**Datensatzstruktur (1\_general\_statistics):**

1. Anzahl der Genres: **260**
2. Gesamtanzahl der Kommentare English: **356.910** (nur englische Kommentare verwendet da Vader und LIWC nicht Englische Wörter versteht)
3. Kommentare pro Genre:  
   a. Die größten Genres haben ca. **3.000 Kommentare**  
   b. Die kleinsten Genres weniger als **100 Kommentare** (wahrscheinlich nicht repräsentativ)

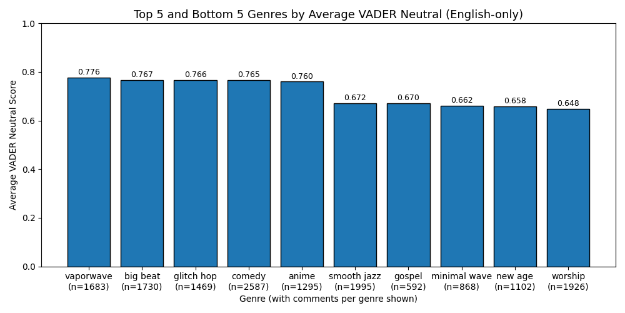
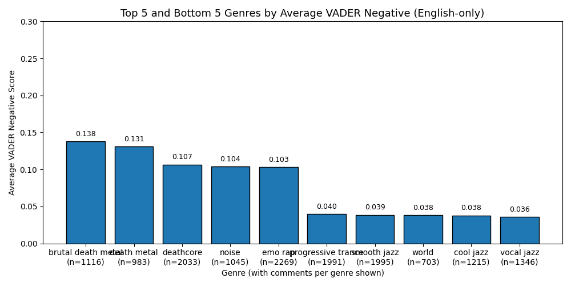
Für analysen und Korrelation habe ich eine Mindestschwelle von Kommentaren von 500 Kommentaren eingeführt, ansonsten wird sie für die Analyse ignoriert.

→ Effektive Anzahl der Genres in der Analyse: **233**.

**VADER-Analyse (2\_vader\_analysis):**

**Fragestellungen:**

1. Wie unterscheiden sich die Genres in den VADER-Scores (*vader\_pos / vader\_neg / vader\_neu / vader\_compound*)?  
   → Gibt es Genres, die systematisch positiver oder negativer sind? (Plots)

A graph of blue rectangular bars

AI-generated content may be incorrect.A graph of blue squares

AI-generated content may be incorrect.

**Welche Genres zeigen die stärkste Polarisierung?**  
→ Das bedeutet: In welchen Genres ist die **Varianz des VADER-Compound-Scores** zwischen den Kommentaren am höchsten?

A graph of blue and orange bars

AI-generated content may be incorrect.

1. Wie hängen **Likes** mit positiven, negativen, neutralen und compound-Sentiments zusammen (auf Genre-Ebene)? Antwort „Nein“

A graph with purple squares

AI-generated content may be incorrect.A graph with red squares

AI-generated content may be incorrect.A graph of blue squares

AI-generated content may be incorrect.A graph with green lines

AI-generated content may be incorrect.

As Hip hop quite a low average has low average compound score which was not expected to me:  
type of hip hop genres(compound rank of all genres) = alternative hip hop(222/233

), experimental hip hop (227/233 ), hiphop(215/233)

|  |  |
| --- | --- |
| 1. checked time component: same as all in all genre’s comments 2. is there one user that skews the data and wrote a lot of comments? Yes “@avalon.as.is1989” 66 out of 5000 comments 3. checked what his average compound score per comment is but was not significant  * Average compound score (all comments): 0.273 * Average compound score (Hip Hop only): 0.150 * Average compound score (Avalon only): -0.019 (n=66) * Average compound score (Hip Hop excl. Avalon): 0.152 (n=5352)  1. check are there any feature that differentiate significantly from the average comments? Only one that stuck out was emojis usage was up by 2. 2. Deepdived into this to maybe find correlation like (more emojis -> lower compound score) 3. Final finding the opposite occurred. People who tend to use more emojis tend to write more positive comments. Conclusion user who write emojis tend to write something positive. Even recalculating compound score without emojis it still show that significantly more comment including emojis are more possitiv the comments with no emojis. The ideal threshold is comments with 1 <= emojis <5 if a comment has more then 5 emojis its still more positive then a average comment with no emojis but less the the comments with 1<= 1 emojis <5 |  |
|  |  |
|  |  |

**LIWC-Analyse:**

**Fragestellungen:**

1. Welche LIWC-/VADER-Features weichen in bestimmten Genres am stärksten vom **Gesamtdurchschnitt** ab (mittels z-Scores)?
2. Welche Genres liegen ganz oben/unten im Vergleich nach **LIWC (emo\_pos − emo\_neg) Balance**?
3. Verteilung verschiedener LIWC-Kategorien über die Genres:
   * **Tone, Health, Religious, Illness, Death, Friend, Social, Family, Swear, Anger, Anxiety, Sadness, Politeness, Conflict, Politics, Ethnicity, Culture**
4. Top 10 high-variance LIWC/VADER features across genres

Interesting findings in those plots:

Hiphop again very low ranked for things like “Tone”, Pop Genres high for Social

**correlation\_between\_scores:**

1. Create correlation heatmap between vader scores and LIWC scores

Ist das bereits “genug” was kann soll ich noch ergenzen? Wie soll ichs dann in der Arbeit schreiben? Einfach ergebnisse also top 5? Hast du