Explicit Content in Songs and Gender

Final Project

Python for Data Science

Group: MMA
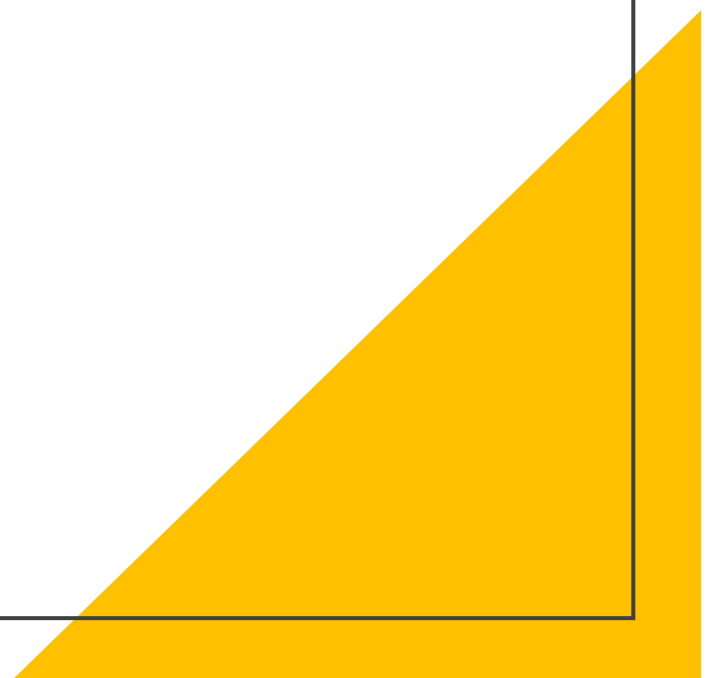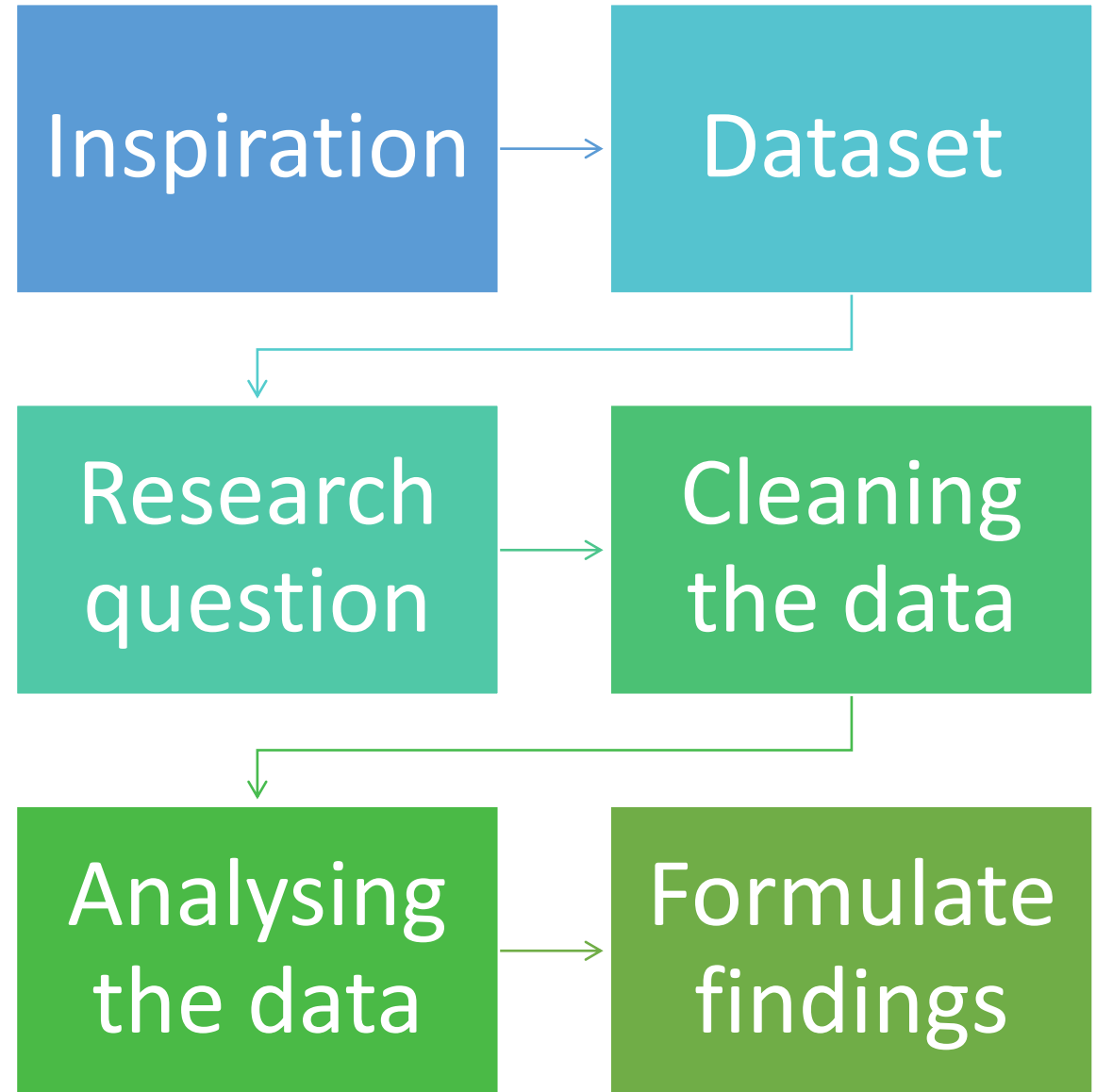
Table of content

