

# **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.

## NOTE

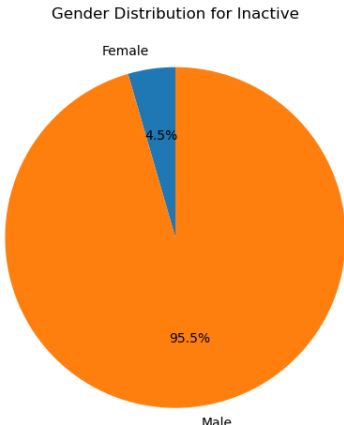
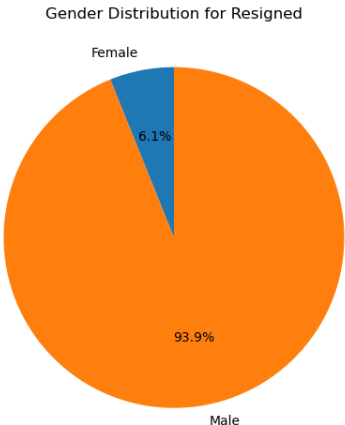
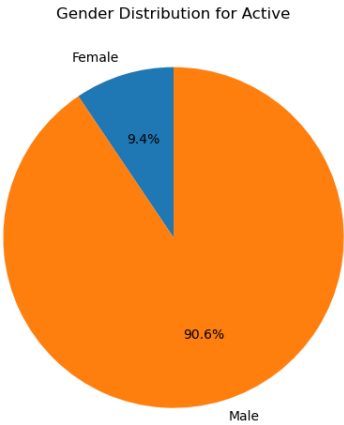
In this analysis we can see mainly 3 types of employees.

1. **Active** - currently working in company.
2. **Resigned** - When a professional resigns, they choose to terminate their employment contract with an organization.
3. **Inactive** - they choose to terminate their employment /Not attempt to work in long time but not informed professionally.

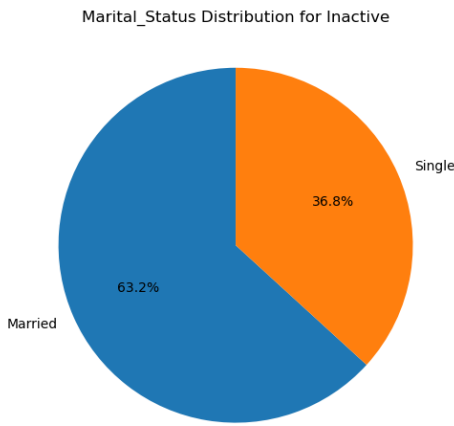
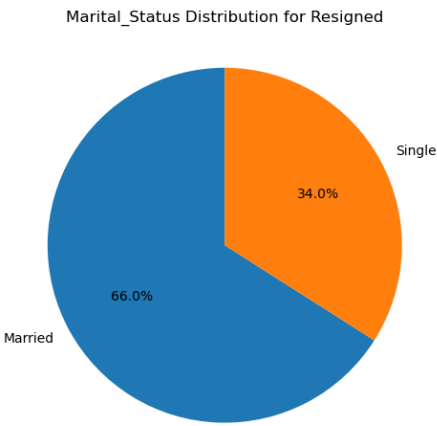
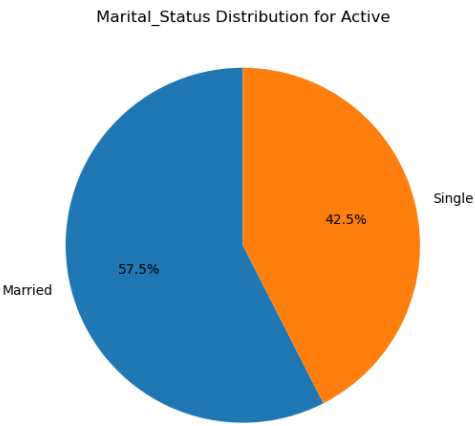
# Insights from data analysis

## Insight 1: Relationship between resignations and Gender, marital status, Religion and Employment Type, Employment category

