including monthly salary working areas, working sites
6. Salary Dictionary: Contains information on salary categories.

Data pre-processing

During the data preprocessing phase, various steps were performed to clean and transform the "Marvelous Construction" dataset. The following operations were applied:

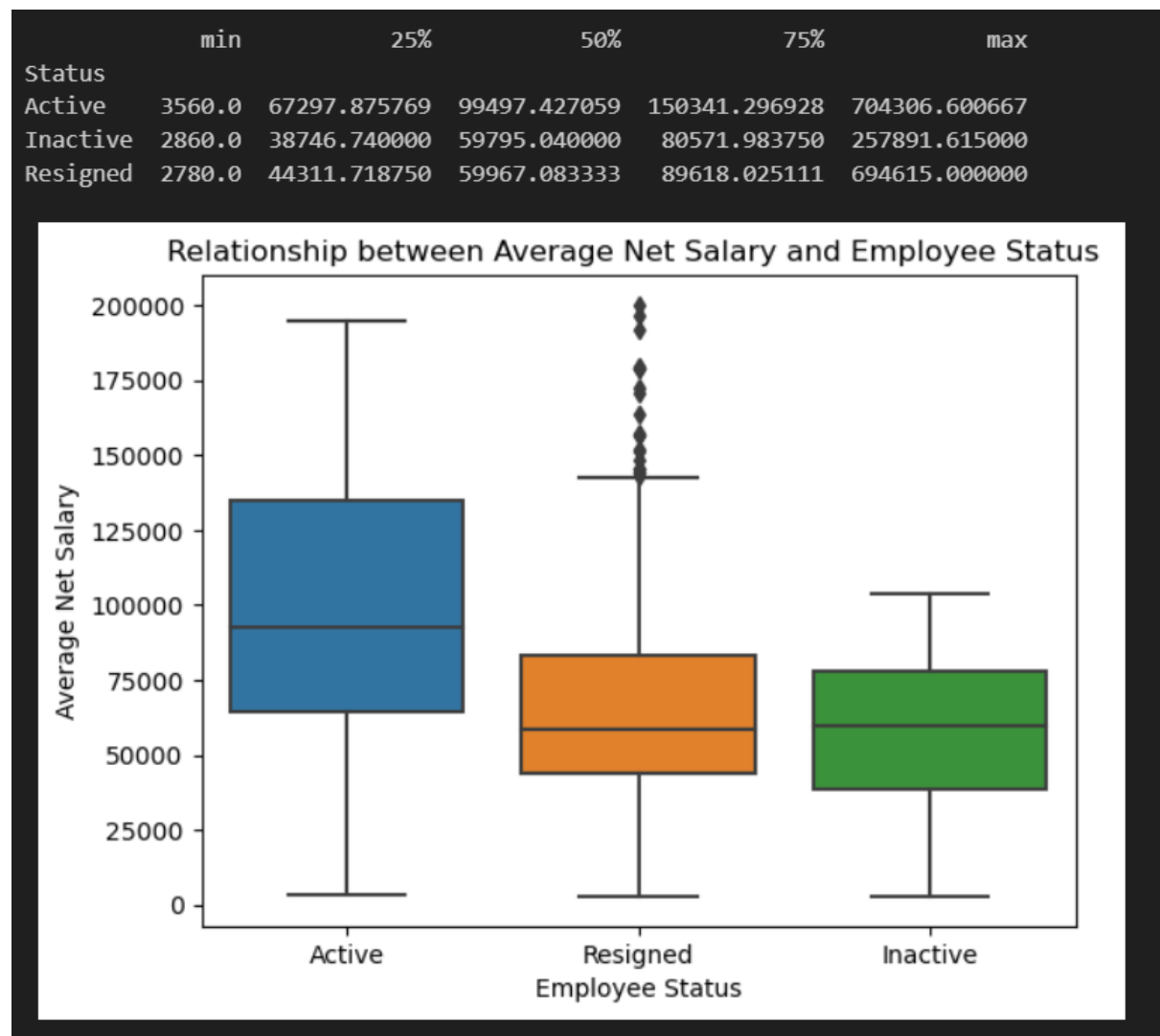
1. Removal of Duplicated Columns: Duplicate columns were identified within the dataset, particularly in the "Religion" feature. One column contained encoded values, while the other column contained corresponding labels. To simplify data processing, the column with labels was dropped, and only the encoded column was retained.
2. Replacement of Encoded Religion Values: Within the encoded religion column, a specific religion was represented by the value "5". For consistency and improved interpretability, this value was replaced with "2" based on the corresponding mapping.
3. Handling of Missing Birth Year Values: Some records in the "Year_of_Birth" column had missing values denoted as NaN. To address this issue, the missing values were replaced with the median value of the column. This approach ensures that the dataset remains representative while appropriately handling missing data. Additionally, the data type of the "Year_of_Birth" column was converted to an integer after filling in the missing values.
4. Formatting of Resignation Dates: The "Date_Resigned" column initially contained dates with inconsistent formatting and NaN values. To resolve this, the column was converted to a proper datetime format using the `pd.to_datetime()` function. Any values that couldn't be converted to datetime (including NaN values) were set to NaN, facilitating easier processing and analysis.
5. Outlier Detection and Removal: Outliers in the "Year_of_Birth" column were detected using the interquartile range (IQR) method. By calculating the IQR and applying a threshold of 1.5 times the IQR, outliers were identified. Given the small number of outliers detected, they were considered insignificant and subsequently removed from the dataset. This step ensures that the data remains reliable and representative for further analysis.

