POLYNOMIAL REGRESSION

This code implements **Polynomial Regression** to predict the length of a product based on its product type ID. Here's a step-by-step explanation of the code:

- First, the necessary libraries are imported - pandas for data handling, PolynomialFeatures from sklearn.preprocessing for feature transformation, LinearRegression from sklearn.linear_model for regression modeling, and mean_absolute_percentage_error from sklearn.metrics for evaluation.

- The training and test datasets are loaded using pandas' read_csv function.The two datasets are concatenated into one using the concat function from pandas. This is done so that any changes made to the data, such as replacing missing values with zeros, are done consistently across both datasets.Any missing values in the full dataset are replaced with zeros using the fillna function from pandas.

- The features and target variables are defined for both the training and test datasets. Here, the feature is the product type ID and the target is the product length.

- The feature is transformed into polynomial features using the PolynomialFeatures function from sklearn.preprocessing. In this case, degree=2 is used to create a quadratic feature.The polynomial features are used to train a linear regression model using the LinearRegression function from sklearn.linear_model.

- The model is used to make predictions on the test dataset.The mean absolute percentage error between the predicted values and the true values is calculated using the mean_absolute_percentage_error function from sklearn.metrics.

- A submission file is created using pandas. The predicted values are stored in a DataFrame along with the corresponding product IDs. The index of the DataFrame is set to the product IDs and the DataFrame is saved as a CSV file.

In summary, this code transforms the product type ID feature into a polynomial feature and trains a linear regression model on it to predict the length of a product. The mean absolute percentage error is used to evaluate the model's performance and a submission file is created with the predicted values.