- Research process
- Findings
- Limitations
- Further research

Research process

Research Process 1/4

Inspiration

- Research by Cambridge Institute

- Reason: increase negative emotional content, decrease positive emotional content
  - Content bias (Negative content has a preference)
  - Prestige bias (Best-selling artists are copied)
  - Success bias (Best-selling songs are copied)

Cultural evolution of emotional expression in 50 years of song lyrics
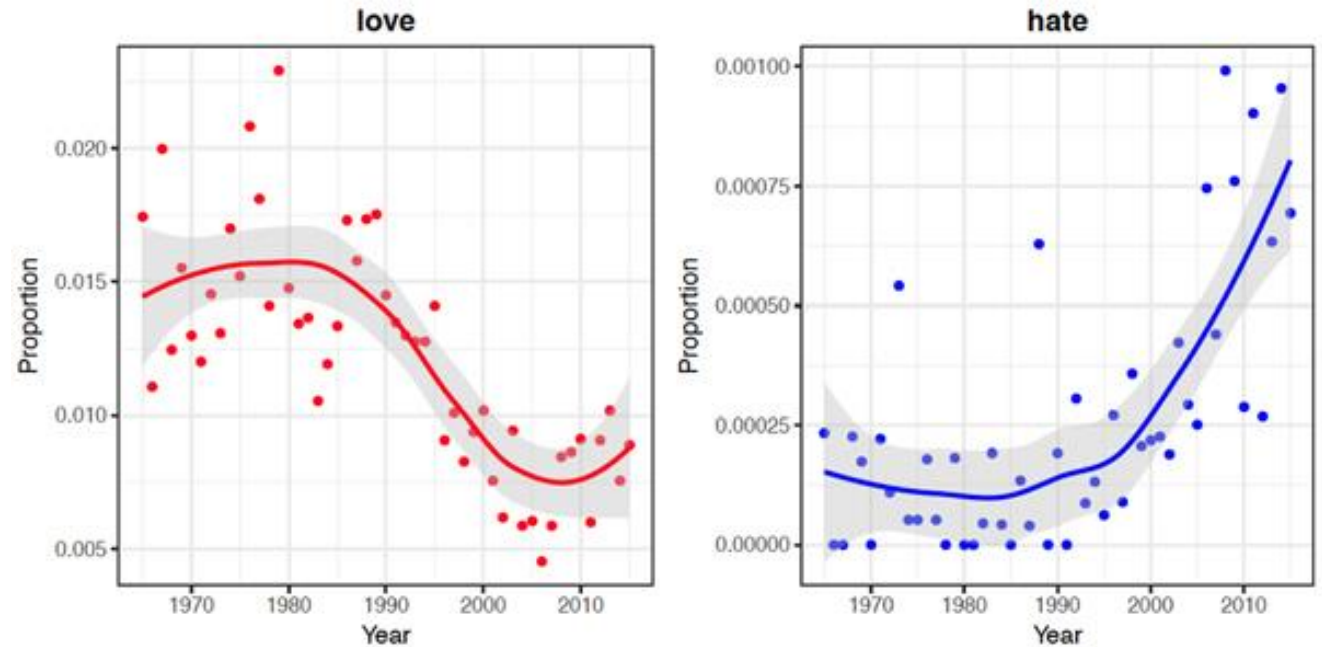
Research process 2/4

