Aline Hirsch, Carlos Moreno Acero, Charles Lee, Quentin Ludet  
CSMC 206 / Group Project  
Prof. Gary Thai  
4 May 2022

**Write-up**

**Summary of your group project, accomplishments and learning experience.**

Our project provides code which is able to gather data from billboard.com and use the API for getGenre.com to create a file with songs data, which includes name of the song, artist, genre, when the given song was on the charts and in what position it was ranked. The group collaborated, learned and shared their research, methods and previous knowledge to build a working Python program.  
 Each of our members contributed technically and creatively to answer questions about the data set gathered. We use Pandas and different graphing tools to show the most popular genres, artists, what artists were popular in different years, and how many times an artist has appeared on the charts, based on user input.  
 Not only did we practice all the concepts learned in class but we also did our own research ahead of the class schedule to be able to complete the project on time. The first step of the project was Web Scraping which was only covered later in the material, so we had to seek that knowledge and share with each other.   
 Our group accomplished most of the goals that were set at the beginning, adding new goals along the way to adapt to our own realities and knowledge.

**What would the group have done differently**

* Improve communication, probably have a secondary means of communication such as group text.
* We could always use more time, so starting research earlier and learning independently so we could implement tools that we were going to use before we studied them in class.

**What would the group do next if you have an additional four weeks**

* Create a website or application to explore results.
* Explore more questions, for instance “What are the most popular songs by a given artist?”, “What genre does a given artist sing?”, “What is the nationality of the 10 top artists?” and so on.
* Find more details related to each song, like language and nationality.
* Generate other types of graphs, such as scatter plot, timeseries, and interactive graphs using Plotly.

**Summary of individual accomplishments and learning experiences (authored by each group member)**

Aline Hirsch

As the group leader, I first organized the group gathering everyone’s contact information and creating a group chat on Discord. I would schedule meetings, keep track of our progress and make sure the communication between the members was happening harmoniously and frequently and submit progress reports.  
 In regards to coding, I, as well as Charles, created a word cloud with the most popular artists, as well as create a graph using pandas and matplotlib to display the most popular genres.I also worked on answering the questions: “What was the most popular artist on a given year?” and “How many times has a given artist appeared in the charts?” with the help of my teammates.  
 For the final submission I was responsible for the creation and design of our PowerPoint presentation, creating and sharing documents with the group and scheduling.

Carlos Moreno Acero

I was responsible for analyzing a generated .json file and by selecting an artist I found all the times it appeared in the chart at any given time. I also helped to coordinate and give ideas for the presentation, powerpoint and the way of communicating through the semester.

Charles Lee

I was responsible for making a bar graph showing the most popular genres and a word cloud of the most repeated artist. I used pandas as well as matplotlib for both graphs.

Quentin Ludet

I was responsible for writing code to scrape the data from the billboard.com website. That involved figuring out how urls were formatted for different weeks, using BeautifulSoup to interpret the html, and saving the results to a json which ended up having over 5MB of data, 250K+ lines of json. I also contacted the developers of api.getgenre.com to get access to the API and find the genres of 1801 different artists. Then, I used pandas and matplotlib to organize the data, and create plots for genre trends and top artists by year.

**Which 6 Python tools did you use?**

We used more than 6 tools. Our project includes the use of Pandas, Matplotlib for graphs, Word Cloud, Data Analysis using lists and dictionaries, APIs and API processing, JSON files, and web scraping.

**What challenges did you encounter?**

A major challenge was formatting the artist data from the billboard website. Since the data from the website is only text, it was a challenge to interpret artist names with featured artists, collaborations, one-off groups, etc. For example, the artist names could be organized as “Drake Featuring Rihanna”, "Machine Gun Kelly x Camila Cabello", or even “Chris Brown Featuring Usher & Gucci Mane”. This mix of formatting was difficult to predict and caused some errors searching artist genres.

Another issue was using the getGenre API; although it worked well, it ran into some hiccups, for example accents and artists with only a single song. Since we didn’t want to leave empty data, these artists that couldn’t be found by the API had to be manually entered

**What would you have done differently?**

If we were to start from the beginning, I would have organized the data in a way that is more easily formatted for Pandas. The json format we used was easy to store and write to, but more difficult to analyze than it could have been. That would have also made it easier to find trends and generate plots.

**What would you do next?**

Next, we would try to collect data about the languages, nationalities, and other metadata about songs so we could have a more diverse perspective about what makes songs popular.