



STEAMVOX

**STEAM REVIEWS: TOPIC MODELLING AND
SENTIMENT ANALYSIS**

Capstone Project by Alfred Tang
Data Science Immersive 8
General Assembly
August 2019

STEAMVOX IN A NUTSHELL

- Steam
 - Online PC game app store
- Vox
 - “Voice” in Latin
 - Voice of the customer

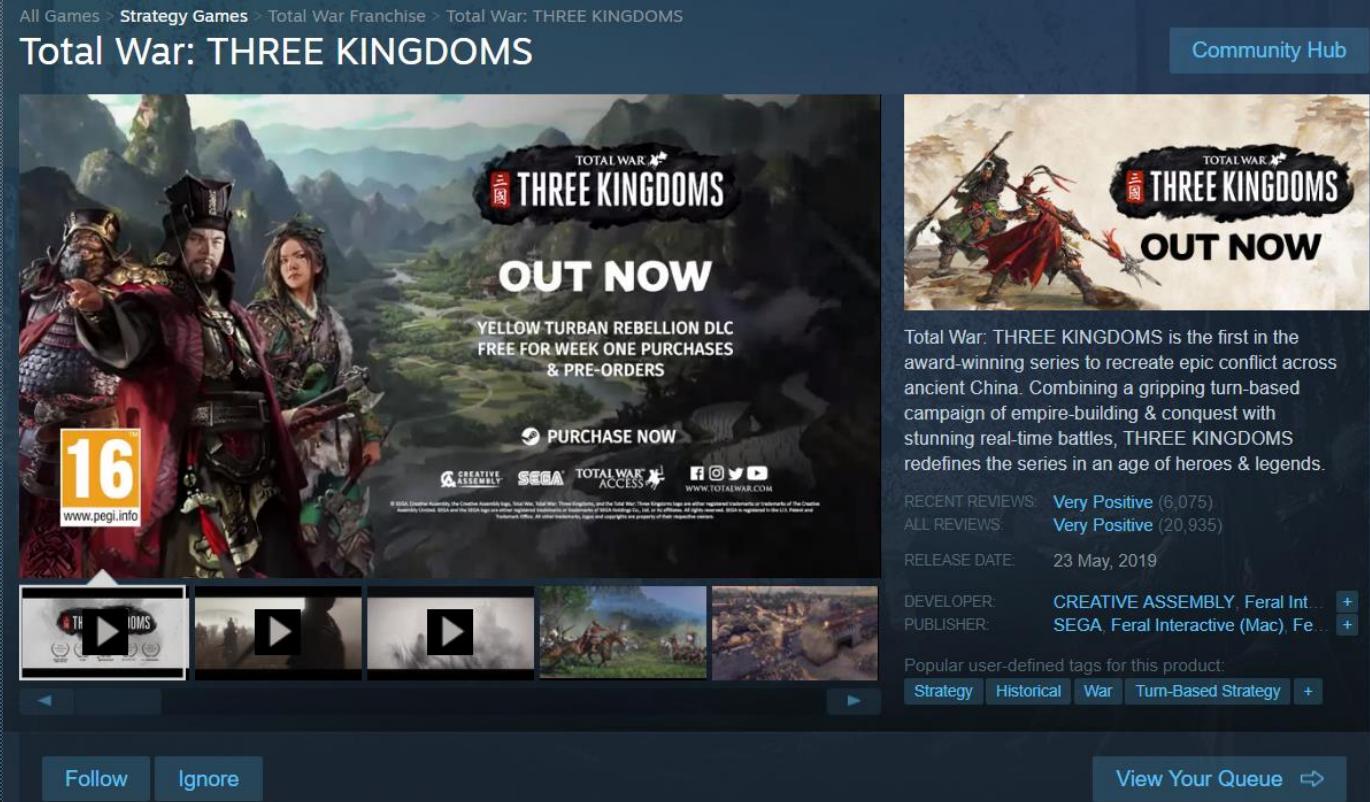
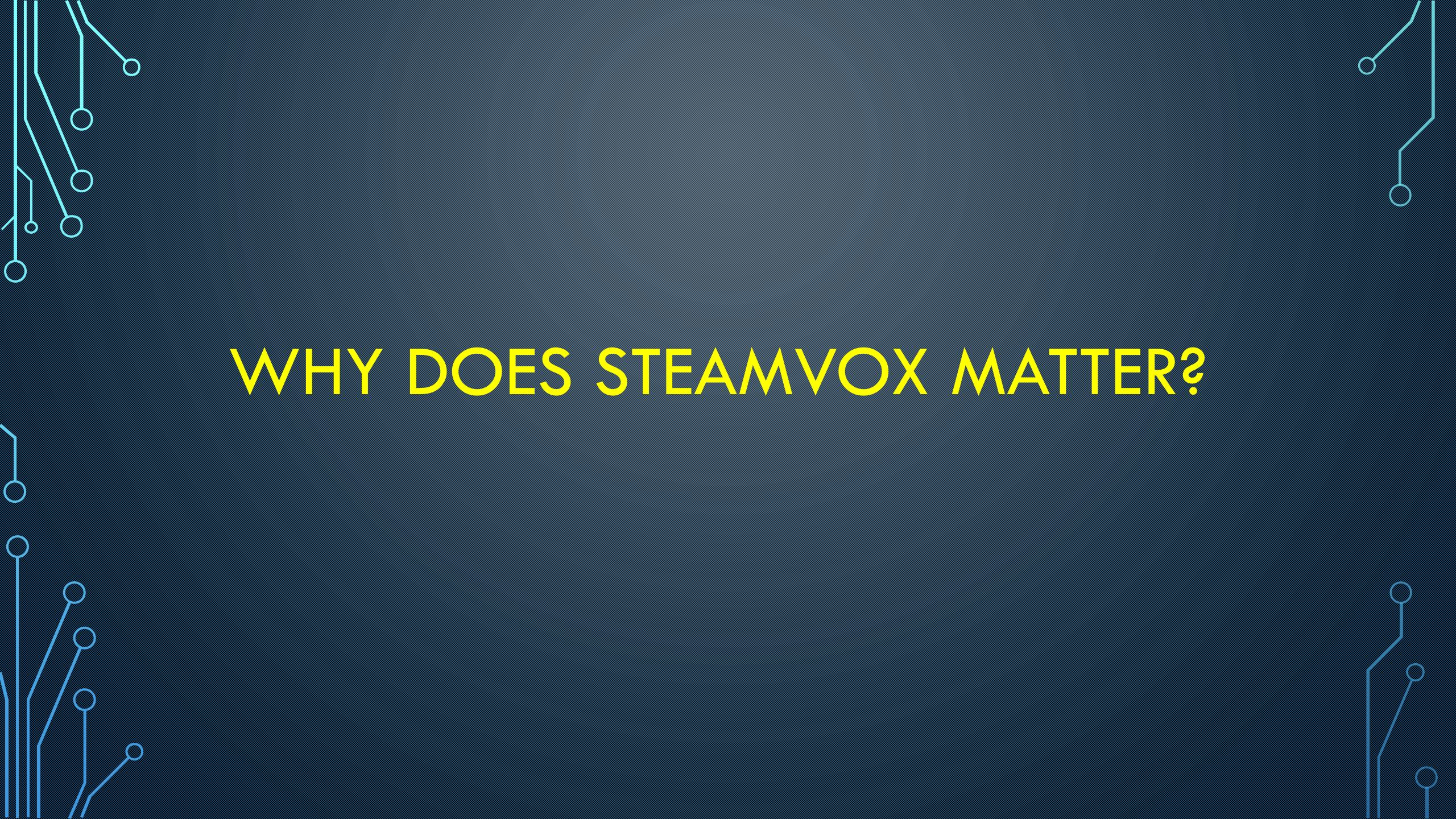


Image source: Valve Corp
<https://store.steampowered.com/>



WHY DOES STEAMVOX MATTER?