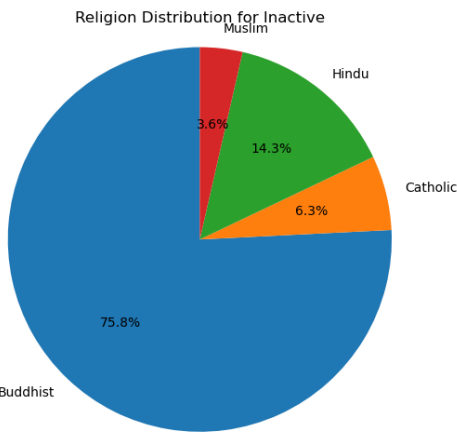
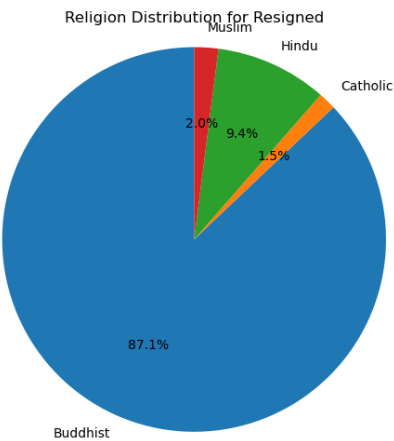
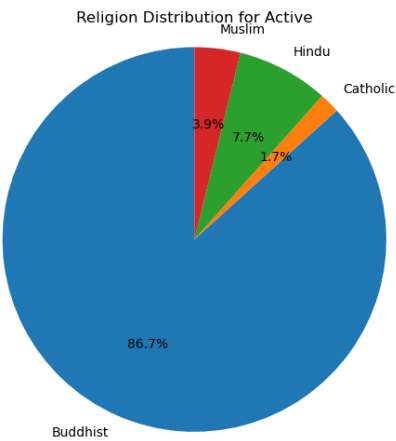
### Gender



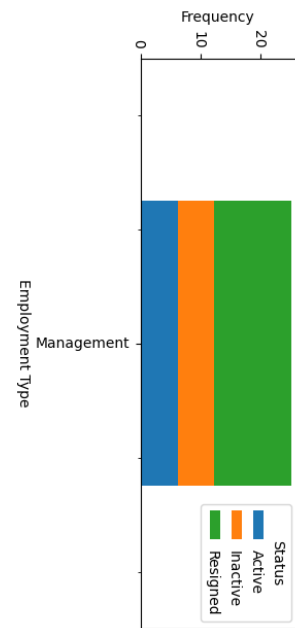
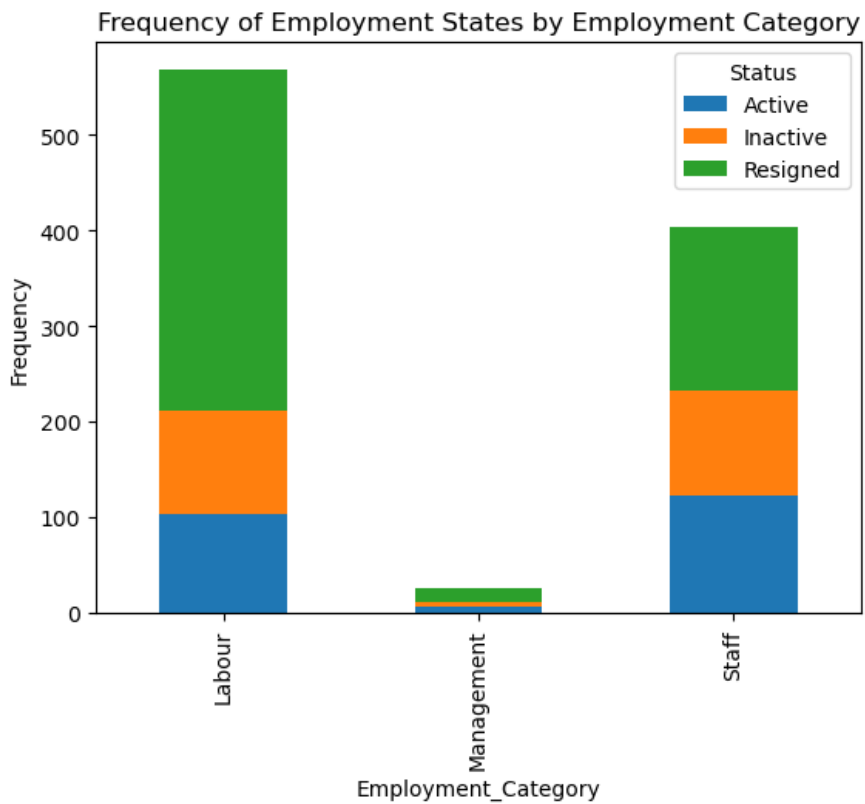
### Marital status



