

Spotify Wrapped 2021 for Different Age Groups

Katrina Baha
COMP 333 - Fall 2021
14 December 2021

Topic

Spotify is a popular streaming app that is generally used for listening to music. One exciting feature of Spotify is that every year around December, they release a personalized summary of data in an engaging format about the music an individual listened to during that year. This is called Spotify Wrapped, and the title is followed by the year that the Spotify Wrapped is focusing on.

I am focusing on collecting and analyzing Spotify Wrapped 2021 data from two different age groups— 18 to 24 year olds and individuals who are 25 years old or older. For the purpose of this project, I will be referring to 18 to 24 year olds as college-aged individuals and people who are 25 years old or older as non-college-aged individuals. I am interested in the differences between the amount of music they listen to on Spotify, their top genres, and their top artists. The amount of music listened to is measured in minutes for the Spotify Wrapped 2021 year, the top genres are the top five most listened to genres for the Spotify Wrapped 2021 year, and the top artists are the top five most listened to artists for the Spotify Wrapped 2021. These are all personal to the listener. This will be fascinating to explore because I feel as if college-aged individuals listen to music much more often than non-college-aged individuals. I want to learn more about this and discover if this is actually true or not. **The question I would be answering is if there exists a statistical significant difference between college-aged individuals and non-college-aged individuals and their listening times in minutes. Additionally, I would be answering if there exists a statistical significant difference between college-aged individuals and non-college-aged individuals and their top genres and artists. I would also answer the question of what the top genres and artists are for these two age groups. These are all based on Spotify Wrapped 2021.** To answer these questions, I conducted a survey to collect Spotify Wrapped 2021 data, performed statistical analyses with the data, and created data visualizations with the data. I hypothesize that there exists a statistically significant difference between college-aged individuals and non-college-aged individuals and their listening times in minutes, their top genres, and their top artists.

Data Gathering

I collected this data by having people fill out a Google form with their Spotify Wrapped 2021 data. I only had people fill out the Google form once. I promoted it via email, text, and social media. The standard message I would send out was “I am collecting data for my final project, which is about comparing Spotify Wrapped 2021 data of different age groups, for my

Human-Centered Data Science class. Please consider filling out my survey by using this link: <https://forms.gle/Hm6WR5tLNNDM63Ca6>. Let me know if you have any questions!"

The data I collected was the individual's age group, whether the individual uses Spotify or not, the individual's top five genres, the individual's top five artists, and the individual's listening time in minutes. If the individual filling out the form selected that they are under 18 years old or do not use Spotify, the Google form forced them to submit without being able to enter their Spotify Wrapped 2021 data. This is because I am not focusing on people who are under 18 years old or people who do not use Spotify.

