Worldwide Trends in Music Streaming According to Spotify

Ruby Imrie and Orlagh Keane

April 11, 2022

1 Overview

Spotify was founded in 2006 in Sweden and ever since has become an increasingly popular way to listen to music across the globe. Spotify is now the most popular music streaming service in the world [4] and is now available in over 170 [7] countries and has changed the way people list to music across the globe.

This report looks at the data collected for songs that reached the daily top 200 in the year 2017. It analyzes how artists and tracks' popularity varies in time across the global stage. In particular, looking at whether regions share the same top ranking songs and if popularity of a track in one region predict its upcoming popularity in other regions. Countries were split into continents and these were the regions that were looked at. After discovering that some regions shared high percentages of the same top ranking songs, this was looked into further by visualising how the ranking of songs differed over time in different regions. The findings show that popularity in some regions encouraged the popularity in others. Logarithmic Regression was used to look at the relationship between ranked position and number of streams for tracks. The study identified that the higher the position the greater the number of streams. In addition the study finds the correlation between a tracks peak position in the chart and how long it charts by using Poisson distribution to derive a probability mass function and by looking at the distribution, it was revealed that the higher the peak position the longer it would chart for. Furthermore the report covers how countries that speak the same language, have a much higher percentage of shared songs rather than those which don't.

2 Introduction

Context and motivation The music streaming industry is one of the world's greatest growing markets as more and more people subscribe to a streaming service every day. Music streaming is the best way to discover new songs and access all genres of music and as allowed countries around the globe to be able to listen to other countries music.

Spotify is the most used streaming service, across the globe, dominating almost a third of the subscription market[4]. Other streaming services such as Apple Music and Amazon Music, which are Spotify's main competitors, have less subscribers when combined than Spotify alone. One of the main reasons it is such an interesting topic to investigate is because, as Spotify users, the data across different regions can depict different global trends and possible explanations as to why some music charts.

Previous work There is several research papers which investigate trends in music streaming across a global stage:

- The Geographic Flow Of Music On Spotify [3], an article which investigates the extent to which sharing language spoken in a regions and geographic distance between regions impacted the similarity in streaming trends in different regions. It found that streaming communities were formed based on the language spoken and geographic distance.

- Cultural Divergence in popular music [5], an article which looks at the diversity of music being listened to across different countries. The article reveals that as now there is a larger variation of songs in the top charts than in previous years national charts have been diversified from a global perspective.
- Local Trends in Global Music Streaming [6], this article looks into whether countries are listening to more local or global music on Spotify. The article shows that in there has been an increase in countries listening to music more local to them and that common language and close geographic distance is a large contributor to what music a country listens to.

Objectives This report will explore how artists and tracks' popularity varies in time across the global stage. Firstly it will explore if regions share the same top-ranking artists or songs. Next the impact of language spoken in countries on the percentage of shared songs will be explored. This will be followed by investigating if the popularity of a track in one region will predict its popularity in other regions. The distribution of ranked position and number of streams, and the relationship between them, will be visualised. Then the impact of the peak position of where a track charts against the number of days it remains on the charts will be examined.

3 Data

Data provenance The data has been obtained from the website kaggle, which was downloaded as a CSV file format and then uploaded to notable for processing. The data was collected using a crawler created by the kaggle author which downloads a CSV file from spotify regional charts every day within a date range and puts it into a readable data frame. As it was an open data set on Kaggle it was within the Kaggle T&C for us to obtain the data for the project[2].

Data description The data came in a table with 7 columns (see 1) and 3441197 rows. The data spans from 1st January 2017 to 9th January 2018 (373 days) and contains the daily charts of the top 200 songs for 53 regions.

Data processing To clean the data set we removed the column "URL" was removed as this column would not be relevant to the study. It was calculated how many rows contained NA values and discovered that 0.019092% of track names were NA values then these were removed, then recalculated this value to ensure that there were no more NA values in any of the columns. This study only looks at data from 2017, as the data for January 2018 was not full, so the extra rows were removed. The 'date' column, which contained string values, was converted into a date time column which contained the date and month as this allowed mathematical operations to be performed on the Date column. The regions were then grouped

Table 1: Data-set Column Variable Description of data from `https://www.kaggle.com/edumucelli/spotifys-worldwide-daily-song-ranking`

Variable	Description	Format
Position	Position on chart	Integer
Track Name	Title of song	String
Artist	Name of musician or group	String
Streams	Number of streams	Integer
URL	link to song on spotify	String
Date	date	String
Region	Country Code	String

into continents. The continents that appeared in the data was Europe, North America, South America, Asia and Oceania. This was added as a column to the data frame. This meant when investigating the data we could treat each continent as the regions being looked at.

4 Exploration and analysis

4.1 Do regions share the same top-ranking songs?

Regions will tend to always share some of the same top ranking songs as there are global hits. This was explored by looking at how many songs each region shared with other regions in their top 100 songs of the year 2017 (see figure 1) .

The mean percentage of songs that a region shares with another region is 45.2%. It can be see from the graph that Oceania and Europe share the most similarity. This confirms that most regions do listen to similar music but that there is a definitely variation in which tracks they listen to. The reasons for this are explored throughout the report.

As can be seen in figure 1 South America shares the least amount of songs with other regions in their top 100 songs. This could be explained by South America mainly being Hispanic counties (9 out of the 13 countries in South America having Spanish as their official language [8]) and Hispanic music is very popular within Hispanic regions meaning they listen to less westernized music which is very popular in the other regions.

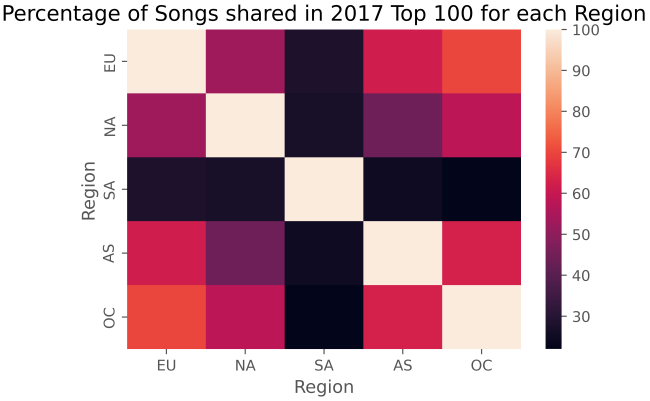

Figure 1: Heat Maps showing the percentage of Shared Songs in Top 100 Positions

4.2 Do countries that speak the same language share the same top-ranking songs?

After noticing that South America shared less songs with the other regions than every other region, and thinking this could be because South America contains a large of Spanish speaking countries this was investigated further by looking into whether countries that speak shared the same top-ranking songs.

As the two most most widely spoken languages in this data frame were English and Spanish, the figure 2 explores the relationship between the percentage of shared songs in countries which speak these two main languages. As can be seen from the map, the first 6 countries, which are English speaking, share a high percentage of songs with each other, and the remaining 14 countries, which are Spanish speaking, share a high percentage of songs with each other. Most importantly, English speaking countries share a very small percentage of music with Spanish speaking countries and vice versa. This reveals that the language a country speaks is a large contributor to what music is most popular in that country.

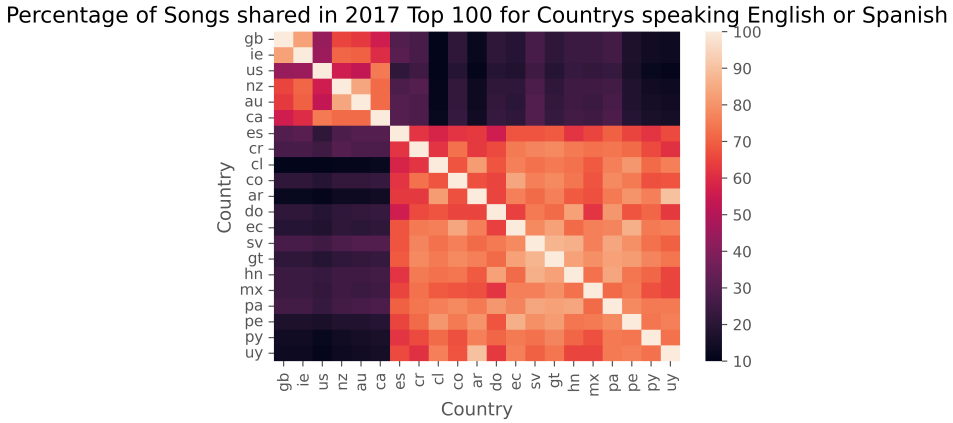

Figure 2: Heat map showing relationship between shared songs in Spanish and English speaking countries.

4.3 Does popularity of a track in one region predict its upcoming popularity in other regions?

After discovering that different regions share some of the same top ranking tracks, positioned in the top 100, we decided to look further in to this. The songs chosen to investigate in the figure 4 were the 4 songs in the top 5 global songs that had the most revealing graphs.These graphs are discussed further section 4.3.2.

Looking at Figure 4 we see that for the positions of each track mainly follow similar journeys. This will be because these are songs popular across the globe so therefore are popular in each region. If songs that weren't so popular globally were looked at this would probably reveal something different. Spotify has a playlists which contains the top 200 for different countries or a global[1] top 200. This means that as a song gets higher in the charts it becomes more and more visible. This will be a large contributor as to why the position journey of songs when it is in the top 5 is very similar in each region. In addition, Spotify recommends users music based on what a user has been listening to and what is popular at that current time which further increases the chances of people in different regions listening to the same music.

4.3.1 The distribution of ranked position and number of streams

Firstly the impact that streams and positions have on each other is explored. The relationship between the distribution of streams and ranked position is a logarithmic one. This is represented by the line of best fit being, the red one in 3, with an equation of $y = a \times log(x) + b$ where a and b are coefficients of the logarithmic equation. The regression model's Correlation coefficient)is $R^2 = 0.76$, to 2 decimal places, which indicates a good fit. This reveals that there is a strong correlation between the position of a track and the number of streams - the higher the position the higher the streams. It can also be seen from the graph for for higher positions (i.e in the top 10) that the number of streams is more varied whereas for lower positions the regression line has a "better fit".

4.3.2 Ranked Positions Overtime

Firstly, the figure 4a, the song 'Something Just Like This', released in February, remained highly positioned in Asia after charting number 1 once released, however in other regions it did not reach as high. An interesting remark is that at the start of December most regions increased the position on the same day however Asia were a day late to this, possibly because of a trend in other regions was later to reach Asia.

Looking at the figure of 4b, the song 'Despacito - Remix' which was released in April and charted immediately at no 1 in South and North America was then followed by the other regions. As it began to drop down the charts there was a sudden jump in the positions in November in North America, which was then followed a month later in the other regions.

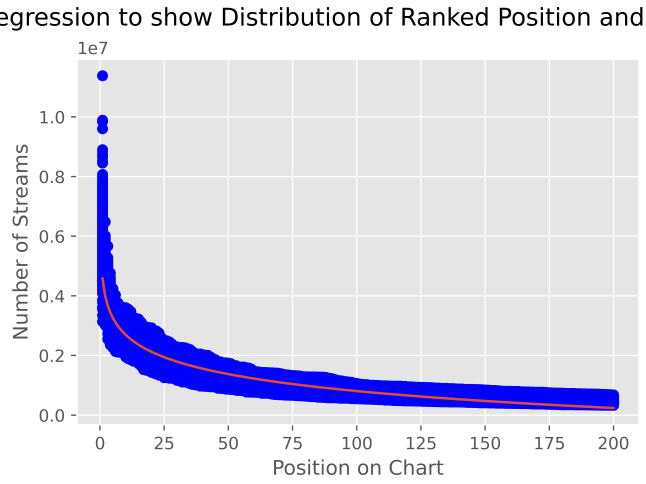

Figure 3

Analysing figure 4c, the track 'Despacito Featuring Daddy Yankee' rose positions in Europe, then North and South America. However it fell swiftly from Europe's rankings until there was a sudden jump in its ranking across all regions in February.

