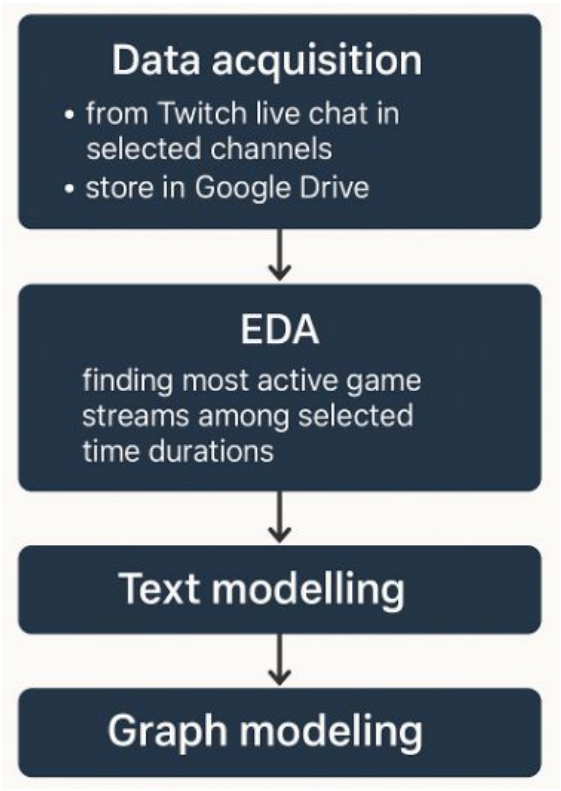
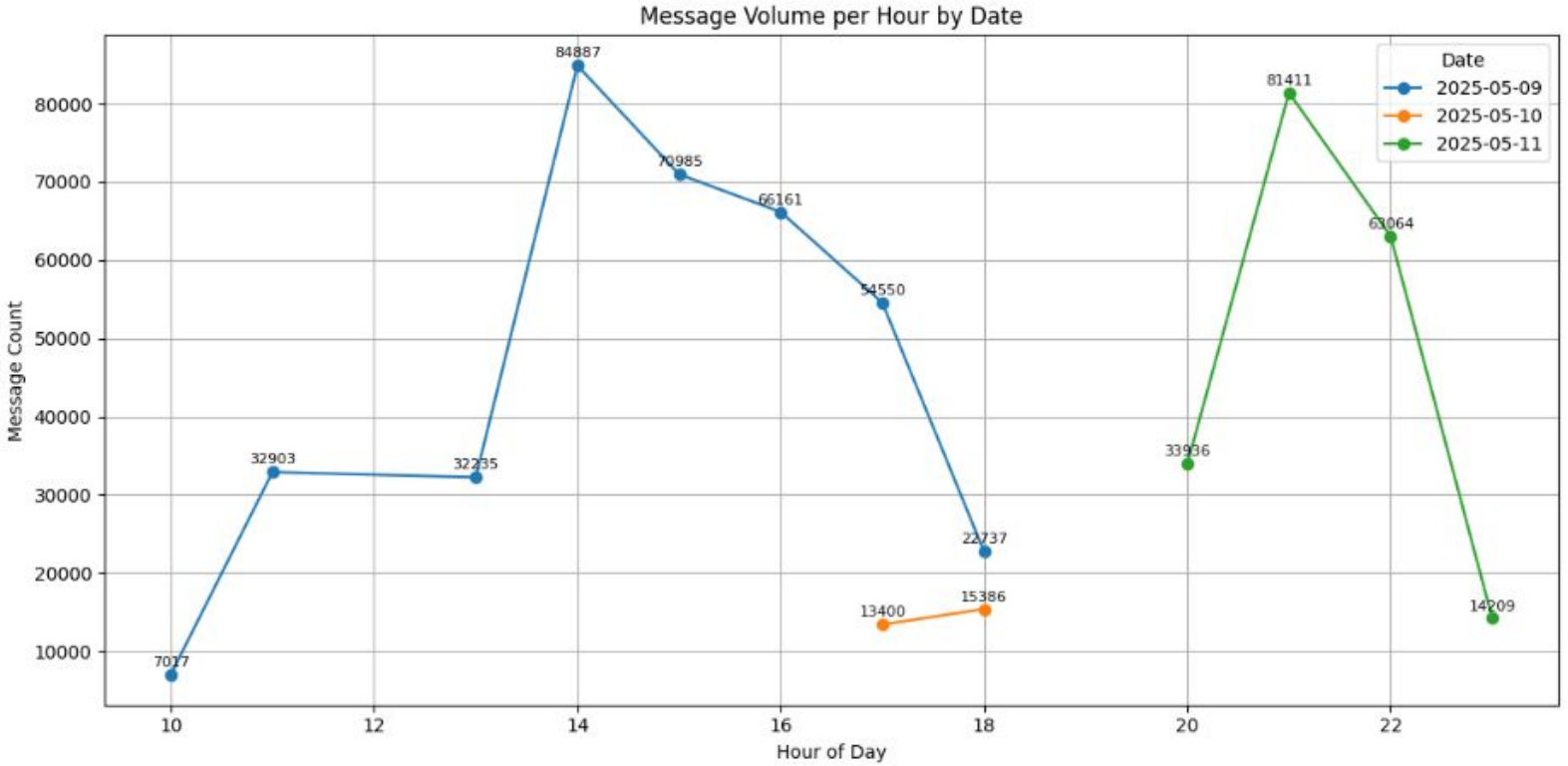
twitch stream chat modeling

