

The Intersection between Country Standing and Consistency of Spotify Charts

Neo Kok, Harish Jaisankar, Dustin Lee

MUgSS Datathon, 3/11/2023



Goal

Identify how a nation's status as a first, second, or third world country correlates with how different Spotify's Top 200 playlist is from the Viral 50 playlist in specific regions.



Background

- Spotify creates a variety of data-driven charts for its users
- This includes the Top 200 and Viral 50
- The Top 200 is decided purely based on a song's Spotify streams
- The Viral 50 is “entirely data-driven” and uses statistics including:¹
 - Stream growth
 - Recency
 - Number of shares



Hypothesis

Third-World countries will see a smaller difference between songs in the Top 200 and Viral 50 Spotify Charts than First and Second World countries due to less online social interactions and therefore less data to create different Viral Charts. In other words, Spotify will have to rely more on number of streams to create Viral Charts rather than data gathered from shares and social media.



Methodology

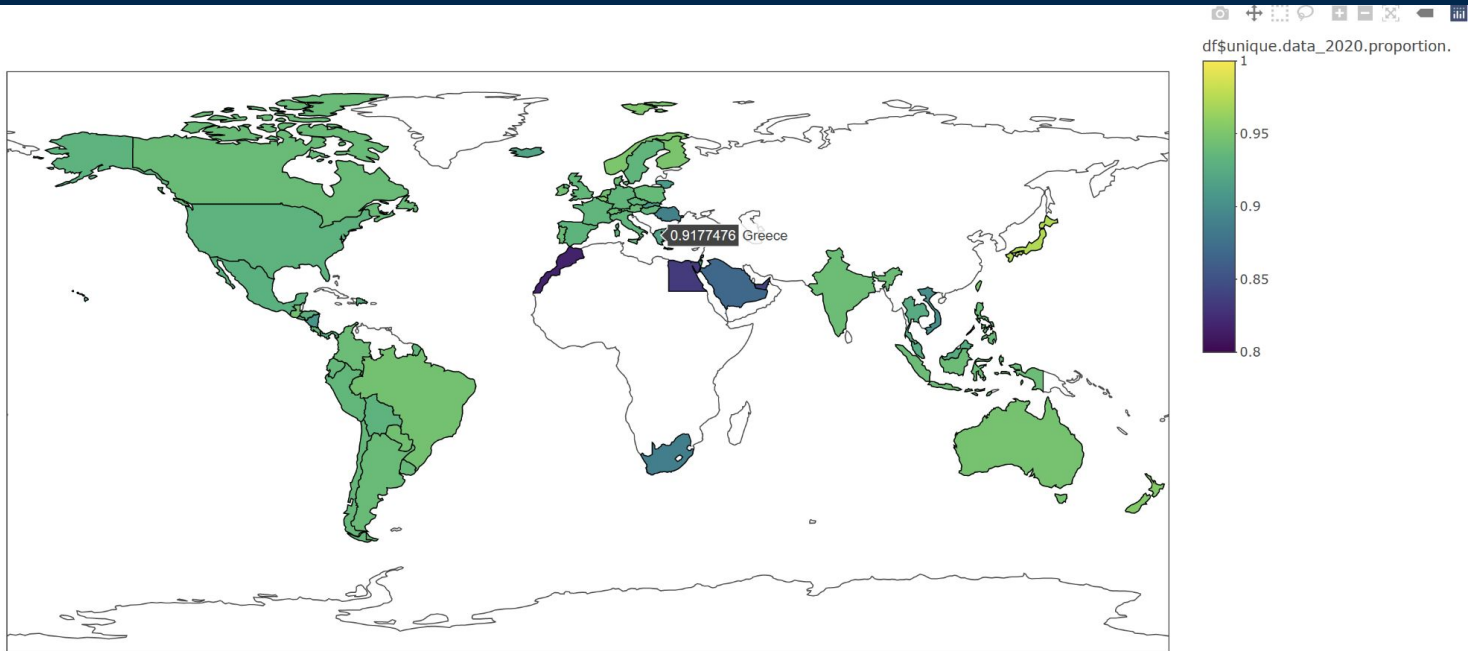
1. Read in large dataset
2. Cut out all rows with rank greater than 50
3. Separate data set by source (Top 200 and Viral 50)
4. Filtered just for 2020
5. Identified First/Second/Third World countries²
6. Calculate percent difference between proportions of Viral 50 and Top 50 for each country on each day
7. Create visualizations to track differences across days



How does data variance
look geographically?



