



# Weight of Evidence

## “Gotcha”

# Weight of evidence

$$WoE = \ln\left(\frac{\text{Proportion of + class}}{\text{Proportion of - class}}\right)$$

- Numerator: sum of + obs per category / total + obs
- Denominator: sum of - obs per category / total - observations

# Weight of evidence: gotchas

$$WoE = \ln\left(\frac{\text{Proportion of + class}}{\text{Proportion of - class}}\right)$$

- Division by zero is not defined
- Logarithm of 0 is inf

# Weight of evidence: solution 1

Numerator:  $(\text{sum of } + \text{ obs per category}) + C / (\text{total } + \text{ obs}) + 2C$

Denominator:  $(\text{sum of } - \text{ obs per category}) + C / (\text{total } - \text{ observations}) + 2C$

Category encoders

# Weight of evidence: solution 2

Group infrequent categories → RareLabelEncoder

**Feature-engine**

# THANK YOU

[www.trainindata.com](http://www.trainindata.com)