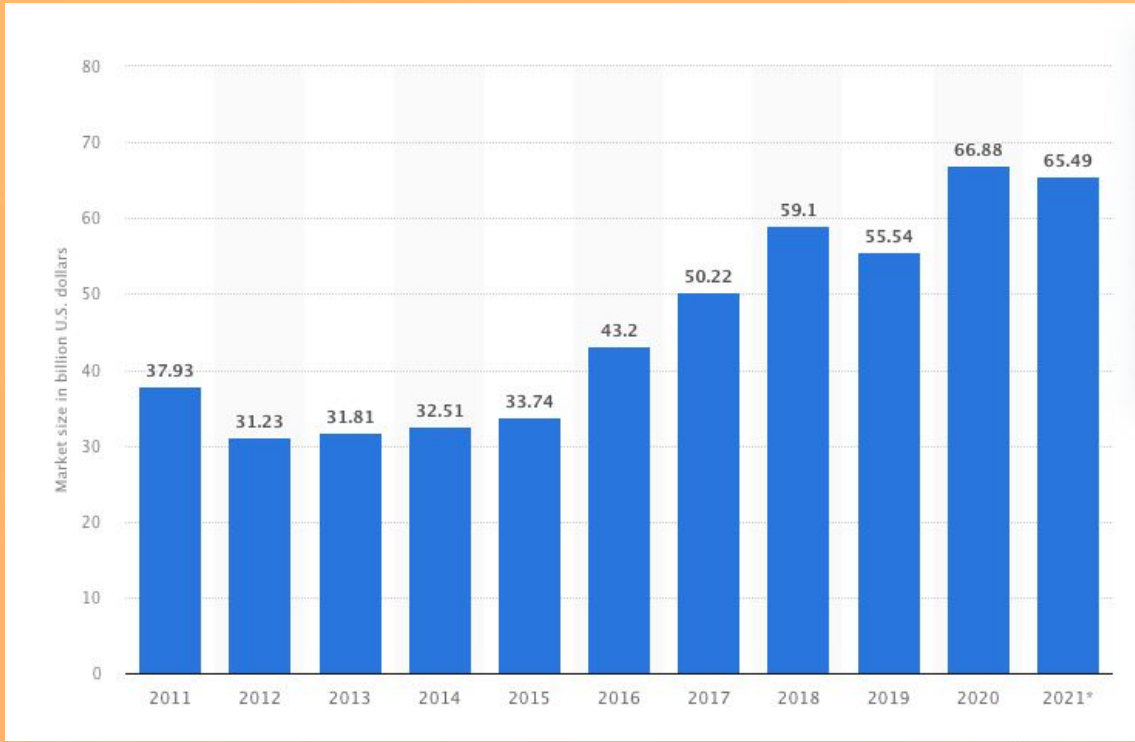
# Video Games

A recommendation system for marketing team to analyze reviews

By Jason W

# Video Games Industry Market Size




Video Games Industry has been rapidly growth over the past decade and almost double the market size from 2011 to 2021

# PROBLEM

## Pricing

Too many similar products on the market with competitive pricing

## Quality

Inconsistencies, glitches, bugs, bad quality assurance

## Lifecycle

Some products are designed for only short term

# GOAL

Build a **Recommendation System** to helps marketing team

Figure out what users most **important needs** for each product category

**Competitive** product features advantage among competitors

Longevity of product **lifecycle** by picking the right materials

# DATA & PRE-PROCESSING

## Dataset

~2M Video Games reviews from Amazon (picking 10k reviews randomly from the dataset)

## Tools

Python
Pandas
Numpy
NLTK
Scikit-learn

## Text Processing

CountVectorizer
Stemming
Lemmatization

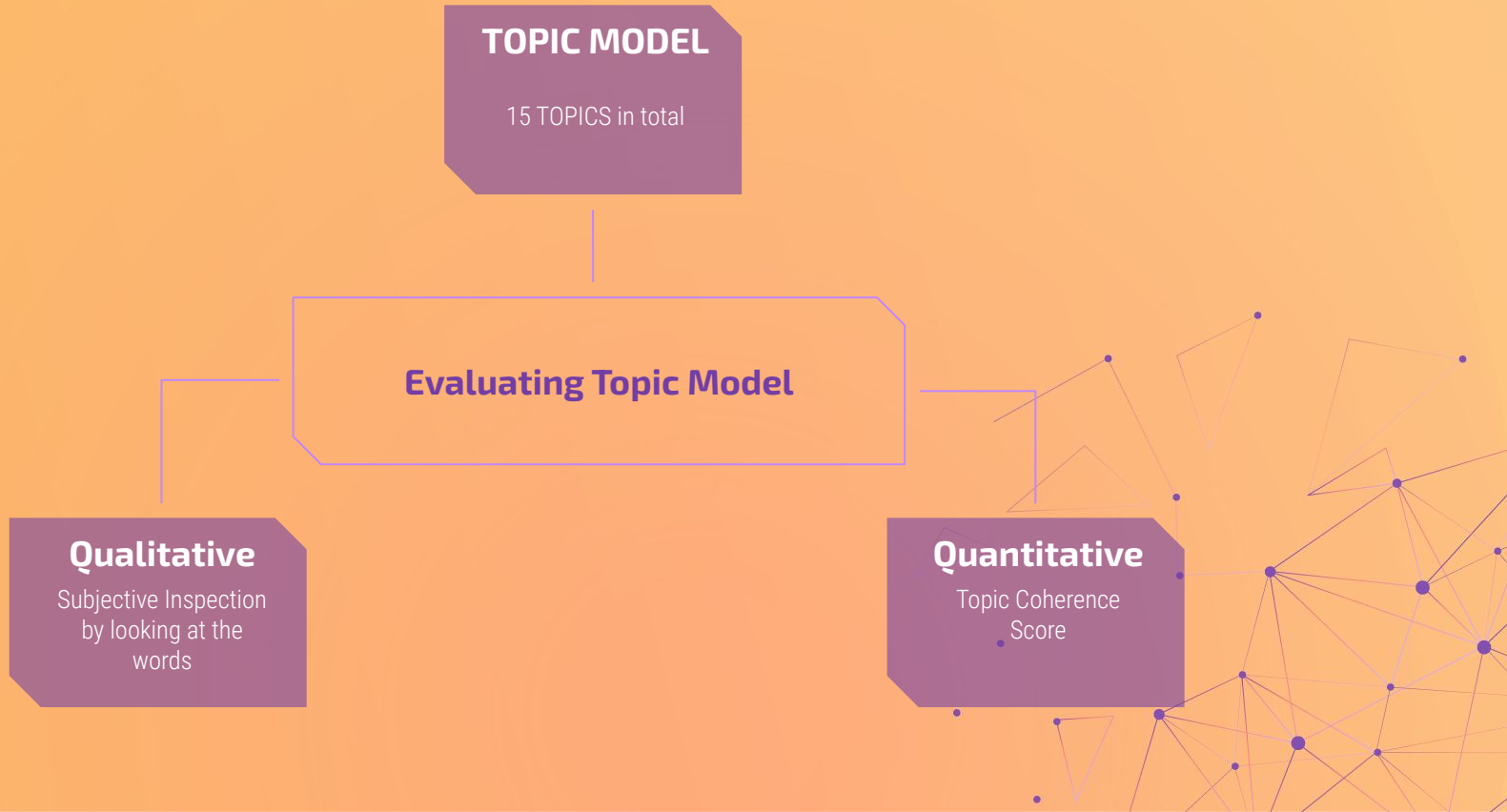
## Topic Model

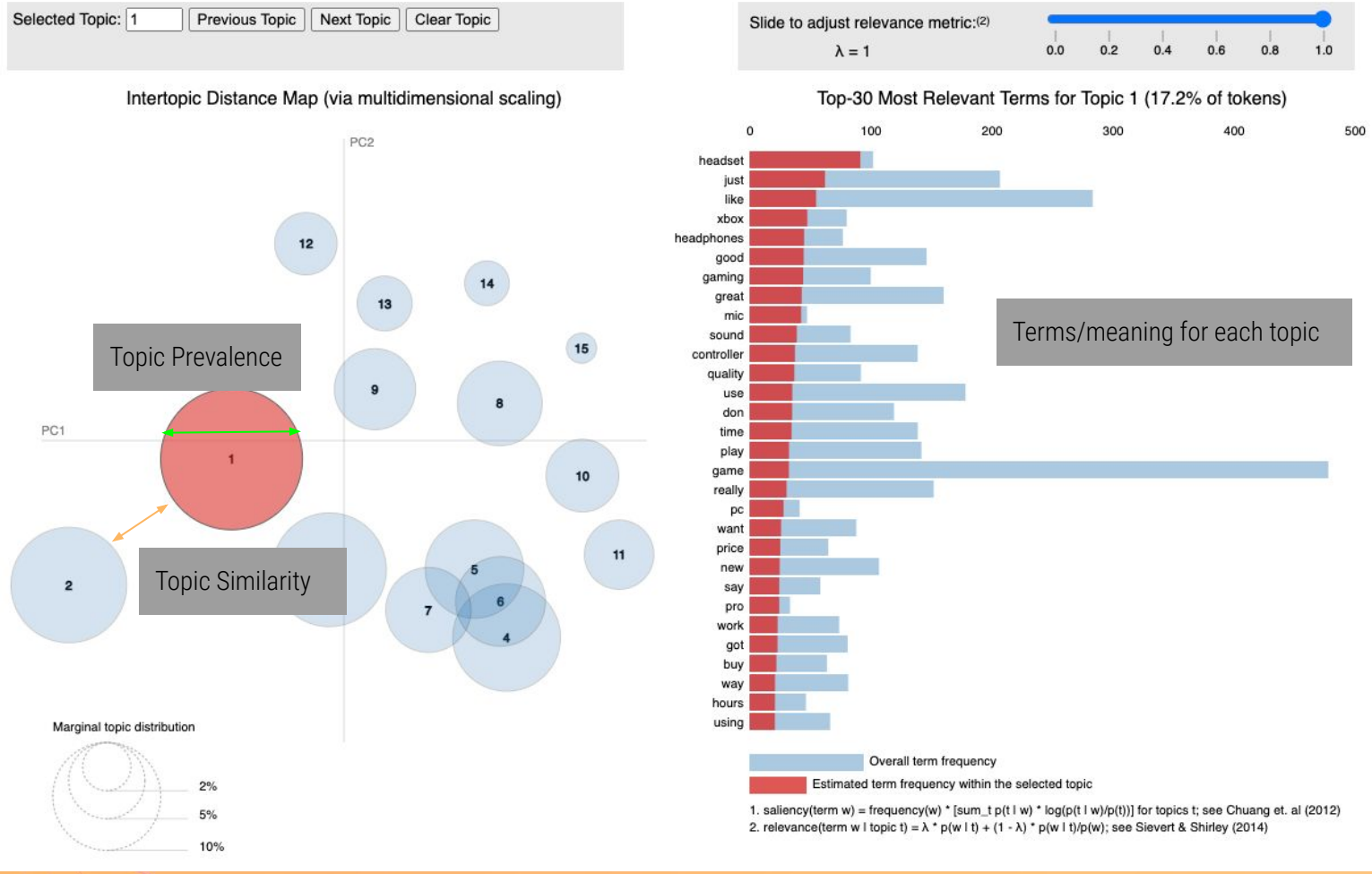
NMF
LDA

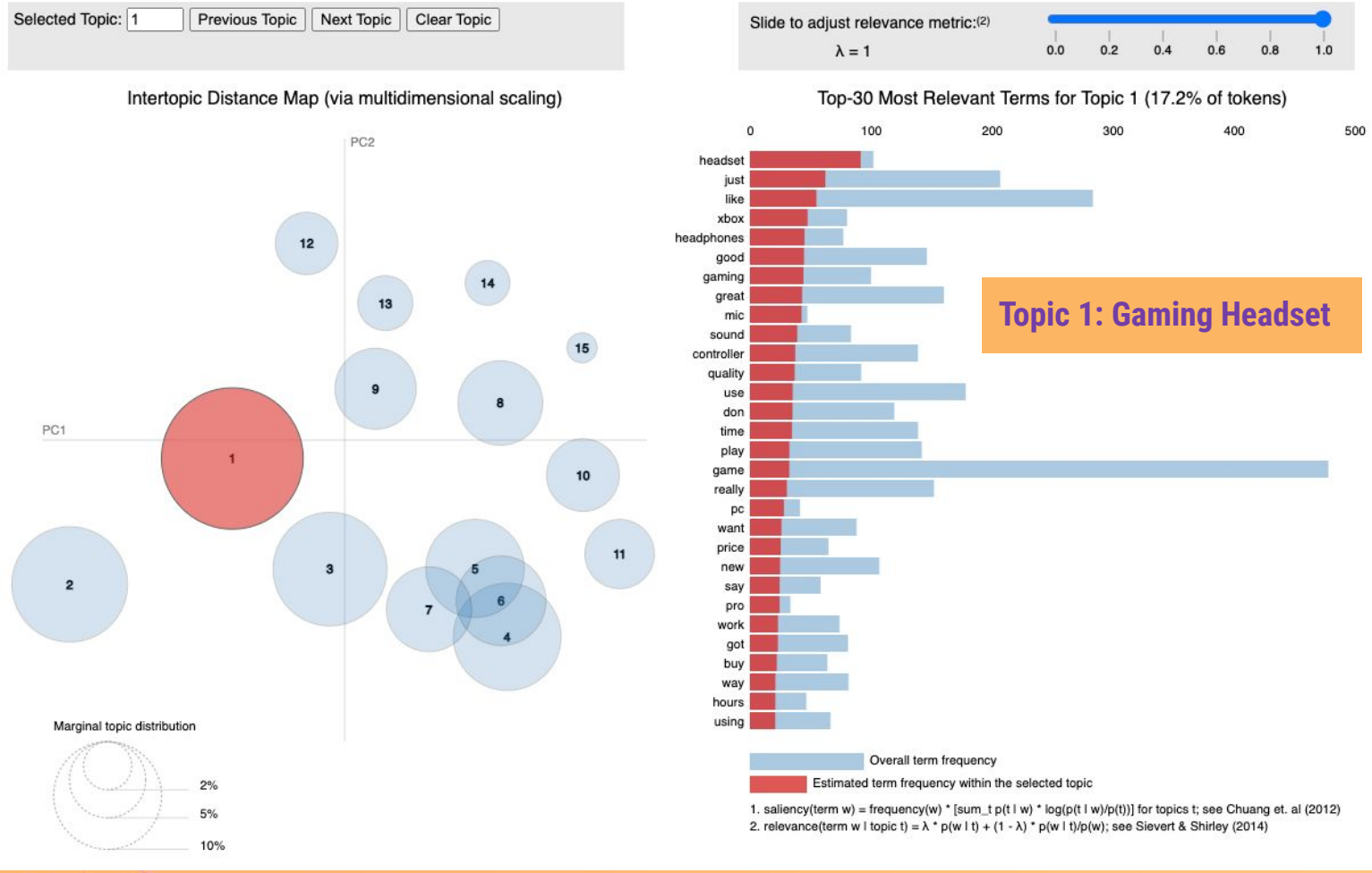
## Eval & Vis

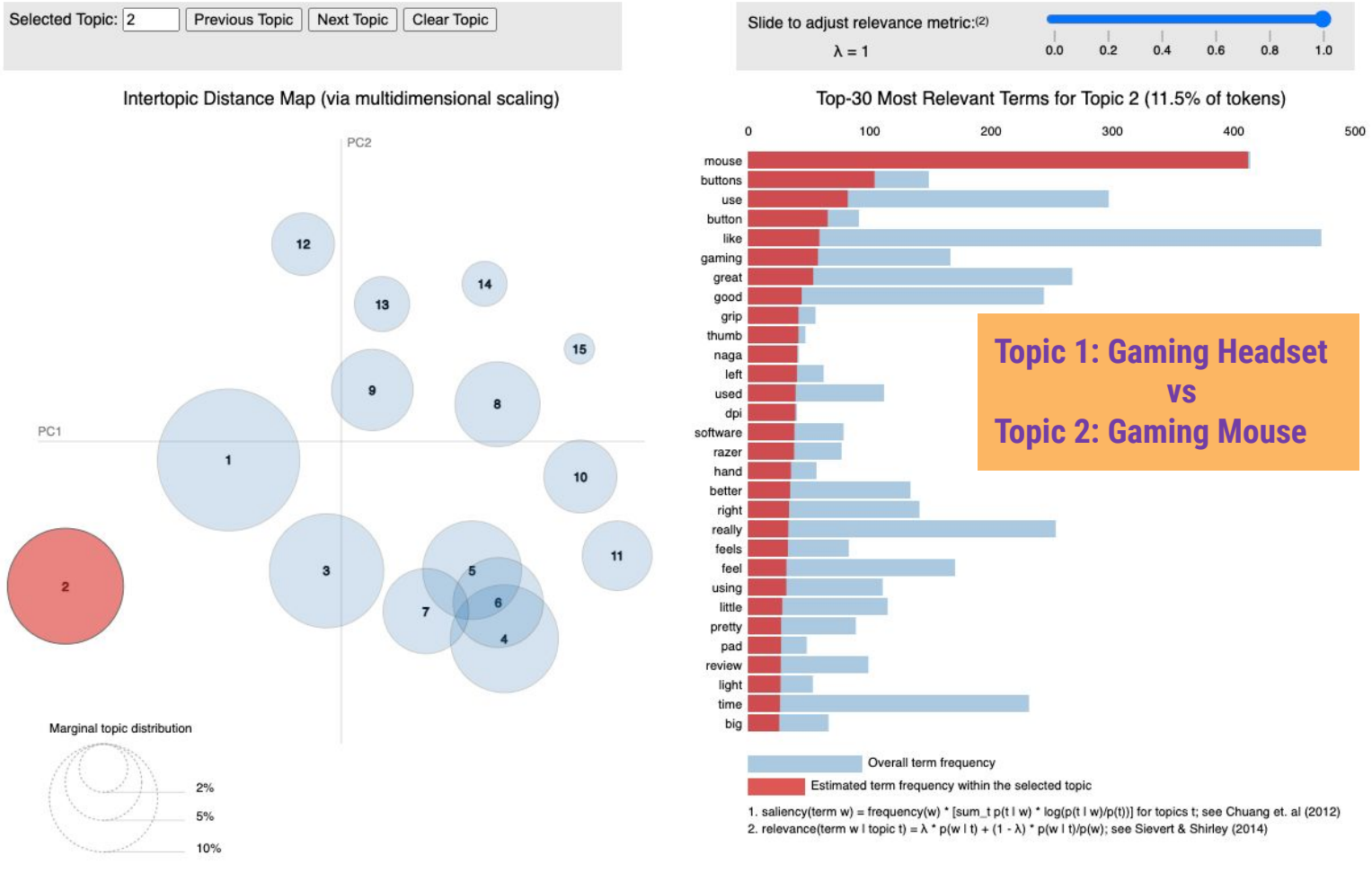
Matplotlib
Seaborn
WordCloud
pyLDAvis

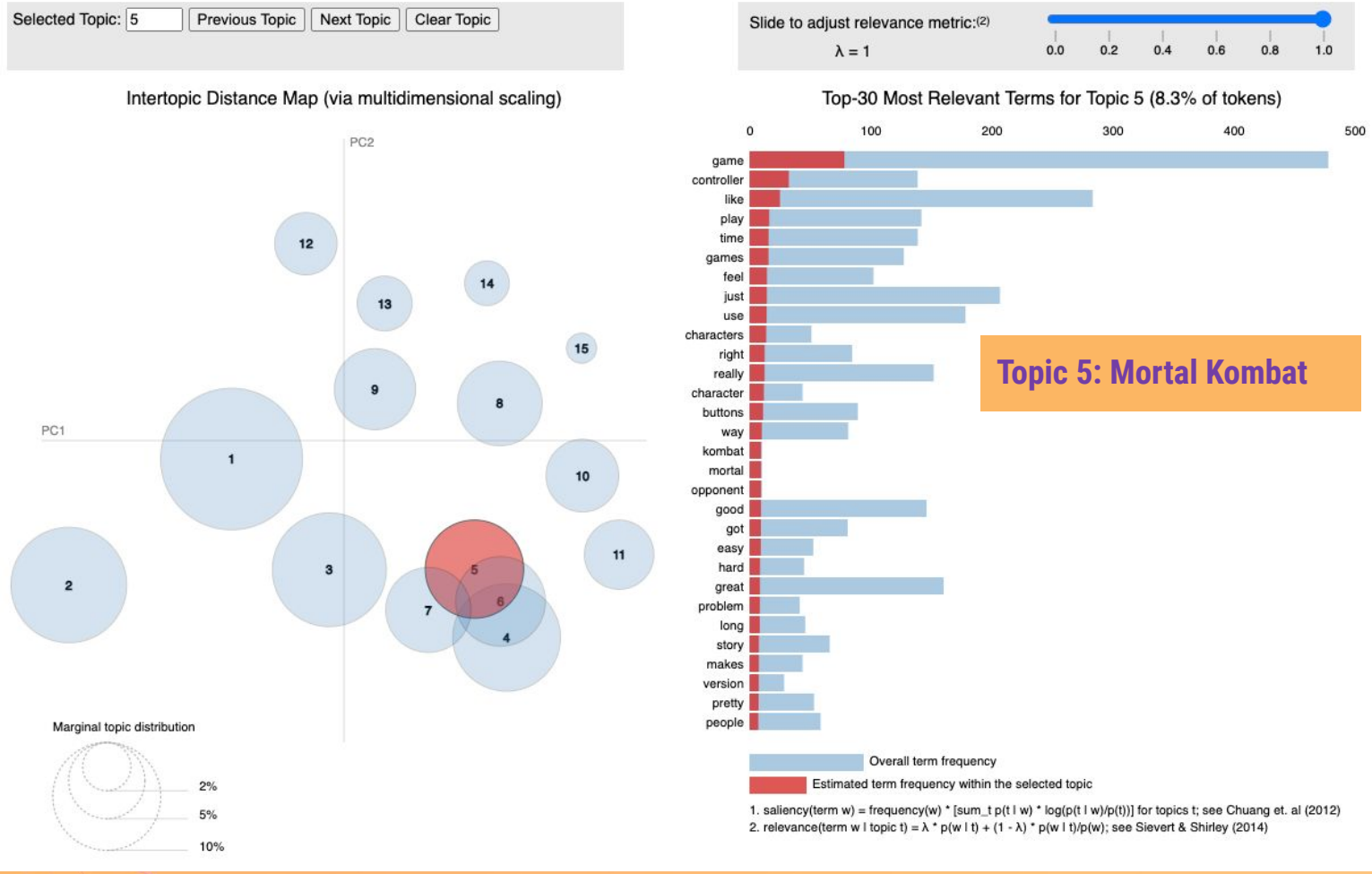
# TOPIC MODEL

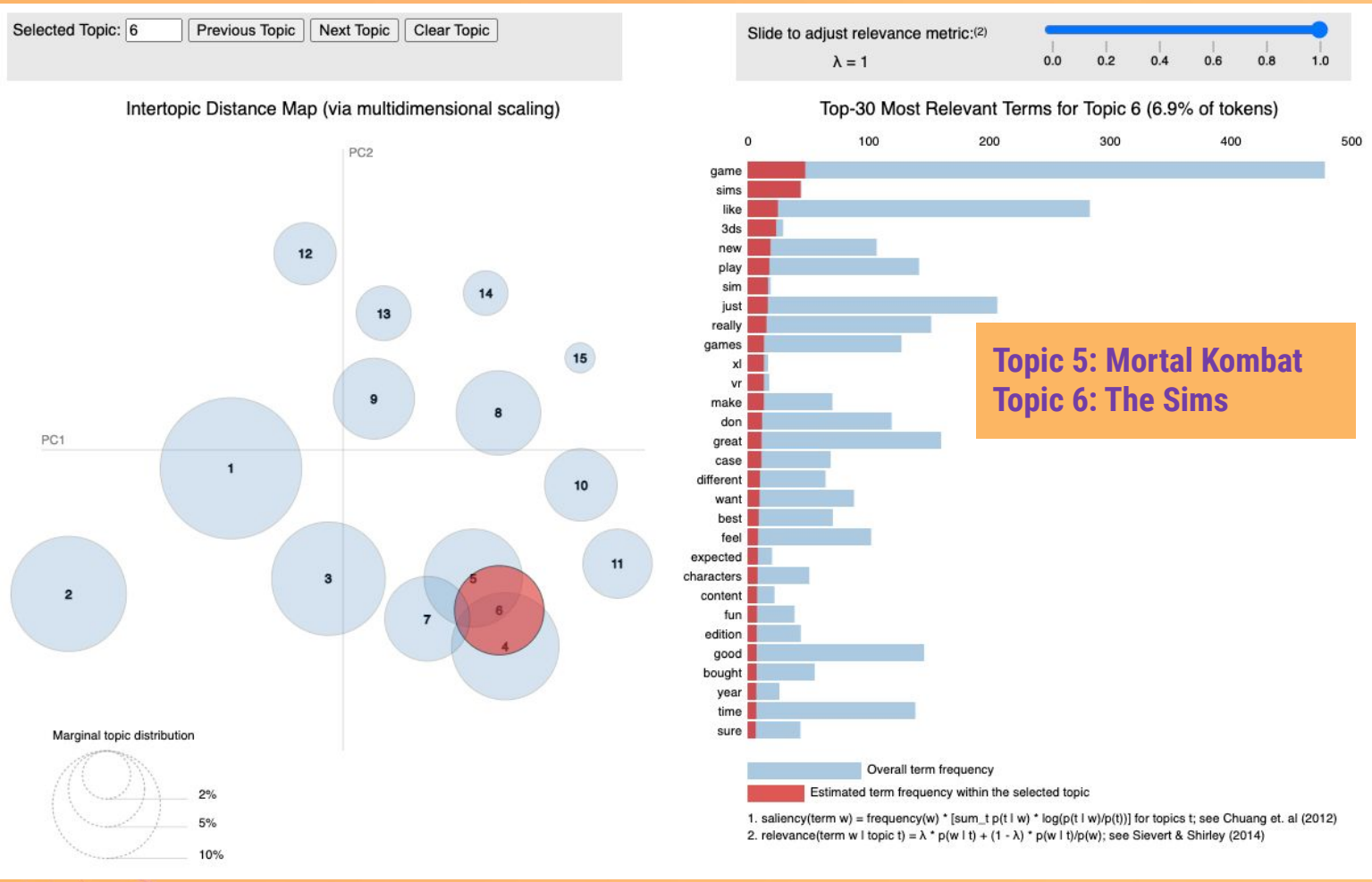

TOPIC MODEL
15 TOPICS in total
Evaluating Topic Model
Qualitative
Subjective Inspection by looking at the words
Quantitative
Topic Coherence Score

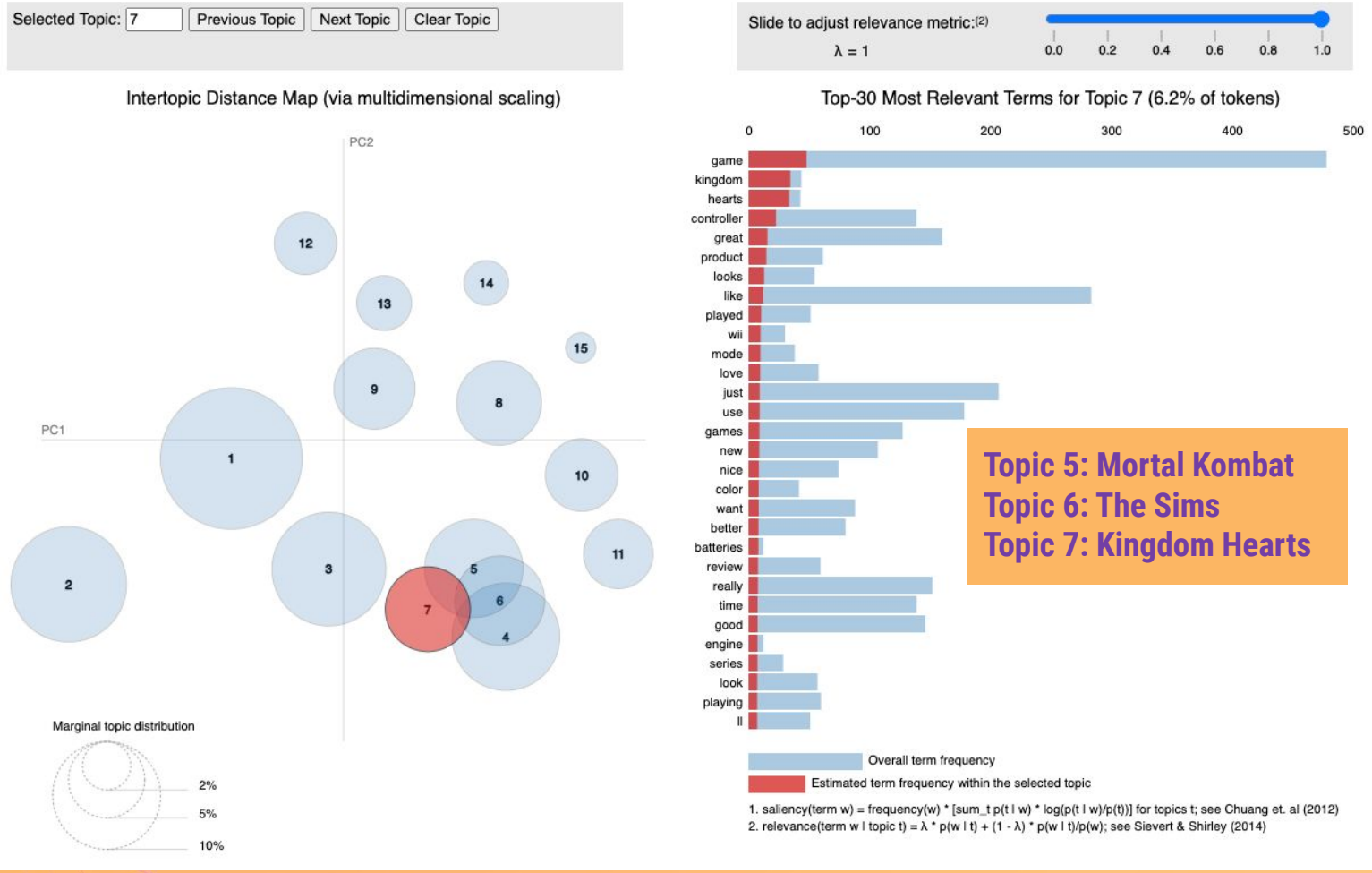

Selected Topic: 1
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
Topic Prevalence
Topic Similarity
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 1 (17.2% of tokens)
0
100
200
300
400
500
headset
just
like
xbox
headphones
good
gaming
great
mic
sound
controller
quality
use
don
time
play
game
really
pc
want
price
new
say
pro
work
got
buy
way
hours
using
Terms/meaning for each topic
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)


Selected Topic: 1
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 1 (17.2% of tokens)
0
100
200
300
400
500
headset
just
like
xbox
headphones
good
gaming
great
mic
sound
controller
quality
use
don
time
play
game
really
pc
want
price
new
say
pro
work
got
buy
way
hours
using
Topic 1: Gaming Headset
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)


Selected Topic: 2
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
12
13
14
15
9
8
1
10
2
3
5
6
7
4
11
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 2 (11.5% of tokens)
0
100
200
300
400
500
mouse
buttons
use
button
like
gaming
great
good
grip
thumb
naga
left
used
dpi
software
razer
hand
better
right
really
feels
feel
using
little
pretty
pad
review
light
time
big
Topic 1: Gaming Headset
vs
Topic 2: Gaming Mouse
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)

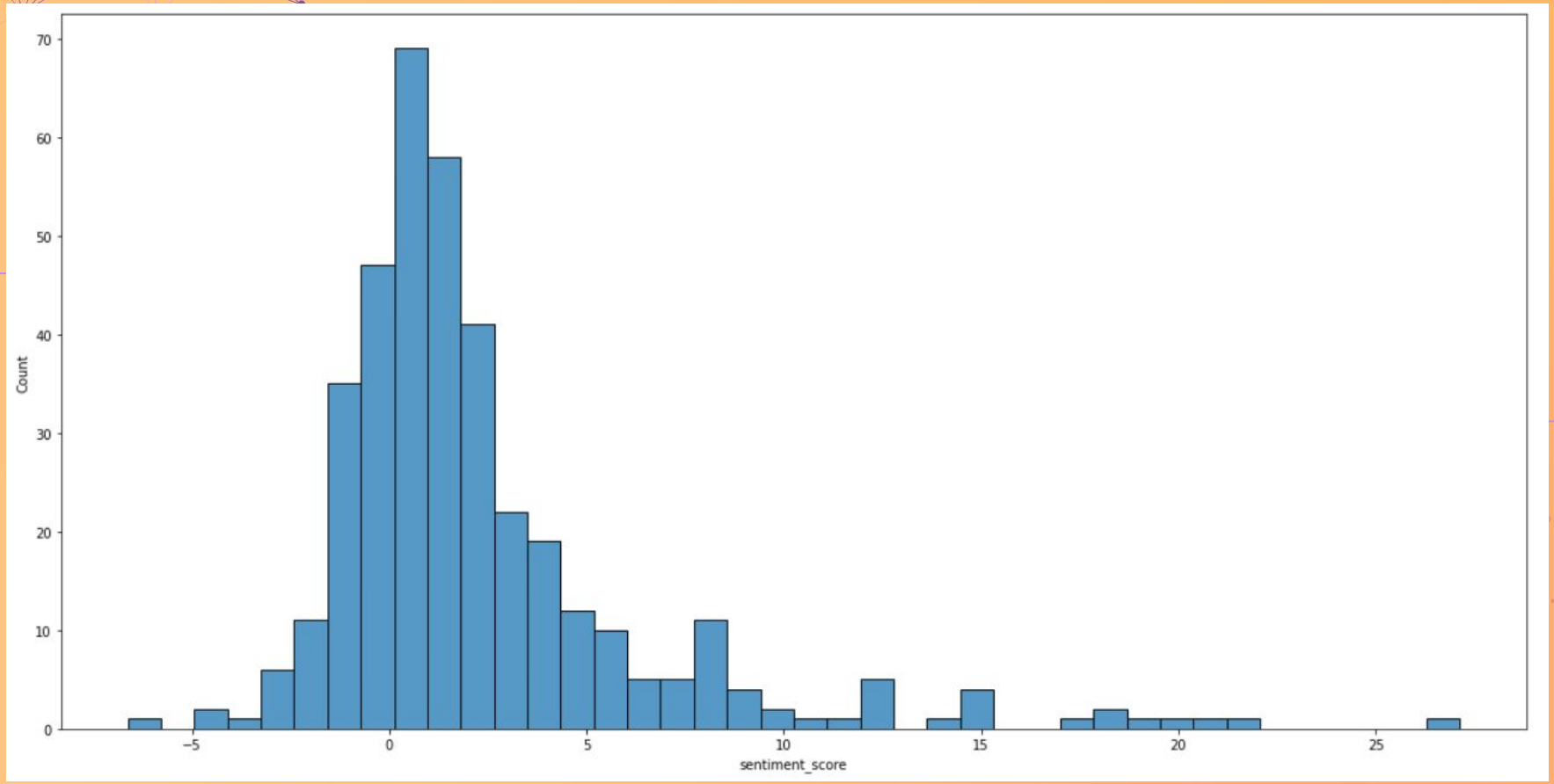

Selected Topic: 5
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 5 (8.3% of tokens)
0
100
200
300
400
500
game
controller
like
play
time
games
feel
just
use
characters
right
really
character
buttons
way
kombat
mortal
opponent
good
got
easy
hard
great
problem
long
story
makes
version
pretty
people
Topic 5: Mortal Kombat
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)