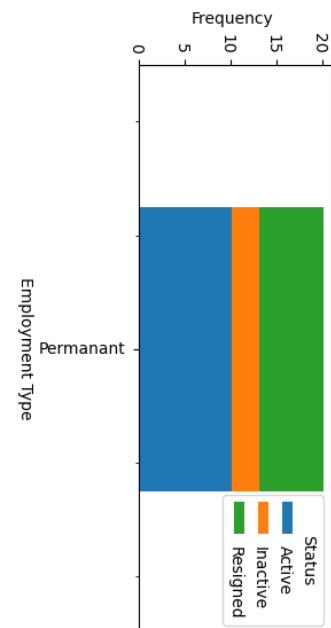
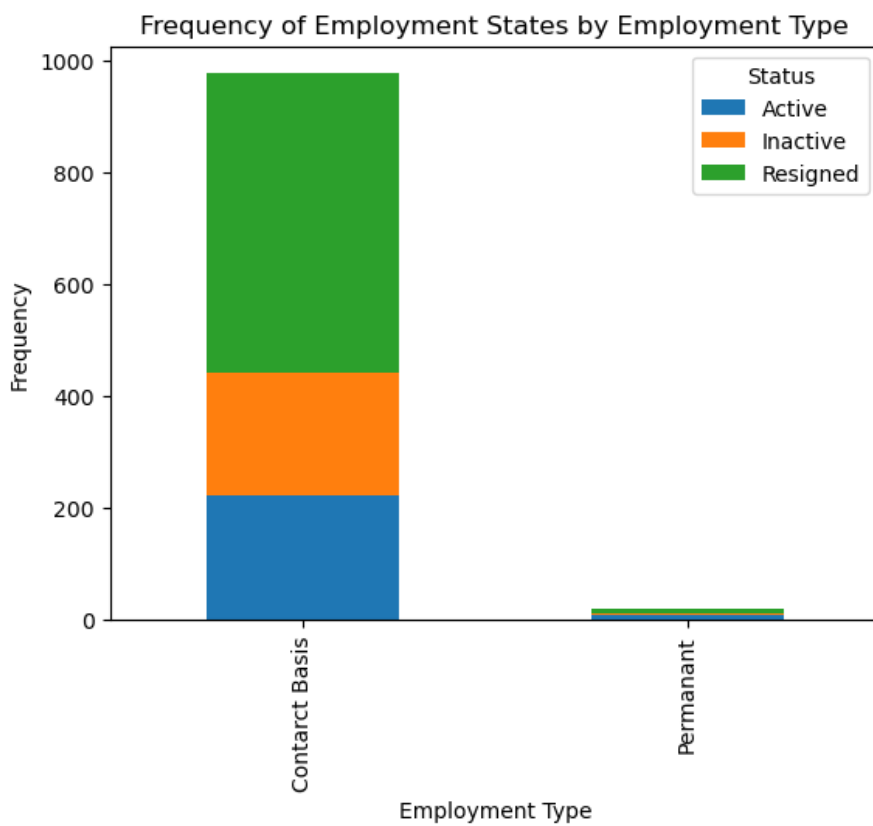
### Religion



