1. useful goal

Coffee shop: features with high rating coffee shops

Type of coffee? Pastries/food

Location hour

Price

Wifi, ambience

What owners can do??

What owners .. on views

1. Clean data

Filter keyword “coffee” ‘tea’ ‘coffehouse’

Word matrix

Filter stopwords: low frequency(>=10) ; ‘he’ ‘she’ ‘and’

Limitation: serve other item; ignore important attributes; foreign language (translate, spanish, french); not all coffee represent

Words not phrase eg. Not great

Stopwords not perfect

Word frequency counted in absolute sense

1. Examine and visualize data

Plot: distribution of stars/word/log words

Specific word: review1 3 times

2 2 times

3 0 times

Mean occurance: (1+1+0)/3=0.66

Stars: x: star y: mean occur of specific word

1.Kinds of coffee------(relate)-------rating

Stars: x: star y: mean occur of coffee-----------suggestion

1. ...environment service..
2. ...

Interpr picture!!!!!

Suggestion on

Type/ ideal menu:

Environment:

He, she, his, her negative!----service!

1. Model

lasso

1. diagnosis