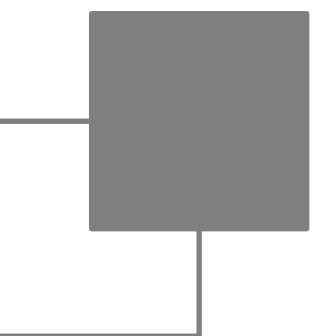


**STAT628 M3**

# **Yelp Reviews Analysis**

Tues Group 8:  
Ruochen Yin  
Jiawen Chen  
Haifeng Liu



1.

# Review

MODELING

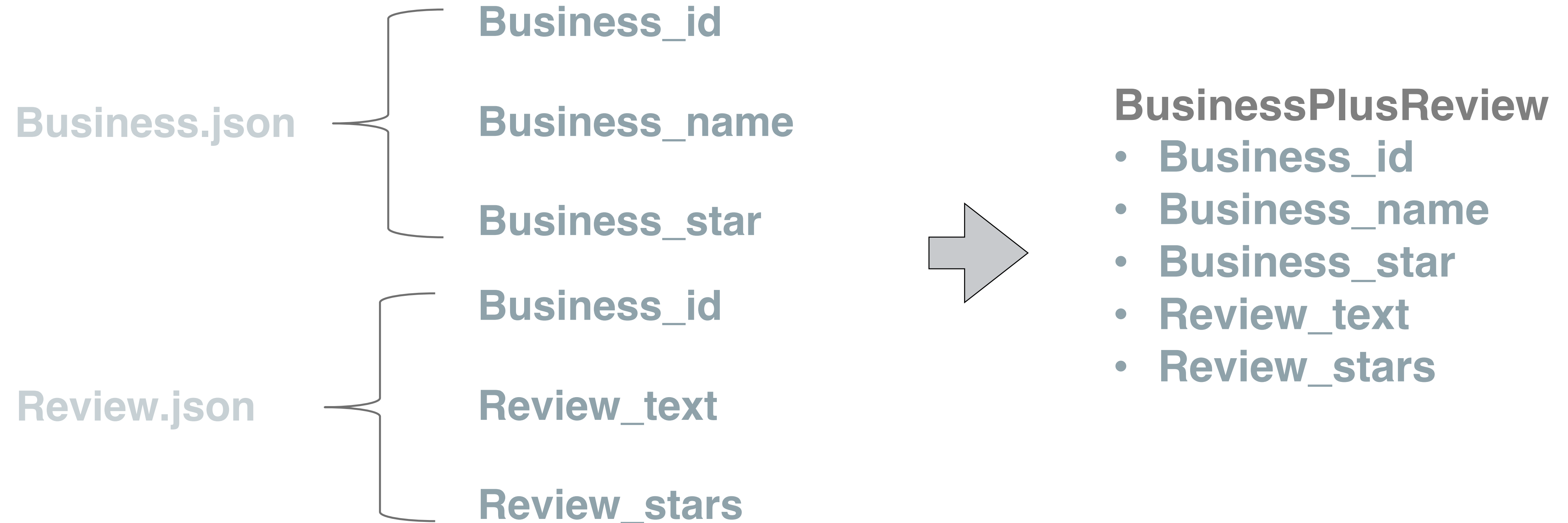
SUGGESTION&EXAMPLE

93488 reviews about restaurants in Wisconsin.