Dataset

- "Top Hits Spotify from 1999-2020"
- "top10k-spotify-artist-metadata"

Research process 3/4

Initial research question

"To what extend has explicit content in songs evolved over the last 20 years?"

Main research question

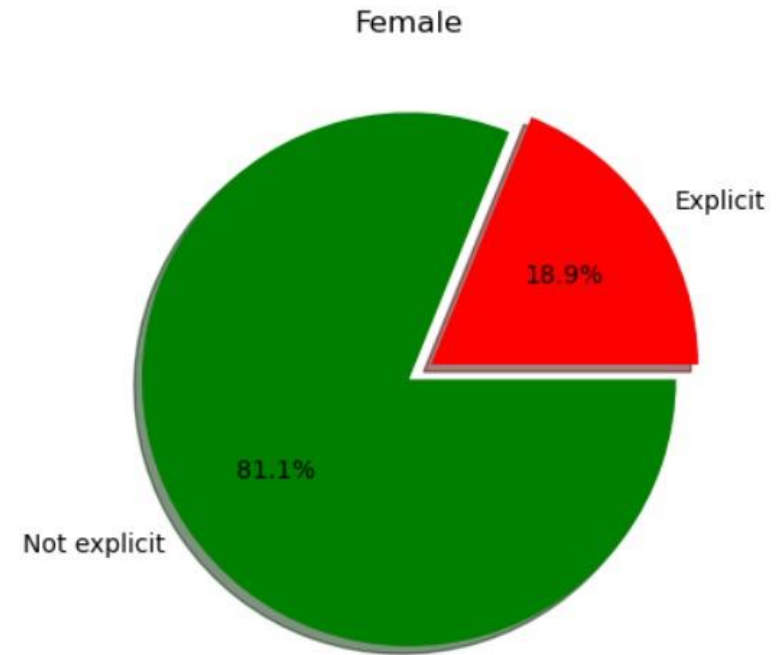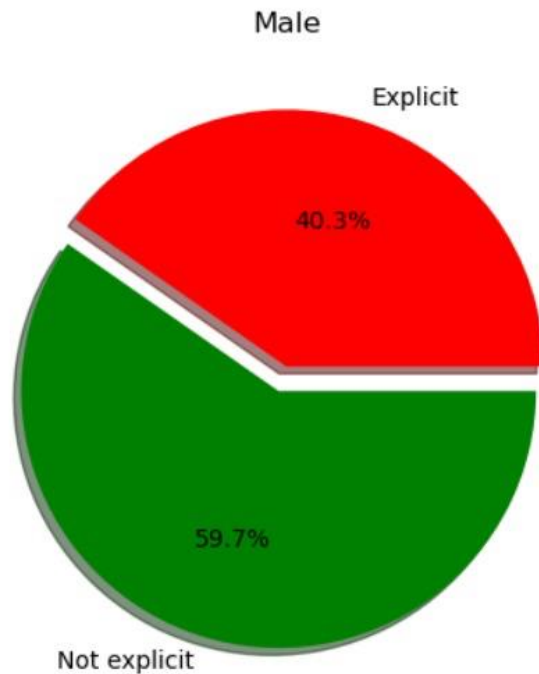"In which way have the evolutionary changes been subject to gender?"

