Analysis Report: Regression Analysis on 'AMT_CREDIT'

Task Type:

I conducted a regression analysis to predict the 'AMT_CREDIT' using appropriate columns from the dataset.

Dataset Overview:

The dataset contains 307511 rows and 122 columns. The goal is to predict 'AMT_CREDIT' using structured data features.

Methodology:

I employed Linear Regression for the analysis. The dataset was divided into training and testing sets for model evaluation.

Model Metrics:

The model's performance metrics are as follows:

- Root Mean Squared Error (RMSE): 67442.582
- R-squared (R^2): 0.979

Feature Insights:

The most influential features affecting 'AMT_CREDIT' prediction are:

- YEARS_BEGINEXPLUATATION_AVG
- CODE_GENDER_XNA
- APARTMENTS_AVG
- YEARS_BEGINEXPLUATATION_MODE
- APARTMENTS_MEDI

Summary (in plain language):

The model can effectively predict 'AMT_CREDIT' based on the dataset features. Lower RMSE and higher R^2 values indicate accurate predictions. Notable features influencing the prediction include 'YEARS_BEGINEXPLUATATION_AVG', 'CODE_GENDER_XNA', 'APARTMENTS_AVG', 'YEARS_BEGINEXPLUATATION_MODE', and 'APARTMENTS_MEDI'.

Technical Summary:

The model was trained using scikit-learn's LinearRegression on standardized and encoded input features.

Limitations and Notes:

- Any limitations or additional notes about the analysis could be included here.

Technical Section for Data Scientists:

Methodology Details:

- Regression Model: Linear Regression
- Data Preprocessing: Standardization and Encoding

Model Evaluation:

- RMSE: 67442.582
- R^2: 0.979

Feature Coefficients:

The coefficients of the most influential features:

- YEARS_BEGINEXPLUATATION_AVG: -694170.0718
- CODE_GENDER_XNA: -363962.3081
- APARTMENTS_AVG: 360868.5837
- YEARS_BEGINEXPLUATATION_MODE: 327357.9474
- APARTMENTS_MEDI: -320577.0265
- LIVINGAPARTMENTS_AVG: -197225.2021
- LANDAREA_MEDI: -186934.7427
- ORGANIZATION_TYPE_XNA: -172956.8661
- NAME_INCOME_TYPE_Pensioner: -172956.8661
- COMMONAREA_MEDI: 154617.8062

Conclusion:

The Linear Regression model shows strong predictive power for 'AMT_CREDIT' with high accuracy and notable feature influences. Further analysis and fine-tuning may enhance the model's performance.