Looking at the figure 4d, the track 'Unforgettable' which was released in April, did not chart at number 1. It firstly increases its charted position in North America and Europe, then followed by other regions. It began to drop positions in Europe and Asia before dropping its ranking in the other regions.

Overall, there is not one region which always sets the trends as such. However most commonly there is a change in the positions firstly in North America and/or Europe which is then followed by other regions. This could be for a number of reasons, one could be that these regions hold the most influence over other regions and possibly start most of the viral trends. Or because these regions have the most variety in languages.

As there is a fall in most of the positions at the end of December the table 2 shows the top 5 global hits on the 25 December 2017. As expected, this table contains mostly Christmas hits which confirms why there was a drop in the charts positions of the top 5 global hits of the year.

Table 2: Top 5 Global Hits on Christmas Day

Position	Track Name	Artist	Date
1	All I Want for Christmas Is You	Mariah Carey	2017-12-25
2	Last Christmas	Wham!	2017-12-25
3	Do They Know It's Christmas? - 1984 Version	Band Aid	2017-12-25
4	Rockin' Around The Christmas Tree - Single Version	Brenda Lee	2017-12-25
5	rockstar	Post Malone	2017-12-25

4.4 Does reaching the top N (e.g., 10, 20, 30) positions have any relationship with the subsequent lifetime of a track in the data?

To explore the question, the data was split into groups of 10 positions, and then plotted in the figure 5. This shows clearly that if a track reached a higher position, it tends to stay on the charts for more days. However, there are sometimes when lower ranked songs stayed on the charts for longer than higher ranked songs. This is true, for example, for tracks reaching positions 31-40, which remained on charts for longer than tracks reaching positions 21-30. This could be because some songs escalate the charts very quickly but then drop off just as quick.

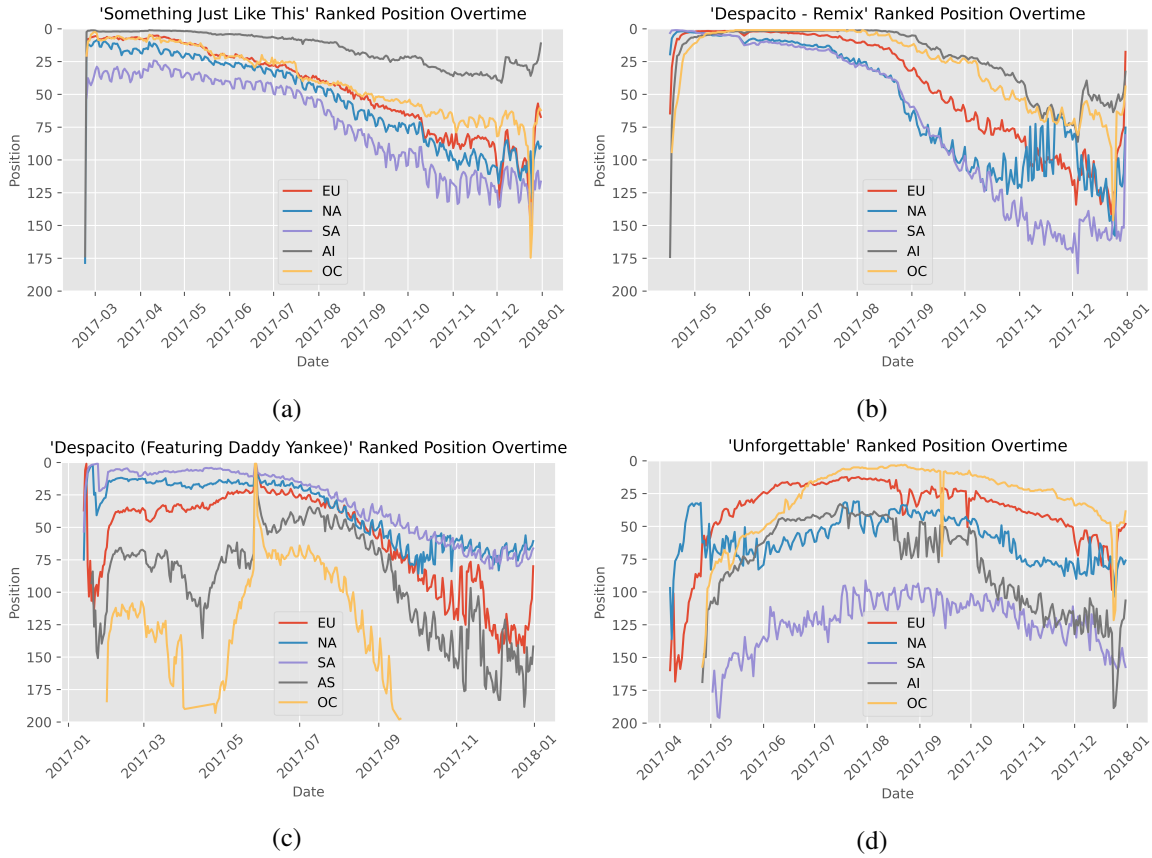

Figure 4: Top 5 global hits showing how the position fluctuates overtime in the different regions.

Poisson distribution was then used to model the number of days a song will chart for by looking at the data in five different sections and creating a PMF for each. Using the formula $\bar{x} = P(x) = \frac{e^{-\lambda} \times \lambda^x}{x!}$ a probability mass function was derived for each section of the dataset being looked at. The discrete random variable being the total number of days in the charts (x) and lambda being the total number of days in the charts for all songs divided by the number of songs in stated section, see Figure 6.

As can be seen from the PMFs. We can see clearly that a peak position in the top20 and between 20 and 40 results in the highest mean number of days being 160 and 115 days respectively. This being compared to peak positions between 80 and 100 having a means of 39 days conveys the higher the peak position the longer it is likely to chart for. However, the "Peak Positions between 80 and 100" in Figure 6b does have a higher probability for reaching its mean which indicates that songs in this category more consistently reach this number of days in the charts. In contrast, a peak position in the top 20 in Figure 6i has the lowest probability of reaching its mean, revealing that these songs less consistently last the mean (160) number of days in the charts. This could be because of the points said above, that some songs may rise the charts quickly, for example Christmas songs as seen in Table 2, which have reached this peak position but only remain on charts for a small number of days. From the PMFs, cumulative distribution functions can be constructed to give us an indication of how likely an interval is to surpass a certain number of days in the charts. An interesting observation is that for the Figure 6b a given song has a 80% chance of charting for longer than 33 days and a 50% chance of charting for 38 days. Whereas songs with a peak position in the top 20 Figure 6j have a 80% chance of a song charting longer than 159 days.

5 Discussion and conclusions

Summary of findings Firstly, regions do share the same top-ranking songs, especially Oceania and Europe. The region which stood out in the visuals was South Africa as it shared the least same top-ranking

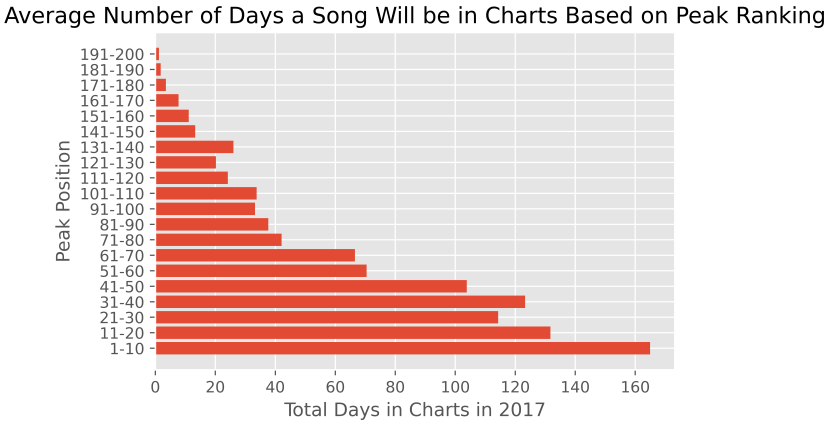

Figure 5: A bar plot showing the mean number of days a song charts for in different peak position categories.