THE PROBLEM(S)

- Game developers always pressed for time
- Game publication is time-sensitive
- Reading reviews is time-consuming

PROBLEM STATEMENT

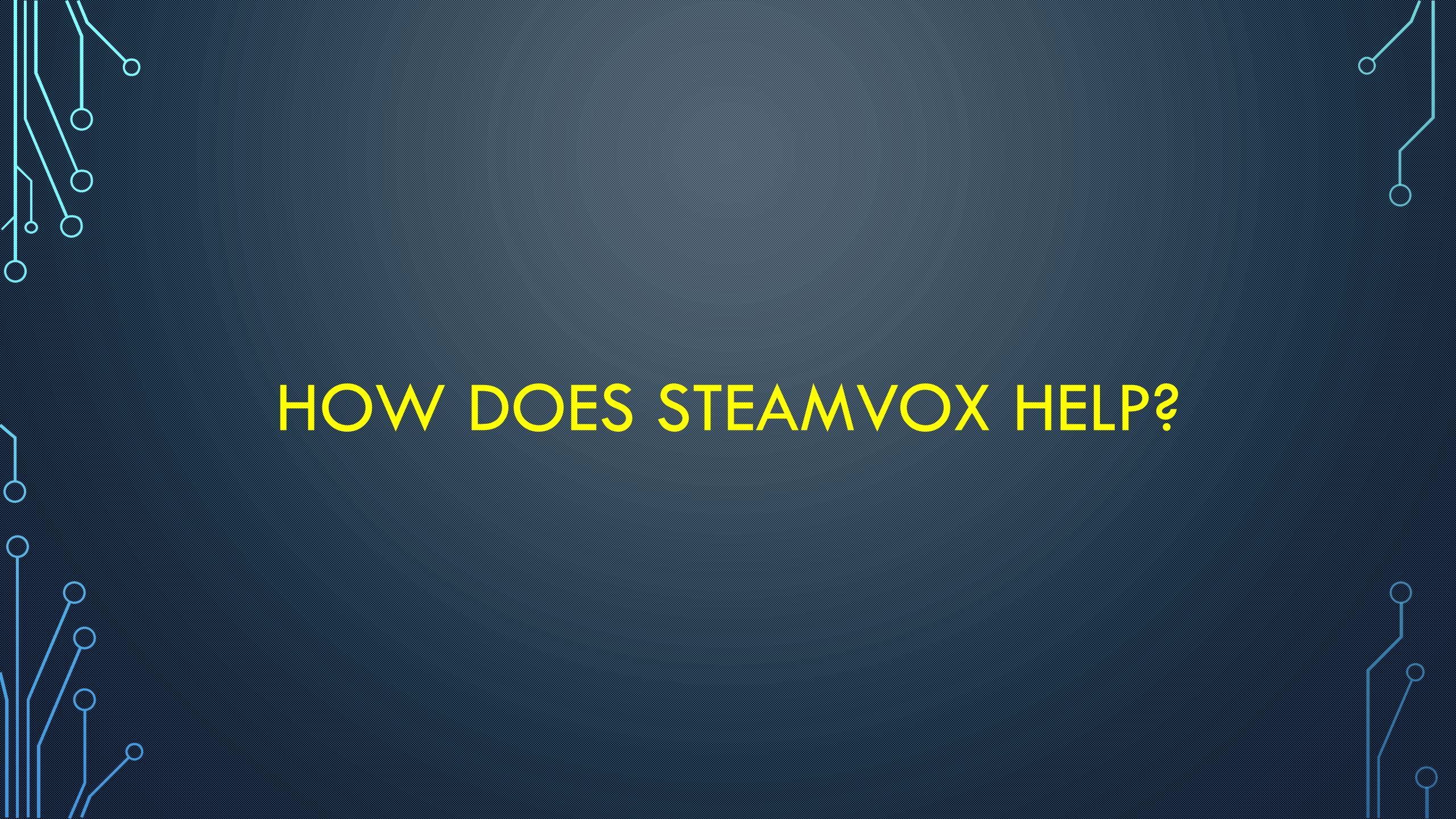
Are the features of my game well-received?

Any urgent issues to fix?

Which features in particular?

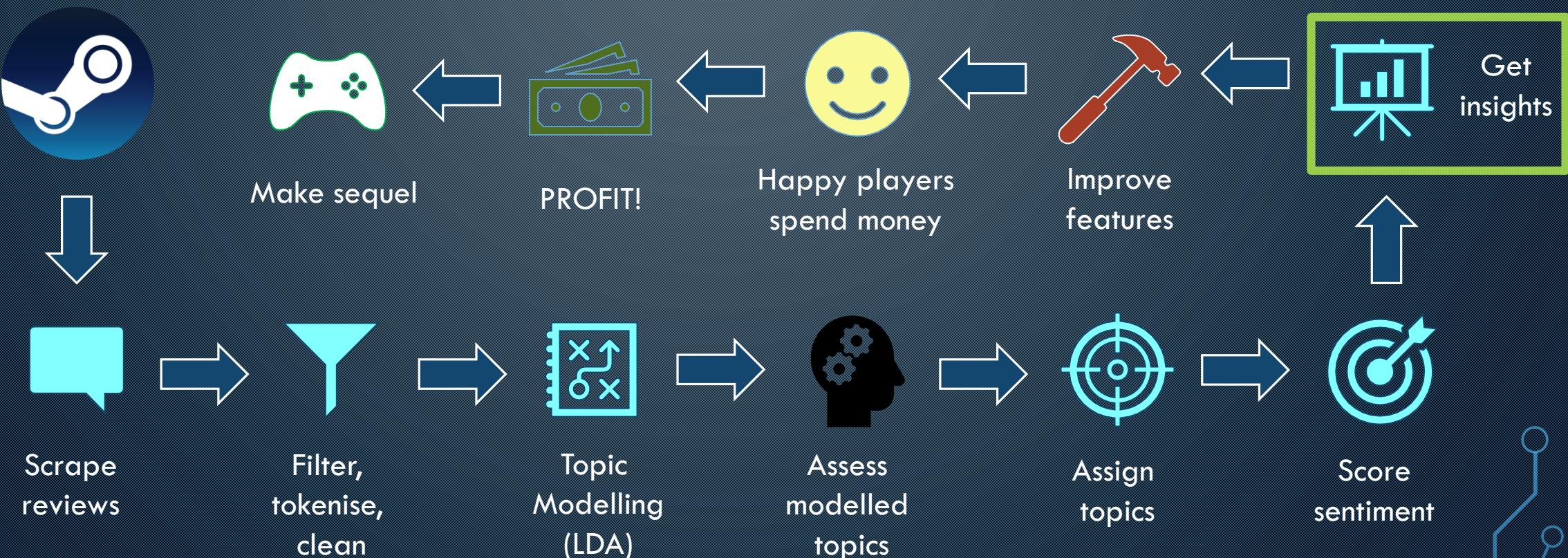
What data supports this conclusion?

How do I obtain supporting data ASAP?



HOW DOES STEAMVOX HELP?

STEAMVOX: AN OVERVIEW



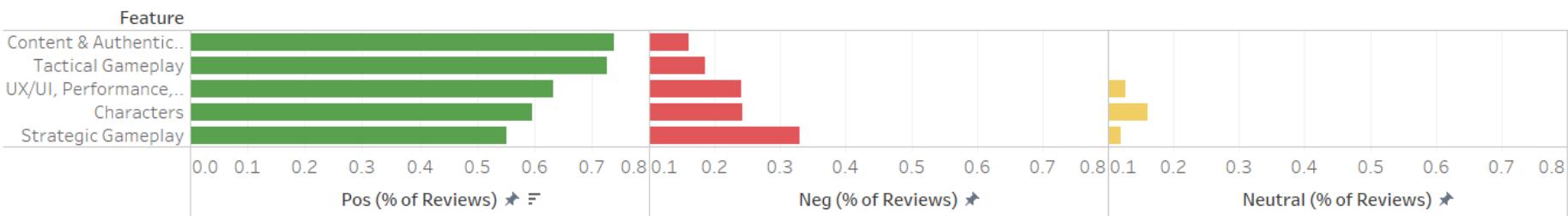
WHAT DIFFERENTIATES STEAMVOX?

