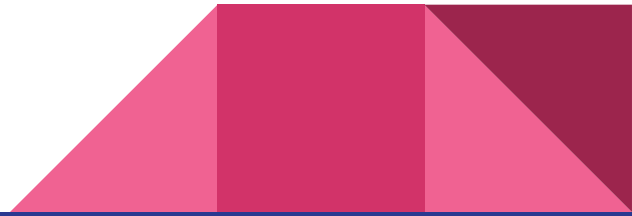
What are your salary expectations, after finishing the bootcamp?

Salary Prediction Tool for U.S-Based Data Science Roles

Project Overview

Team

- Cristian Llanes (Square Role)
- Maria Sevillano (Triangle Role)
- Alejandra Villarreal (Circle Role)
- Sharof Abdoollayev (X Role)



Project Overview

Objective

The purpose of this project is to build a resource for job-seekers to predict the salary of a given career field, Data Science, based on set variables.

- Answer the "What Are Your Salary Expectations?" question that a hiring manager might pose during an interview process.
- Determine if they should accept or decline a job offer.

Project Overview

Data Source

Original data set : Levels_Fyi_Salary_Data.csv

[https://www.kaggle.com/datasets/jackogozaly/data-science-and-stem-salaries?resource=download].

Project Overview

Technologies Used

- Pandas
- Postgres
- Amazon AWS
- SciKitLearn
- Tableau

Project Overview

Questions Data Set Will Answer

- Will salary for Data Science jobs continue to experience growth in the future?
- Based on the selected set of variables, what is the expected salary range?
- Determine salary trends based on specific factors.

Preliminary Machine Learning Model

Data Preprocessing

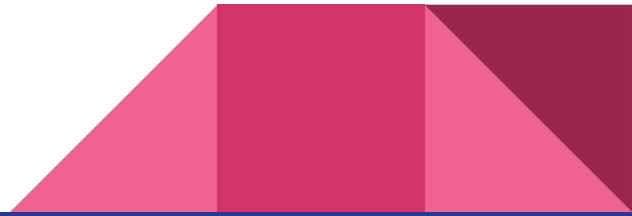
Preprocessing will involve the followings:

- Checking and handling imbalanced datasets.
- Performing initial exploratory analysis, including scatter plotting and correlation.
- Removing non-beneficiary columns.
- Preparing the data by working with any missing values, scaling the data, and converting categorical variables by using the one-hot encoding scheme.

Preliminary Machine Learning Model

Splitting the dataset

The dataset will be split into training and testing sets using the 80/20 Pareto principle resulting in a test size of 20%.



Preliminary Machine Learning Model

Supervised Machine Learning Model

We will use a supervised machine learning model since we are looking to predict a value.There are different models we can use:

- **Regression**
- **Classification / Ensemble Methods**

Preliminary Machine Learning Model

Regression

- Apply a Linear Regression to predict salary.
- We will also explore applying a Multilinear Regression Model to add other factors that might influence the salary prediction.

Preliminary Machine Learning Model

Classification / Ensemble Methods

We could use Random Forest Regression to discover the connection between the target and independent variables to determine a continuous value. This connection can then be used to predict salaries of data science jobs..

Preliminary Machine Learning Model

Model Evaluation

We will evaluate the models based on:

- **Explained Variance Score:** Similar to the R^2 score, with the notable difference that it does not account for systematic offsets in the prediction.
- **Model Score:** Returns the mean accuracy on the given test data.