songs. The findings also show some signs that if a song becomes more popular in some regions, such as Europe and North America, then this popularity will follow in other regions. This also follows the opposite way, if a song position drops in these two regions, it then becomes less popular in the other regions. The relationship between ranked position and number of streams is a logarithmic one, such that as a track's position decreases, the number of streams decreases. Another finding was that if countries speak the different languages, in particular English speaking countries and Spanish speaking countries, they do not share a large percentage of songs where as if they share a language the percentage of songs they share is high.. The evidence shows that if a track charts highly, it is more likely to remain on the charts longer.

Evaluation of own work: strengths and limitations A strength of our study was that the R^2 value for 3 was 0.76 and this indicated a good fit. In addition, the visuals in this report are clear and intriguing and strongly back up any conclusions drawn. The data set only contained data for 53 countries, doesn't contain any African countries this limited the way we could look at the data on a global scale as it didn't contain a wide enough variety of countries or regions. When looking at whether a tracks popularity in one region predicted its popularity in other regions. The top global songs of 2017 were looked at in figure4 and showed that for these songs the positions followed similar paths, this did not reveal as much as we had hoped. This will be because 4 global songs were looked at were popular in each region.

Comparison with any other related work The article "The Georgraphic Flow of Music on Spotify"[3] found that streaming communities were often based on the language spoken, this study also found that countries that shared the same official language were listening to more of the same songs.The study "Cultural Divergence in Popular Music" revealed that in 2017 there was in increase in diversity across regions in the music they were listening to compared to previous years. This study did see that there was differences in what different regions were listening to but there was still a large number of songs that were being listened to in all regions(as seen in 2. The study "Local Trends in Global Music Streaming" also revealed that common language and geographic distance was a large contributor to what music was being listened to in that country, something which this study also proved.

Improvements and extensions Future would include obtaining a larger data set including more countries, which would allow for a more detailed investigation.And in addition to language spoken, the geographic proximity to other regions could be looked at.