## Employment category



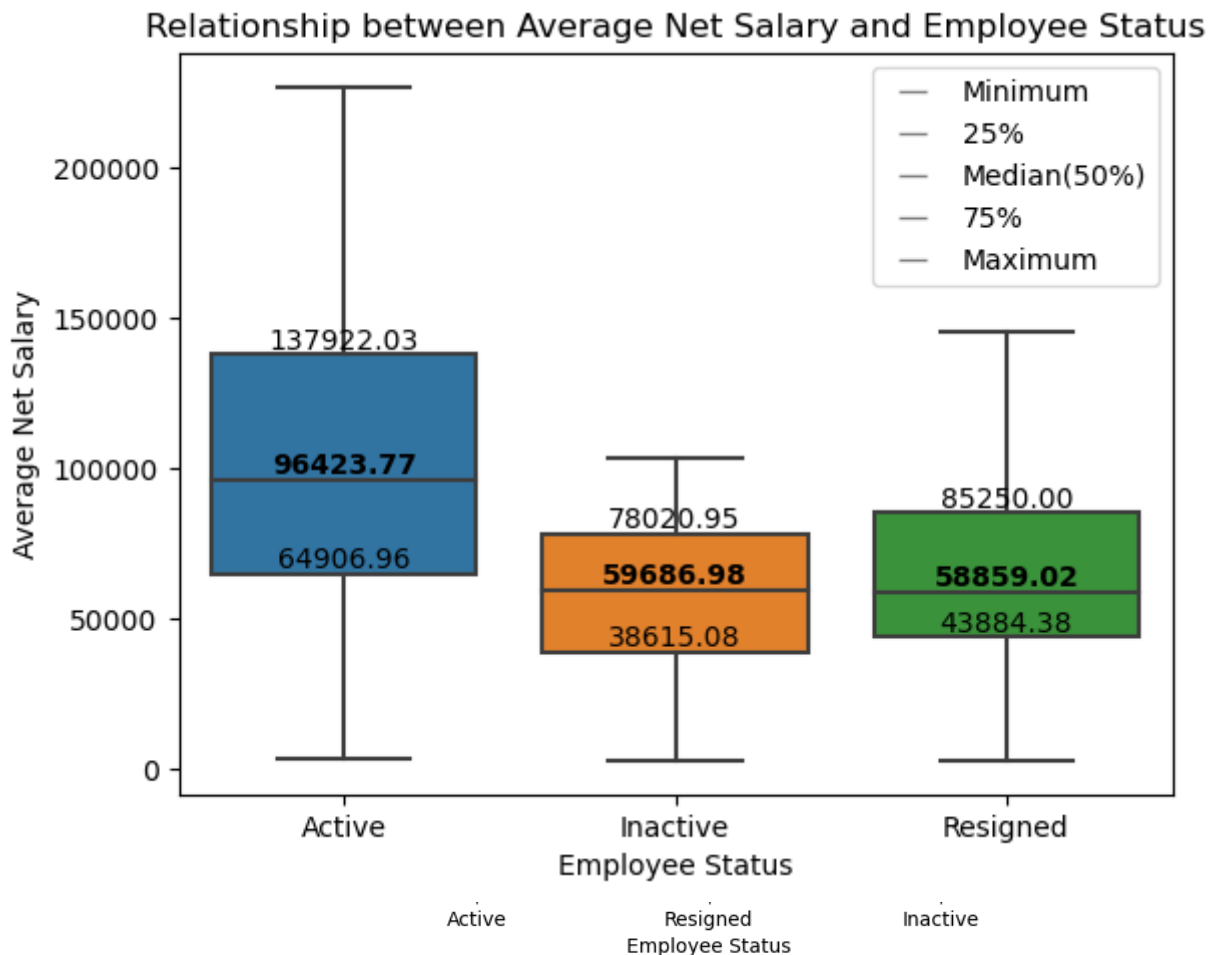
## Employment Type



- Gender, Marital State and Religion have same distribution among all states. So, they are not a key factor in this analysis.
- Employment Type and Employment category have same distribution among all states. So, they are not a key factor in this analysis.

