Contained within this repository are the graphic visualizations of group 9 as well as their individual Jupyter Notebooks containing the individual code necessary to articulate our research. This readme has the aggregate findings of group 9's project 1.

Using Kaggle our team identified a dataset (Top 200 Spotify Charts 2020-2021) that we found both interesting and containing enough numerical/statistical information for us to infer analytical queries and implications while making meaningful deductions to support our presentation.

The parameters we set out to achieve are as follows:

1. Each member selected an artist found in the dataset. a. Sewit – Doja Cat b. Audrey – The Weeknd c. Jason – Ed Sheeran d. Ashwin – Drake
2. What are the total Number of followers for each selected artist? (Code pulled from Jason Jupyter Notebook)

![Graphical user interface, application, Word

Description automatically generated]()

1. What are the total number of songs each artist has had in the top 200 and did the artist get #1 charting position? (Code pulled from Jason Jupyter Notebook)

![Graphical user interface, text, application, email

Description automatically generated]()

1. What were the total number of streams per song? Then aggregate the complete total number of streams and display a bar graph for total number of streams per song.

![Chart, bar chart

Description automatically generated]()

1. Utilizing an overall data comparison, we established the following:
   * 1. Which artist had the greatest number of streams, and then we visualized this with a pie chart and a horizontal bar graph for the individual artists selected.

![Chart, bar chart

Description automatically generated]()

b. Total number of times each individual artist has charted compared to one another and then displayed via bar graph and ejected pie chart (ejecting the lowest charting artist)

![Chart, bar chart

Description automatically generated]()

c. Which artist has the highest and lowest number of followers and then displayed via scatter plot.

![Chart, scatter chart

Description automatically generated]()![Chart, scatter chart

Description automatically generated]()

Summary Analysis Conclusion:

While our research did not contain an overarching core question, we did answer all of the questions we set out to solve. This project (and the supporting files contained herein) establish our understanding of the concepts we have learned thus far in the course (Aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis), with our visualizations further enforcing our comprehensions.

While the scope of our research is limited to identifying specific numeric evaluations, the benefits of accessing the contents of the entire dataset can be found in; utilizing genre motifs and patterns to ascertain popular music trends, ranking artists popularity by total stream count and follower count to identify optimal ratios to project future successes of new artists.

It is also worth noting that we discovered that when looking at the data, the highest charting artist did not have the most amount of streams and when comparing different data points among the artists we were able to evaluate popularity using differing metrics.

Some additional limitations that were identified during our research of the dataset are as follows:

* The data provided information regarding the highest charting position for each song as well as the number of times each song has charted, however we were unable to view how long each song maintained its highest charting position.
* The observational limitations of the data are set by our choice of four individual artists and not representative of the larger set of data. This limitation prevents a more detailed analysis of the top 200 streamed songs on Spotify from 2020-2021