Statistical Binning of Numeric Risk Factors

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

Statistical Binning in Credit Risk

Binning is not a mandatory step in model development but does offer several advantages over the use of continuous risk factors.

The benefits mainly include:

- the reduction of outliers
- the ability to justify the empirical relationship between the risk factor and the target variable
- the involvement of business judgment in the binning process.

Principles of Good Binning Process

The most commonly used principles of the good binning process:

- each bin should contain at least 5% of the observations
- each bin should contain at least one bad case
- adjacent bins should have different riskiness levels
- risk level of the bins should have either a monotonic or U-shape trend
- number of bins should not be greater than ten.

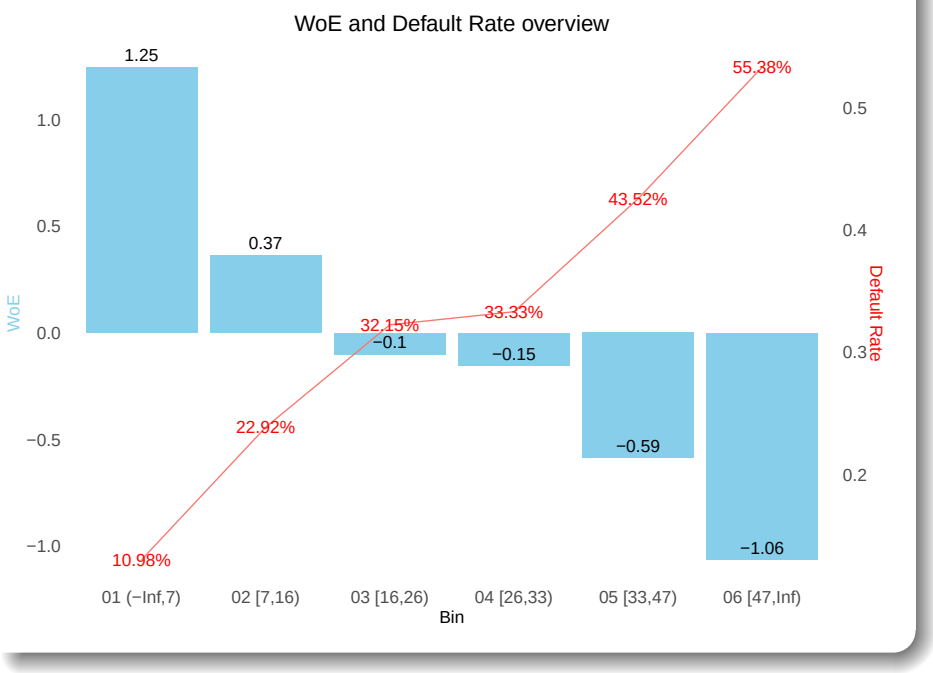
Do we have to follow all the principles?

How does binning link with the other modeling steps (e.g., encoding or multivariate analysis)?

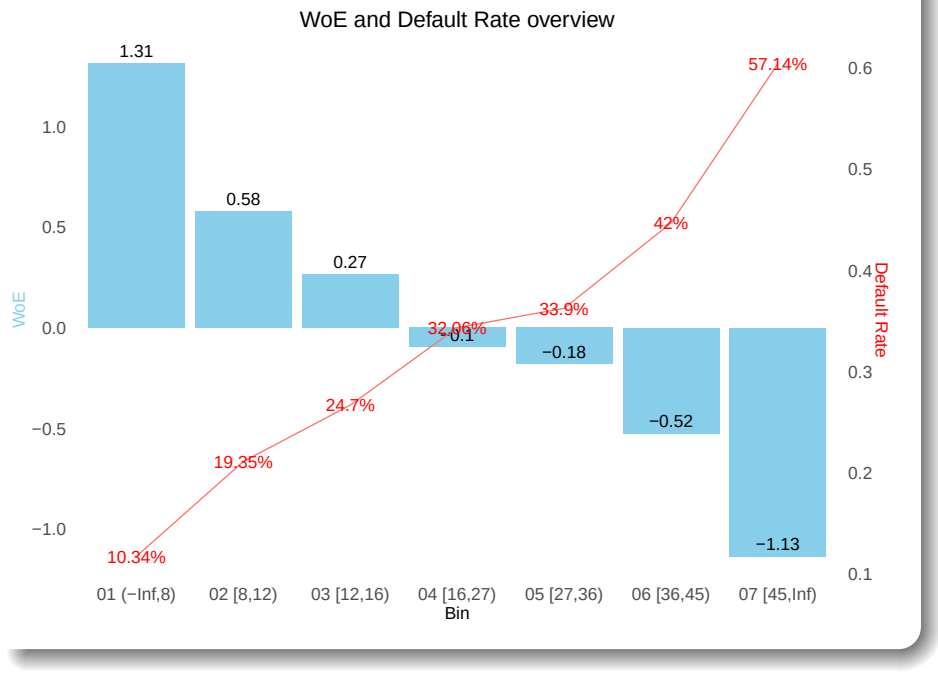
Monotonic Binning

Different binning algorithms exist and some of them are readily available in `R` and `Python` (e.g., monobin and monobinpy).

Monotone Adjacent Pooling Algorithm (MAPA)



Isotonic Regression



U-shape Binning

U-shape trend is sometimes a desired property of the relationship between the risk factor and the target variable. This relationship is present if we find a point in our risk factor (usually called an inflection point), after which the relationship changes direction.

Binning process

- Determine inflection point (expertly, statistically or both)
- Perform binning for values before and after inflection point.

B-spline basis functions can be convinient for testing and determining the inflection point.

Testing and U-shape binning available in `R` package PDtoolkit via functions `ush.test` and `ush.bin`.

U-shape Binning based on Isotonic Regression