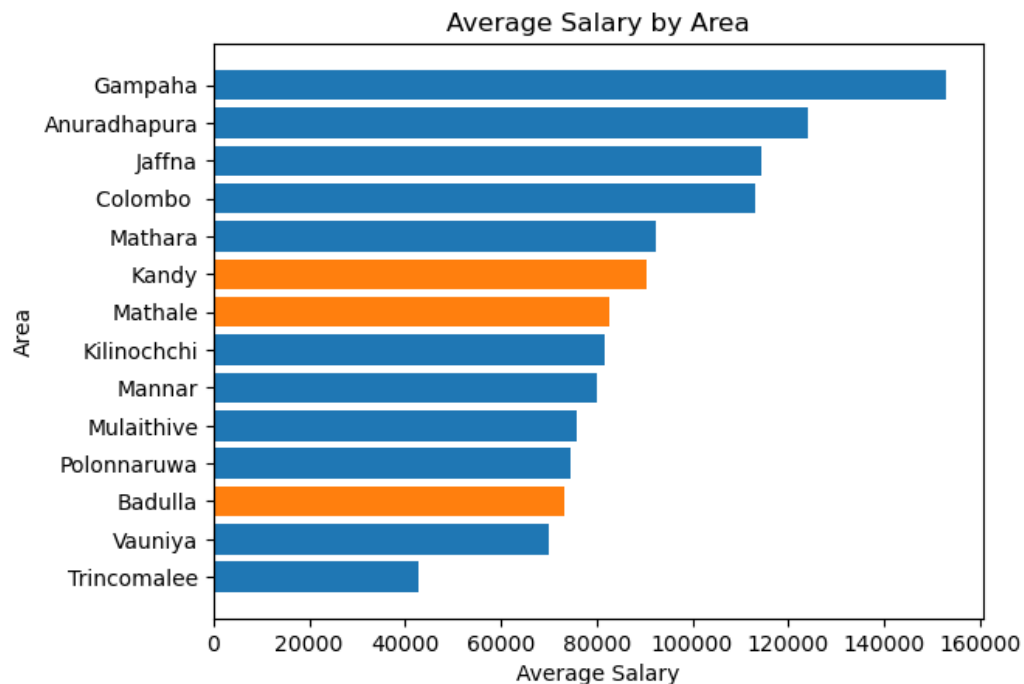
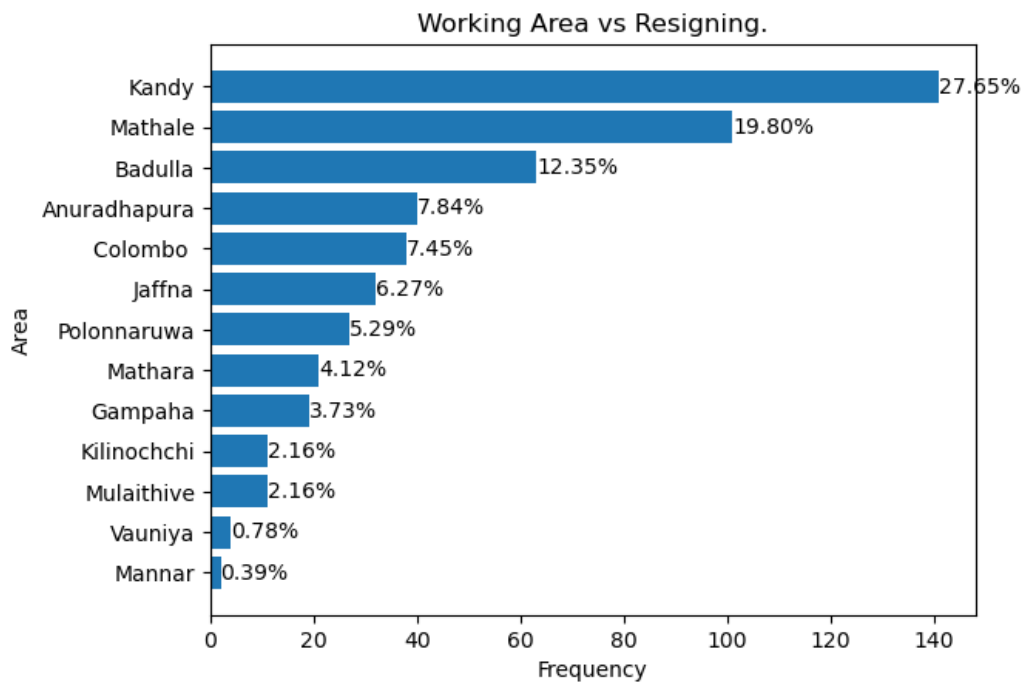
- Based on this observation, it suggests that these factors may not be significant contributors to the differences in resignation rates among states. Other factors, such as working conditions, job satisfaction, management practices, or individual preferences, may have a more substantial impact on resignation rates.