☆: 9195, ☆☆: 9531, ☆☆☆: 13595, ☆☆☆☆: 27924, ☆☆☆☆☆: 33243

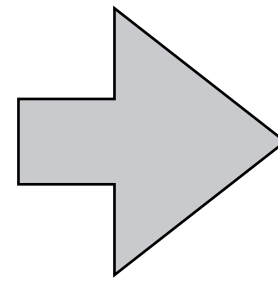
# MODELING

- Create a new dataset

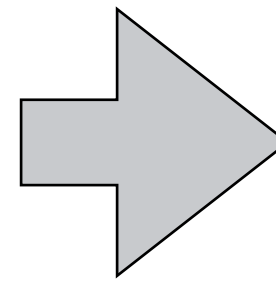


# MODELING

- **Create frequency matrix**



- **Linear regression**

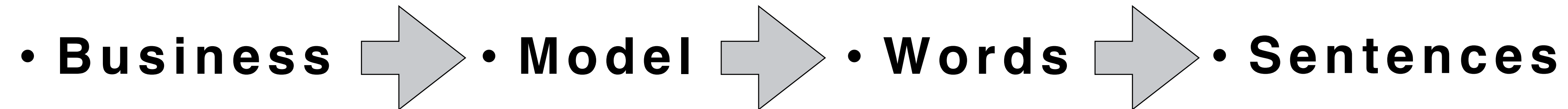


- **Sort words**

# SUGGESTION

- **Give unique advice to different business owners**

We pick out some important sentences in the reviews which will affect the stars of reviews.



# SUGGESTION

- **Example**

**Business name: Ancora Coffee**

**Location: Madison**

**These sentences will affect the increase of business stars:**

1. I've tried at least 3 different versions and they were all fantastic!
2. The coffee is great (love the house dark roast) and the grilled cheese sandwiches are worth every calorie!
3. This always struck me as the most Madison of all Madison Coffee shops.

**These sentences will affect the decrease of business stars:**

1. The beans are a little on the pr.
2. The line was relatively short, which was good because I was ready for my coffee!
3. Ancora makes their own beans, too, which are a regular standby in my house.

2.

# Business

Data processing

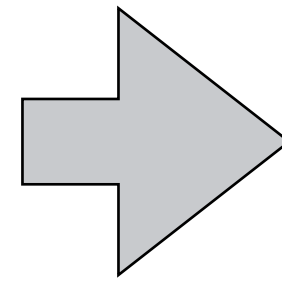
GUIDE

Linear regression

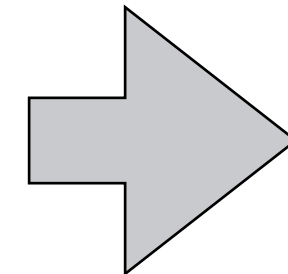
1703 restaurants in Wisconsin.

