ANALYZING POPULAR MUSIC: SPOTIFY AND VISUALIZATION

Arianna Schmid, Ellen Khachatryan, Marcos Micheloni, Margaret Pires-Fernandes
Group 157

INTRODUCTION

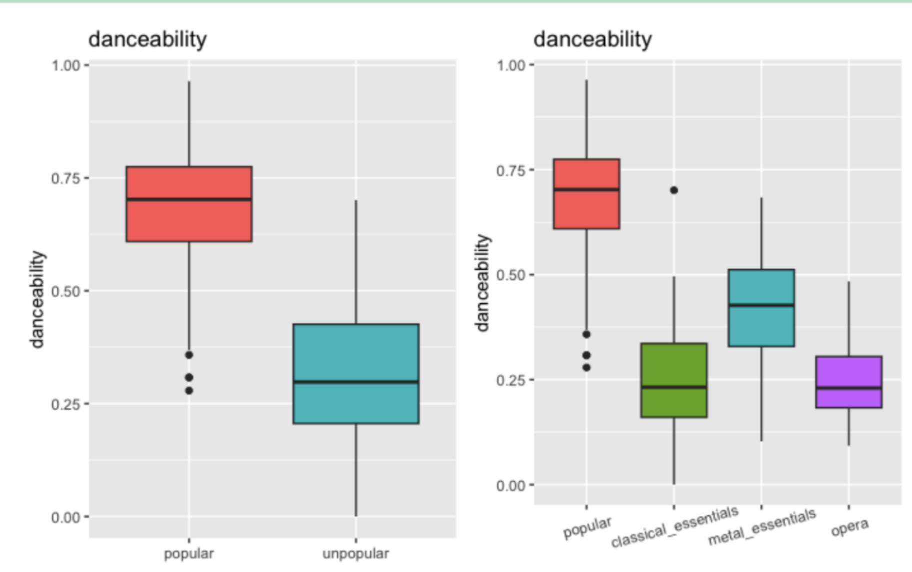
- Most music is digital. Big data + prediction tools are now essential for artists, labels, and streaming platforms to make money.
- We aimed to build an interactive platform that lets you:
  - listen to songs
  - explore their musical attributes and components
  - predict their likelihood of becoming popular
- There aren't many platforms with the capability to both listen and explore the nature of what you are listening to.
- This platform can be used for research, entertainment, or analysis. It is intended for music enthusiasts aspiring artists, and those in the music business to better understand what makes music popular.

DATA METHODOLOGY

- The popular music data frame was built by using webscraping techniques as well as creating spotify playlists.
  - BeautifulSoup and Python used to scrape the top 50 Billboard songs from 2012 to 2022.
  - Created spotify playlists containing these songs for each year. This allowed us to access their attributes (tempo, danceability, etc.) through the spotify API.
  - Our final data frame for popular music contained 1096 rows and 17 columns (year, rank, name, artist, URI, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, and duration).
- Literature suggested classical and heavy metal generally rank low on popularity so we aggregated classical music, opera, heavy metal, and other playlists in order to create our frame for unpopular music with 261 rows.
-

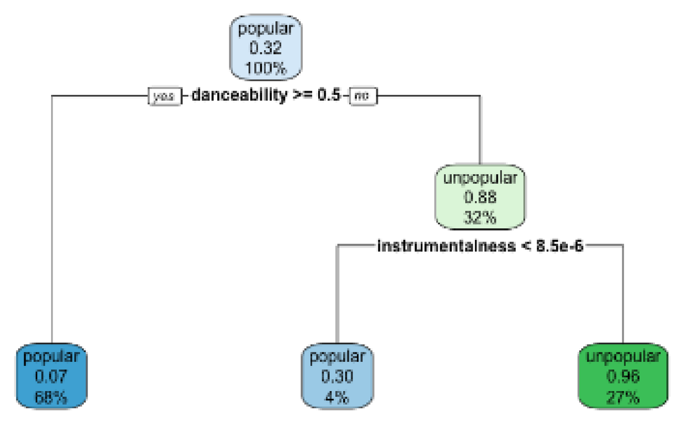
year <int>	rank <int>	name <chr>	artist <chr>	danceability <dbl>	energy <dbl>	key <fctr>	loudness <dbl>	speechiness <dbl>
2012	1	Somebody That I Used To Know	Gotye	0.864	0.495	0	-7.036	0.0370
2012	2	Call Me Maybe	Carly Rae Jepsen	0.783	0.580	7	-6.548	0.0408
2012	3	We Are Young (feat. Janelle Monáe)	fun.	0.378	0.638	10	-5.576	0.0750
2012	4	Payphone	Maroon 5	0.739	0.756	4	-4.828	0.0394
2012	5	Lights	Ellie Goulding	0.703	0.731	8	-6.283	0.0311