6. Data Type Transformation: Certain columns required specific data types for appropriate analysis. The "Religion_ID," "Year_of_Birth," and "Designation_ID" columns were transformed to integers using the ``astype()`` function, ensuring consistency and compatibility with subsequent operations.

Overall, these data preprocessing steps were implemented to improve data quality, consistency, and suitability for subsequent descriptive, exploratory, and predictive analysis tasks.

Insights from data analysis

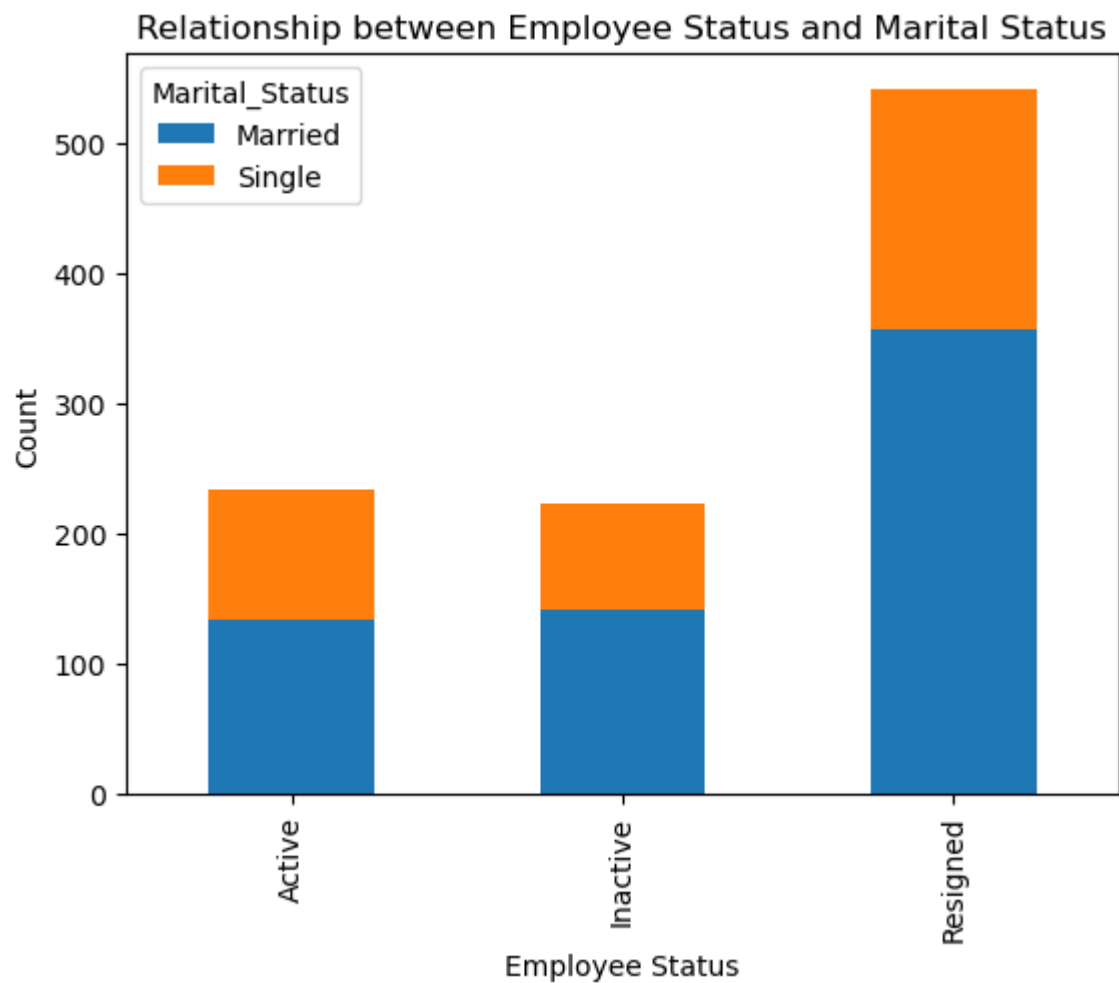
Insight 1: Relationship between resignations and average net salary



Active State employee have around 100000 mean net salary.

Resigned and Inactive state employee have around 60000 equant mean.

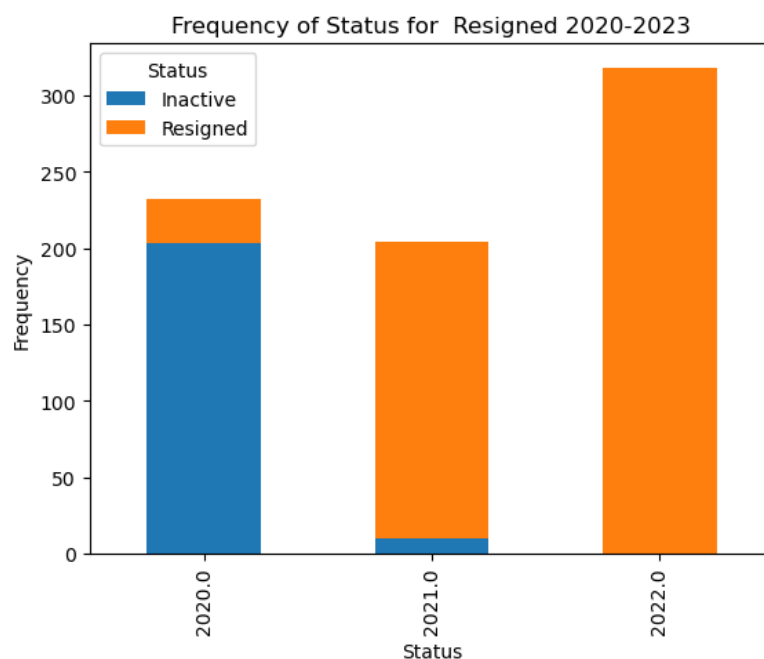
Insight 2: Relationship between resignations and marital status



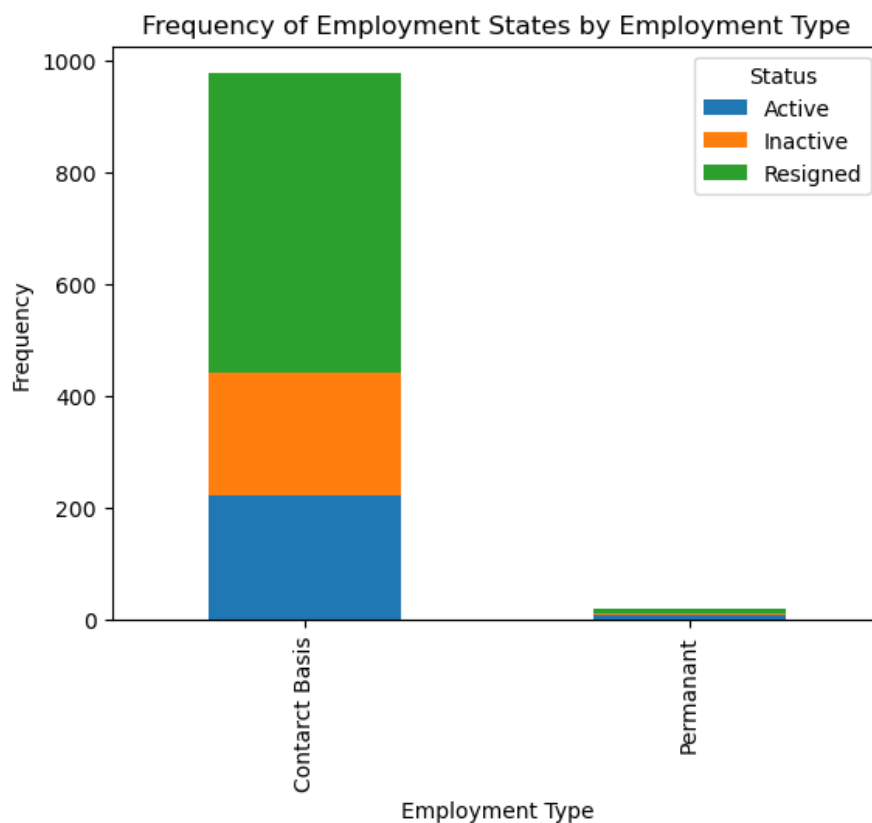
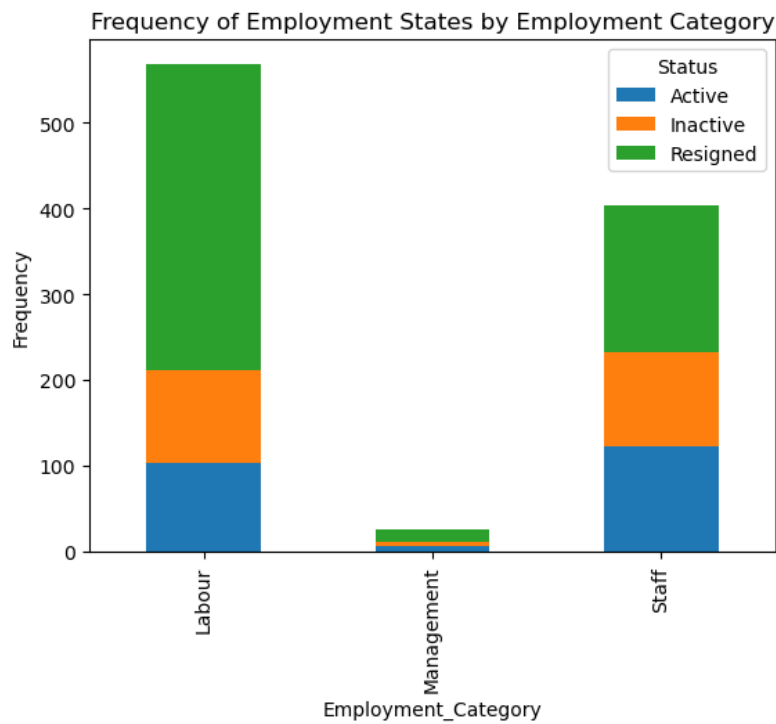
It clearly shows that married employees tend to resign more.

