

PreFlight-ML Preparation Report

Dataset Shape: 400 rows × 11 columns

Target: churned

Severity Counts

CRITICAL: 0

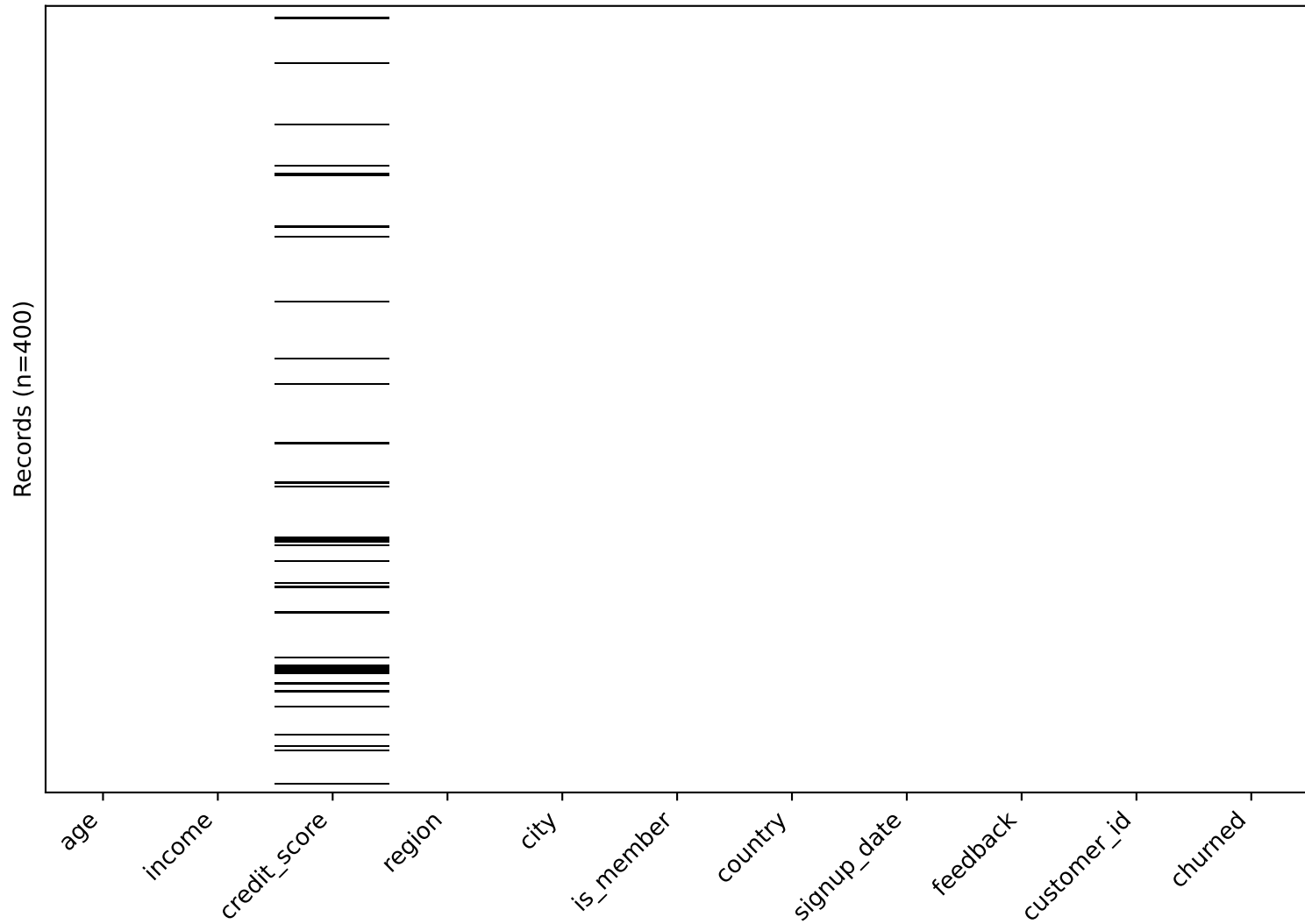