Jia Jia
Shameek

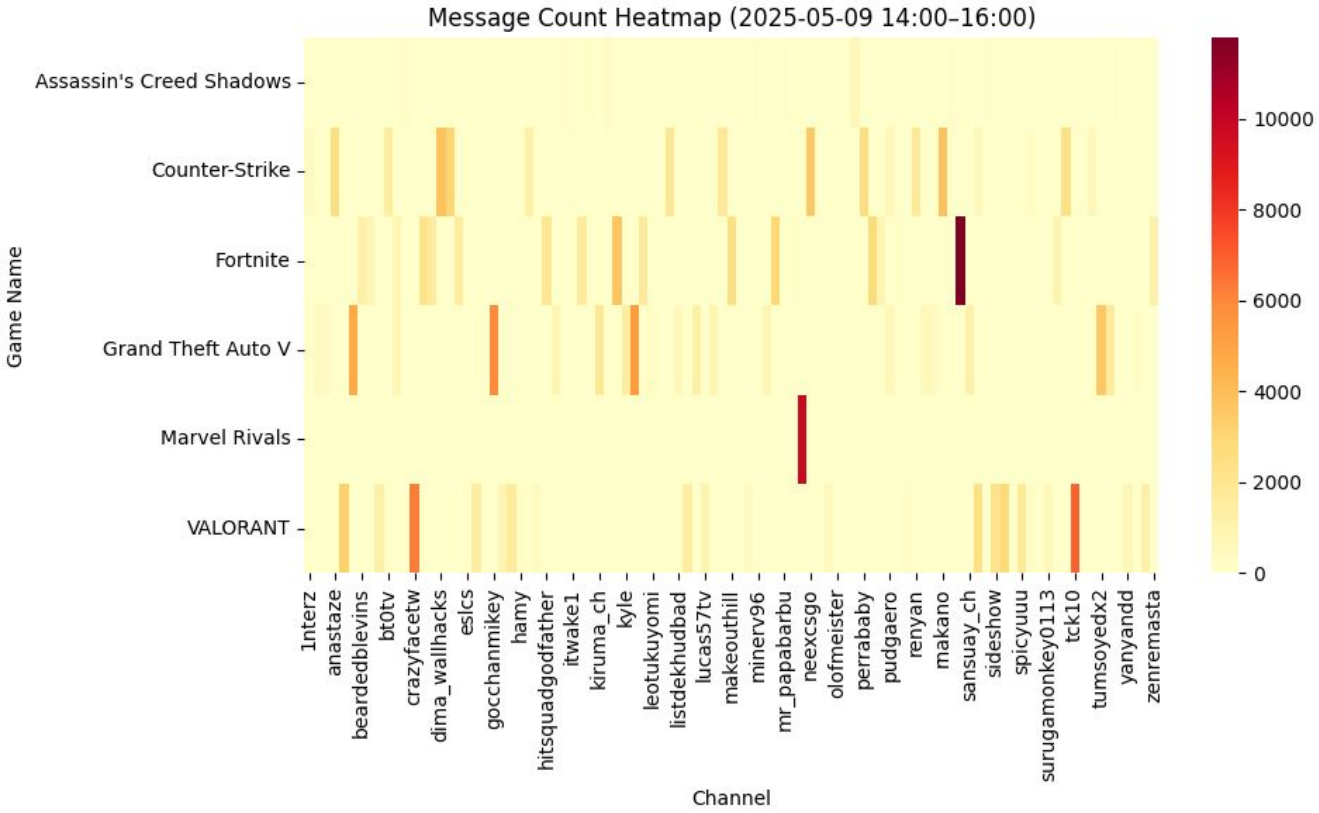
Data Project Process



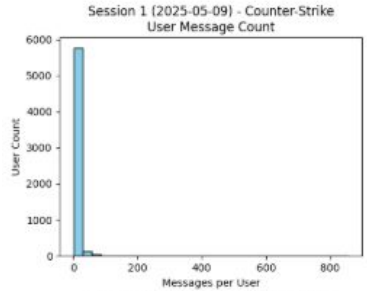
EDA

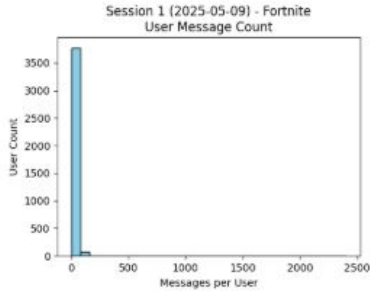

Message Volume per Hour by Date
Message Count
Hour of Day
Date
2025-05-09
2025-05-10
2025-05-11
7017
32903
32235
84887
70985
66161
54550
22737
13400
15386
33936
81411
63064
14209

