Linear Regression Analysis Summary

This project involves performing a linear regression analysis to predict the number of medals won by teams based on the number of athletes and previous medals won. The analysis is done using a combination of manual calculations and the LinearRegression model from the scikit-learn library.

Data Preparation:

1. **Importing Libraries**: The necessary libraries for data manipulation and numerical calculations are imported.
2. **Loading Data**: The dataset is loaded from a CSV file named teams.csv into a pandas DataFrame for easier data handling.
3. **Defining Features and Target Variable**: The features used for prediction are the number of athletes and previous medals won by each team. The target variable is the number of medals won.

Manual Linear Regression Calculation:

1. **Adding Intercept**: An intercept term is added to the feature set to account for the baseline effect in the linear model.
2. **Calculating Coefficients**: The coefficients of the linear regression model are calculated using the normal equation method. This involves matrix operations to find the best-fit line that minimizes the sum of squared residuals.
3. **Making Predictions**: Predictions are made using the calculated coefficients, applying them to the feature set.
4. **Evaluating Model**: The model's performance is evaluated by calculating the sum of squared residuals (SSR) and the total sum of squares (TSS). The R-squared value is derived to measure how well the model explains the variability of the target variable.

Using Scikit-learn for Linear Regression:

1. **Importing the Model**: The LinearRegression model from the scikit-learn library is imported for an alternative method of performing linear regression.
2. **Training the Model**: The model is trained using the features and target variable from the dataset.

3. **Model Coefficients**: The intercept and coefficients are obtained from the trained model, which can be used to make predictions and understand the influence of each feature.

Summary:

The linear regression analysis demonstrates how to predict the number of medals based on team attributes. Both manual calculation and scikit-learn's LinearRegression model yield similar results, showcasing the consistency and reliability of the linear regression technique.

This project serves as a basic example of linear regression and can be extended with more features or different types of regression models for improved predictions. The analysis provides a clear understanding of how linear regression works and its application in real-world scenarios. By following the steps outlined in the Jupyter notebook, users can replicate the analysis or adapt it for other datasets and use cases.

The project highlights the importance of data preparation, the process of fitting a linear regression model, and the evaluation of model performance. It serves as a practical guide for anyone looking to understand or implement linear regression analysis in Python.