(Definition of Explicit according to Spotify: "Songs which can contain coarse language, reference to abuse/violence etc., glorification of bad/criminal behaviour, discrimination, sexual preference of all of the above")
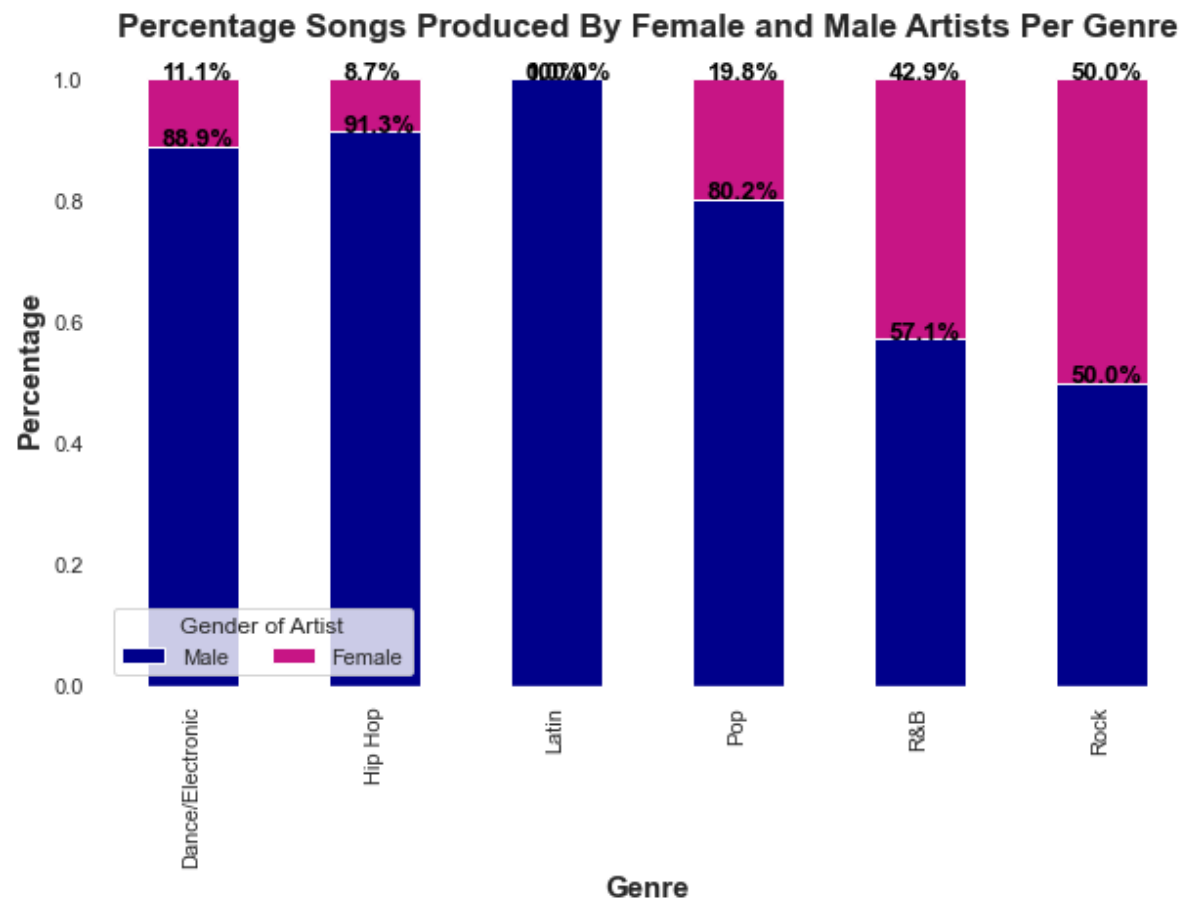
Research process 4/4
