

Assignment : Statistics and Trends

Aim: Acquire and explore a dataset, identify potential relationships through various visualizations, and present these findings along with relevant statistics in a clear and concise report, demonstrating your proficiency in data analysis and visualization

Github Link : <https://github.com/Biradarkaran/Statistics-and-Trends->

Data Description:

Dataset: HR Metrics and Analytics

Kaggle Dataset link : <https://www.kaggle.com/datasets/rhuebner/human-resources-data-set/data>

Why I choose this dataset:

I choose the HR Metrics and Analytics dataset for my assignment due to its alignment with the required visualizations (histograms, bar charts, scatter plots), descriptive statistics (describe), and correlation analysis (corr). Additionally, its comprehensiveness with various employee information facilitates the exploration of relationships and insights through data analysis techniques. This dataset is designed for educational purposes in an HR Metrics and Analytics course. It simulates real-world HR data.

Features of Dataset:

Personal information: Name, Date of Birth (DOB), Age, Gender, Marital Status

Employment details: Date of Hire, Reasons for Termination (if applicable), Department, Active/Terminated Status, Position Title, Pay Rate, Manager Name

Performance: Performance Score, Most Recent Performance Review Date

Engagement: Employee Engagement Score (recent addition)

Additional: Absences (recent addition)

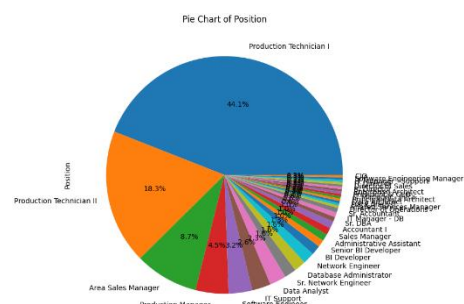
Code Description:

The code performs Exploratory Data Analysis (EDA) on a human resources (HR) dataset. It starts by importing necessary libraries and loading the data. Then, it preprocesses the data by selecting relevant columns. Various visualization functions are defined to create different plots like histograms, bar charts, and scatter plots to explore the data visually. Additionally, advanced visualizations like heatmaps and corner plots are used to identify potential relationships between features. Finally, descriptive statistics and the correlation matrix are calculated to understand the central tendency, spread, and relationships between numerical features in the dataset.

Graphs and their Interpretations:

1. Pie Chart:

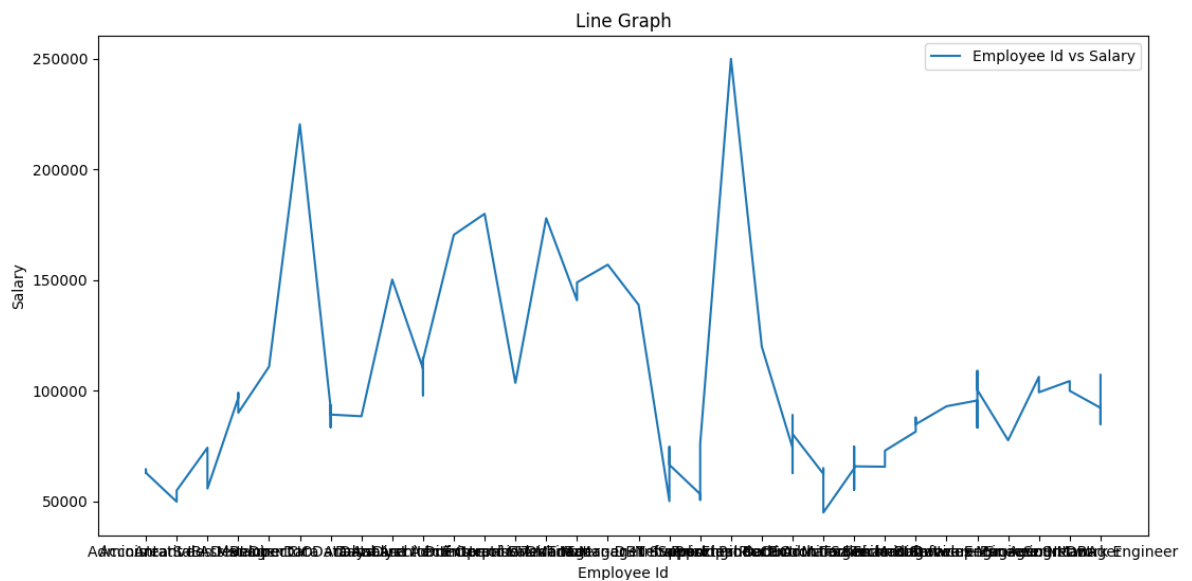
The pie chart showcases the distribution of employees across various positions in the HR dataset's. Each slice represents a specific position, and its size reflects the proportional number of employees holding that role. Notably, the largest slice corresponds to "Production Technician I",



indicating it has the highest number of employees compared to other positions. This observation suggests the company might prioritize production operations or have a larger workforce dedicated to entry-level technical roles.

2. Line Graph:

This line graph visualizes the salary distribution across various positions within the company. While a clear trend might not be evident due to the categorical nature of positions, it's noteworthy that the highest salary of \$250,000 corresponds to the "President & CEO" position. This observation aligns with expectations, as leadership roles typically command higher compensation within an organizational structure. However, further analysis, like comparing salary distributions across different positions using other visualizations, could offer deeper insights into salary patterns within the company.



3. Heatmap:

The heatmap depicts a complex web of correlations between various employee attributes in the HR dataset. While individual data points might require further investigation, the color intensity reveals potential relationships. Warmer colors suggest positive correlations, like higher salaries potentially aligning with certain positions. Conversely, cooler colors indicate negative correlations, which could exist between specific marital statuses and genders (though interpretation needs caution due to potential societal biases). Overall, this heatmap provides a starting point for exploring deeper connections within the data, potentially leading to further analysis and interpretation to gain a more comprehensive understanding of the employee landscape.