NO SIMILAR TOOL AVAILABLE ON
STEAM OR STEAMSPY (YET)

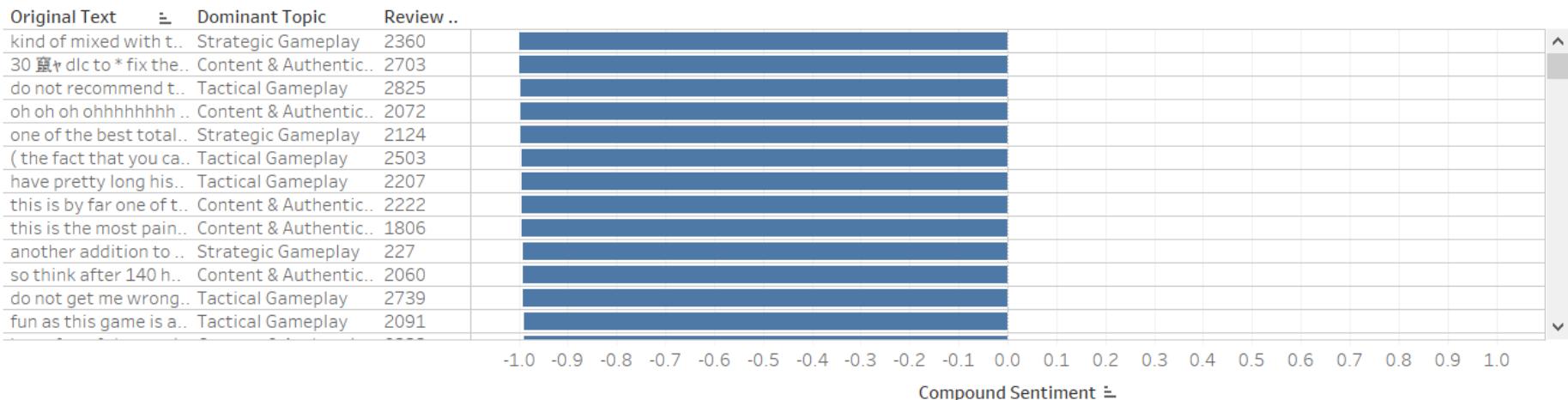
STEAMVOX ON TABLEAU

SteamVox Snapshot -- Total War: Three Kingdoms

Score Summary



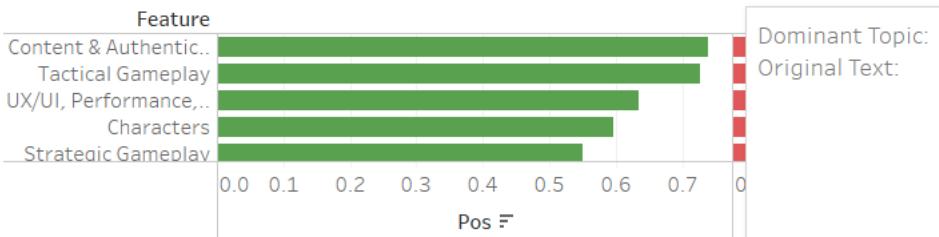
Reviews, Topics, Compound Sentiment Scores



STEAMVOX ON TABLEAU

SteamVox Snapshot -- Total War: Three Kingdoms

Score Summary



Dominant Topic:
Original Text:

Content & Authenticity
three kingdoms would say is two parts of total war smashing into each other. would recommend this game but not going lie its hard sell. its not like warhammer as food plays huge part in the game and planning settlement is hugely important , where warhammer does not need much brain power to plan your settlement. if your settlement has not been planned you will be handicapping yourself. the campaign side of things play more like the historical side of total war. on the flip side the battles play lot more like warhammer where having beef general can make the difference in battle (fyi think recorded mode plays more like old the total war games). the big different is the duel mode which personally think is good things and bad thing. good as it is cool feature and love watching them but bad as sometimes they have duels in the worst possible places eg. behind enemy lines which either makes them route or worst kills them. would suggest feature where after the fight the general has time of invincibility so the general can get back to his own line. another thing would wish is like dynasty warriors empire where you can choose the starting of game such as starting after the battle of red cliff , just suggestion. overall like the game and would recommend

Reviews, Topics, Compound Sentiment Scores

Original Text	Dominant Topic	Review ..
kind of mixed with t..	Strategic Gameplay	2360
30 章+ dlc to * fix the..	Content & Authenticity	2703
do not recommend t..	Tactical Gameplay	2825
oh oh oh ohhhhhh..	Content & Authenticity	2072
one of the best total..	Strategic Gameplay	2124
(the fact that you ca..	Tactical Gameplay	2503
have pretty long his..	Tactical Gameplay	2207
this is by far one of t..	Content & Authenticity	2222
this is the most pain..	Content & Authenticity	1806
another addition to ..	Strategic Gameplay	227
so think after 140 h..	Content & Authenticity	2060
do not get me wrong..	Tactical Gameplay	2739
fun as this game is a..	Tactical Gameplay	2091
huge fan of the total..	Content & Authenticity	2333
three kingdoms wou..	Content & Authenticity	3349

Review Number:
Compound Sentiment: -0.989



STEAMVOX V0.1 DATASET

- Steam API
- Review Counts
 - 8,235 English Steam reviews
 - 8,150 unique reviews
 - 3,661 determined usable
 - > 100,000 words

STEAMVOX ANATOMY

- 1. Scrape Steam Reviews**
- 2. Filter, Clean, Tokenise**
- 3. Identify topics (using topic model)**
- 4. Sentiment Analysis**
- 5. Generate snapshots**

2. FILTER, CLEAN, TOKENISE

- Objectives:
 1. Get meaty, coherent reviews

2. Clean reviews so LDA Model can process them

2. CLEAN & TOKENISE

ORIGINAL TEXT

As a Chinese player and a fan of Three Kingdoms, I've never thought that a foreign company can make such a fantastic Three Kingdoms game. The Creative Assembly did its best to make the game immersive and close to the history. So shut up and take my money.

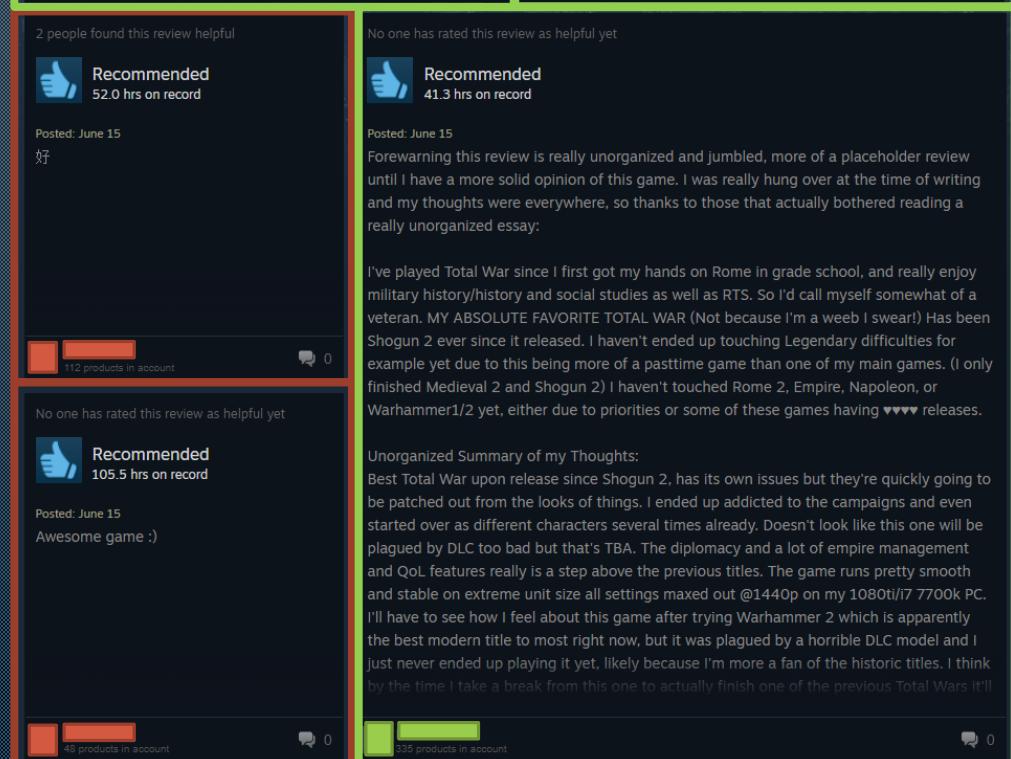
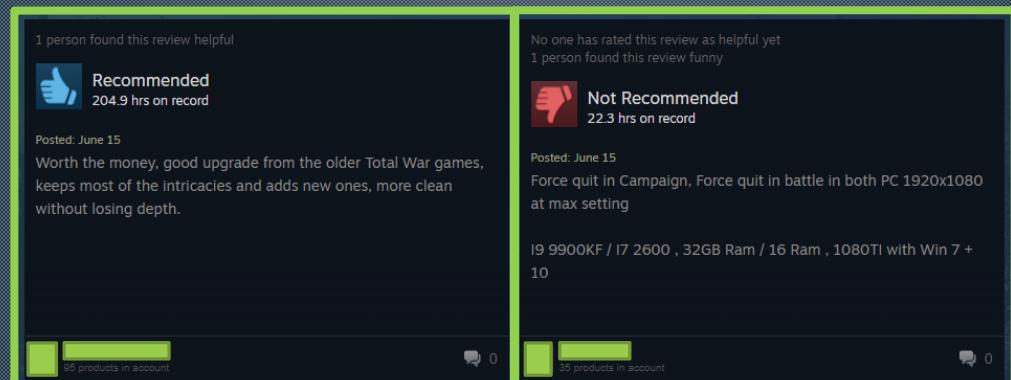
CLEAN, TOKENISED TEXT

```
['chinese', 'player', 'fan', 'three', 'kingdoms',  
 'thought', 'foreign', 'company', 'fantastic',  
 'three', 'kingdoms', 'creative', 'assembly',  
 'best', 'immersive', 'close', 'history', 'shut',  
 'money']
```