WARNING: 3

INFO: 16

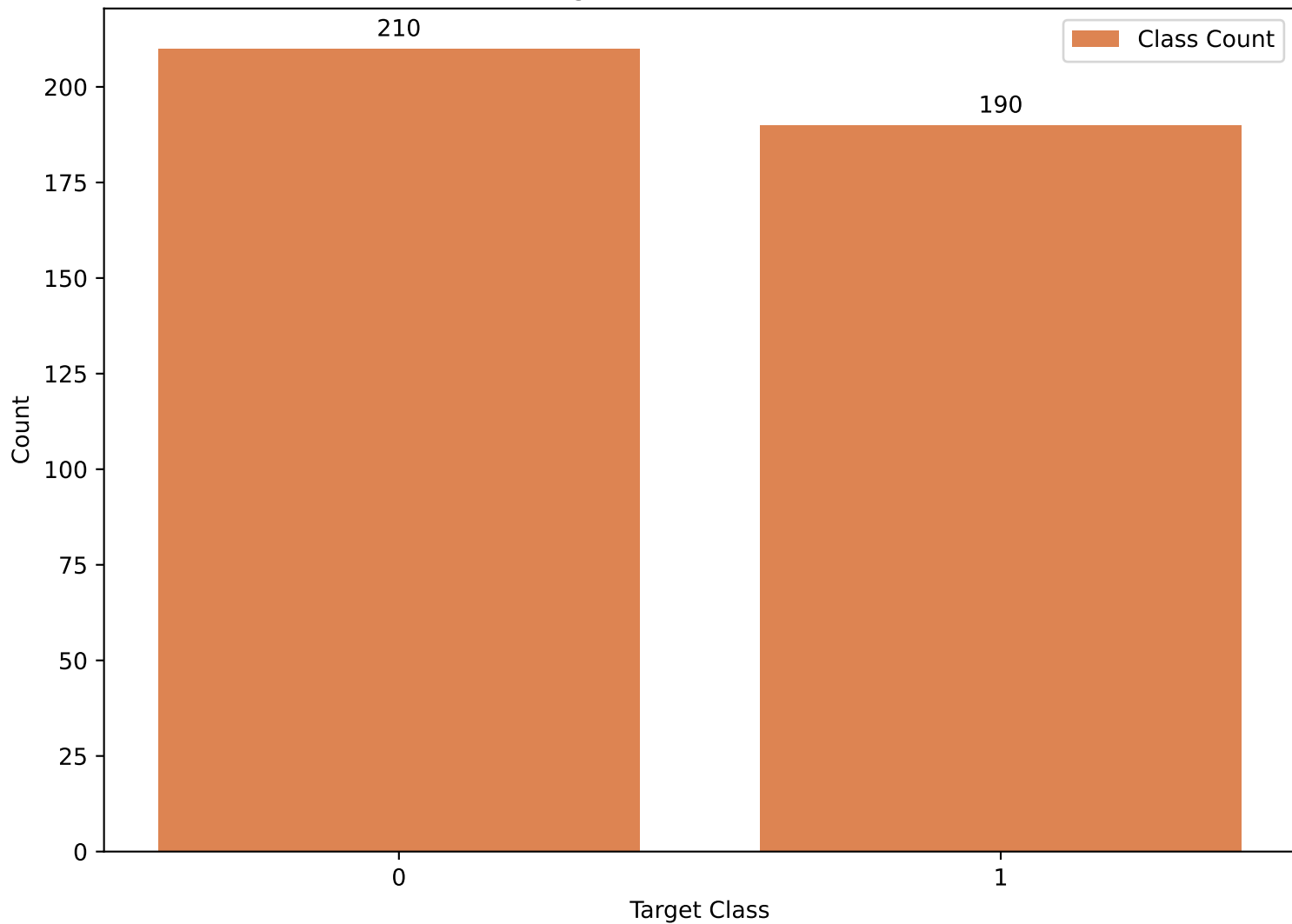
Stage: Profiler

4 decisions logged

Missingness Heatmap (Black = Missing)



