

Bootcamp T5

WRITTEN DESCRIPTION FOR EDA PROJECT

Nasser Alqahtani | EDA | 17/11/2021

Abstract

The main purpose of this analytical report is to provide clear insights into financial management about how to manage the company's budget and include nominations and rewards for employees. The first stage of analyzing the HR data is to clean, prepare, and handle data. The second stage is to apply the feature selection to the HR data frame and to reduce the compositional processing in such a way that we achieve the goals of this analysis. Lastly, provide a clear visualization to make the data easier to understand and pull insights from.

Data Sets for Human Resources. The datasets are generated through random logic in VBA. This data set can be categorized under "Human Resources" category. The dataset from E for Excel.

Design

This project originated from our client needs to know the HR budget, to see the prosperity of the company, and compare the company's extension across the states. will be in project data collecting, pre-process, and Visualization for hiring in all years from 1979 to 2017, Comparison the mean salaries of each year from 1979 to 2017, The mean of Age in Company (Years) and Salary from 1979 to 2017, The percentage of Male & Female in the company, top 10 States in hiring, and compare of Gender & Years in the company for all States.

Data

The dataset contains 10,000 employees with 37 features for each, 23 of which are categorical. A few feature highlights include Emp ID, Name Prefix, First Name, Last Name, gender of the employee, year of Joining, a quarter of Joining, Age in Company (Years), Salary, and State.

Algorithms

- 1-Performed data cleaning.
- 2-Visualized the data.
- 3-Aggregate the data by Year of Joining and ordered by Highest and Lowest year of employment.
- 4-Comparing the mean salaries of each year from 1979 to 2017.
- 5-The mean of Age in Company (Years) and Salary from 1979 to 2017.
- 6-The percentage of Male & Female.
- 7-Statics of Gender & Years in company for all States.
- 8-Visualized the resulted from clean data.

Tools

- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting