A Streaming Odyssey

Danish James

Objective

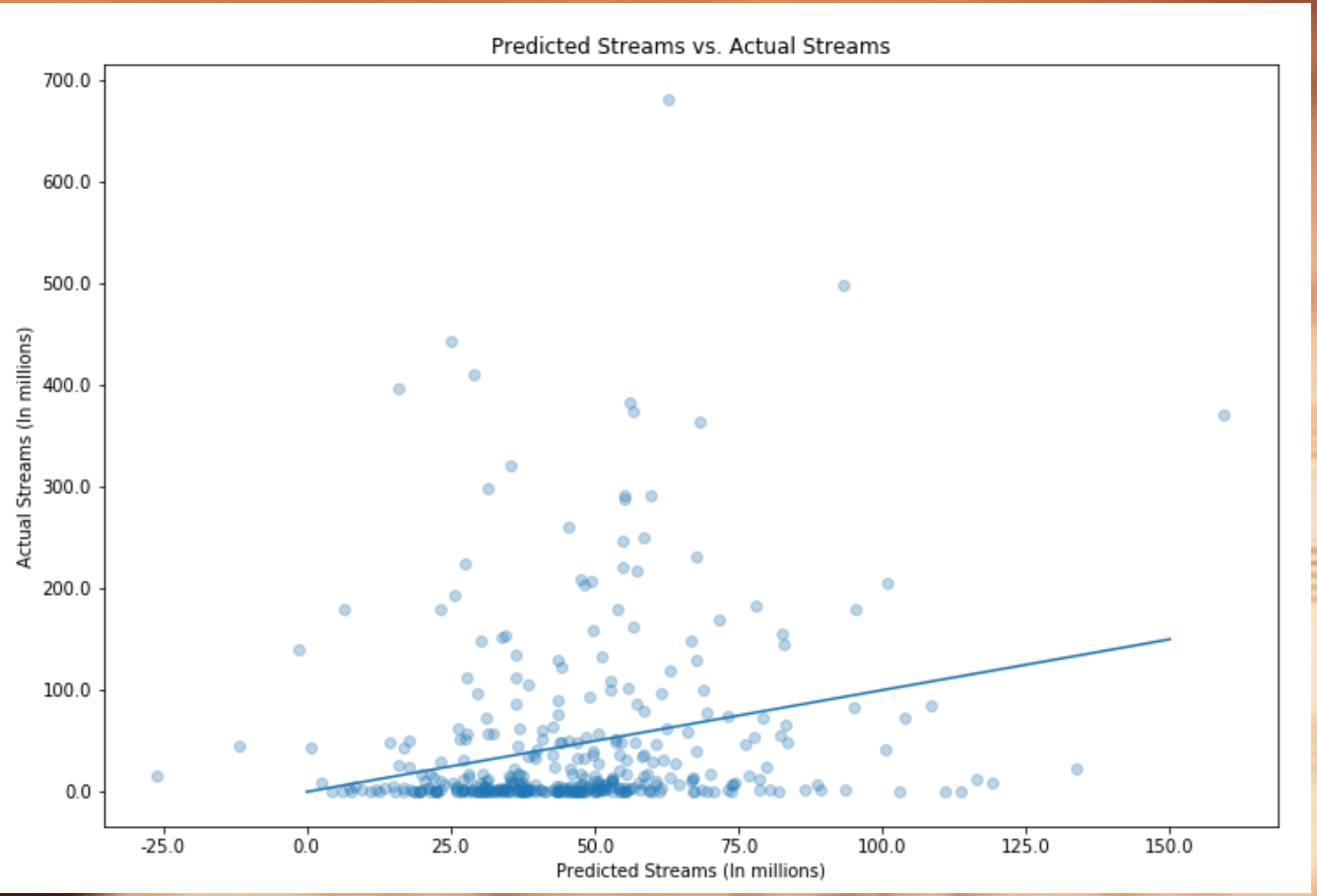
- Music Streaming increasing in popularity every year
  - Most common method of listening to music
- What factors lead people to listening to a specific song?
- Analyze data from Spotify
  - Identify what intrinsic values of a song lead to people listening to them more (BPM, Mode, etc)
  - Can we design songs specifically to get more streams?

Methodology

- DATA
  - Scraped Streaming Data from Spotify
  - Audio Features pulled from Spotify API
- TOOLS
  - Python, Pandas, Numpy
  - BeautifulSoup, sklearn, Spotipy
  - Matplotlib

Results

- LASSO Regression Model
- R2 Value of 0.0239
- Low Correlation
- In general model underpredicts stream counts greatly
  - Lot of discrepancy between songs that only showed up once on charts verses songs that lasted on the charts for months
  - Very high outlier streams



Conclusion

- No strong correlation found between intrinsic factors and streams
- Songs that share the same features are not guaranteed to perform the same
  - Additional outside factors affect how songs are received
- Lots of songs overlap on intrinsic factors which also caused issues

Further Work

- Implement more features that are not dependent on the song
  - Artist information
  - Seasonal/Musical trends
- Utilize other services besides Spotify
  - Might be inherent bias towards specific songs based on Spotify's audience.
  - Trends may be different on services that are fully paid vs freemium

Appendix

- https://developer.spotify.com/documentation/web-api/
- https://spotifycharts.com/regional/global/daily/latest
- Daily chart data pulled over the course of a year (2018-10-02 to 2019-10-01)
- Utilized global stream data, potentially could be better if focused on specific regions