2. FILTER

- Steps:
 1. Filter reviews by conditions
 2. Tokenise reviews into lists of important terms
 3. Identify phrases (n-grams)
 4. Lemmatise each token and retain only nouns

FILTER REVIEWS



2. FILTERING CONDITIONS

Review Attribute	Condition
Player Playtime (when written)	<ol style="list-style-type: none">1. Total \geq 3 hours2. Recent playtime \geq 10 min
Review Length	\geq 5 words
Review uniqueness	Duplicates from same user NOT allowed
Language	EN only

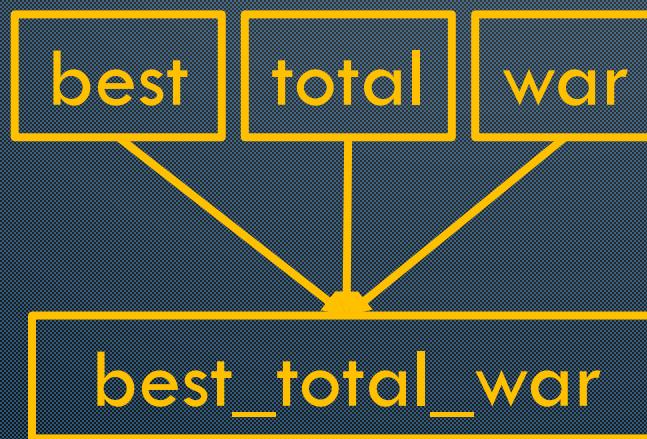
2. CLEAN AND TOKENISE

What is cleaned?	What is it?	Before	After
BBCode	Markup language used by Steam	[B]I love this game![/B]	I love this game!
Contractions	Informal shorthand	Can't, won't, ain't	Cannot, will not, is not
Punctuation	Anything not alphanumerical	Very gud game. Is balans))))))))))))	Very gud game is balans
Numerals (to Words)	Arabic & Roman numerals	Shogun 2	Shogun Two
Spam words	Unwanted, incoherent words	EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE LU BU OP	LU BU OP

2. CLEANING AND TOKENISATION

3. Identify phrases (n-grams)

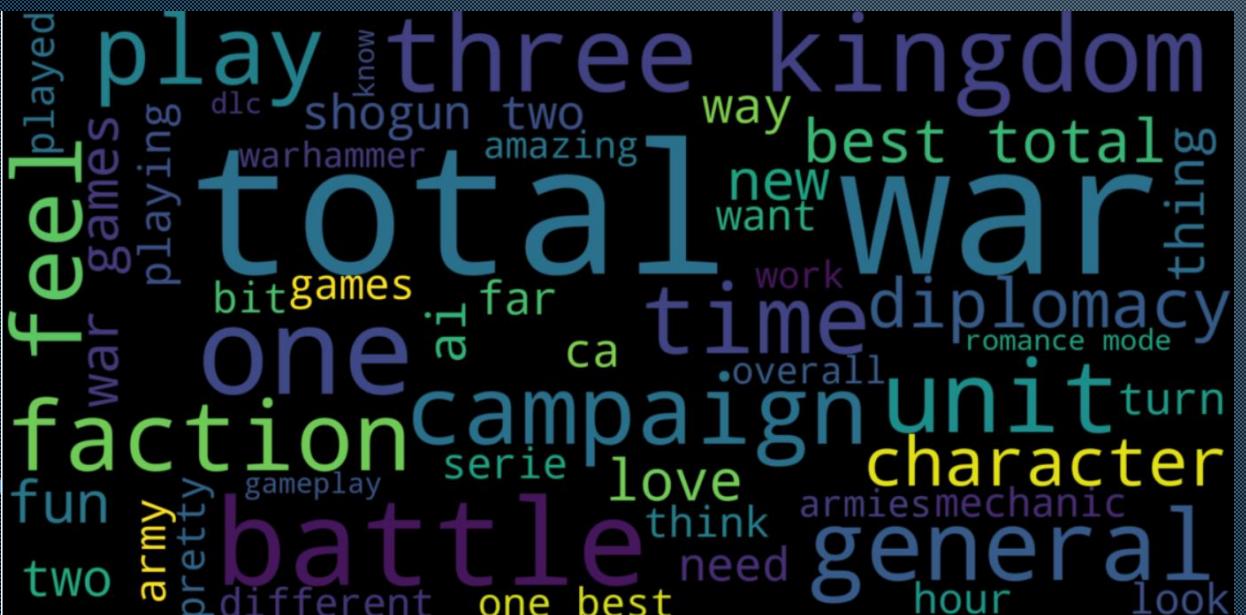
- Tokens that frequently occur together form phrases



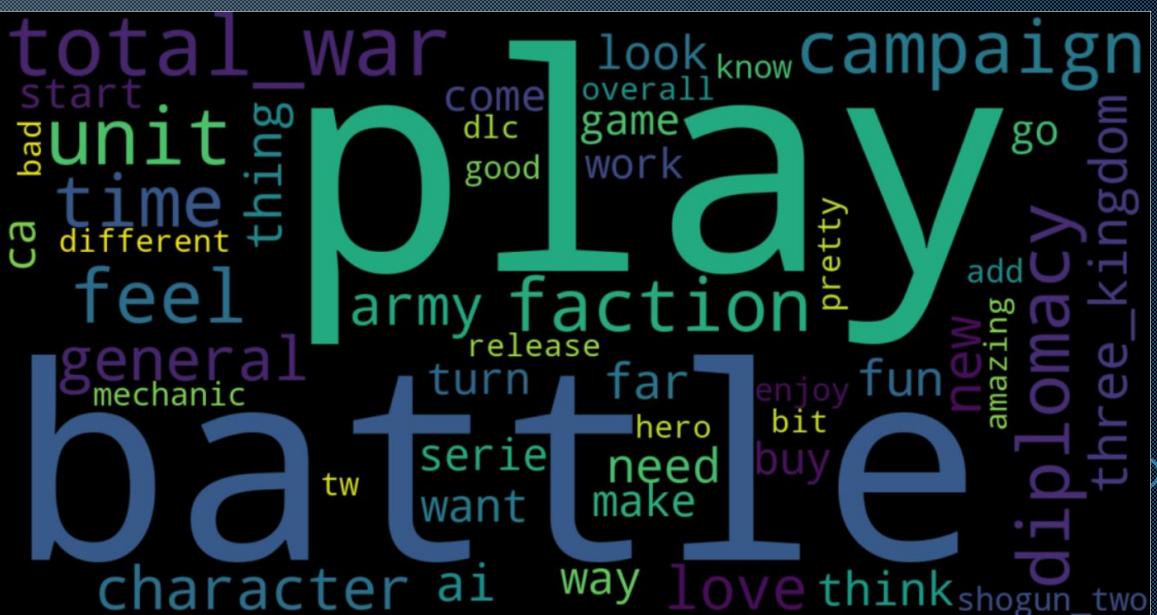
2. CLEANING AND TOKENISATION

3. n-grams increase coherence

1-grams only



n-grams included



2. CLEANING AND TOKENISATION

4. Lemmatisation

- Reduce inflected words to common dictionary form

Word	Lemma
Playing	Play
Plays	Play
Played	Play



3 counts of 'Play'

2. CLEANING AND TOKENISATION

4. Lemmatisation (Cleaning)

- Retain only phrases and nouns

Word	Class	Drop?	Lemma
playing	Verb	Drop	
exciting_battles	Noun (even if 2 non-nouns)		exciting_battles
exciting	Adjective	Drop	
battles	Noun		battle
cao_cao	Noun		cao_cao

