Feature Engineering Recommendations

Feature: `loan_amnt`

- **Reason:** Higher loan amount may indicate higher risk as borrower may struggle to repay larger sums.

- **Suggested Transformations:** `Log transformation to handle skewed distribution. Interaction with installment to create a debt-to-income ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `installment, dti`

Feature: `installment`

- **Reason:** Higher monthly payment may strain borrower's finances, increasing default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship. Interaction with loan amount to create a debt-to-loan ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `loan_amnt, dti`

Feature: `emp_length`

- **Reason:** Longer employment history may indicate stability and lower default risk.

- **Suggested Transformations:** `Interaction with annual_inc to create a stability-to-income ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `annual_inc`

Feature: `annual_inc`

- **Reason:** Higher income may imply better ability to repay debts and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with dti to create a debt-to-income ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `dti`

Feature: `dti`

- **Reason:** Higher debt-to-income ratio may indicate financial strain and higher default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship. Interaction with annual_inc to create a debt-to-income ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `annual_inc`

Feature: `bc_util`

- **Reason:** High credit utilization may suggest financial stress and higher default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with total_rev_hi_lim to create a credit utilization ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `total_rev_hi_lim`

Feature: `fico_range_high`

- **Reason:** Higher FICO score may indicate lower default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `-`

Feature: `total_rev_hi_lim`

- **Reason:** Higher total revolving credit limit may indicate better credit management and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with bc_util to create a credit utilization ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `bc_util`

Feature: `purpose`

- **Reason:** Different loan purposes may have varying default risks (e.g., debt consolidation vs. small business).

- **Suggested Transformations:** `One-hot encoding to convert categorical variable.`

- **Relationship Type:** Threshold-based

- **Suggested Interactions:** -

Feature: `total_bc_limit`

- **Reason:** Higher bankcard credit limit may indicate better credit management and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with percent_bc_gt_75 to create a credit utilization ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `percent_bc_gt_75`

Feature: `acc_open_past_24mths`

- **Reason:** Higher number of trades opened may indicate credit-seeking behavior and potential financial instability.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** -

Feature: `mo_sin_old_rev_tl_op`

- **Reason:** Longer history of revolving accounts may indicate credit experience and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with mo_sin_old_il_acct to create a credit history ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `mo_sin_old_il_acct`

Feature: `tot_hi_cred_lim`

- **Reason:** Higher total credit limit may indicate better credit management and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with mort_acc to create a credit diversity feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `mort_acc`

Feature: `mo_sin_old_il_acct`

- **Reason:** Longer history of installment accounts may indicate credit experience and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with mo_sin_old_rev_tl_op to create a credit history ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `mo_sin_old_rev_tl_op`

Feature: `mths_since_recent_bc`

- **Reason:** Longer time since recent bankcard account opening may indicate stability and lower default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `-`

Feature: `earliest_cr_line`

- **Reason:** Longer credit history may indicate credit experience and lower default risk.

- **Suggested Transformations:** `Interaction with mo_sin_old_rev_tl_op to create a credit history ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `mo_sin_old_rev_tl_op`

Feature: `pct_tl_nvr_dlq`

- **Reason:** Higher percentage of trades never delinquent may indicate good credit behavior and lower default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** -

Feature: `percent_bc_gt_75`

- **Reason:** Higher percentage of bankcard accounts with high utilization may indicate financial stress and higher default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship. Interaction with total_bc_limit to create a credit utilization ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `total_bc_limit`

Feature: `bc_open_to_buy`

- **Reason:** Higher open-to-buy on revolving bankcards may indicate better credit management and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with total_rev_hi_lim to create a credit utilization ratio feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `total_rev_hi_lim`

Feature: `total_acc`

- **Reason:** Higher total number of credit lines may indicate credit experience and lower default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** -

Feature: `mort_acc`

- **Reason:** Higher number of mortgage accounts may indicate stability and lower default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship.  Interaction with tot_hi_cred_lim to create a credit diversity feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `tot_hi_cred_lim`

Feature: `total_il_high_credit_limit`

- **Reason:** Higher total installment credit limit may indicate better credit management and lower default risk.

- **Suggested Transformations:** `Log transformation for skewed distribution. Interaction with mo_sin_old_il_acct to create a credit diversity feature.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** `mo_sin_old_il_acct`

Feature: `mths_since_recent_inq`

- **Reason:** Longer time since recent inquiry may indicate stability and lower default risk.

- **Suggested Transformations:** `Square transformation to capture non-linear relationship.`

- **Relationship Type:** Nonlinear

- **Suggested Interactions:** -