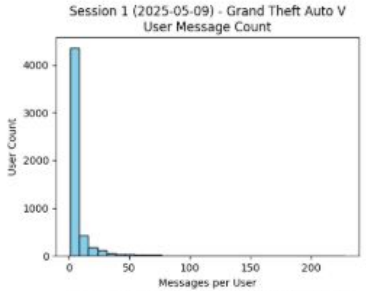
EDA: session1

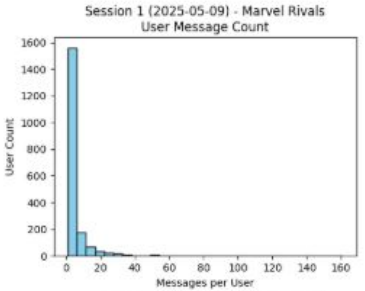


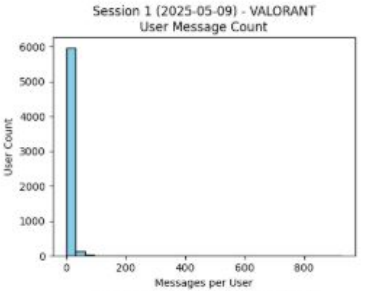
EDA: session1

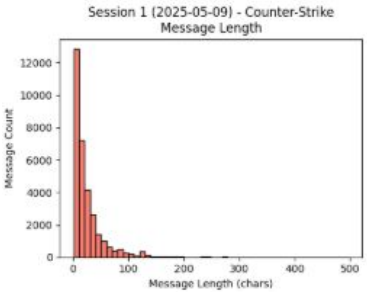


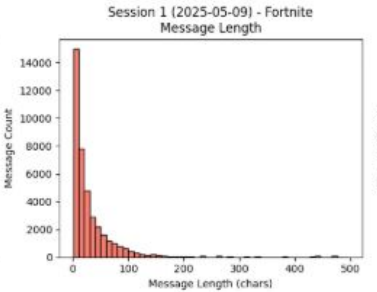


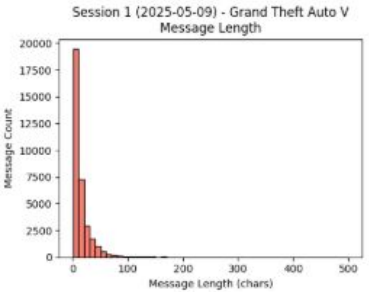