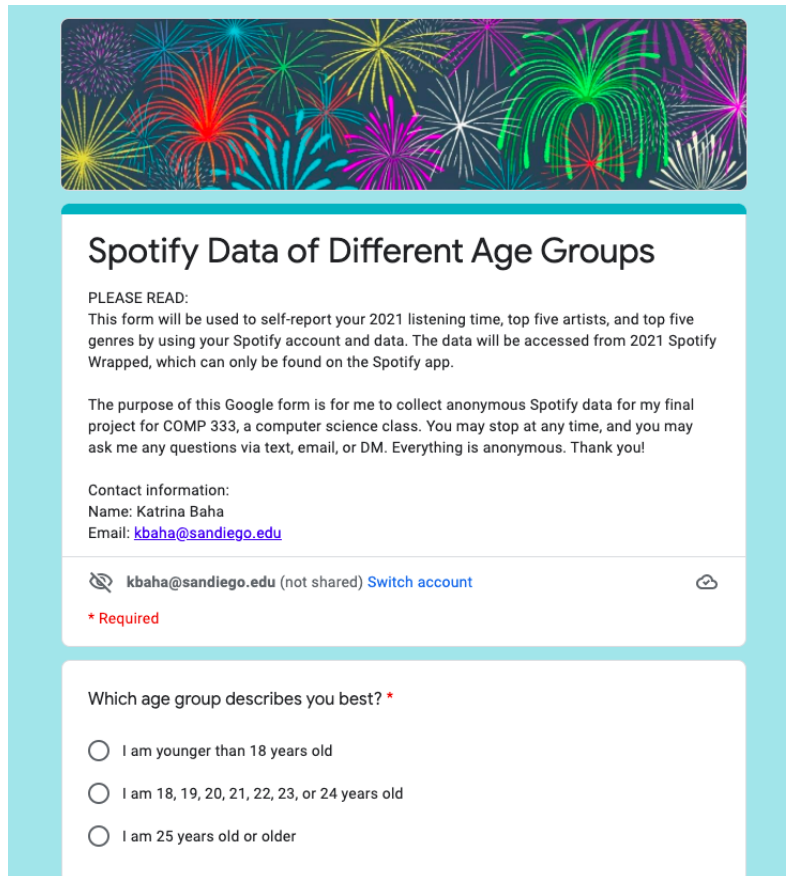
The image shows a screenshot of a Google Form titled "Spotify Data of Different Age Groups". At the top, there is a decorative header with a pattern of colorful fireworks. Below the title, there is a "PLEASE READ:" section. The text in this section states: "This form will be used to self-report your 2021 listening time, top five artists, and top five genres by using your Spotify account and data. The data will be accessed from 2021 Spotify Wrapped, which can only be found on the Spotify app." It then explains the purpose: "The purpose of this Google form is for me to collect anonymous Spotify data for my final project for COMP 333, a computer science class. You may stop at any time, and you may ask me any questions via text, email, or DM. Everything is anonymous. Thank you!" Contact information is provided: "Name: Katrina Baha" and "Email: kbaha@san Diego.edu". Below this, there is a section for the user's account, showing "kbaha@san Diego.edu (not shared)" with a "Switch account" link. A red asterisk indicates a required field. The first question is "Which age group describes you best? *", with three radio button options: "I am younger than 18 years old", "I am 18, 19, 20, 21, 22, 23, or 24 years old", and "I am 25 years old or older".

Figure 1: The first page of the Google form

The Google form is accessed through this link: <https://forms.gle/2TXH9KaEqC1JZjS7A>.

I also created and covered two concepts for the Google form questions.

Concepts:

1. Concept 1: Background
 - a. Sub-concept a: Age Group
 - b. Sub-concept b: Spotify User Status
2. Concept 2: Top 5 Most Listened To Data For Spotify Wrapped 2021
 - a. Sub-concept a: Top 5 Artists Most Listened To For Spotify Wrapped 2021

- b. Sub-concept b: Top 5 Genres Most Listened To For Spotify Wrapped 2021
- c. Sub-concept c: Listening Minutes

Google Form Questions:

1. Which age group describes you best?
 - a. Concept 1a
2. Do you use Spotify?
 - a. Concept 1b
3. What is your first top genre?
 - a. Concept 2a
4. What is your second top genre?
 - a. Concept 2a
5. What is your third top genre?
 - a. Concept 2a
6. What is your fourth top genre?
 - a. Concept 2a
7. What is your fifth top genre?
 - a. Concept 2a
8. What is your first top artist?
 - a. Concept 2b
9. What is your second top artist?
 - a. Concept 2b
10. What is your third top artist?
 - a. Concept 2b
11. What is your fourth top artist?
 - a. Concept 2b
12. What is your fifth top artist?
 - a. Concept 2b
13. What is your listening time in minutes?
 - a. Concept 2c

Data Analysis

The listening time for the year in minutes is a continuous variable while the top five artists and top five genres would both be categorical variables. The age of the users is categorical because I am comparing between the age groups of 18 to 24 years old (college-aged individuals) and individuals that are 25 and older (non-college-aged individuals). If I want to see the relationship between the age groups of the users and their listening times for the Spotify Wrapped 2021 year, I would use an independent t-test.

I will address the age groups and the differences between their top genres and top artists through data visualizations instead of using a Pearson chi-square because some of the genres or artists only had one person who listed them as their top genre or top artist. The chi-square would reveal more statistical information if there were at least five individuals per genre or artist.

There were 74 responses to the survey. 62 of the responses were from college-aged individuals, 10 of the responses were from non-college-aged individuals, and 2 of the responses were from individuals younger than 18 years old. 5 of the 74 responses indicated that they did not use Spotify. 1 response did not fill out the survey correctly while another response left some questions unanswered. So, after I clean the data, there will be 67 lines of data total, with 60 of these responses being from college-aged individuals and 7 of these responses being from non-college-aged individuals. One college-aged individual's line of data will have 2 top genres empty and 4 top artists empty. One non-college-aged individual's line of data will have 5 top genres empty.

Data Cleaning

There was a lot of variety in how people entered their Spotify Wrapped 2021 data into the Google form. For the top genres and top artists, people would either write in all lowercase letters, only the first word first letter capitalized, or all the words in the phrase with the first letters capitalized. Since Google form responses can be created into a Google sheet, I replaced the genres and artists written as all first letters capitalized by using the =PROPER function in Google sheets. The listening times in minutes also varied in response styles because some people used commas before the thousands place while other people did not. This I fixed by removing all commas in the Google sheet by highlighting the minutes and changing the format to automatic.

After this, I removed any data that was irrelevant. For example, if a person that was under 18 or did not use Spotify answered the survey, those responses were still recorded in the Google sheet. So, I deleted these rows in Google sheets. I also had one person that put their top songs instead of their top genres, so I removed the top songs but kept all their other data in the Google sheet.

In addition to cleaning the main data file called Spotify_Cleaned, I created a reformatted data file called Spotify_NewFormat. The reformatted file had the columns labeled as Age Group, Order, Genre, Artist, and Listening Time. I created this file for any statistical analyses or data visualizations that require all the genres or all the artists.

Statistical Analysis

First, I checked the normality of the listening time data by creating a histogram. I found that it was not normally distributed because it is skewed right with the tail end reaching the higher values.

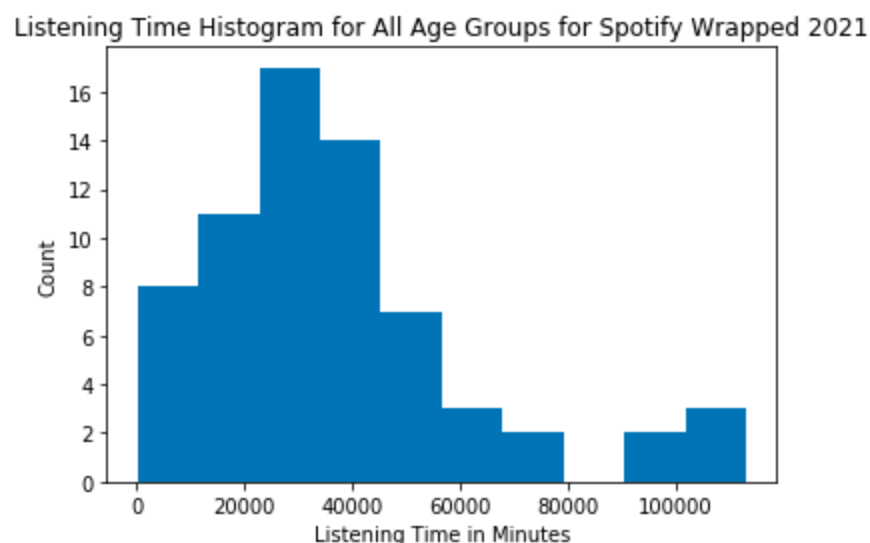


Figure 2: The shape of the listening time data is not normally distributed

I also counted the number of genres and artists that are represented in the data by using the `=COUNTUNIQUE` function in Google sheets. There are 125 different genres and 232 different artists represented in the data.

Next, I performed an independent t-test to determine whether the age group of an individual related to their listening time in minutes for Spotify Wrapped 2021. The outcome variable is continuous due to listening times in minutes being continuous. The predictor variable is categorical since I am treating the age groups as categories. I failed to reject the null hypothesis because the p-value (~ 0.335) is greater than 0.05. This means that my results are not significant.

To view all my analyses in depth, please see the attached Jupyter notebook.

Data Visualization

I developed a story in Tableau as my data visualization. My intended audience for the story are people who are interested in Spotify Wrapped 2021. I want my intended audience to be able to see the similarities and differences between the different age groups in relation to their Spotify Wrapped 2021 information.

The story starts with the average listening times in minutes per age group for Spotify Wrapped 2021. This is shown with a bar graph. Here the viewers can see that there is not a significant difference between average listening times in minutes between age groups.