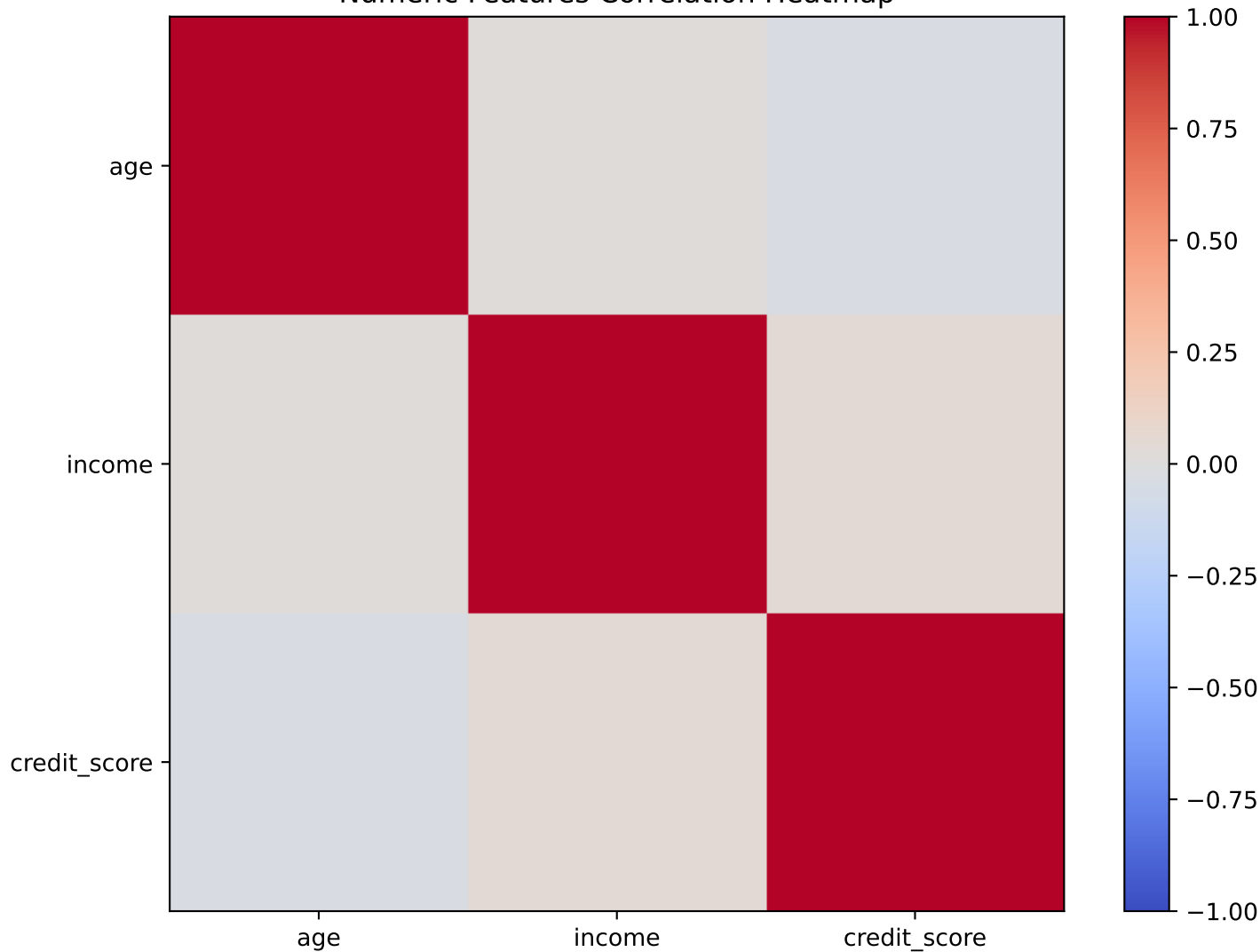
Target Class Distribution



Stage: Cleaner

10 decisions logged

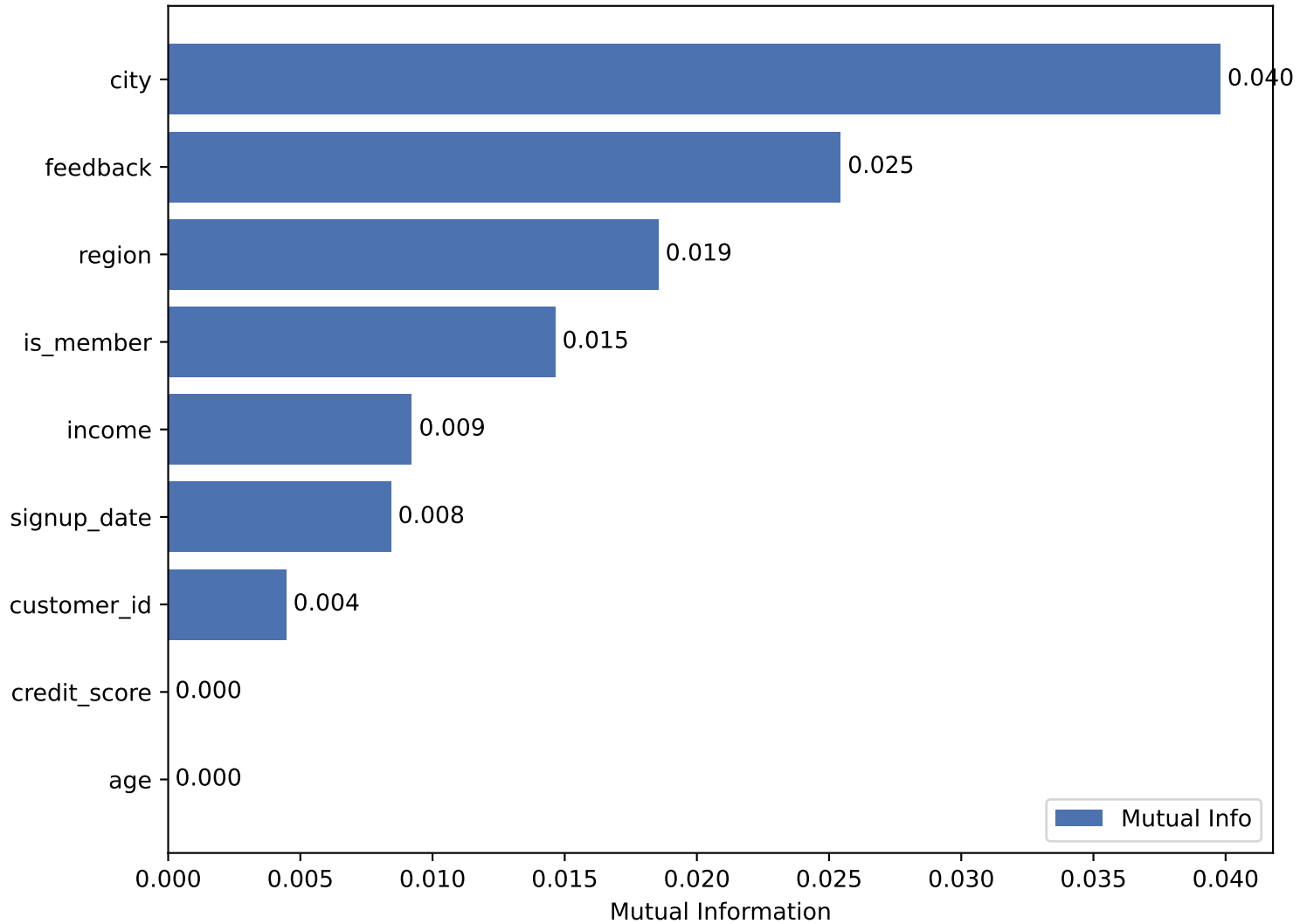
Numeric Features Correlation Heatmap



Stage: Engineer

5 decisions logged

Mutual Information with Target



Appendix: All Report Entries

SEV	COLUMN	ACTION	RATIONALE
WARN	city	flagged_high_cardinality	Cardinality is 60, >= 20
INFO	feedback	flagged_text_column	Column detected as text (high cardinality, high average l...
WARN	customer_id	flagged_high_cardinality	Cardinality is 400, >= 20
INFO	customer_id	manual_semantic_type_o...	column 'customer_id' forced to CATEGORICAL_HIGH (auto-inf...
WARN	country	drop_column	Column is constant
INFO	income	winsorize_outliers	Winsorized 7 outliers via iqr
INFO	credit_score	add_missing_indicator	Missing rate 8.75% > 0
INFO	credit_score	median_impute	Missing rate 8.75% < 30%
INFO	region	normalize_categories	Lowercased and stripped strings
INFO	city	normalize_categories	Lowercased and stripped strings
INFO	city	group_rare_categories	Grouped 4 rare categories
INFO	signup_date	coerce_datetime	Parsed string dates to datetime64
INFO	customer_id	normalize_categories	Lowercased and stripped strings
INFO	customer_id	group_rare_categories	Grouped 400 rare categories
INFO	region	ordinal_encode	Applied ordinal encoding for tree model_hint.
INFO	city	target_encode_cross_fit	Applied 5-fold cross-fit target encoding to prevent leaka...
INFO	customer_id	target_encode_cross_fit	Applied 5-fold cross-fit target encoding to prevent leaka...
INFO	feedback	generate_text_features	Generated char_length, word_count, and has_text for TEXT ...
INFO	feedback	generate_text_tfidf	Generated TF-IDF features for top 5 terms for 'feedback'.