## Insight 2: Relationship between Resignations and Average Net Salary



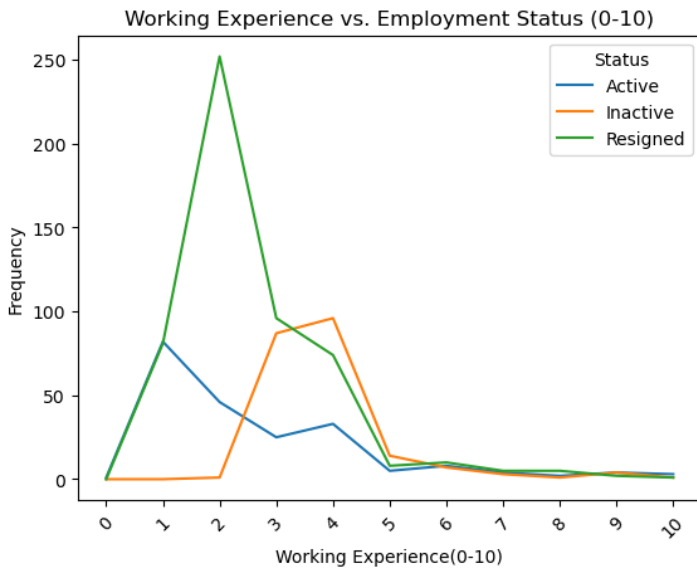
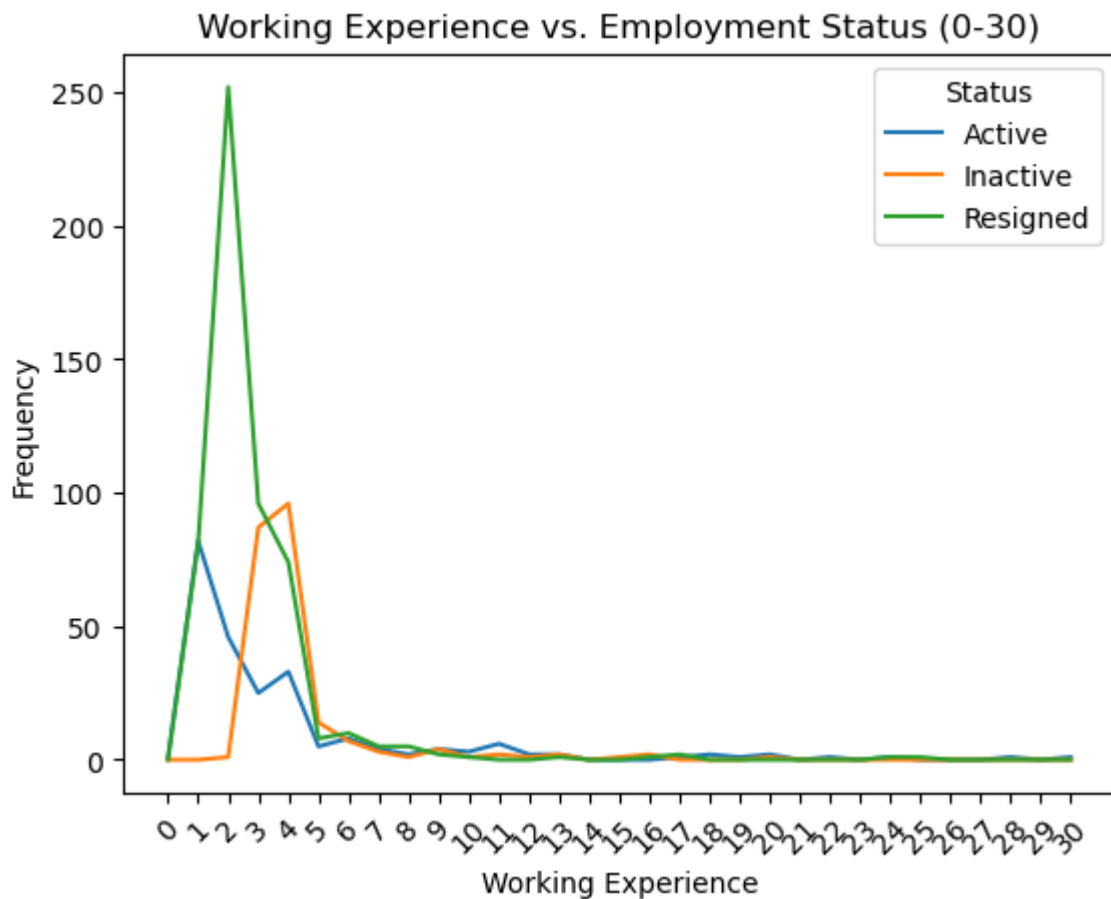
- The discrepancy in mean net salary between active state employees (around 100,000) and resigned/inactive state employees (around 60,000).
- From this summary, we can observe that active employees tend to have higher salaries compared to both inactive and resigned employees. The median salary of active employees is higher than the median salary of resigned employees. However, it's important to note that salary alone may not be the sole factor influencing employee attrition.
- It is essential to conduct a comprehensive analysis considering factors like job responsibilities, experience, working areas, Bones for Special vacation, and performance to identify specific reasons and provide targeted suggestions for addressing the salary discrepancy.