Average Listening Time for Different Age Groups for Spotify Wrapped 2021

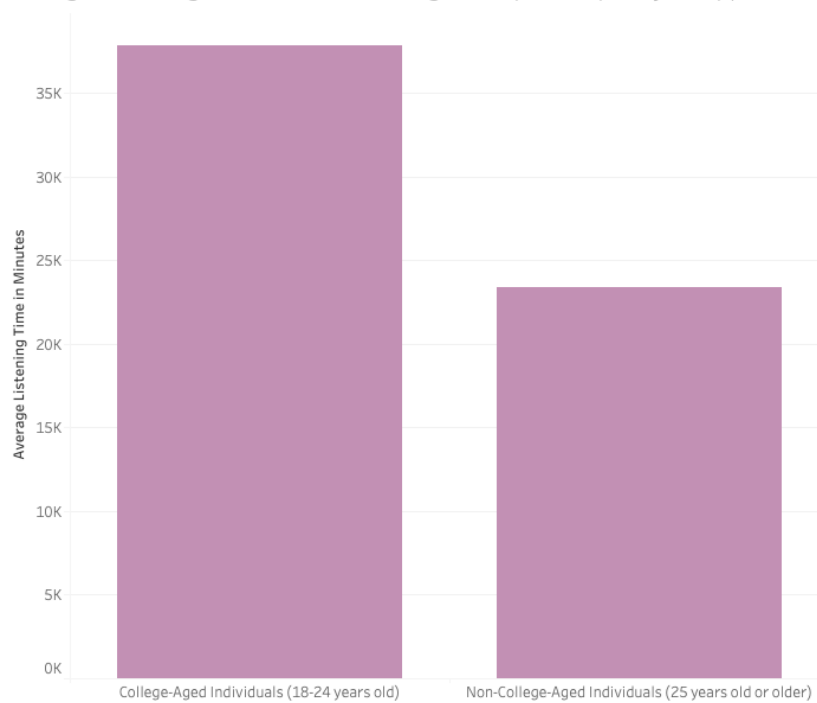


Figure 3: Average listening time for different age groups for Spotify Wrapped 2021

After this, the story shows the top genres and top artists of each age group and for both age groups so that viewers can compare these visualizations by themselves. The top genres and top artists are visualized with word clouds. Here the viewers can see why there is a statistical significant difference between the two age groups and their top artists and top genres for Spotify Wrapped 2021. There is a clear difference between college-aged individuals and non-college-aged individuals and their top genres and top artists, so there exists a statistical difference between the two age groups and their top genres and top artists.

Top Genres Across All Age Groups for Spotify Wrapped 2021



Figure 4: Top genres across all age groups for Spotify Wrapped 2021

Top Genres for 18 to 24 Year Olds for Spotify Wrapped 2021



Figure 5: Top genres for 18 to 24 year olds for Spotify Wrapped 2021

A word cloud of music genres. The words are arranged in a circular pattern around the center. The largest word is 'Dance Pop' in a light grey font. Other prominent words include 'Alternative Metal' in orange, 'Indie Pop' in pink, 'J-Pop' in purple, 'Conscious Hip Hop' in teal, 'Modern Rock' in blue, 'J-Rock' in purple, 'Indie Rock' in grey, 'Motown' in blue, 'Indiesoul' in pink, 'Country Road' in grey, 'Glitchcore' in teal, 'Stomp And Holler' in teal, 'Neosoul' in yellow, 'Slaphouse' in green, 'Soundtrack' in yellow, 'Pop Punk' in grey, 'Metalcore' in purple, 'Indie Folk' in pink, and 'J-Pop' in purple. The words vary in size and color, creating a vibrant and dynamic visual.

Top Artists Across All Age Groups for Spotify Wrapped 2021



Figure 7: Top artists across all age groups for Spotify Wrapped 2021

[illegible]

Top Artists for 25 Year Olds and Older for Spotify Wrapped 2021



Figure 9: Top artists for 25 year olds and older for Spotify Wrapped 2021

Spotify Wrapped 2021 for Different Age Groups



Figure 10: First page of the *Spotify Wrapped 2021 for Different Age Groups* Tableau story

Conclusions

Overall, I found that there are statistically significant differences between age groups and their top genres and top artists for Spotify Wrapped 2021. There are not statistically significant differences between age groups and their listening times in minutes for Spotify Wrapped 2021. However, I would like to acknowledge that I only had 7 non-college-aged individuals represented in this project, so the analyses might have benefitted from more data from that group. If I were to continue working on this project, I would try to collect more data from all the age groups. I think if this were to be done, there could be better analyses and patterns emerging.