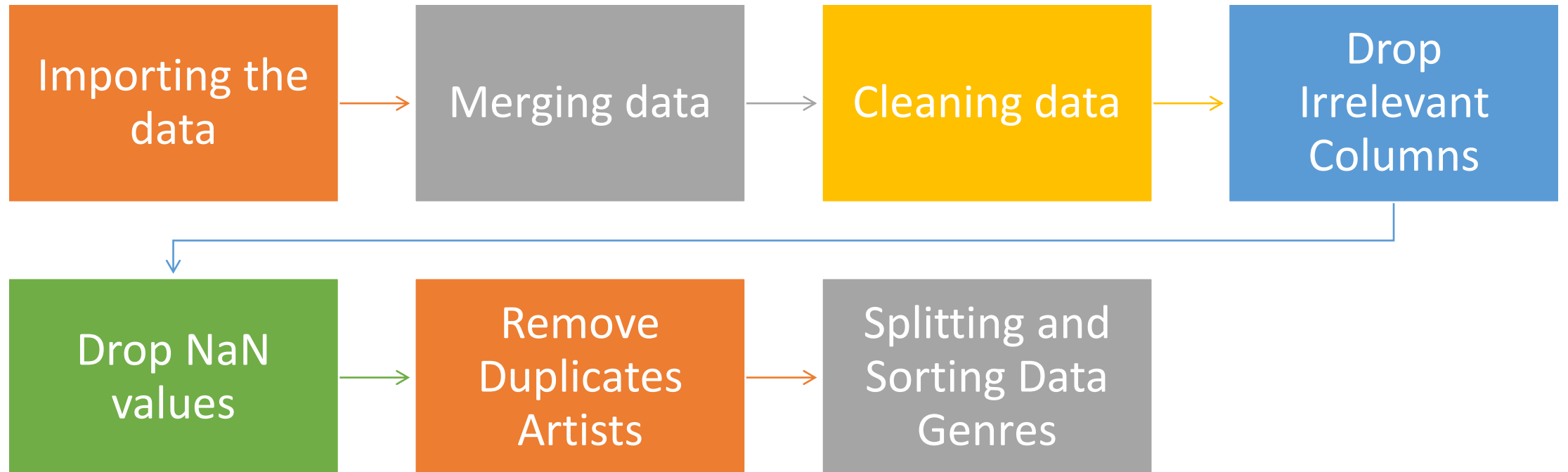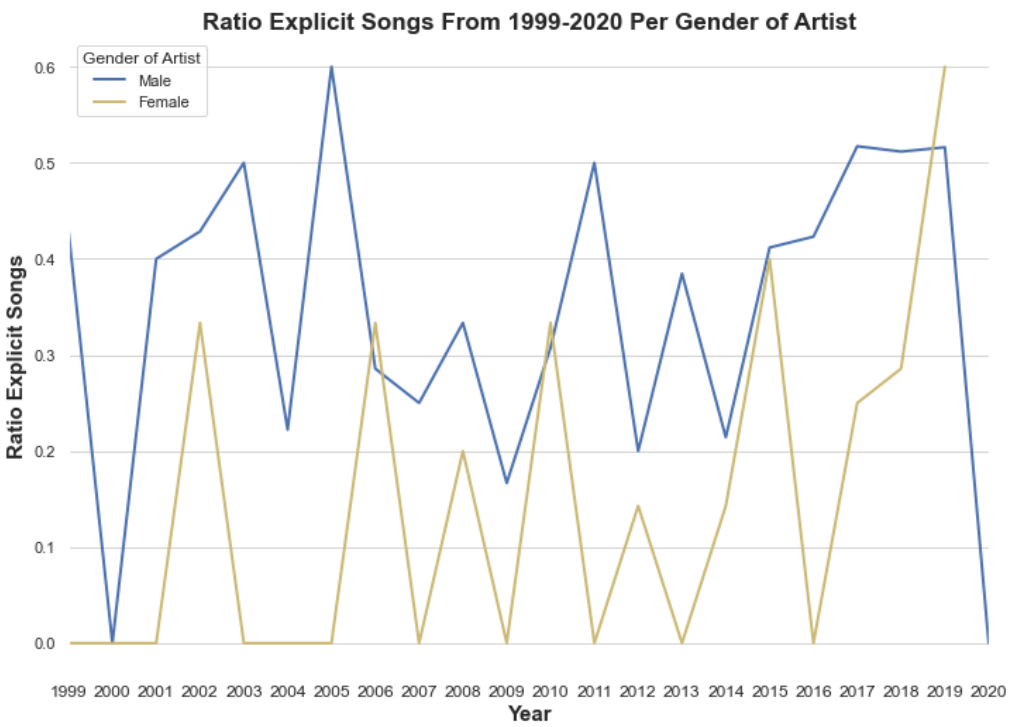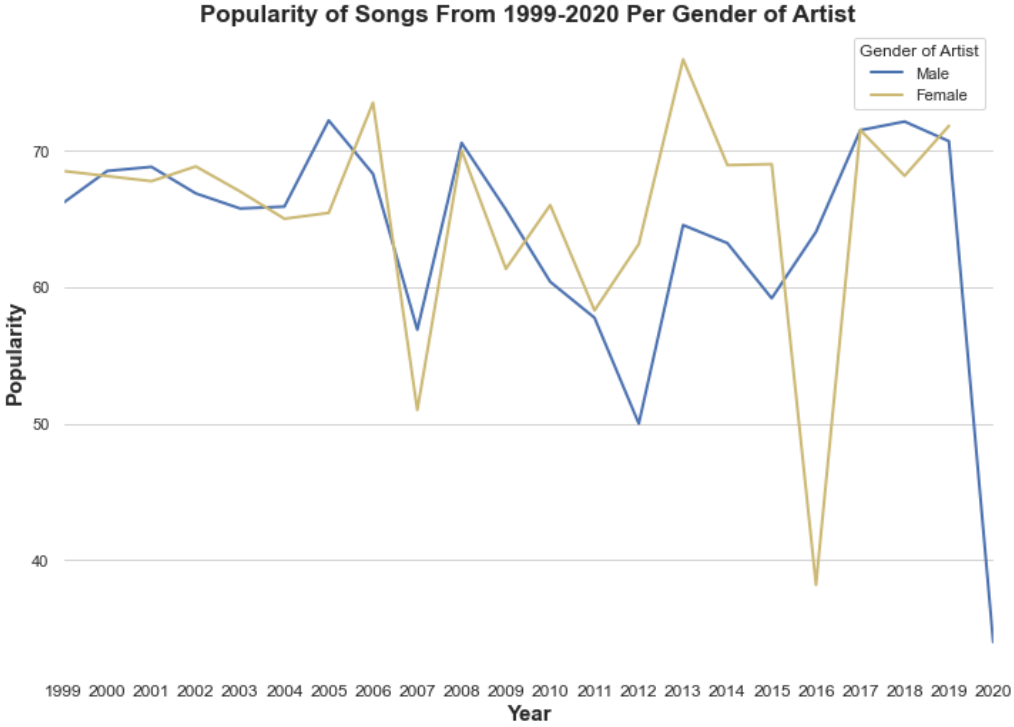
Findings 1/4