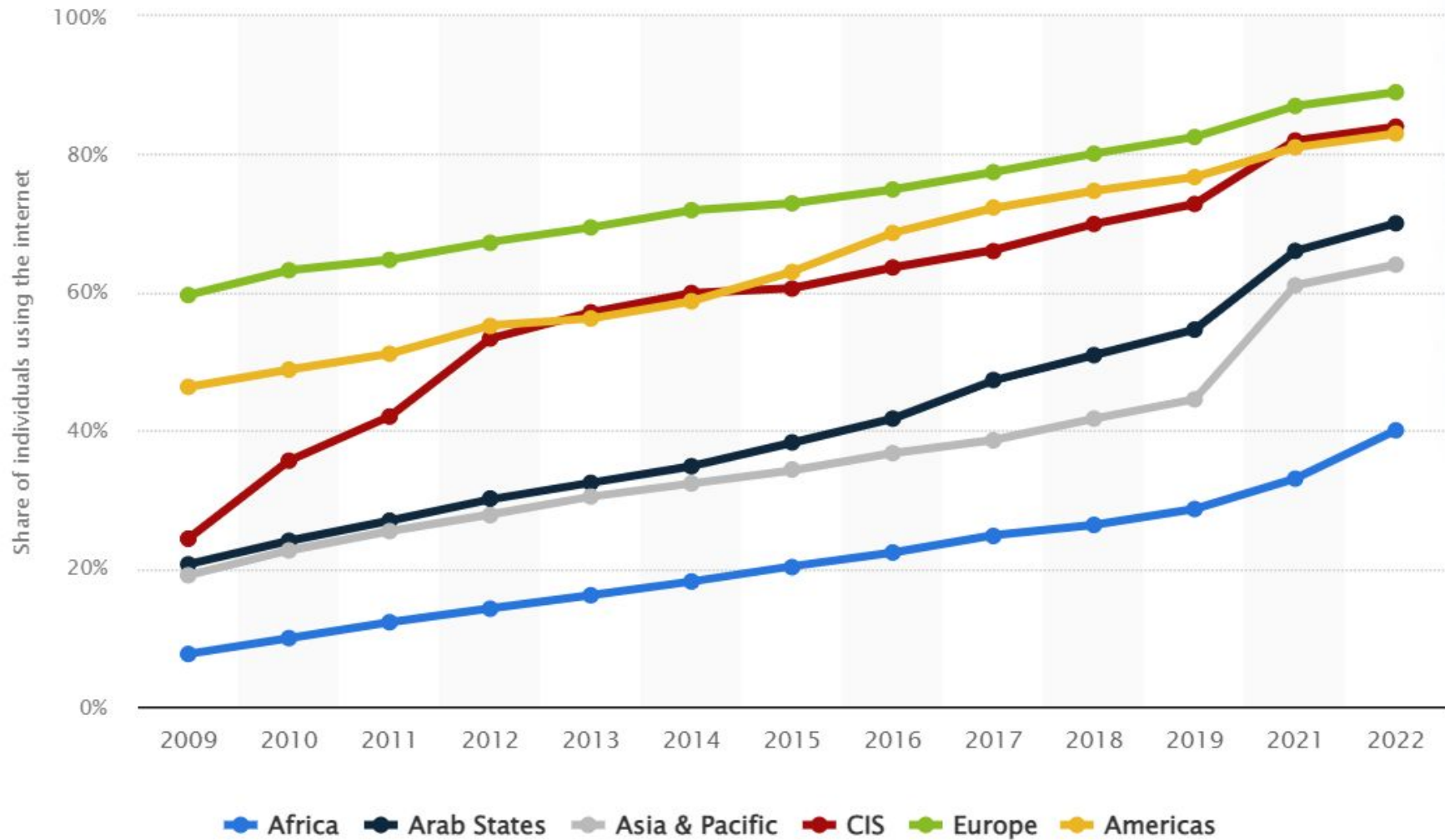
World Map



Play

df\$date: 2020-01-01

2020-01-01 2020-01-29 2020-02-26 2020-03-25 2020-04-22 2020-05-20 2020-06-17 2020-07-15 2020-08-12 2020-09-09 2020-10-07 2020-11-04 2020-12-02 2020-12-30





Do First-World countries have a significantly different variance in their Spotify data-generated charts?



Results and Findings

We ran a proportion test between the difference in proportions of songs that were in the Viral 50 Charts but not in the Top 50 charts for First World countries and Third World countries. Although there was a difference in proportions between the two types of countries (0.9302 to 0.91), there was not a significant difference at the 0.05 level ($p = 0.5989$).



Conclusion

In conclusion, though the correlation we noticed was not statistically significant, it still lends itself to useful insights regarding internet access around the world and the prevalence of social networking in different regions.

Investopedia's definition of a first-world country does not highlight internet usage and focuses more on economic factors (GDP, GNP) and the HDI, but perhaps it's time to start considering internet access and usage as a real deciding factor in the differentiation of “developed” and “developing” countries.²



References

1. <https://artists.spotify.com/en/help/article/how-viral-charts-work>
2. <https://www.investopedia.com/terms/f/first-world.asp>
3. <https://www.statista.com/statistics/265149/internet-penetration-rate-by-region/>