### Insight 3: Working Area vs Resigning.



- It appears that the area of "Kandy" has the highest resignation rate, but its average salary falls in the middle range. This indicates that there might be other factors influencing the resignation rate besides salary and working area.
- When considering the highest three resignation rates, it is interesting to note that they do not correspond to the areas with the highest average salary. This suggests that there may be other factors contributing to the decision to resign, such as job satisfaction, work-life balance, career growth opportunities, or other personal factors.
- To gain a deeper understanding, we can conduct surveys or interviews with current and resigned employees to gather qualitative data about their experiences, motivations, and reasons for leaving the company. This information can help identify patterns or common themes that contribute to the high resignation rates in certain areas.

#### Insight 4: Working Experience vs Resigning.



- Our findings indicate that many employees tend to leave within the first few years of employment, with the initial three years being particularly critical. This indicates the importance of focusing on employee satisfaction and providing competitive salaries, especially during this early stage of employment, to improve retention rates.
- During the onboarding and orientation process, it is essential to ensure that new employees feel welcomed, supported, and engaged. Providing clear expectations, comprehensive training programs, and opportunities for professional growth can contribute to a positive employee experience and increase the likelihood of retention.

- In terms of salary, it is crucial to offer competitive compensation packages that align with industry standards and reflect the value and skills of new employees. Conducting salary benchmarking exercises can help ensure that Marvelous Construction remains competitive in attracting and retaining talent.

#### Insight 5: Resigning vs Age of Employee.



- it appears that the age ranges from 28 years to 33 years and the age range from 45 years to 50 years have higher resignation percentages compared to other age ranges. There could be several reasons why the age ranges from 28 to 33 years and 45 to 50 years have higher resignation percentages. Some possible reasons could include:
  - Career Growth: Employees in these age ranges may seek better career growth opportunities and feel that their current organization does not offer sufficient advancement prospects.
  - Work-Life Balance: Individuals in these age groups may prioritize work-life balance and may be more likely to leave if they feel their current job does not provide a suitable balance.
  - Job Satisfaction: Factors such as job satisfaction, fulfilment, and alignment with personal values may play a significant role in employee retention. If employees feel dissatisfied with their work or do not find their roles fulfilling, they may choose to resign.
  - Retirement Planning: Individuals nearing retirement age (45 to 50 years) may decide to leave the workforce to focus on retirement planning or pursue other personal interests.
  - Changes in Personal Circumstances: Life events such as marriage, starting a family, or caring for elderly parents may prompt employees in these age ranges to reassess their work commitments and consider resigning.