DATA ANALYSIS

GALACTIC WELL-BEING INDEX PREDICTION

AUTHOR: NASHON OKUMU

PROBLEM DEFINITION

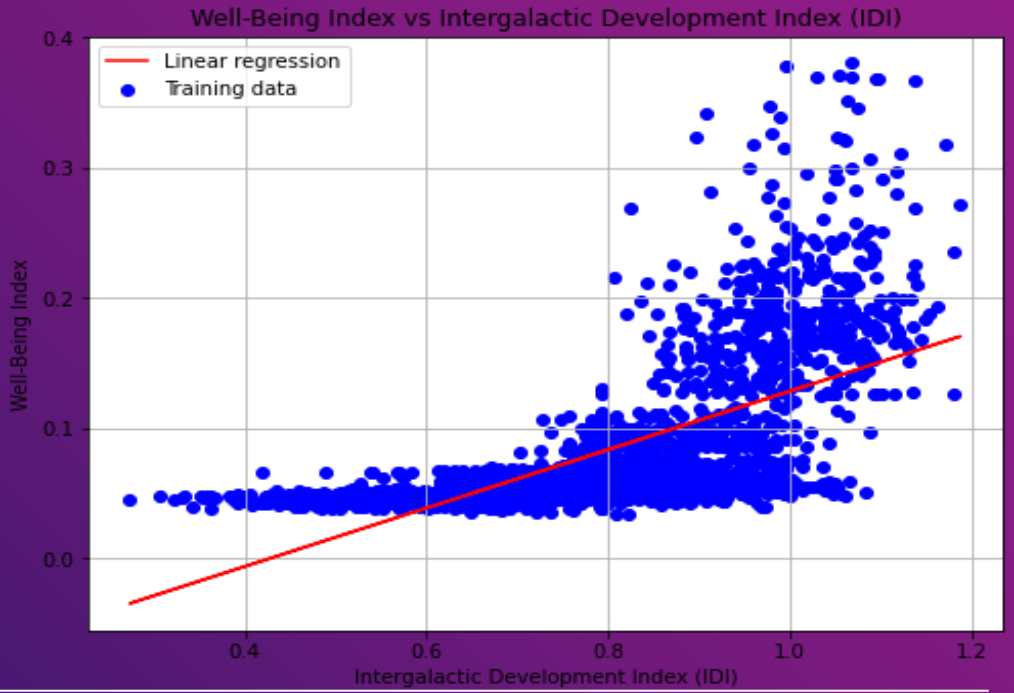
- The goal of this project was to analyze and predict the well-being index of 181 galaxies over a period of up to 26 years based on a dataset containing 80 demographic and socio-economic variables. The project aims to address two key challenges:
  - Identify Influential Variables: Determine which demographic and socio-economic variables best explain the variance in the well-being index.
  - Predict Future Well-Being: Develop and evaluate regression models that accurately predict future values of the well-being index for the galaxies.
- This helps anticipate future trajectories of well-being for these celestial bodies.

DATA ANALYSIS

- A correlation analysis indicated a high correlation between these top five variables and the target variable (Well-being index).
  - Intergalactic Development Index (IDI), Rank
  - Intergalactic Development Index (IDI)
  - Education Index
  - Expected years of education (galactic years)
  - Income Index

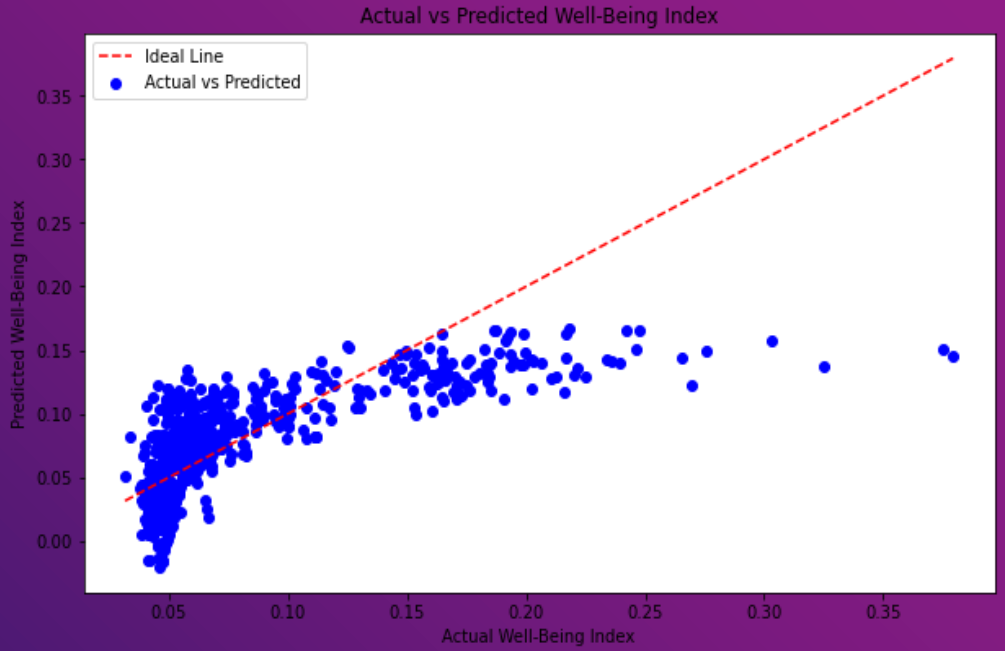
SIMPLE LINEAR REGRESSION MODEL

- SLR Produced an RMSE of 0.0393555493792426O6.
- In this case it means that, on average, the model's predictions are off by approximately 0.0393 units from the actual values.



MULTIPLE LINEAR REGRESSION MODEL

- MLR Produced an RMSE of 0.037945376887868414.
- In this case, it means that, on average, the model's predictions are off by approximately 0.0379 units from the actual values.



CONCLUSION

- These findings of Data Analysis suggest that socio-economic factors, such as education level, income, and overall development index, play a crucial role in determining the well-being of galaxies.
- Based on the RMSE metric, the Multiple Linear Regression model is preferred for predicting the Well-Being Index of galaxies due to its slightly higher prediction accuracy compared to the Simple Linear.
- The MLR model demonstrated improved performance compared to SLR, indicating that considering multiple features simultaneously enhances the predictive power of the model.