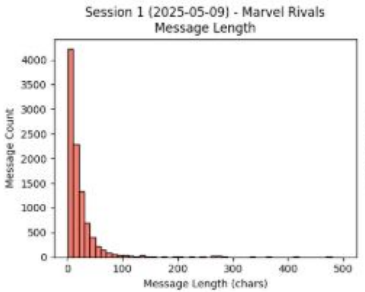


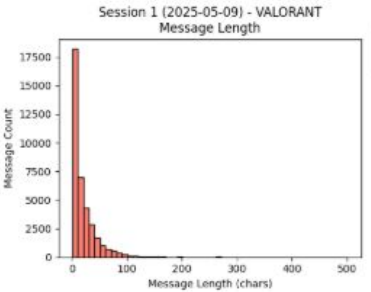




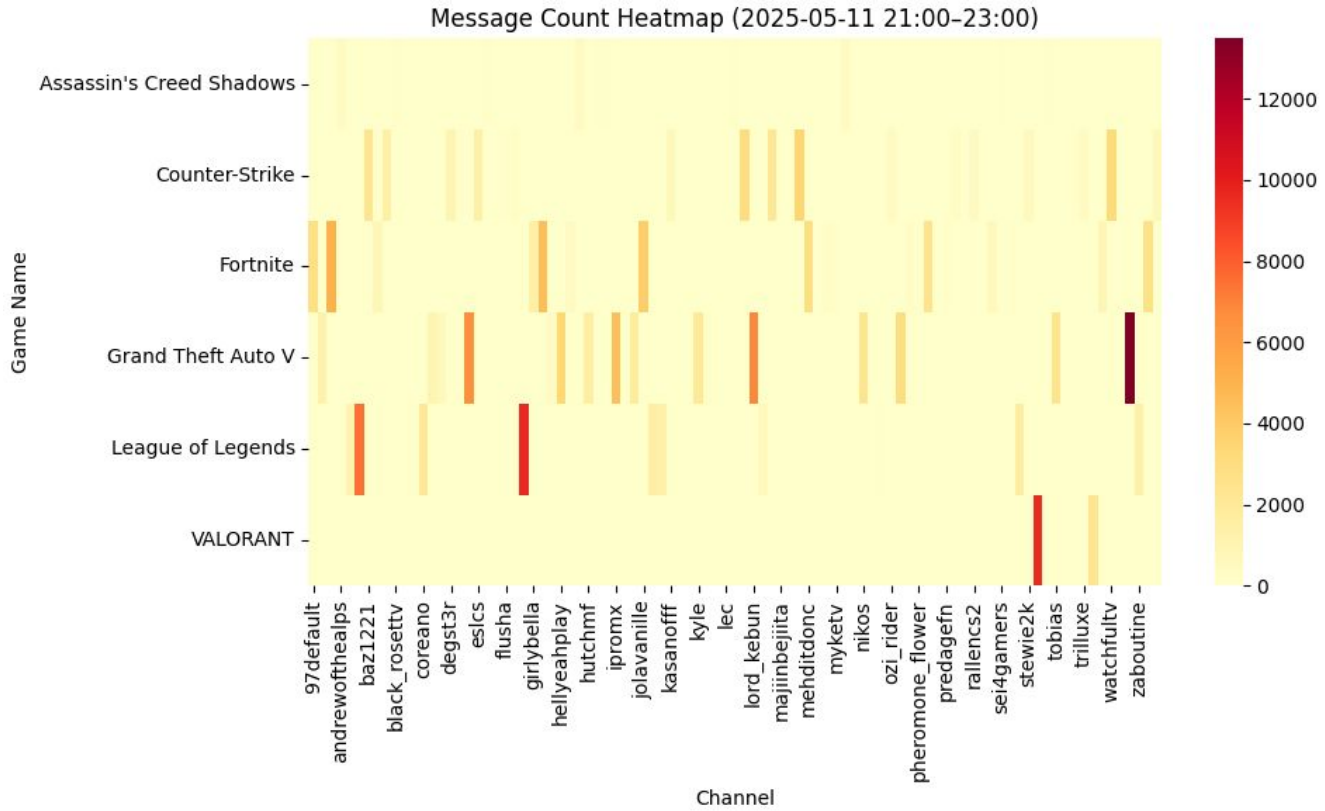




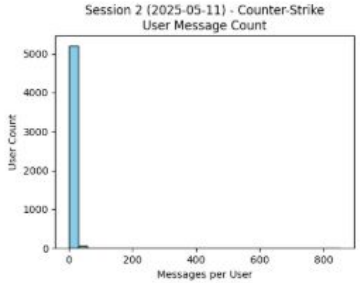


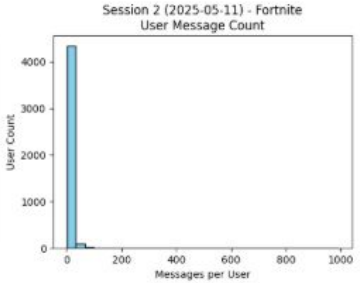


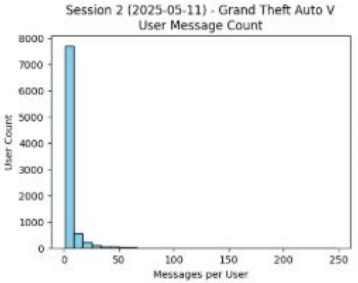
EDA: session2

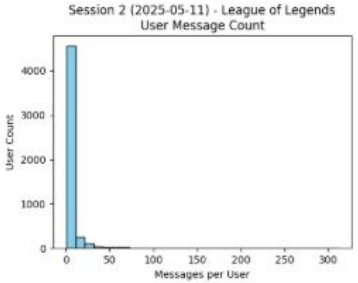