Insight 3: Inactive vs Resigned.

More employees Resigned in 2022

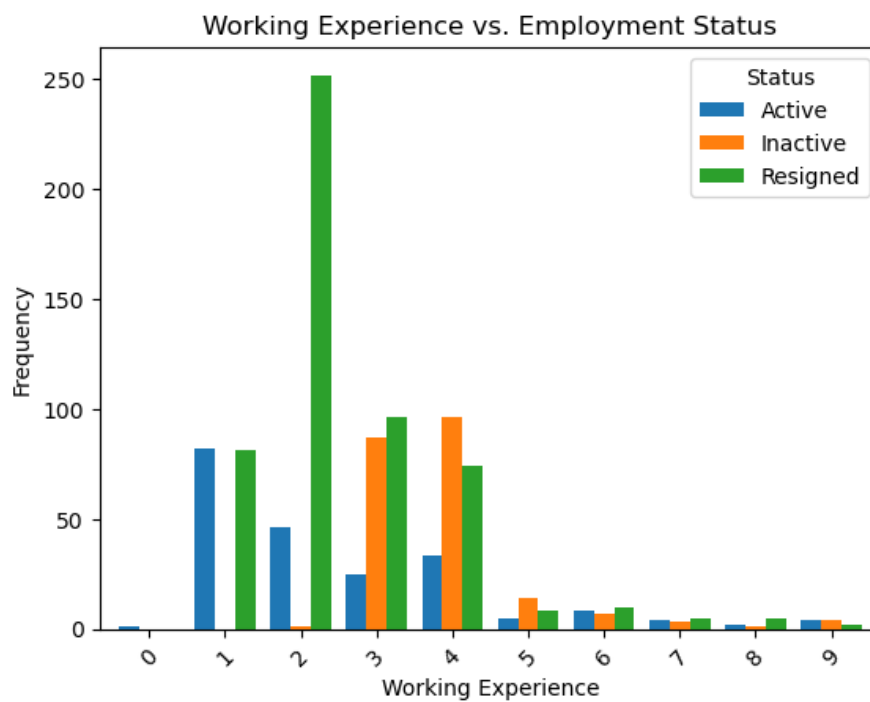


Insight 4 : Frequency of Employment States by Employment Type and Employment category



We can see similar distribution with employee Category and Type. So employee Category and Type are not important in this analyses

Insight 5 : Working Experience vs Resigning.



According to this we can see clearly more employees tends to resigning before 5 year experience