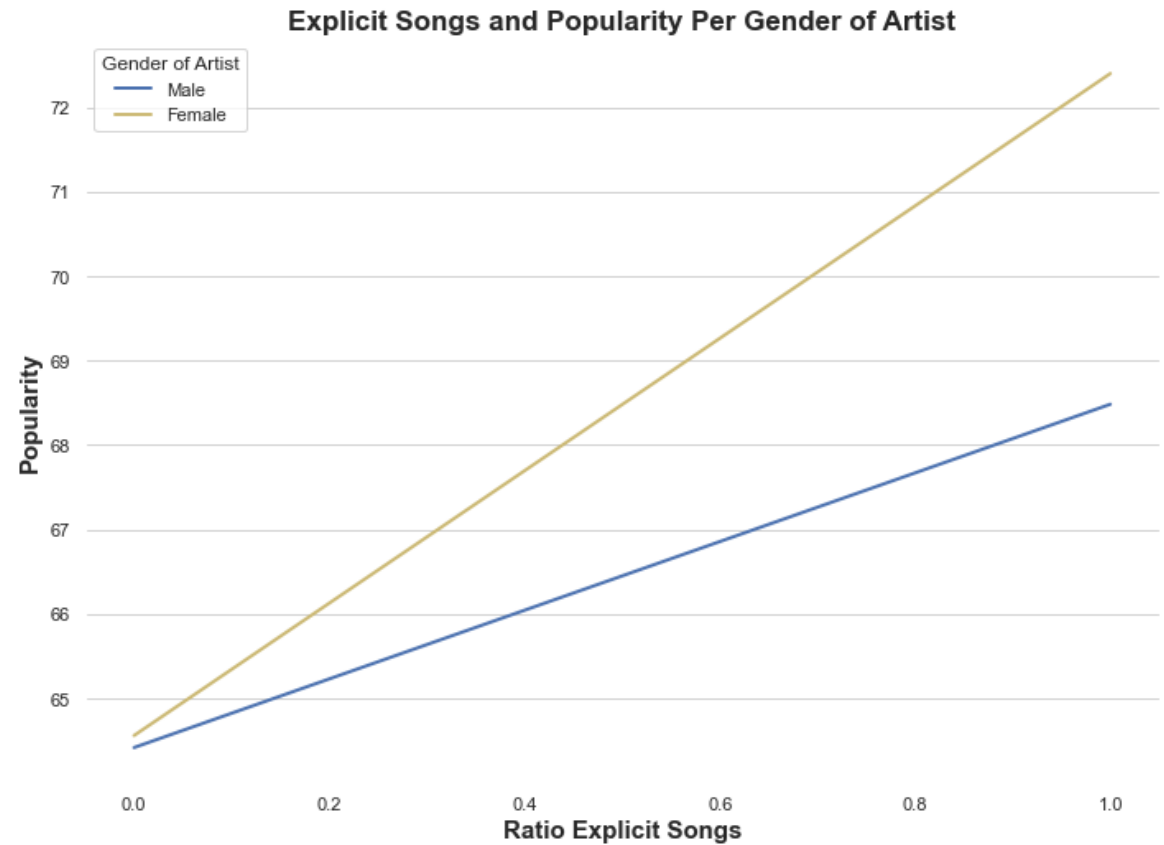Ratio explicit/non-explicit per gender

Findings 2/4

Percentage Songs Produced By Female and Male Artists Per Genre

Note: Women in Latin are underrepresented

Findings 3/4

Popularity of Songs From 1999-2020 Per Gender of Artist

Ratio Explicit Songs From 1999-2020 Per Gender of Artist

Findings 4/4

Similar to the Cambridge studies, the explicit content has a preference

Explicit Songs and Popularity Per Gender of Artist

Limitations

- Merging of datasets dramatically decreased the number of artists
- Less data for women
- No graph for total songs men women in genres
- Some genres have less data than others
- We did not do a Z or T test for everything

Further research

- More research on timeline

- More data across different years to clearly see the change

- Further research could more accurately analyze the differences between genders, more data needed

- More research of the importance and correlation of non-conventional genres, as we did not have much data in this

- More research on significant analysis, T tests Z tests