# Attribute

- **Dictionary to flat list**



- **Unify key word and value**

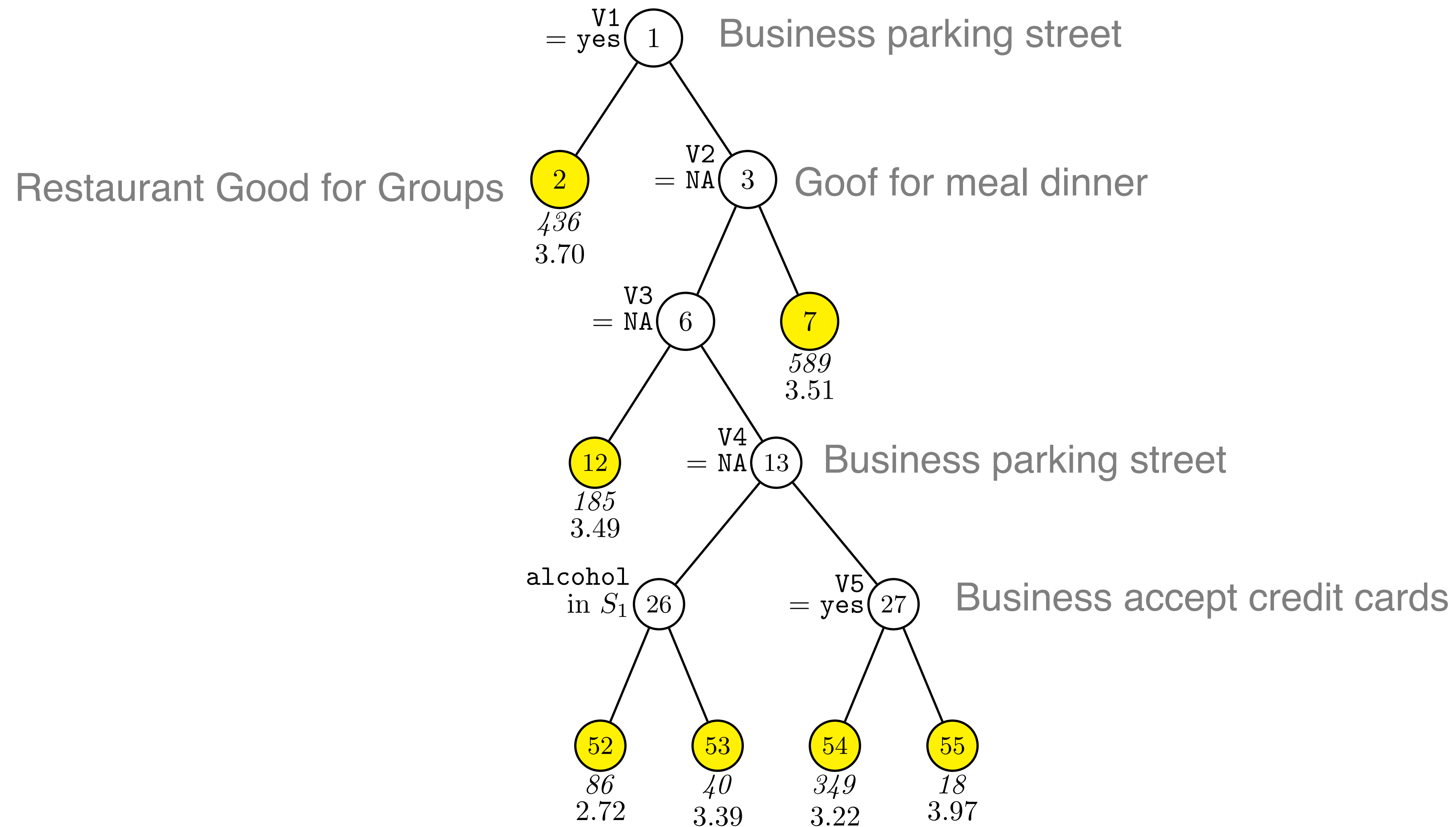


- **Add NA value**

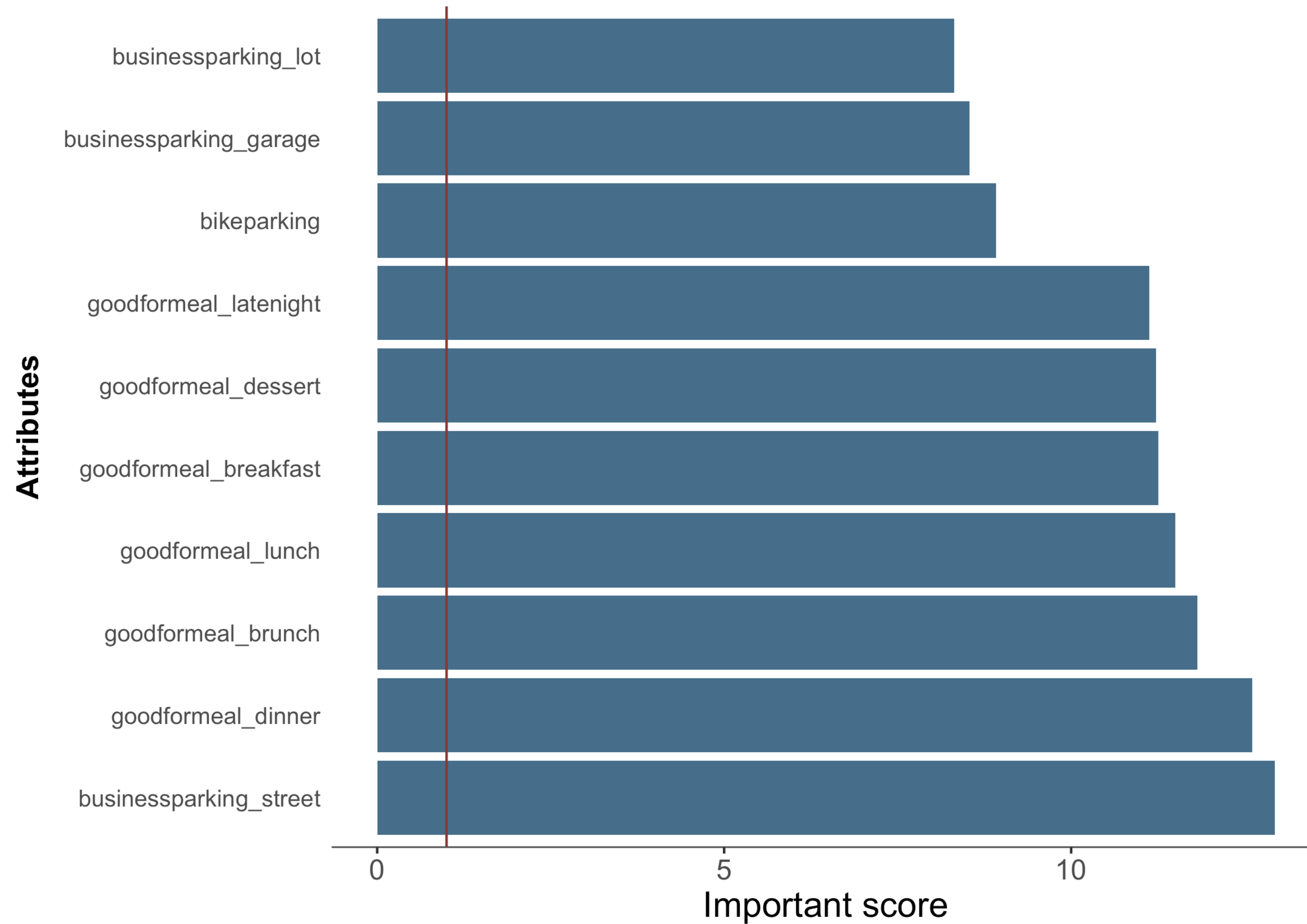
WIFI: free paid yes non NA



# Regression tree



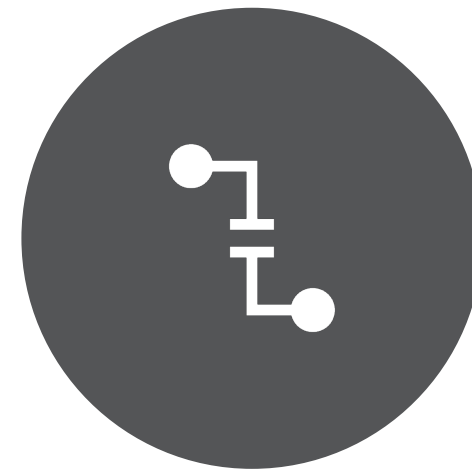
# Importance score



# Linear regression



REPLACE NA  
WITH NEW LEVEL



DELETE  
VARIABLES TO  
PREVENT MULTI-  
COLLINEARITY



FIT MODEL

Linear regression result

Attribute:Level	Estimate	P value
businessparking_street: yes	0.43	5.22E-09
restaurantsattire: new_level	0.42	5.52E-08
noiselevel: very_loud	-0.60	9.30E-06
drivethru: yes	-0.64	1.20E-05
goodformeal_dinner: yes	0.22	1.77E-04
businessparking_street: none	0.24	1.98E-04
goodformeal_dinner: none	0.17	1.15E-03
noiselevel: quiet	0.18	1.53E-03
bikeparking: yes	0.18	1.64E-03
restaurantstableservice: none	0.19	4.33E-03

# Advice

noiselevel: very_loud	-0.60	9.30E-06
-----------------------	-------	----------

"The noise level in your bussiness is too loud  
and it decrease your bussiness review about 0.6 stars. ",

"Your customers are satisfied with the noise level in your  
bussiness. "

Generate a paragraph for each business

# Website

<https://vast-harbor-08232.herokuapp.com>