**Salary Prediction for Data Scientists & STEM (Science Technology Engineering Mathematics) Employees in Different Countries.**

**Project Summary: -**

This project is designed to predict the salaries of the employees in regards to their work positions and number of years of work experience in varied companies around the globe.

The dataset consists of 62000 salary records from the top companies.

1st column – Timestamp – Records from year June 2017 to April 2021

2nd column – Company – list of top companies like Google, Microsoft, Amazon, Apple, etc.

3rd column – Level – level of position the employee is working in that company.

4th column – Title – The designation of the employee ex: Product Manager, Software Engineer, Data scientist, Business analyst, etc

5th column – Total Yearly Compensation – the annual compensation amount

6th column – Location – location of the company in which employee is working.

7th column – Years of Experience – number of years of work experience of the employee.

8th column – Years at Company – number of years since the employee is working for that company.

9th column – Tag – defines whether the system is full stacked, distributed or any other.

10th column – Base salary – starting salary of the employee in that position, in that company.

11th column – Stock Grant Value – stock grant value of the company.

12th column – Bonus – the extra salary/perk amount of an employee in that organisation.

13th column – Gender – gender details of an employee

14th column – Other details – other details of the company

15th column – CityID – the location ID of the company

16th column – dmaid

17th column – rowNumber

18th column – Masters Degree – whether the employee has master’s degree or not

19th column – Bachelor’s Degree – whether the employee has bachelor’s degree or not

20th column – Doctoral Degree – whether the employee has doctoral degree or not

21st column – High School – whether the employee went to high school or not

22nd column – Some\_college – whether the employee went to college or not

23rd column – Race\_Asian – whether the employee is of Asian race or not

24th column – Race\_White – whether the employee is of the White race or not

25th column – Race\_two\_or\_more – whether the employee is of two/more races or not

26th column – Race\_Black – whether the employee is of black race or not

27th column – Race\_Hispanic – whether the employee is of Hispanic age or not

28th column – Race – determines the race of an employee

29th column – Education – defines the qualification level of the employee

Given below is the dataset link:-

**Dataset Link**: [*https://www.kaggle.com/datasets/jackogozaly/data-science-and-stem-salaries*](https://www.kaggle.com/datasets/jackogozaly/data-science-and-stem-salaries)

From the given dataset, we need to make changes in an already existing application to view the salary prediction of the employees in different companies at different locations with reference to their number of years of work experience. We will analyze the data, clean it and build the model accordingly using regression model and scikit learn tools in Python. I have decided to include two web pages wherein first one would be predicting the salaries and second page would be defining the data design of the dataset to get a clear view of the data. The prediction and exploration strategies applied in the application will improve it not only visually but also would be immensely helpful for the younger generation and people who are looking for a career/job switch to get an idea about the career roles and its aesthetic demand in the industry.