3. TOPIC MODELLING

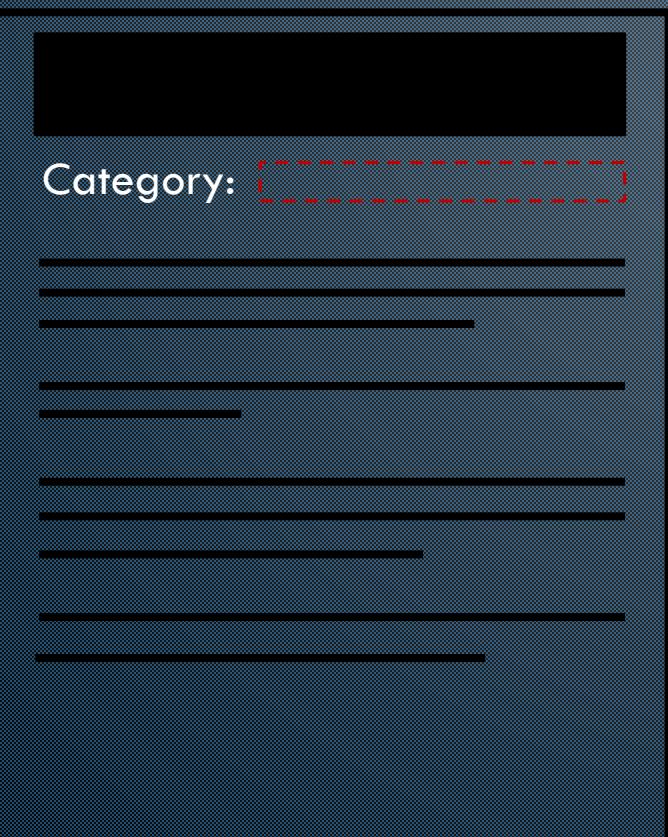
Latent

Dirichlet Allocation

Hidden

Probability Distribution

3. HOW LDA WORKS



Missing category!



Read Text



Identify
keywords



Identify Category

3. TOPIC MODELLING

Iteration	Word Types	Keyword Count	Topics Specified	Topic Identification	Classification Accuracy
1	All types	Any number	10	<ul style="list-style-type: none">Incoherent	NA
2	<ul style="list-style-type: none">NounsVerbsPhrases	Any number	5	<ul style="list-style-type: none">Incoherent	NA
3	<ul style="list-style-type: none">NounsPhrases	Any number	5	<ul style="list-style-type: none">CoherentClear separation	~85%
Final	<ul style="list-style-type: none">NounsPhrases	At least 2	5	<ul style="list-style-type: none">CoherentClear separation	~87%

3. TOPIC MODELLING – TOPICS IDENTIFIED

Topics	Relevant Keywords
Content & Authenticity	three_kingdoms, series, fan, diplomacy, battle, love, best_total_war, campaign...
Characters	china, war, cao_cao, lu_bu, liu_bei, yuan_shao, army, character, man, emperor, warlord...
Strategic Gameplay	faction, army, diplomacy, vassal, war, player, people, way, city, problem, campaign...
Tactical Gameplay	unit, battle, faction, army, general, character, diplomacy, campaign, hero, building, duel...
UX/UI, Performance, Stability	campaign, hour, issue, battle, bug, crash, multiplayer, review, rome, fix, play, patch, work...

TOPIC 1: CONTENT & AUTHENTICITY



Keywords:

three_kingdoms, series, fan, diplomacy, battle, love, best_total_war, campaign...

TOPIC 2: CHARACTERS



Keywords:

china, war, cao_cao, lu_bu, liu_bei, yuan_shao, army, character, man, emperor, warlord...

Images: [Total War Forums](#)
© Creative Assembly

TOPIC 3: STRATEGIC GAMEPLAY



Image: [StrategyGamer](#)

Keywords:

faction, army, diplomacy, vassal, war, player, people, way, city, problem, campaign...

TOPIC 4: TACTICAL GAMEPLAY



Image: [Total War: Three Kingdoms on Steam](#)

Keywords:

unit, battle, faction, army, general, character, diplomacy, campaign, hero, building, duel...

TOPIC 5: UX/UI, PERFORMANCE, STABILITY



Your PC ran into a problem and needs to restart. We're just collecting some error info, and then we'll restart for you. (0% complete)

Image:
[HowToGeek](#)

Keywords:

campaign, hour, issue, battle, bug, crash, multiplayer, review, rome, fix, play, patch, work...

4. SENTIMENT ANALYSIS

Anatomy of a Review

- Multiple topics
- Each topic has its own sentiment

 Recommended
65.3 hrs on record

Posted: May 29

I've played almost every TW game since the original Shogun, and this one blows all the others away. I was skeptical how the character-centric approach would work with Total War's usual formula, but it's implemented brilliantly. Inter-character relationships are now far more intricate and important while still being accessible, particularly when it comes to diplomacy.

Whereas previous TW games made diplomacy almost feel like an afterthought, building coalitions, alliances, and vassals is extremely vital. The introduction of faction specific resources like Cao Cao's ability to manipulate other factions to like or dislike one another or even start proxy wars really adds a unique layer that makes the personality of that faction really stand out.

The decision to make commanderies have specific exploitable resources is also a good refinement over the previous games like Rome 2, where specific resources appear pretty sparingly throughout the map. While it may not sound like an enormous change, it is; having each county in a commandery have a specific, extremely useful resource makes you prioritize your plans more carefully and realistically, and incentivizes players to pay more attention to what they're building to get the most synergy bonuses possible.

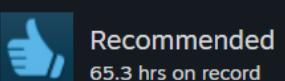
The addition of duels between characters is a nice touch as well, particularly in the game's Romance mode, which replaces retinues with single, uber-powerful characters that make rewatching replays feel cinematic.

Content & Authenticity

Strategic Gameplay

Tactical Gameplay

4. SENTIMENT ANALYSIS – SCORING REVIEWS



Recommended
65.3 hrs on record

Posted: May 29

I've played almost every TW game since the original Shogun, and this one blows all the others away. I was skeptical how the character-centric approach would work with Total War's usual formula, but it's implemented brilliantly. Inter-character relationships are now far more intricate and important while still being accessible, particularly when it comes to diplomacy.

Whereas previous TW games made diplomacy almost feel like an afterthought, building coalitions, alliances, and vassals is extremely vital. The introduction of faction specific resources like Cao Cao's ability to manipulate other factions to like or dislike one another or even start proxy wars really adds a unique layer that makes the personality of that faction really stand out.

The decision to make commanderies have specific exploitable resources is also a good refinement over the previous games like Rome 2, where specific resources appear pretty sparingly throughout the map. While it may not sound like an enormous change, it is; having each county in a commandery have a specific, extremely useful resource makes you prioritize your plans more carefully and realistically, and incentivizes players to pay more attention to what they're building to get the most synergy bonuses possible.

The addition of duels between characters is a nice touch as well, particularly in the game's Romance mode, which replaces retinues with single, uber-powerful characters that make rewatching replays feel cinematic.