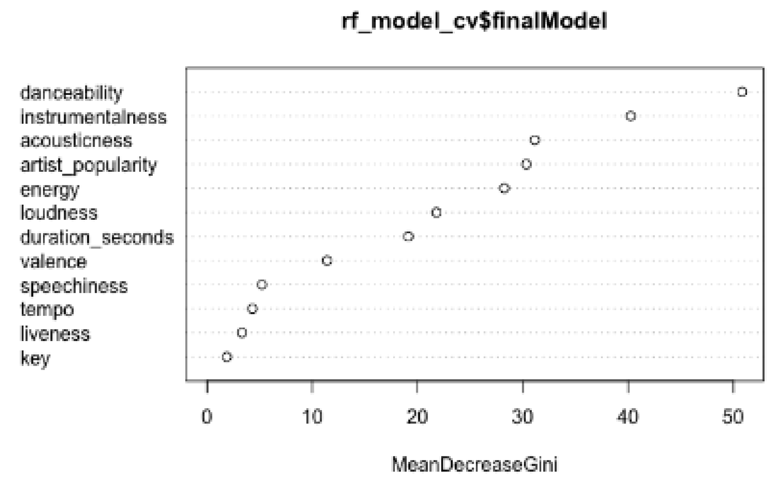
POPULAR VS UNPOPULAR MUSIC



These boxplots demonstrate how danceability varies between popular and unpopular music (left), and how it varies between the different genre's of music (right).