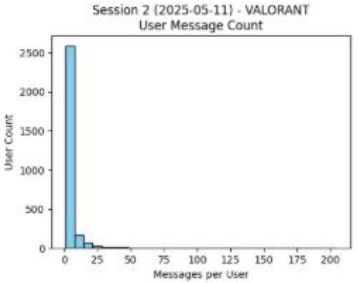
EDA: session2

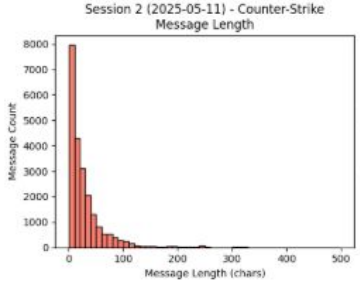


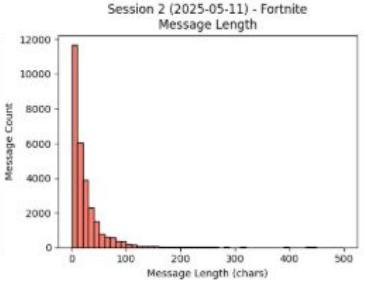


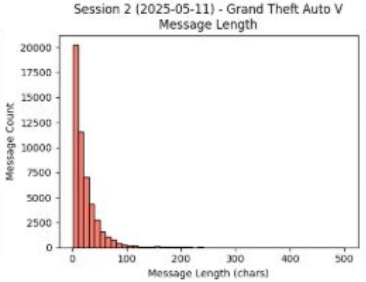


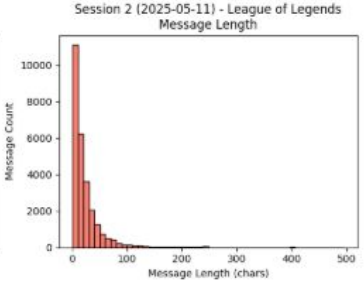


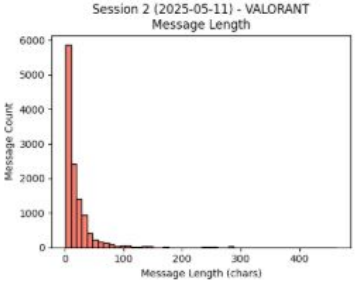




