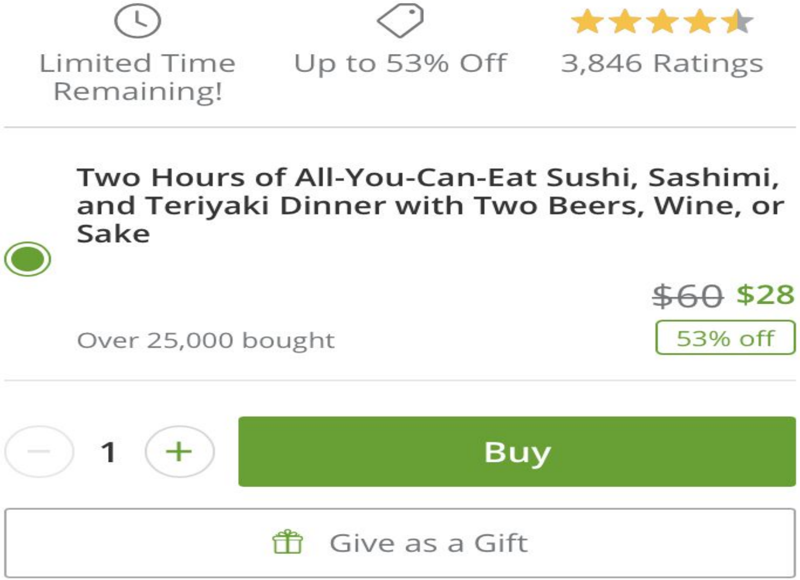
Groupons: Understand and Group

By: Nicholas Maloof

Introduction

What is Groupon?

A website that publishes coupons known as “groupon's” from different venders that allow you to receive a discounted price for certain activities, food purchases, or etc.



Limited Time Remaining!

Up to 53% Off

★★★★★ 3,846 Ratings

● Two Hours of All-You-Can-Eat Sushi, Sashimi, and Teriyaki Dinner with Two Beers, Wine, or Sake

~~$60~~ $28

Over 25,000 bought

53% off

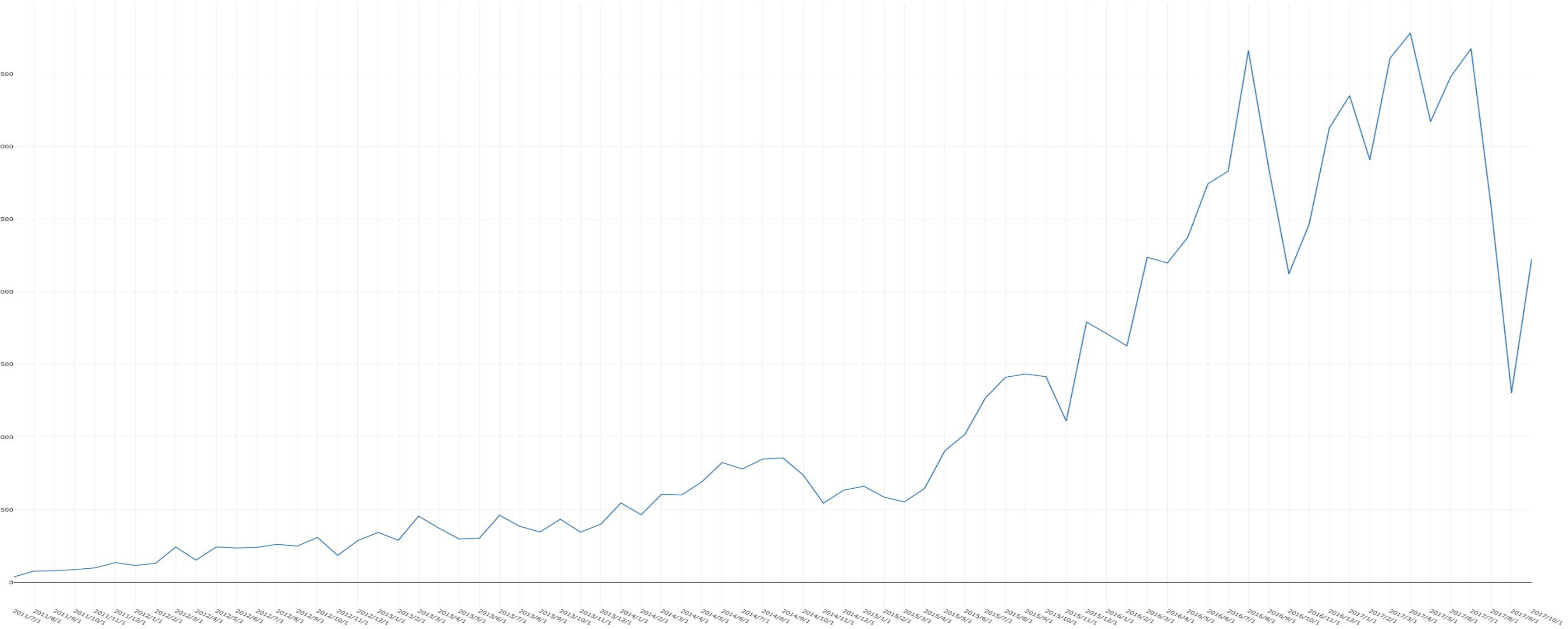
Scraping

Scrapy the Fast and Selenium the Slow

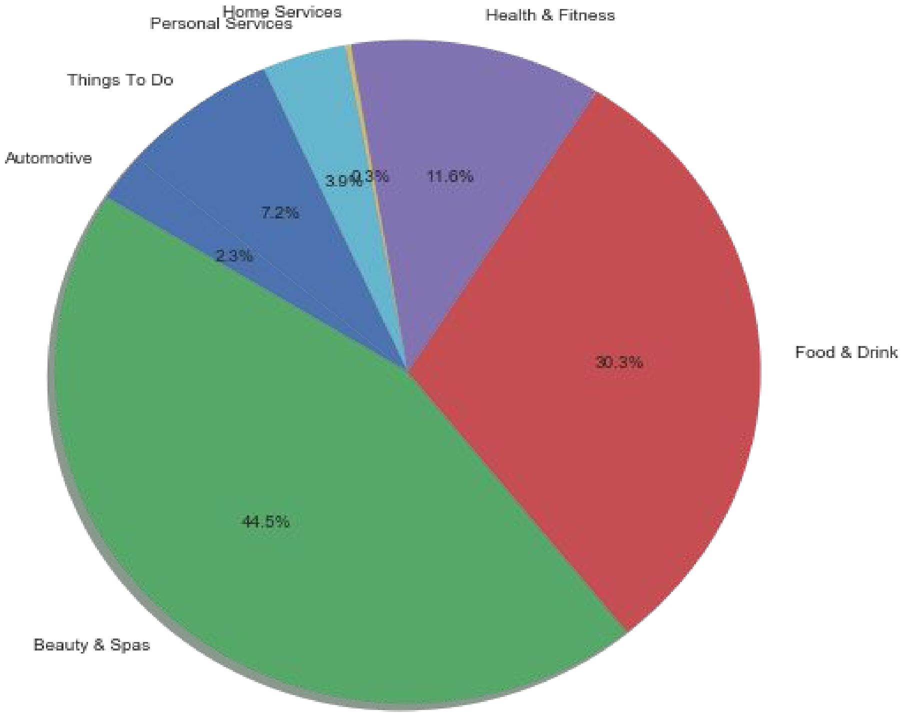
- Scrapy
  - Scraped all the available local groupon data present at the time.
  - Scraped everything important from the individual groupon
  - Compiled a list a urls
- Selenium
  - Using urls, went into each groupon and clicked on the read reviews button
  - Scraped every review and continued along the pages until the “end”

Data Analysis

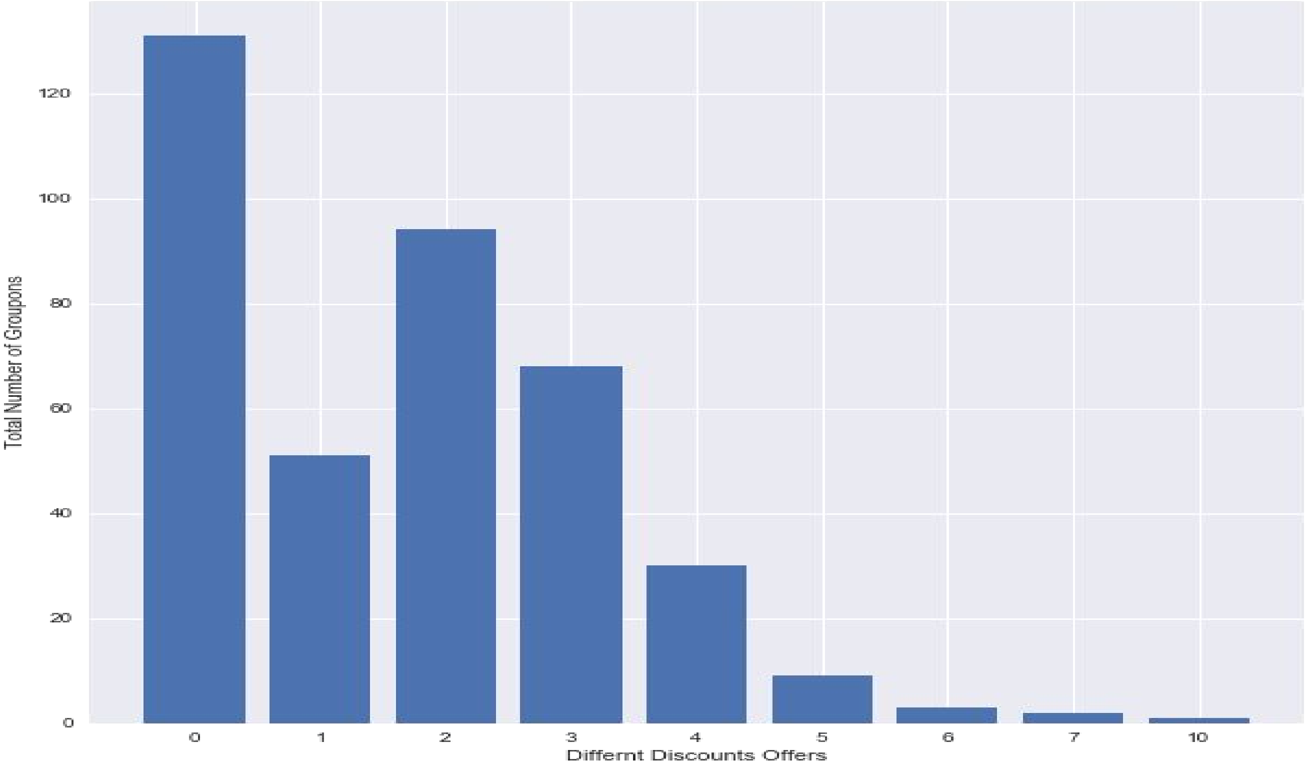
More People are Using Groupon



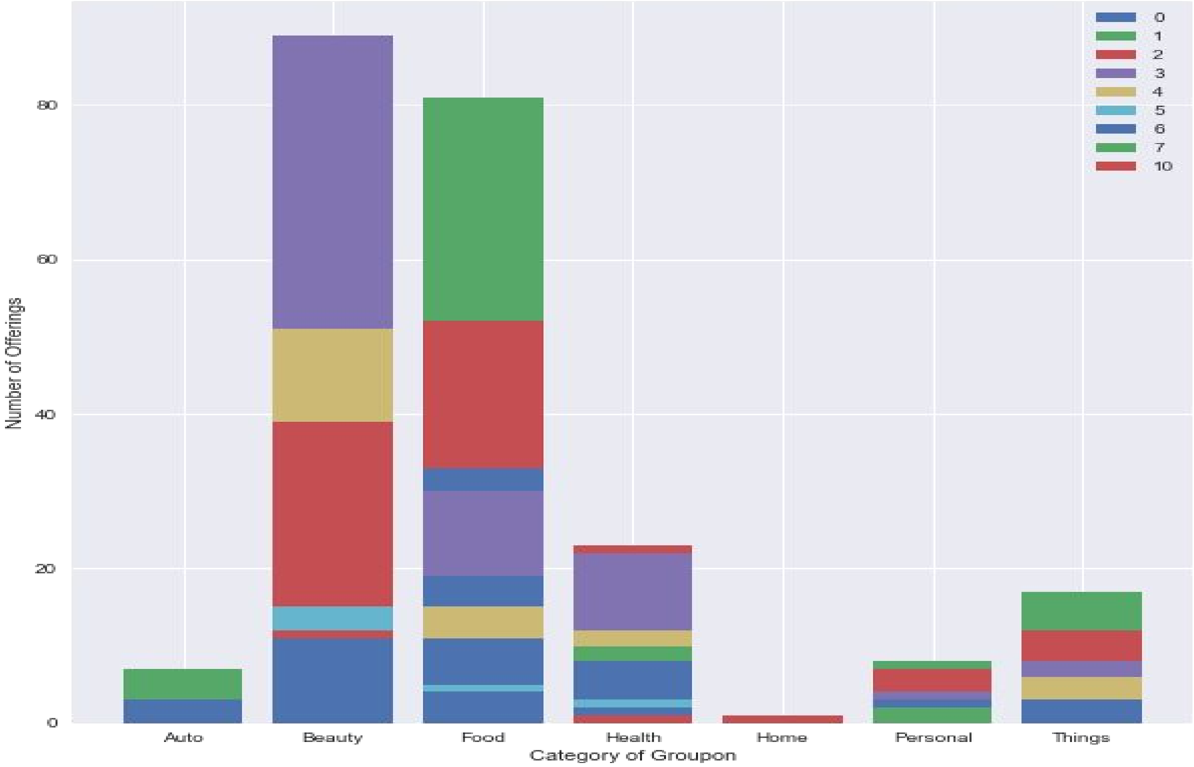
Distribution of Categories



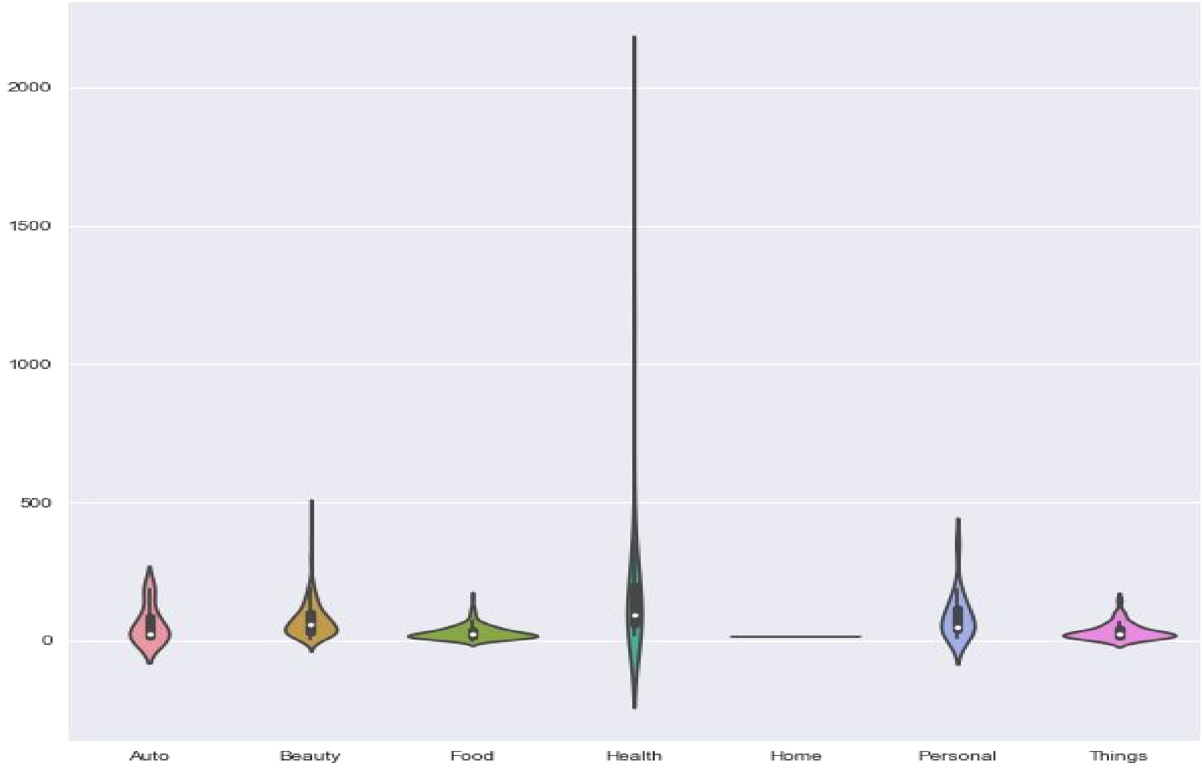
Groupon’s Have Differing Deal Offers

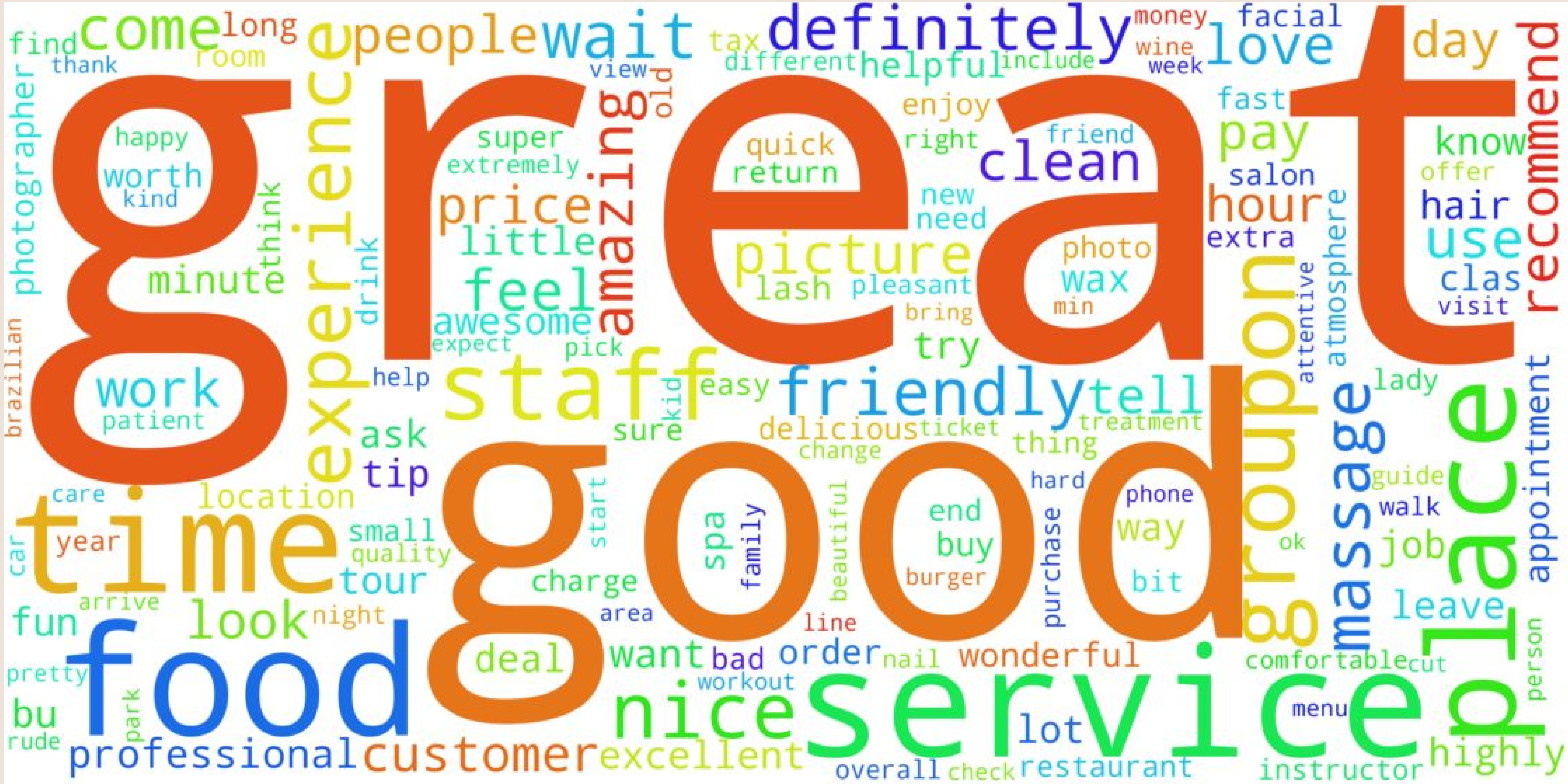


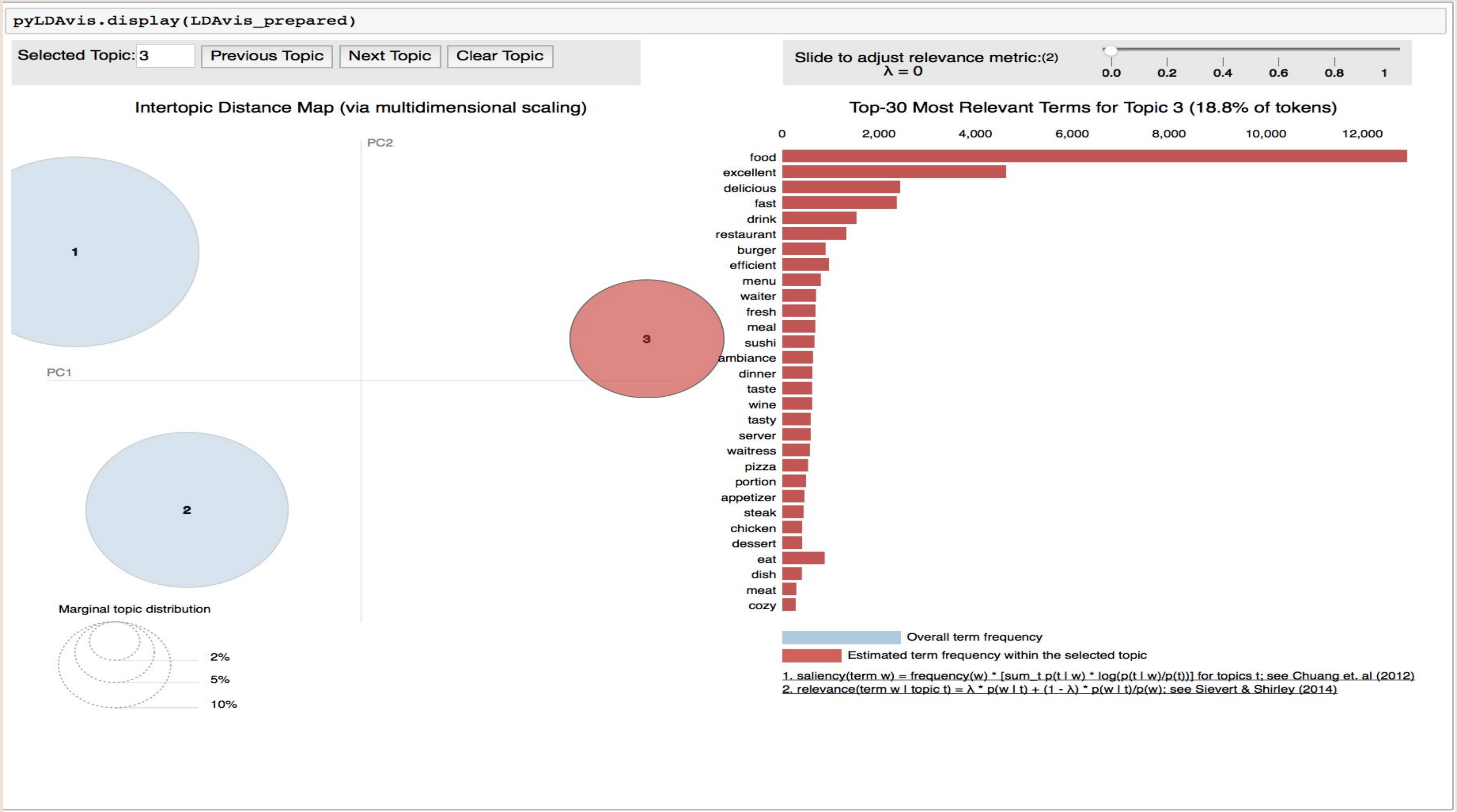
Offering Counts vs. Category



Distribution of the Savings




great
good
service
place
food
time
staff
experience
definitely
come
long
people
wait
love
day
money
wine
facial
week
find
thank
room
tax
different
helpful
include
view
old
photographer
recommend
groupon
massage
friendly
nice
customer
professional
excellent
picture
clean
amazing
price
feel
hour
pay
use
hair
know
fast
friend
salon
offer
extra
clas
visit
atmosphere
attentive
appointment
happy
worth
kind
super
extremely
quick
return
enjoy
right
new
need
little
minute
think
drink
awesome
expect
pick
lash
pleasant
photo
wax
bring
min
try
work
patient
help
kid
easy
tell
lady
ask
sure
delicious
ticket
treatment
thing
change
tip
location
care
small
quality
year
car
tour
charge
spa
family
beautiful
end
buy
purchase
way
hard
phone
guide
walk
job
ok
bit
burger
leave
arrive
fun
look
night
area
line
brazilian
start
deal
want
bad
order
nail
wonderful
comfortable
cut
person
pretty
workout
bu
park
menu
lot
rude
overall
check
restaurant
instructor
highly



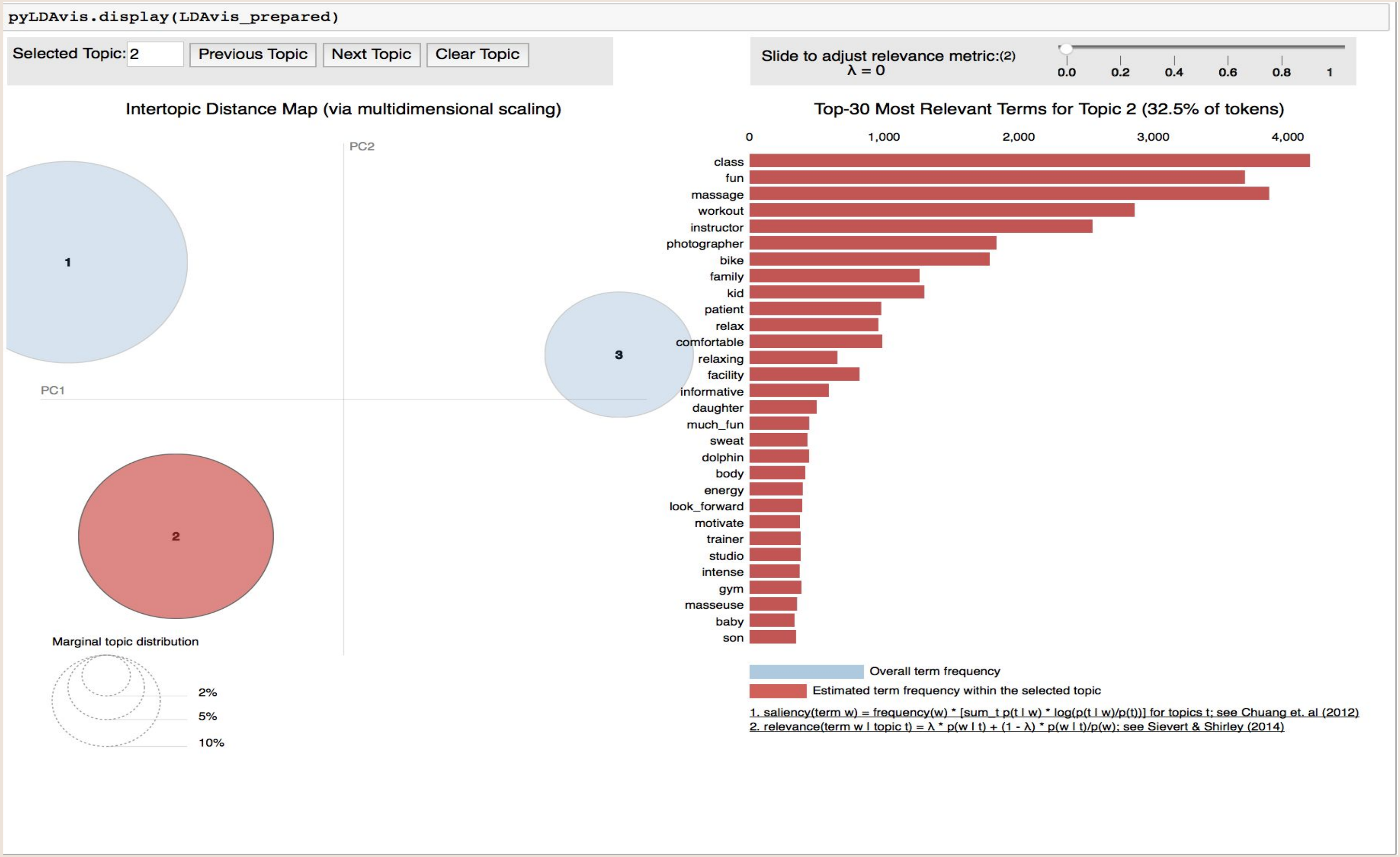
```
pyLDAvis.display(LDAvis_prepared)
```

Selected Topic: 3

Slide to adjust relevance metric:(2) λ = 0

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

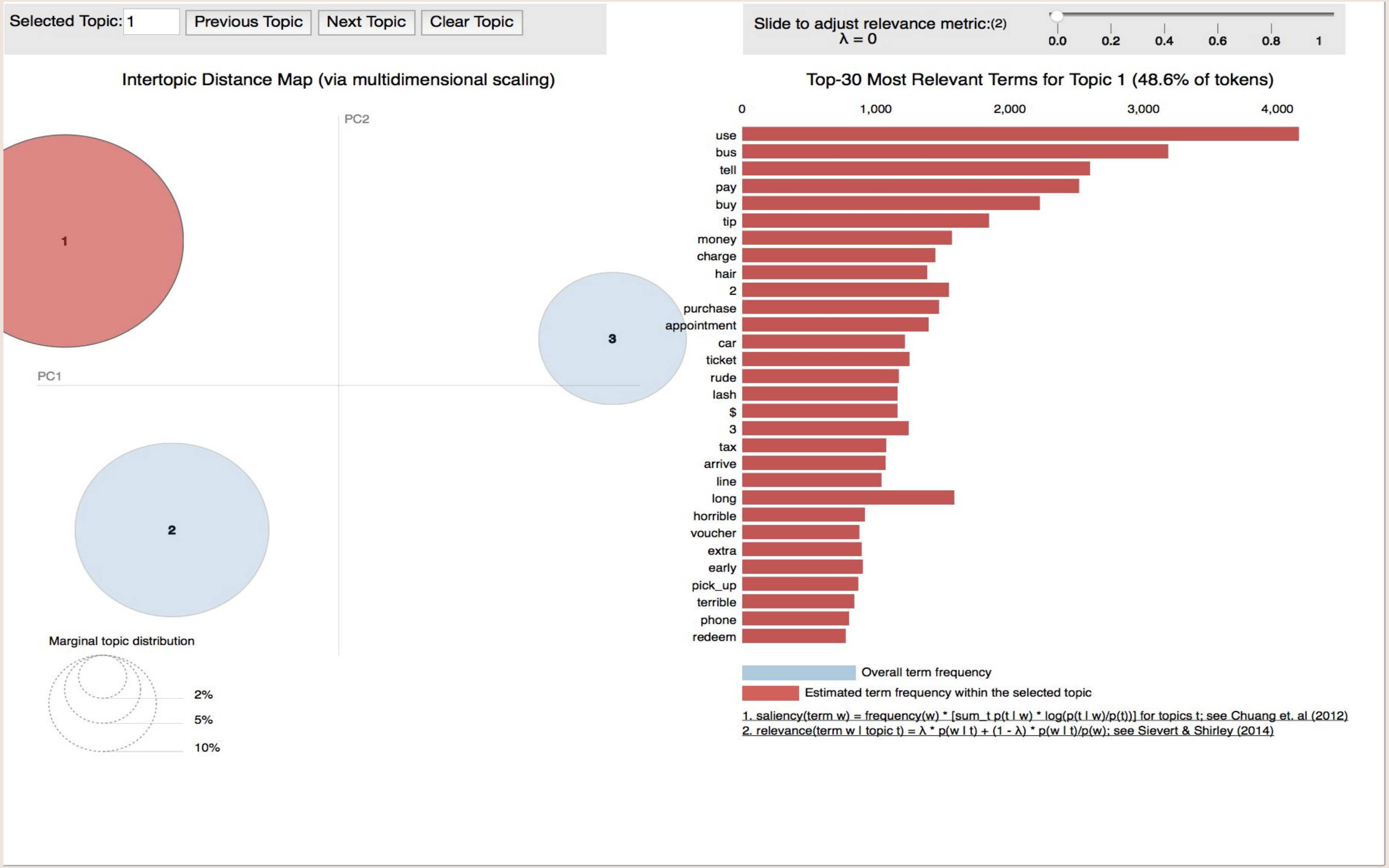


```
pyLDAvis.display(LDAvis_prepared)
```

Selected Topic: 2

Slide to adjust relevance metric:(2)
λ = 0

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)



Selected Topic: 1

Slide to adjust relevance metric:(2)
λ = 0

Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 1 (48.6% of tokens)

🟦 Overall term frequency

🟥 Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)