Content & Authenticity



Strategic Gameplay



Strategic Gameplay



Strategic Gameplay

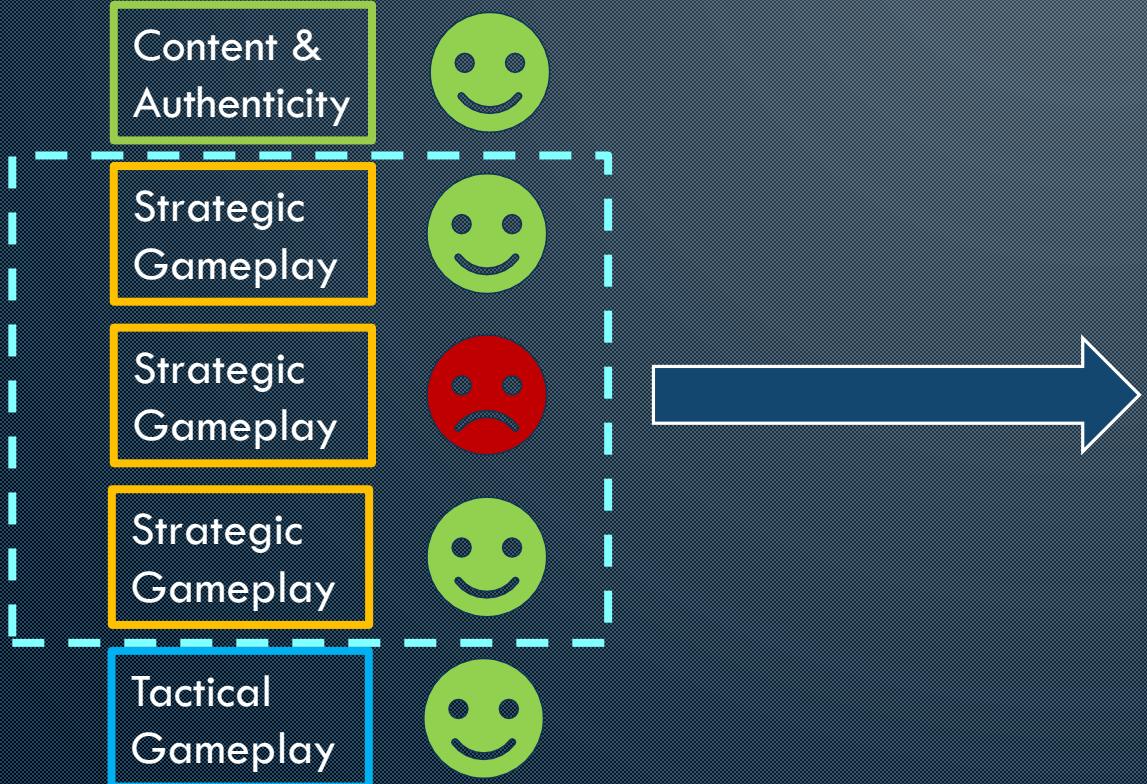


Tactical Gameplay



Topic	Sentiment
Content & Authenticity	
Strategic Gameplay	
Tactical Gameplay	

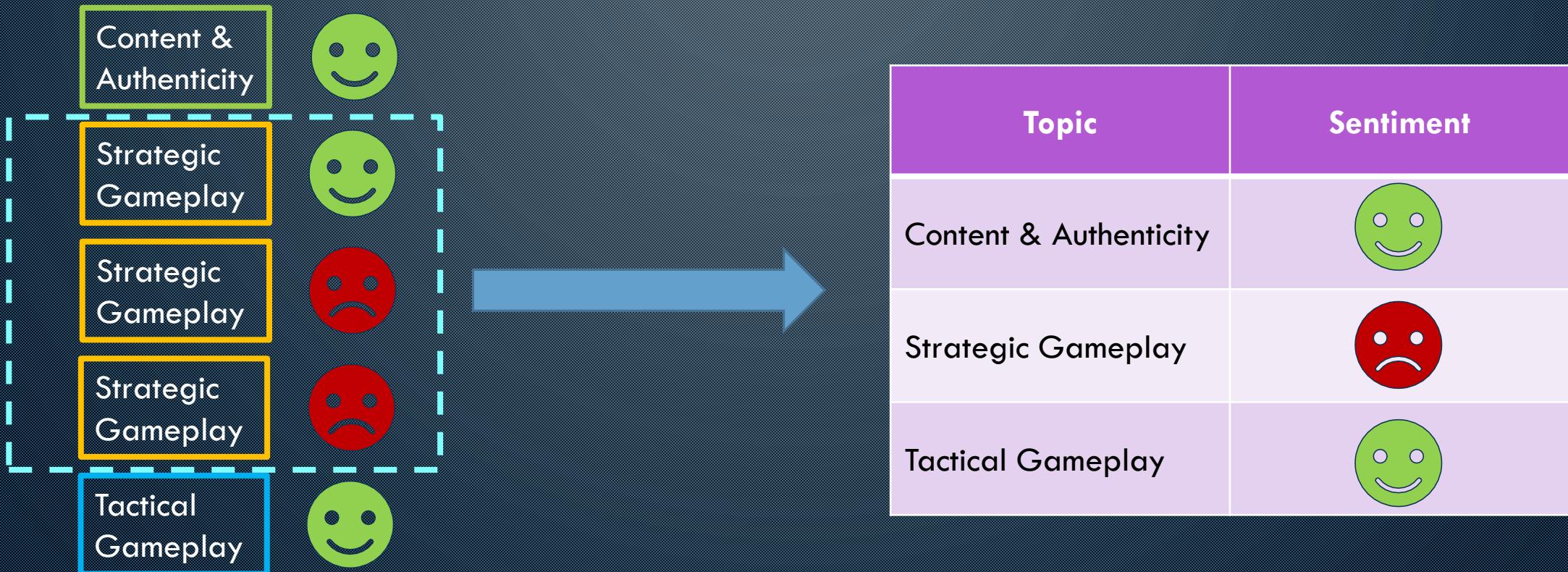
4. SENTIMENT ANALYSIS – POSITIVE



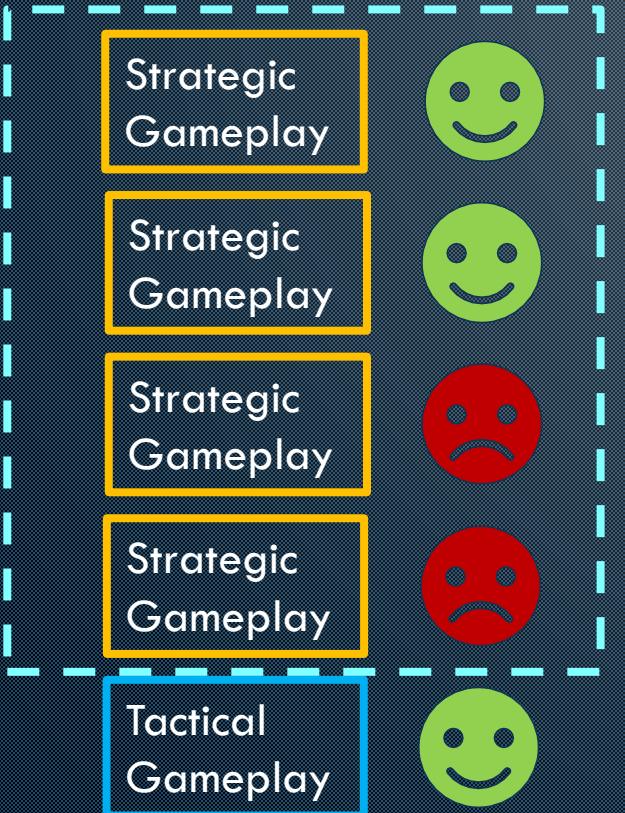
Topic	Sentiment
Content & Authenticity	😊
Strategic Gameplay	😊
Tactical Gameplay	😊

It's all about dominant sentiment!

4. SENTIMENT ANALYSIS – NEGATIVE

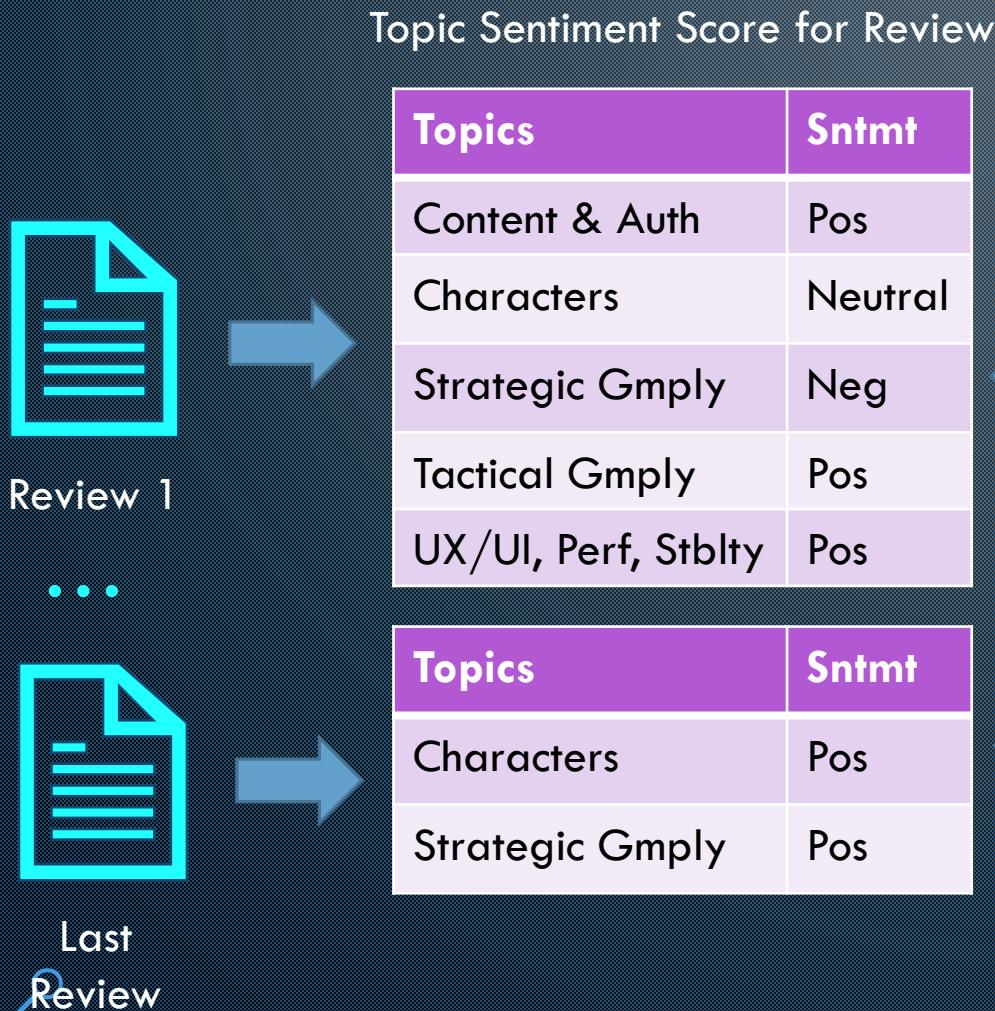


4. SENTIMENT ANALYSIS – NEUTRAL



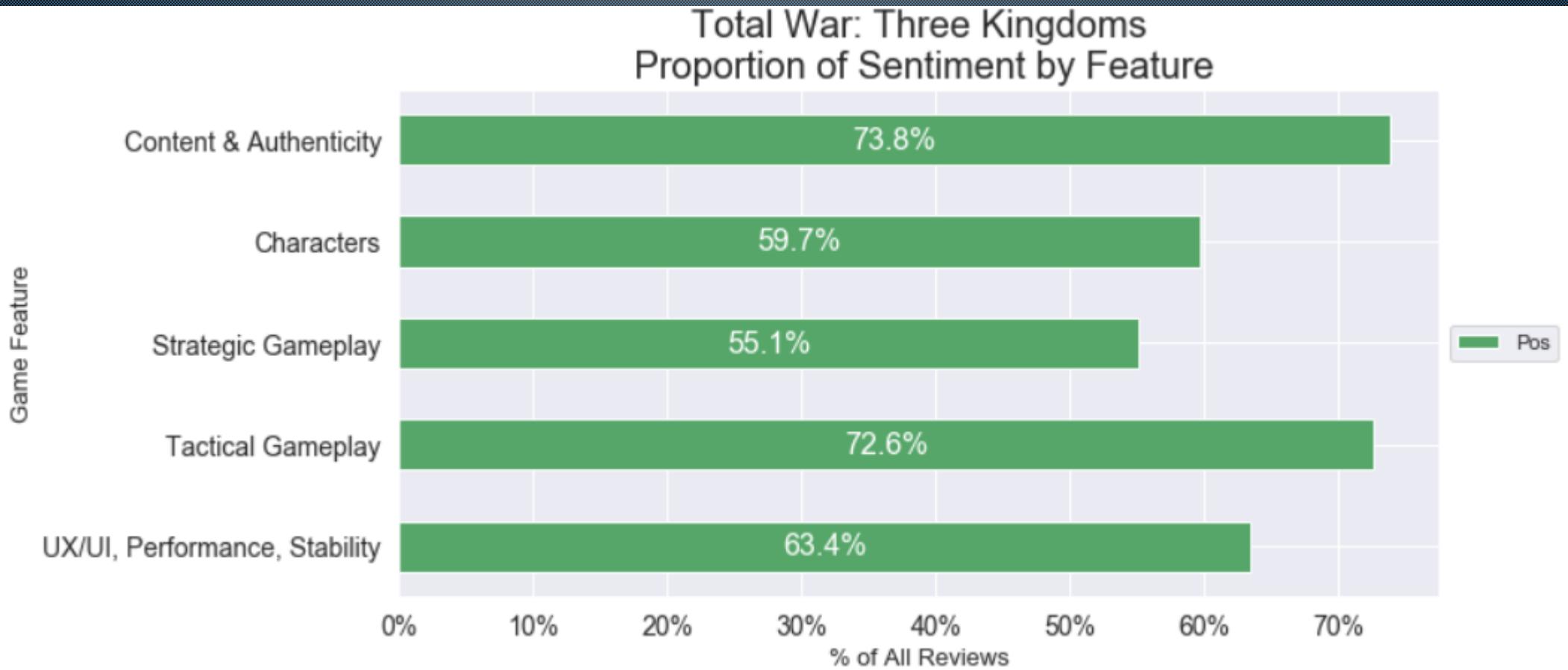
Topic	Sentiment
Strategic Gameplay	
Tactical Gameplay	

4. SENTIMENT ANALYSIS – AGGREGATION

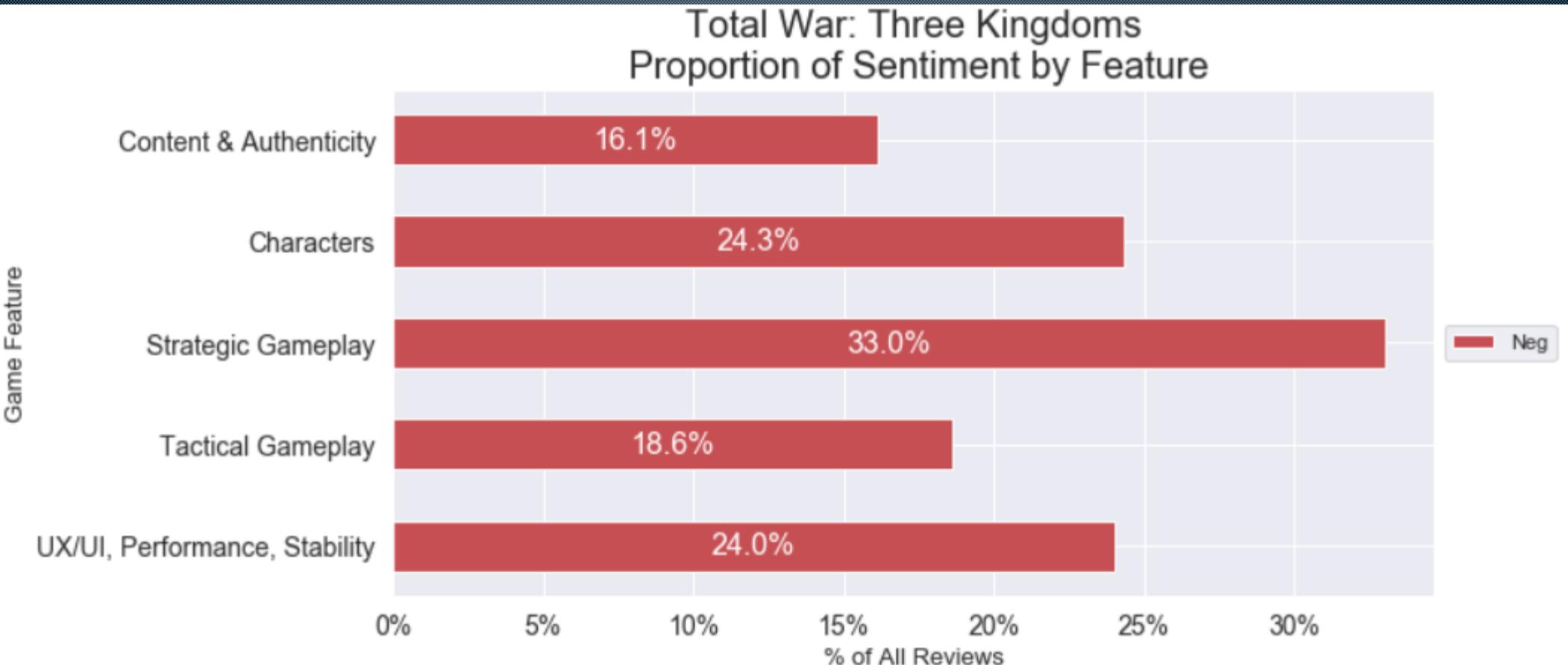


Topics	Pos %	Neutral %	Neg %
Content & Authenticity	73.8	10.1	16.1
Characters	59.7	16.0	24.3
Strategic Gameplay	55.1	11.9	33.0
Tactical Gameplay	72.6	8.9	18.6
UX/UI, Performance, Stability	63.4	12.6	24.0

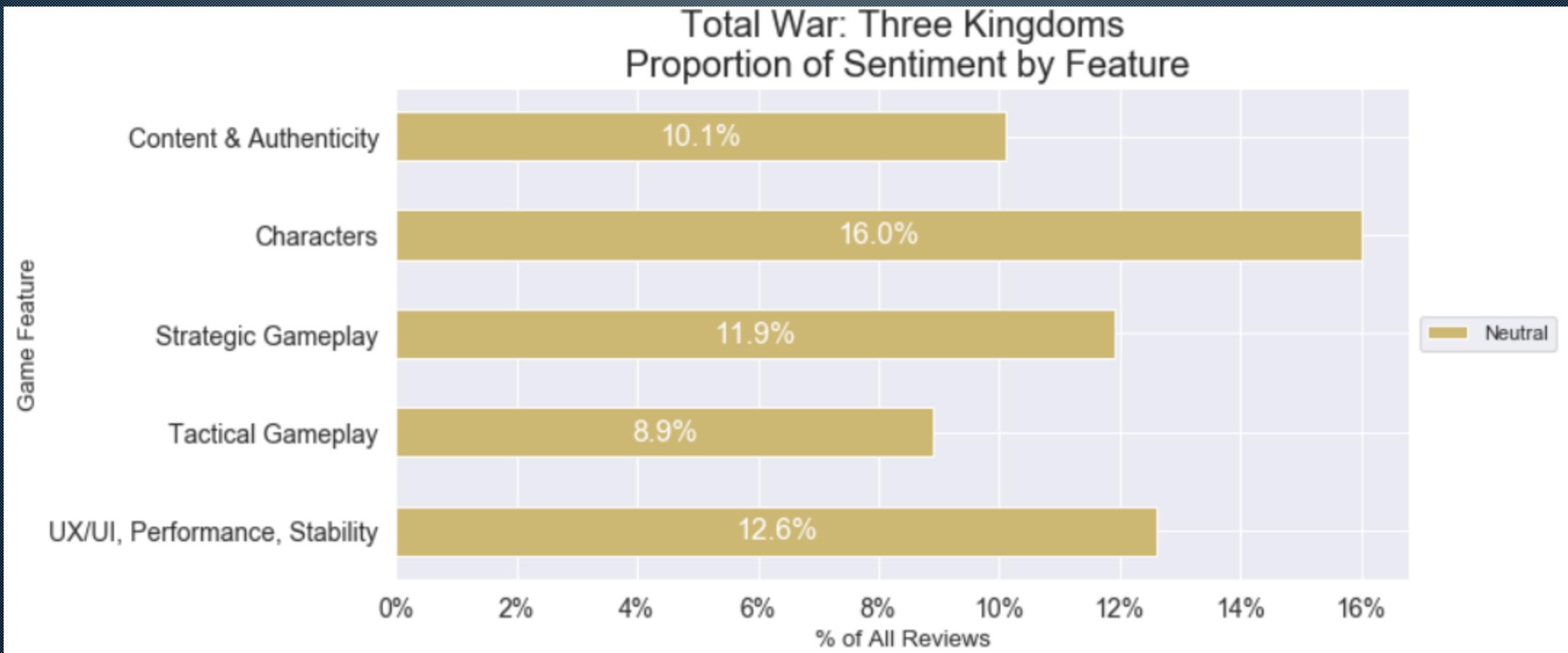
5. STEAMVOX SNAPSHOT (1)



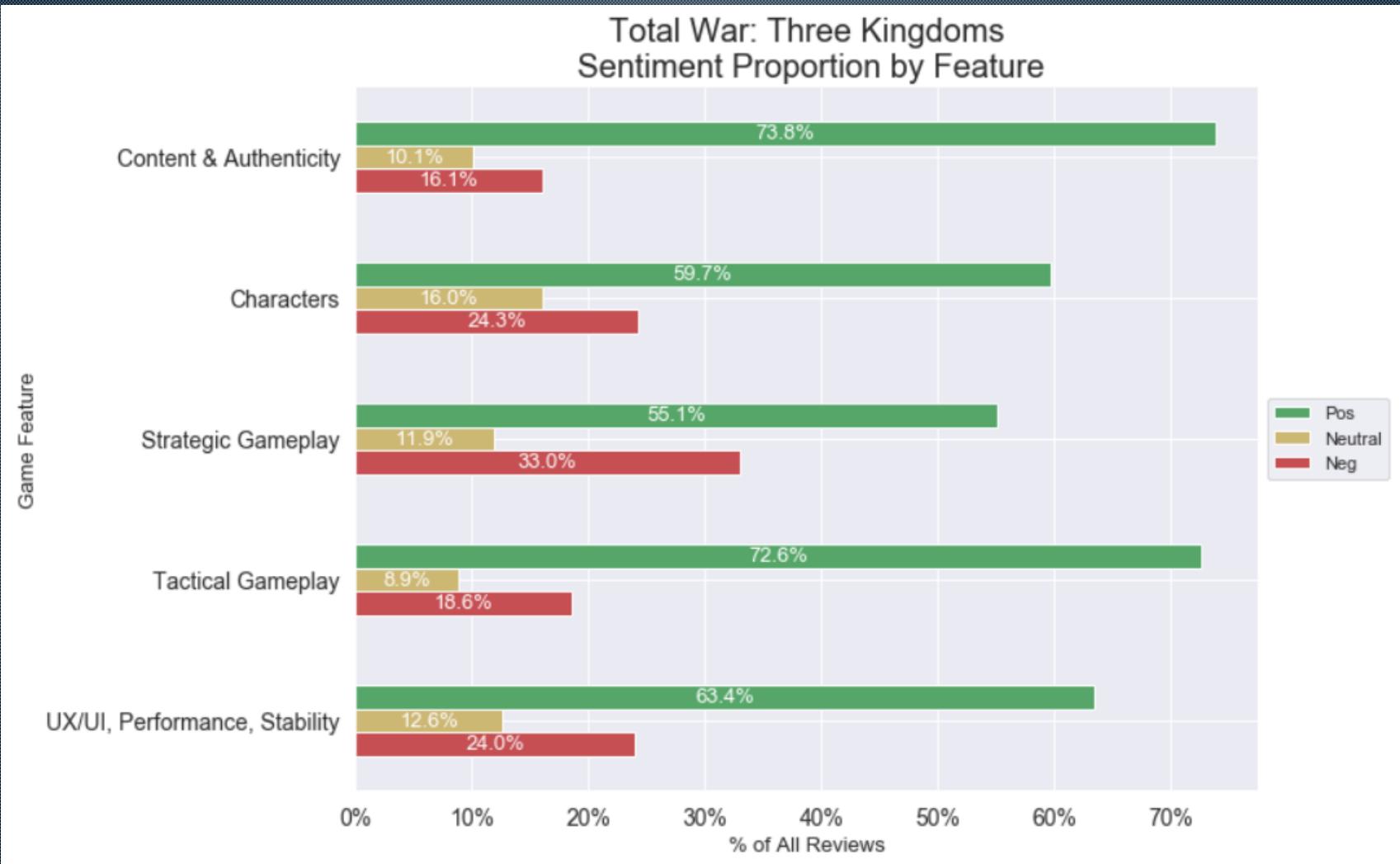
5. STEAMVOX SNAPSHOT (2)



5. STEAMVOX SNAPSHOT (3)



5. STEAMVOX SNAPSHOTS (4)



STEAMVOX ROUNDUP

- SteamVox fulfils its intended purpose
- Identifies coherent topics
- Assigns accurate sentiment scores

NEXT STEPS

- Make SteamVox generally applicable across most games on Steam
 - Train SteamVox on larger Steam dataset
 - Tweak LDA Model further
- Explore paragraphing methodology
 - Much more room to improve



THANK YOU!

REFERENCES

- DoctorSnapshot by nuwapi @ <https://github.com/nuwapi/DoctorSnapshot/tree/master/src>
 - <https://blog.insightdatascience.com/topic-modeling-and-sentiment-analysis-to-pinpoint-the-perfect-doctor-6a8fdd4a3904>
- steamreviews scraper by wocitezuma @ <https://github.com/wocitezuma/download-steam-reviews>
- expandContradictions by nealrs @ <https://gist.github.com/nealrs/96342d8231b75cf4bb82>
- syntok by fnl @ <https://github.com/fnl/syntok>
- Guide to Topic Modelling with Gensim @ <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/>
- SpaCy documentation @ <https://spacy.io/api/tokenizer>