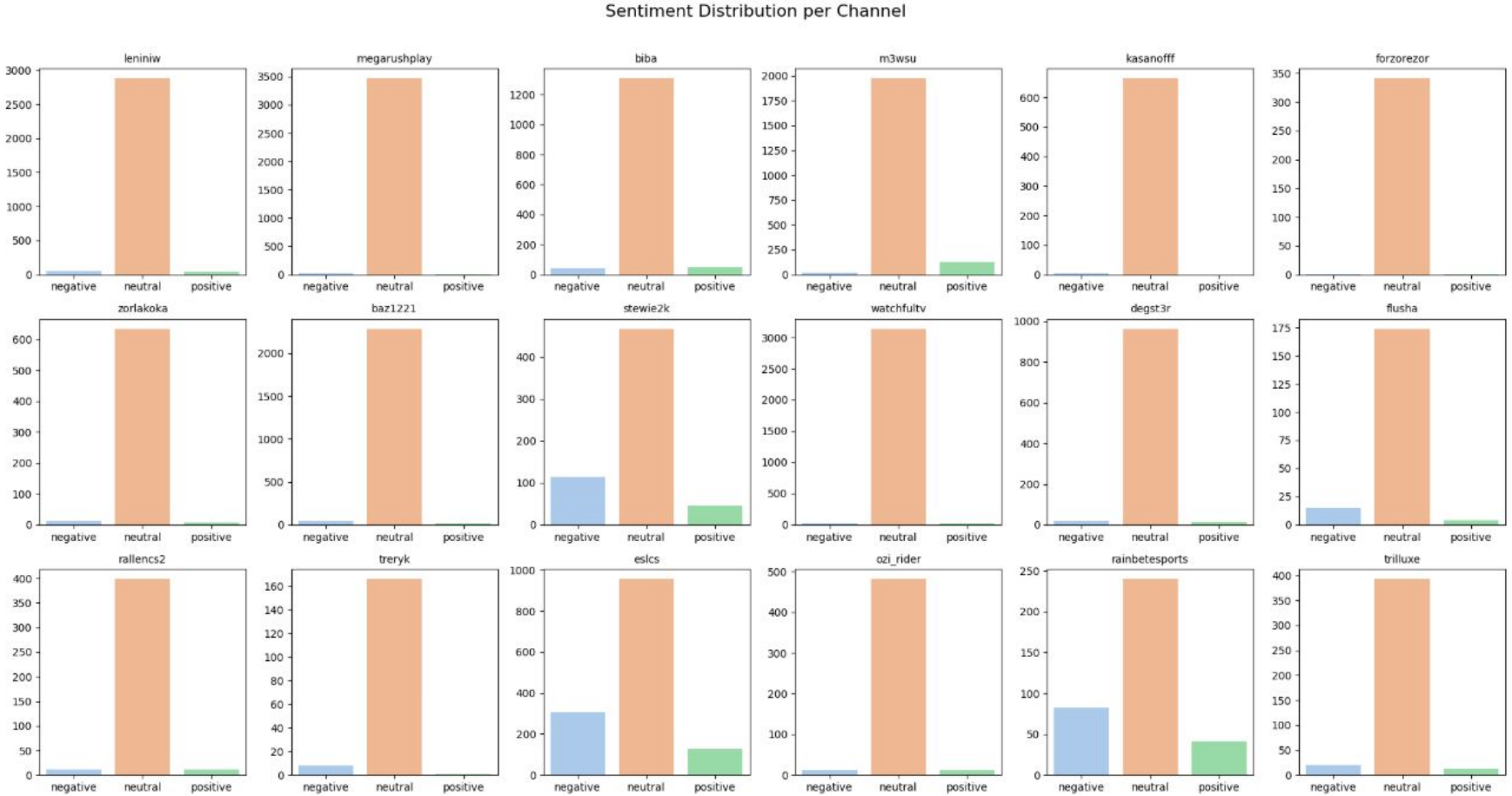






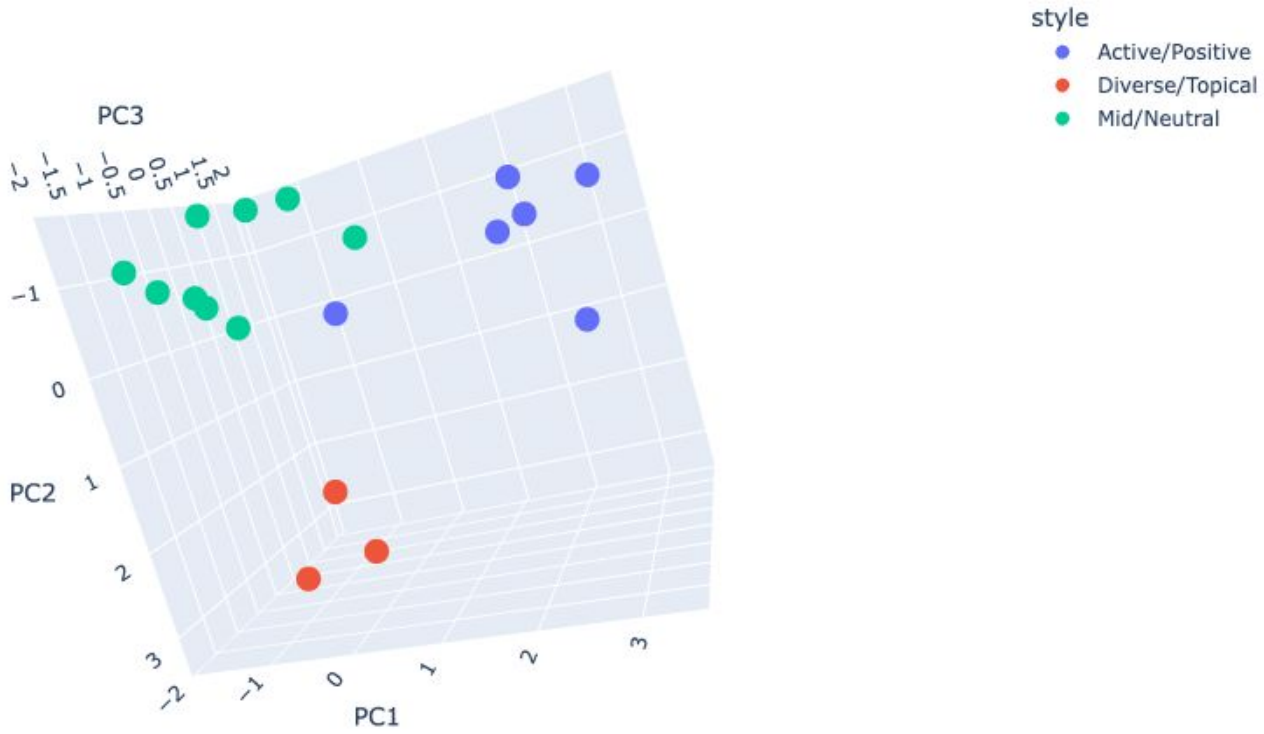


Text Modeling: Counter-Strike in Session2

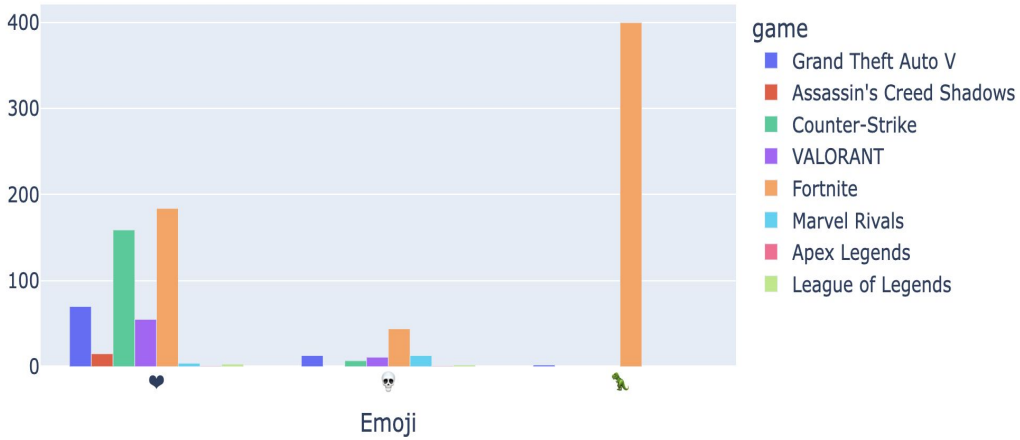


Text Clustering: Counter-Strike in Session2

🎮 Channel Style Clustering (3D PCA)



Emoji Preferences Across Games

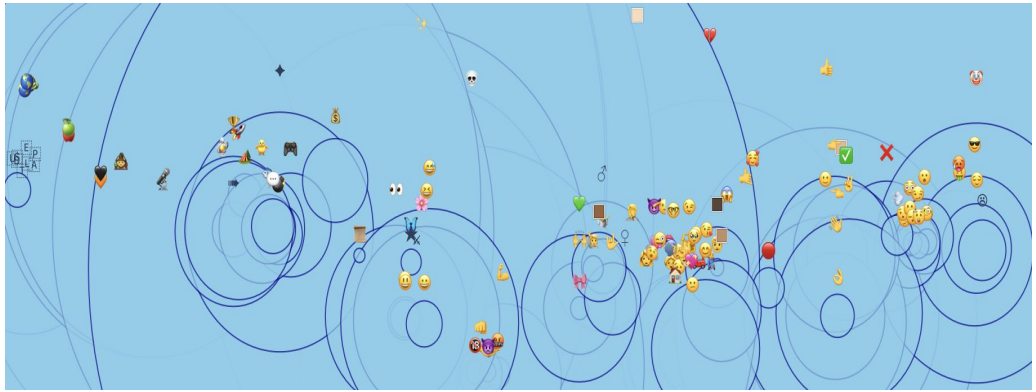


Before diving into semantic embeddings, let’s first look at how emoji usage varies by game

t-SNE of Emoji Embeddings

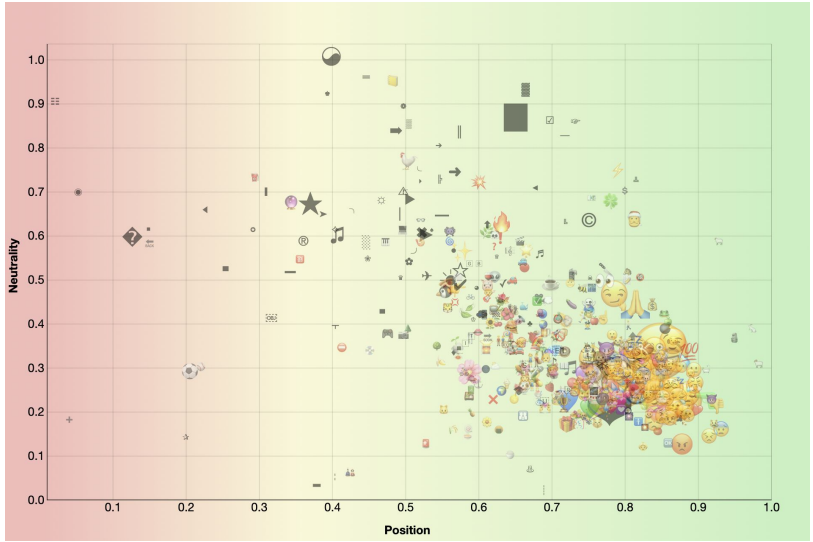
Some dense regions correspond to tactical game callouts, while others are driven by reactions, memes or inside jokes

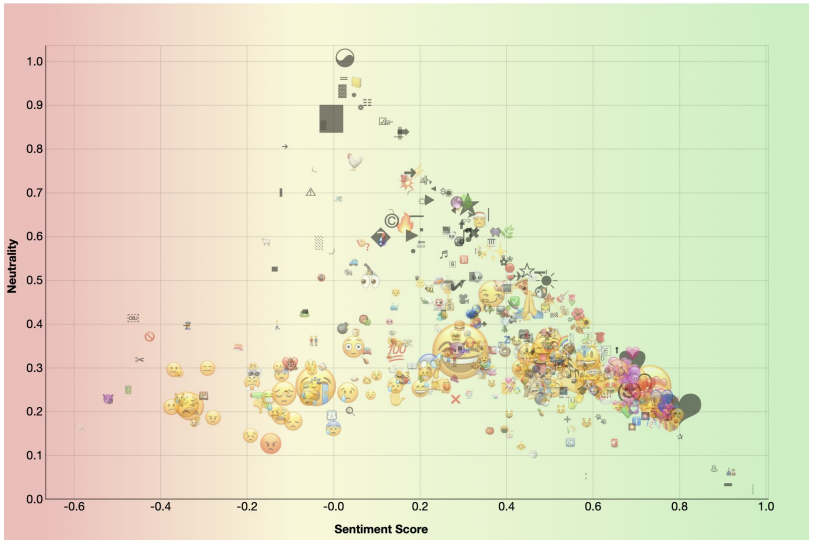
Even without labels, structure emerges, showing how shared language, repetition, and cultural references organise chat semantically



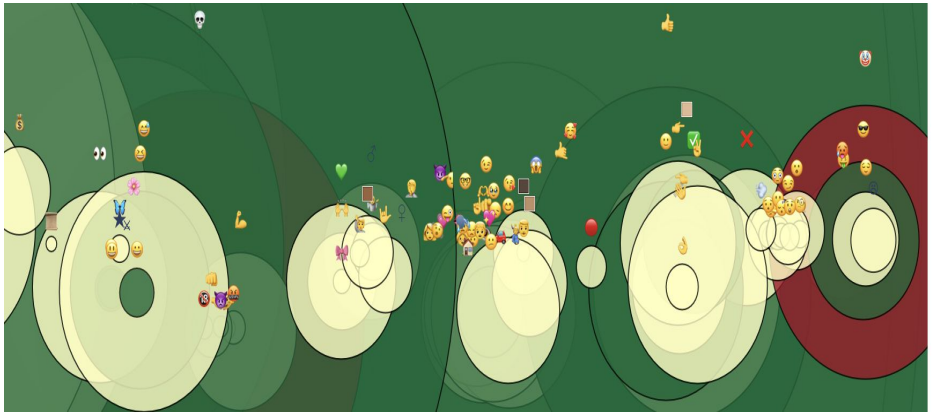
We move from words to emojis: Emoji Sentiment Score

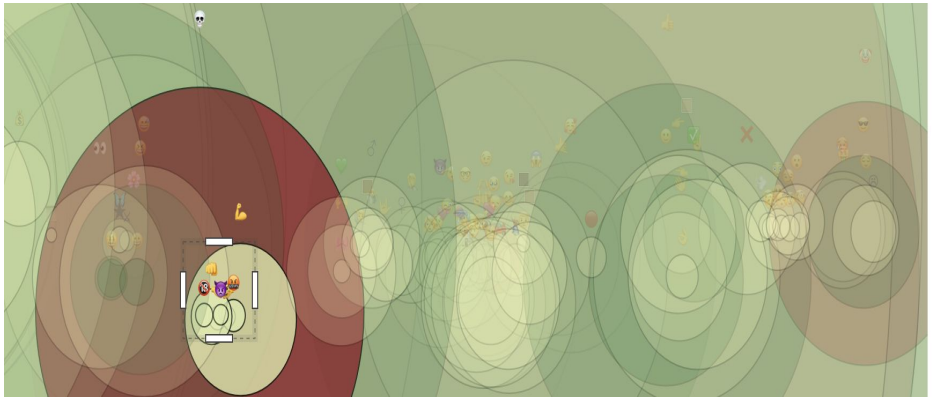
Using a combination of external sentiment rankings (like Emoji Sentiment Ranking from Škrlj et al.) and custom mappings





Emoji Co-occurrence Embeddings





Emojis that appear together often are embedded closer using Node2Vec.

We then color them using a sentiment gradient red for negative, green for positive, gray for neutral.

What emerges is a kind of emotional map of chat