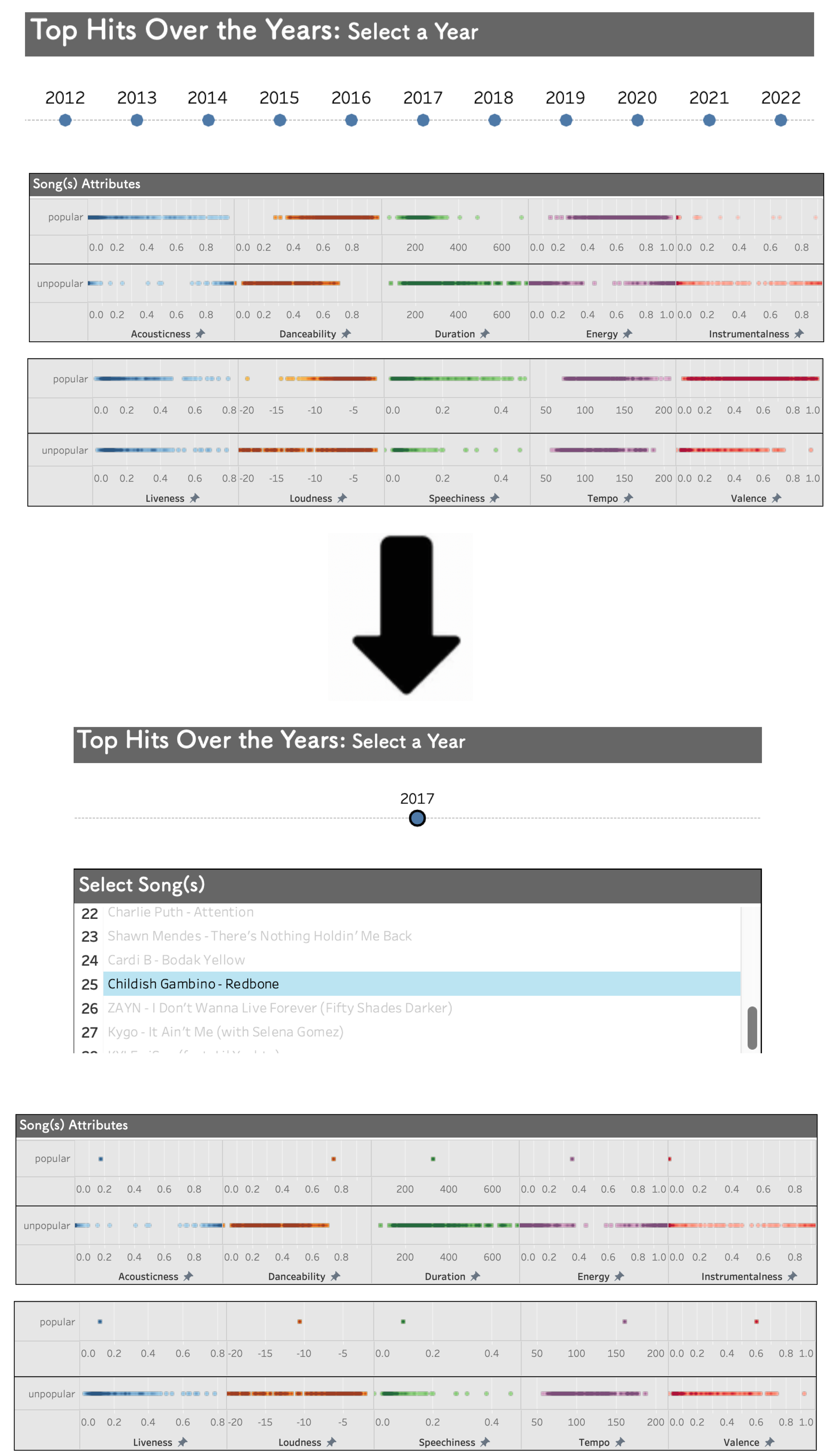
MODELS





- Logistic regression, SVM, KNN, decision trees, and random forest models were used. All models were cross validated.
- Decision tree and random forest results above helped us conclude that danceability is the most influential factor for a song becoming a pop hit.
- By reading song input and classifying it as popular or unpopular, our models can perform predictions using a given song's corresponding attributes. Future project extensions should incorporate this prediction component in the interactive visualization

TABLEAU DASHBOARD



REFERENCES

- [1] Omar Hujran (2020) Big Data and its Effect on the Music Industry, The 3rd International Conference on Software Engineering and Information Management, DOI:10.1145/3378936.3378957 [2] Bauer, C., & Schedl, M. (2019). Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems. PLOS ONE, 14(6). https://doi.org/10.1371/journal.pone.0217389 [3] Gupta, Shubhanshu. (2014). Music Data Analysis: A State-of-the-art Survey. Dhirubhai Ambani Institute of Information and Communication Technology [4] Sciandra M, Spera IC. A model-based approach to Spotify data analysis: a Beta GLMM. J Appl Stat. 2020 Aug 10;49(1):214-229. doi: 10.1080/02664763.2020.1803810. PMID: 35707796; PMCID: PMC9042099 [5] Kim, J. (2021, October 29). Music popularity prediction through data analysis of Music's characteristics. International Journal of Science, Technology and Society. https://www.sciencepublishinggroup.com/journal/paperinfo?journalid=183&doi=10.11648%2Fj.ijsts.2021 0905.16 [6] Chryssanthacopoulos, J. (2020, December 15). Advanced Music Analytics using machine learning. Medium. Advanced music analytics using machine learning [7] Interiano, Myra (2018), Musical trends and predictability of success in contemporary songs in and out of the top charts, The Royal Society, v.5(5). PMCID: PMC5990848; PMID: 29892348 [8] Rasmus Fleischer (2021) Universal Spotification? The shifting meanings of "Spotify" as a model for the media industries, Popular Communication, 19:1, 14-25, DOI: 10.1080/15405702.2020.1744607 [9] Sun P. Music Individualization Recommendation System Based on Big Data Analysis. Comput Intell Neurosci. 2022 Jul 5;2022:7646000. doi: 10.1155/2022/7646000. PMID: 35837215; PMCID: PMC9276508. [10] Khulusi, R., Kusnick, J., Meinecke, C., Gillmann, C., Focht, J. and Jänicke, S. (2020), A Survey on Visualizations for Musical Data. Computer Graphics Forum, 39: 82-110. https://doi.org/10.1111/cgf.13905 [11] K. E. Barkwell et al., "Big Data Visualisation and Visual Analytics for Music Data Mining," 2018 22nd International Conference Information Visualisation (IV), Fisciano, Italy, 2018, pp. 235-240, doi: 10.1109/iV.2018.00048 [12] Spotify Developer Policy. https://developer.spotify.com/policy [13] Taren-Ida Ackermann, "Rationales and functions of disliked music: An in-depth interview study," 2021 Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany. PMCID: PMC8