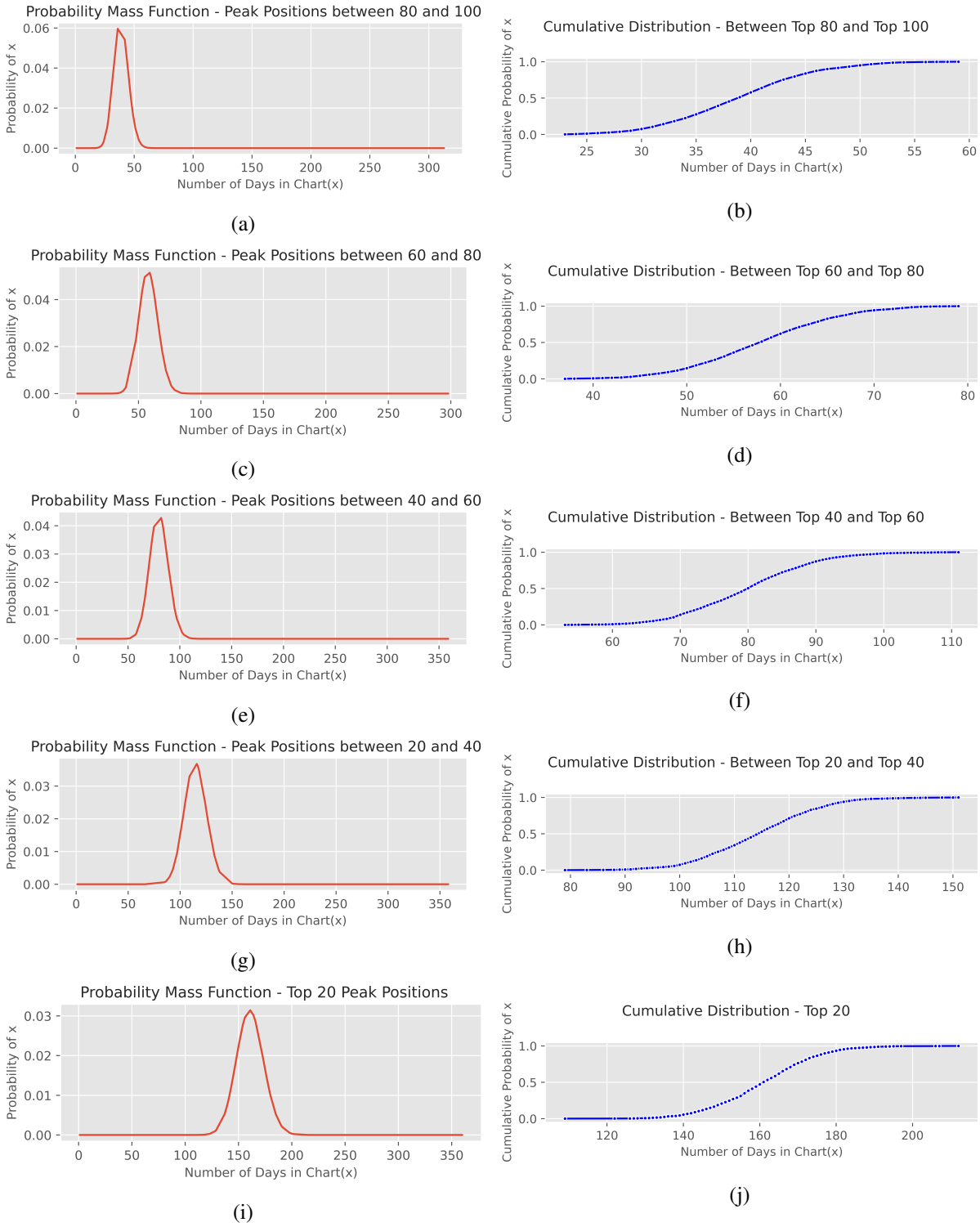

Figure 6: Probability Mass Function and Cumulative Density Frequency for the number of days a track remains in the top 100 global chart using a Poisson distribution

References

[1] URL: https://open.spotify.com/playlist/7h9otsUlMJUmtWJBEQKcrJ.

[2] *Kaggle Terms of Use*. Last accessed 10 April 2022. 2022. URL: https://www.kaggle.com/terms.

[3] Lidija Jovanovska Miroslav Mirchev. "The Geographic Flow Of Music On Spotify". In: (2019). Retrieved on 10 March 2022. URL: https://www.academia.edu/39275850/The_Geographic_Flow_Of_Music_On_Spotify?auto=citations&from=cover_page.

[4] Mark Mulligan. *Music subscriber market shares*. Last accessed 17 March 2022. 2022. URL: https://www.midiaresearch.com/blog/music-subscriber-market-shares-q2-2021.

[5] David Garcia Pablo Bello. "Cultural Divergence in popular music:" in: (2021). Retrieved on 10 March 2022. URL: https://www.nature.com/articles/s41599-021-00855-1.

[6] Henriette Cramer Samuel F. Way Jean Garcia-Gathright. "Local Trends in Global Music Streaming". In: (2020). Retrieved on 10 March 2022. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7336/7190.

[7] Sam Shead. *Spotify plans to launch in over 80 more countries -CNBC*. Last accessed 17 March 2022. 2021. URL: https://www.cnbc.com/2021/02/23/spotify-plans-to-launch-in-over-80-more-countries.html.

[8] *Which Spanish speaking countries are in South America?* Last accessed 6 April 2022. 2021. URL: https://pvillage.org/archives/103864.