Selected Topic: 6
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 6 (6.9% of tokens)
0
100
200
300
400
500
game
sims
like
3ds
new
play
sim
just
really
games
xl
vr
make
don
great
case
different
want
best
feel
expected
characters
content
fun
edition
good
bought
year
time
sure
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)
Topic 5: Mortal Kombat
Topic 6: The Sims

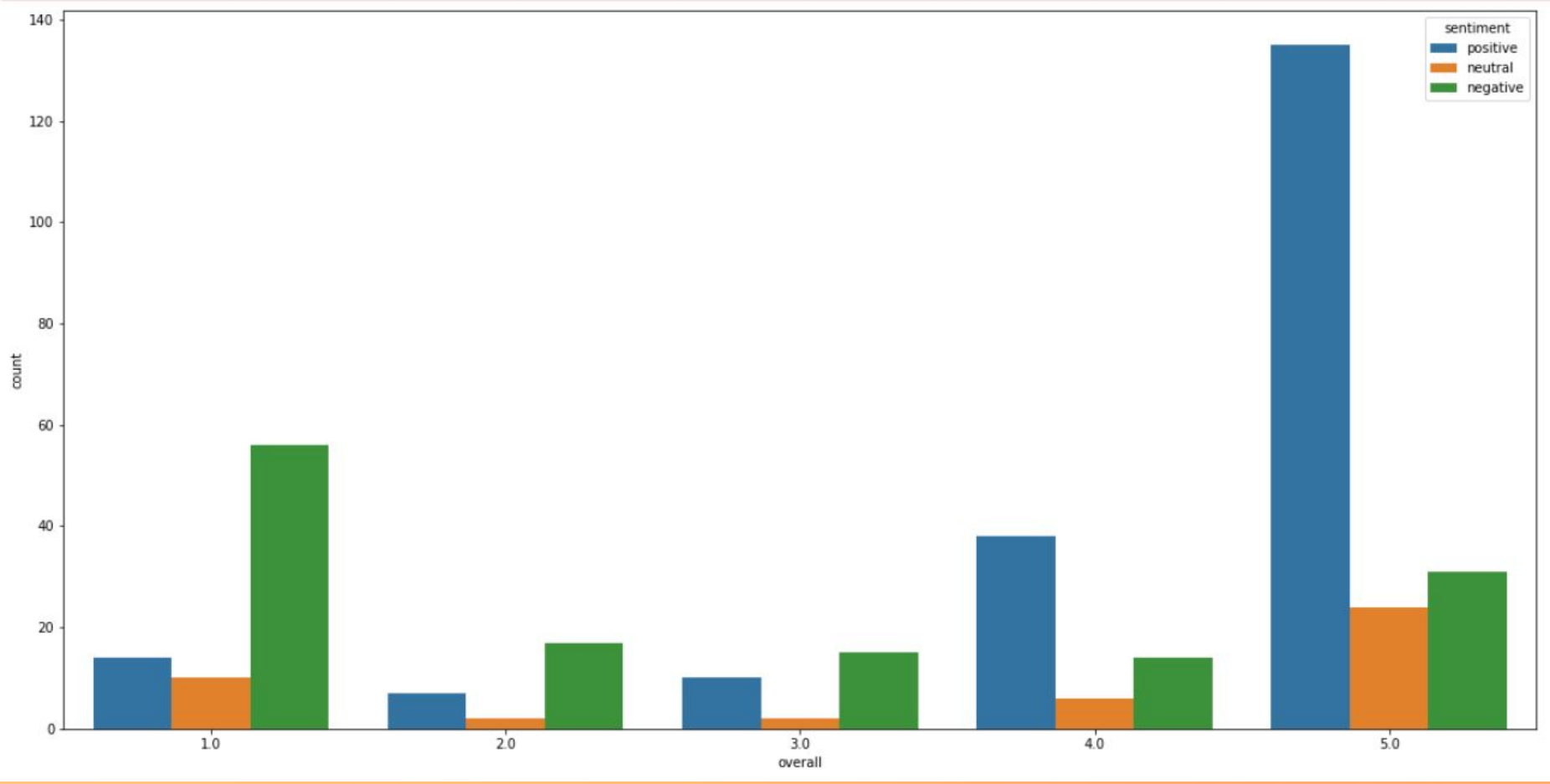



**Topic 5: Mortal Kombat**
**Topic 6: The Sims**
**Topic 7: Kingdom Hearts**

# SENTIMENT



# SENTIMENT


sentiment
positive
neutral
negative
count
overall
0
20
40
60
80
100
120
140
1.0
2.0
3.0
4.0
5.0

# FUTURE ANALYSIS

## NLP PROCESSING

More hyperparameters

## MODEL OPTIMIZATION

Different model, # of topics

## SEARCH & RECOMMENDATION SYSTEM

Query search and recommendation

# THANKS