

Project Description

Business Context

In today's rapidly evolving tech industry, understanding salary trends is crucial for job seekers, employers, and policymakers. Compensation varies significantly based on job roles, company size, experience, and location, making it essential to analyze patterns that impact salary structures.

This project leverages job postings data from Glassdoor.com (2017) to predict salaries for various tech job positions. The dataset includes key features such as job title, company size, and job location, which can be used to identify salary trends, compare roles across industries, and provide valuable insights into compensation expectations.

Dataset Description

The Glassdoor Jobs Dataset contains job postings from Glassdoor.com from 2017-2018. This dataset includes features such as job title, salary estimate, job description, rating, company name, location, headquarters, size, founded, type of ownership, industry, sector, revenue, competitors compile a list of the most important features in this dataset. , hourly wage minima and maxima,, avg_salary,, num_comp,,sdesc_len If you are looking for a job in the data science field then this is the dataset for you to explore!

Business Objectives

- For Job Seekers: Helps professionals make informed career decisions by understanding expected salary ranges for different roles.
- For Employers: Assists companies in setting competitive salaries to attract and retain top talent.
- For Analysts & Researchers: Provides data-driven insights into salary trends based on industry, experience, and geography.
- For Recruiters: Aids in benchmarking salaries and ensuring fair compensation practices.

Problem Statement

- How does salary vary by job position (e.g., Data Scientist vs. Software Engineer vs. DevOps Engineer)?
- What is the impact of company size on salary levels?
- How do salaries differ by location (e.g., San Francisco vs. Austin vs. New York)?
- Can we build a predictive model to estimate salaries based on job attributes?

By analyzing this dataset, we can predict salary ranges, uncover market trends, and provide insights to tech professionals and organizations.

Main Libraries to be used:

- Pandas for data manipulation, aggregation
- Matplotlib and Seaborn for visualization and behavior with respect to the target variable. Use at least 5 different visualizations.
- NumPy for computationally efficient operations

Project Architecture:

