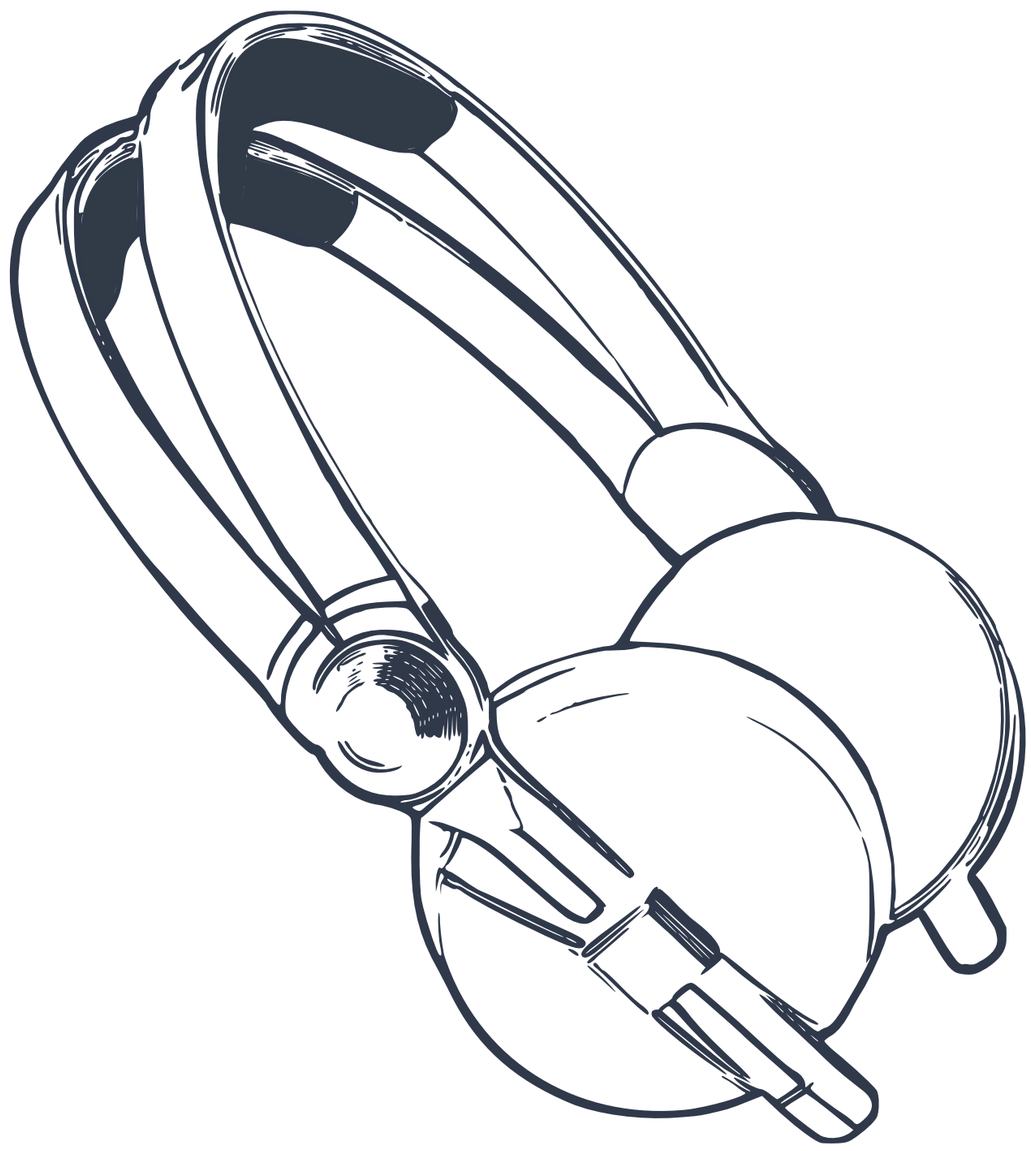
GEOG 187
SEMESTER PROJECT



A deep dive into my listening and work habits

University of Waterloo
Vincent Fong
November 24, 2023

Table of Contents

I. Introduction 1
II. Method 2
    a) Data Collection 2
III. Results 4
IV. Reflections 9

Introduction

What is Spotify?

Spotify is a music streaming service that allows users to listen to music on desktop or mobile.

ABOUT

My project is a data analysis and collection project on my Spotify listening activity. I aim to analyze my how my listening activity (time of day, duration of listening, mood, etc.) compares with a performance metric of my “productivity” or tangible work that I produce. These metrics include words typed per session, or break efficiency. Additionally, the project aims to identify what types of songs I listen to the most and how that reflects my work performance indicators; songs can be categorized by their audio features such as tempo, or energy which Spotify is able to analyze using their developer API.

INSPIRATION

My project was inspired by the change of my study habbits. Before coming into university, I didn’t listen to much music, and really only listened sparingly during long trips on the road, or while doing chores. However, one of the few things I’ve connected with some people over was my preference in music (or lack thereof). I was also introduced to a bunch of new artists and genres I never thought I would get into, such as rap, or rock to name a few. I’ve since started actively listening to music, and doing so more regularly. This seeped into my work and study habits, and soon enough I couldn’t find myself working without music in the background. Despite my newfound companion in late nights, and stale afternoons, I couldn’t help but feel that music was taking away from the work I was putting on assignments or projects. As such, one of my major goals coming into this project was to see if I really was doing as much work as I felt I was while listening to music.

Method

DATA COLLECTION PRECONDITIONS

Before diving into my process of data collection, I first must specify a few preconditions that determined how, and what I collected.

CONDITION 01

First, I only listen to music on Spotify, which not only records my most listened tracks, but can also give audio features for any of these given tracks using their developer API. My work would only start being tracked as soon as I started playing something on Spotify.

CONDITION 02

Second, I only tracked the sessions where I would be writing a file on my computer (which includes documents and code/programs). This meant that listening to music while I was exercising, walking or doing chores was NOT recorded. Additionally, I also listen to different genres of music while working compared to exercising or other activites. In order to prevent the posibility that Spotify would mix my most listened “working” tracks with the other tracks, I decided to create another Spotify account and exclusively use that account to listen to my “non-working” music.

CONDITION 03

Third, the “productivity” variable measures the changes in the number of words per session; the productivity variable does not account for work that would be in different forms from writing such as reading or problem solving (computation). However, since I do not typically listen to music while I do either of these activities, my original question of “How productive am I while listening to music?” was readjusted to “How much writing can I produce while listening to music?”.

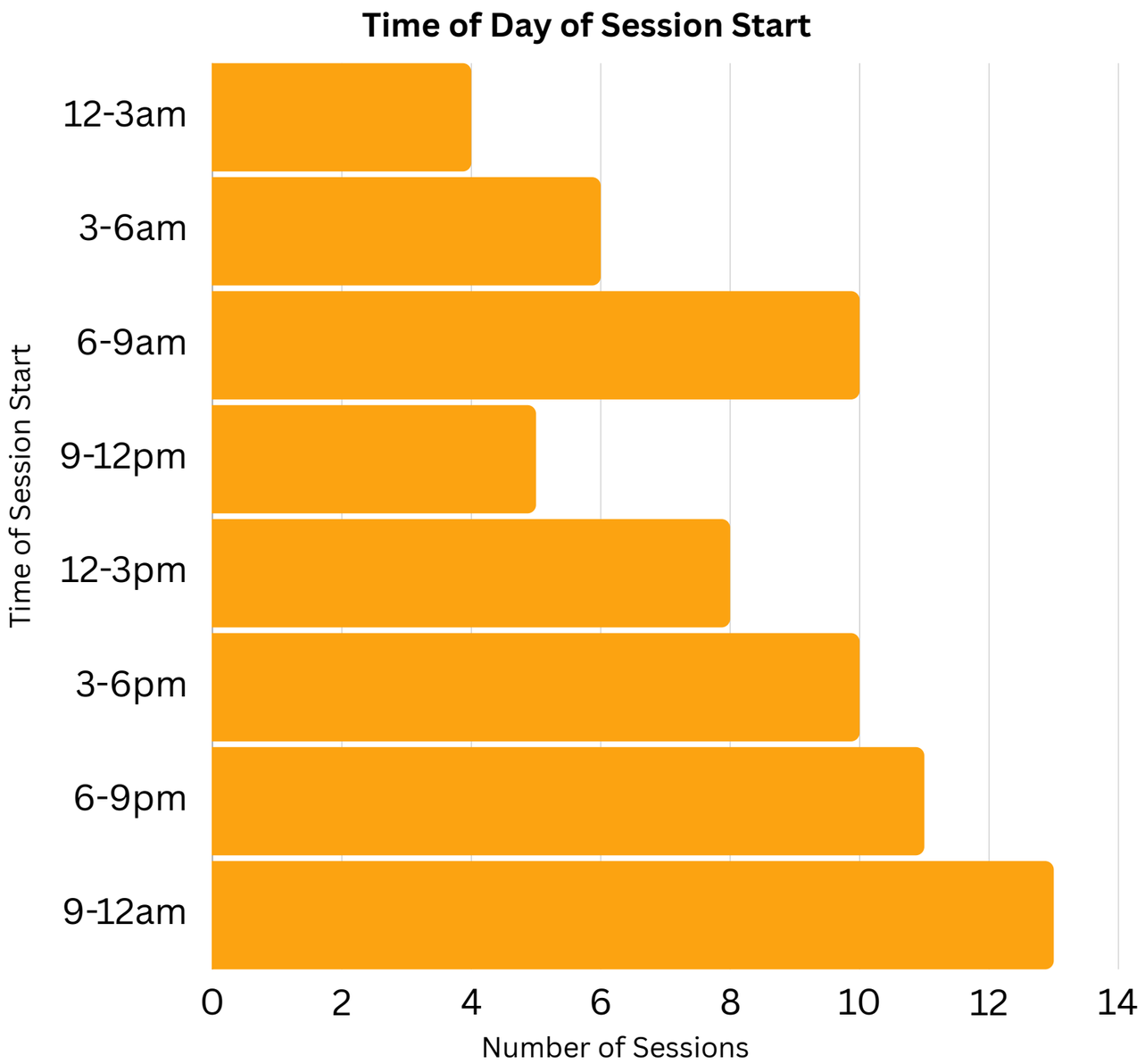
Method

THE SCRIPT

One of the key components of my data collection process was creating a bash script that reads the MPRIS D-Bus interface by piping the output of a shell command from the playerctl package into my script, which would a) get and update the current "state" of the *Documents* folder – which would represent the number of words that are currently different since the last time the script was ran. b) listen for changes on the D-Bus interface for when the Spotify client is paused/played. Using these two principles, along with a short Rust script to write to a csv file, I was able to create an automated script that tracks how much work a user did since starting to listen to a song on Spotify. The workflow is designed so that the process would look something like this: the user starts the script and opens Spotify. The user would then work as usual until Spotify is paused, where the user will be prompted to enter some variables (fatigue, mood, heart rate, number of distractions, duration of focus), which the script will write as a row to a csv.

Code can be found on my github: https://github.com/kaisic1224/produce-check

Results



Ignoring the fact, that I am awake at basically all hours of the day, and my sleep schedule is erratic, you can find that the majority of the time I listen to music to do work will be sometime at night or afternoon!

AVERAGE DURATION OF FOCUS

53.44 mins

AVERAGE BREAK LENGTH

10.17 mins

AVERAGE DURATION OF SESSION

86.39 mins

AVERAGE NUMBER OF DISTRACTIONS

3.24 per session

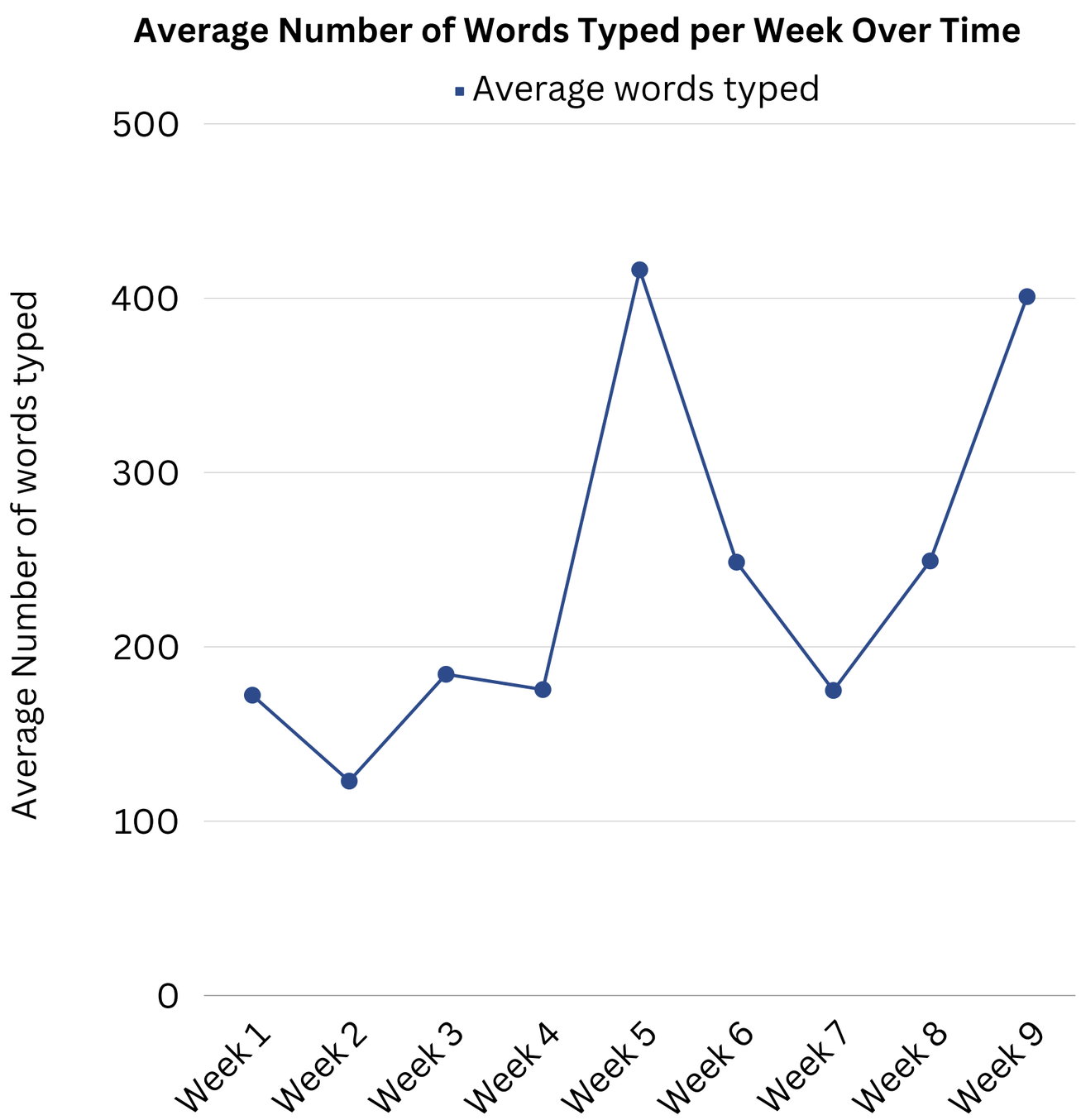
AVERAGE WORDS PER SESSION

3.18 WPM

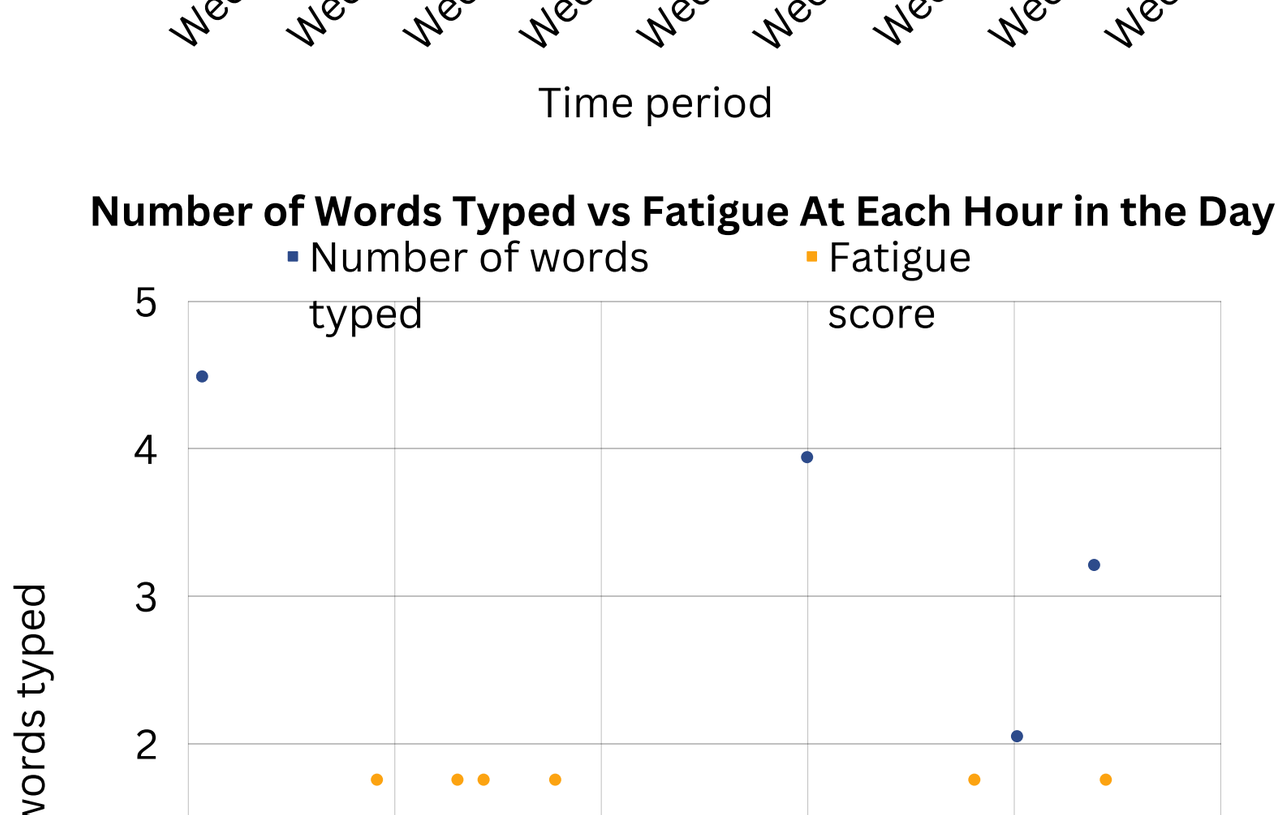
AVERAGE WORDS PER FOCUS MINUTE

5.14 WPM

Results

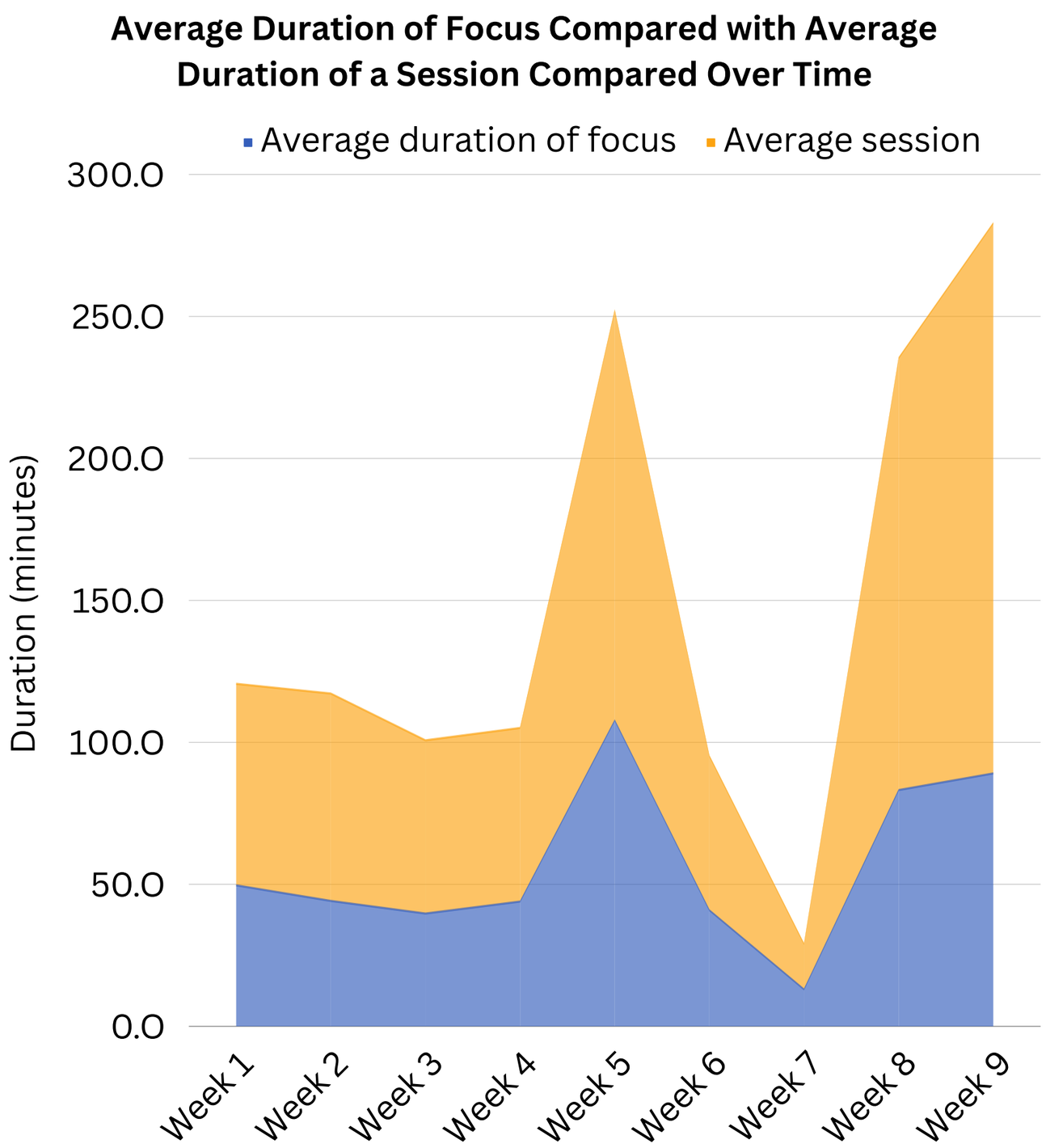


The spike at weeks 4-5 can be explained by the coming of many of my midterms around that time which also explains the increase in session length and average number of words being typed as I was spending more time reviewing and practicing. This would contradict the upcoming weeks, where I fell sick shortly after and consequently, barely listened to any music while doing work (between week 6 and week 7 I only had a total of 4 records!). However, we can see that a typical average word count for a week will hover between 175-235 words/week.

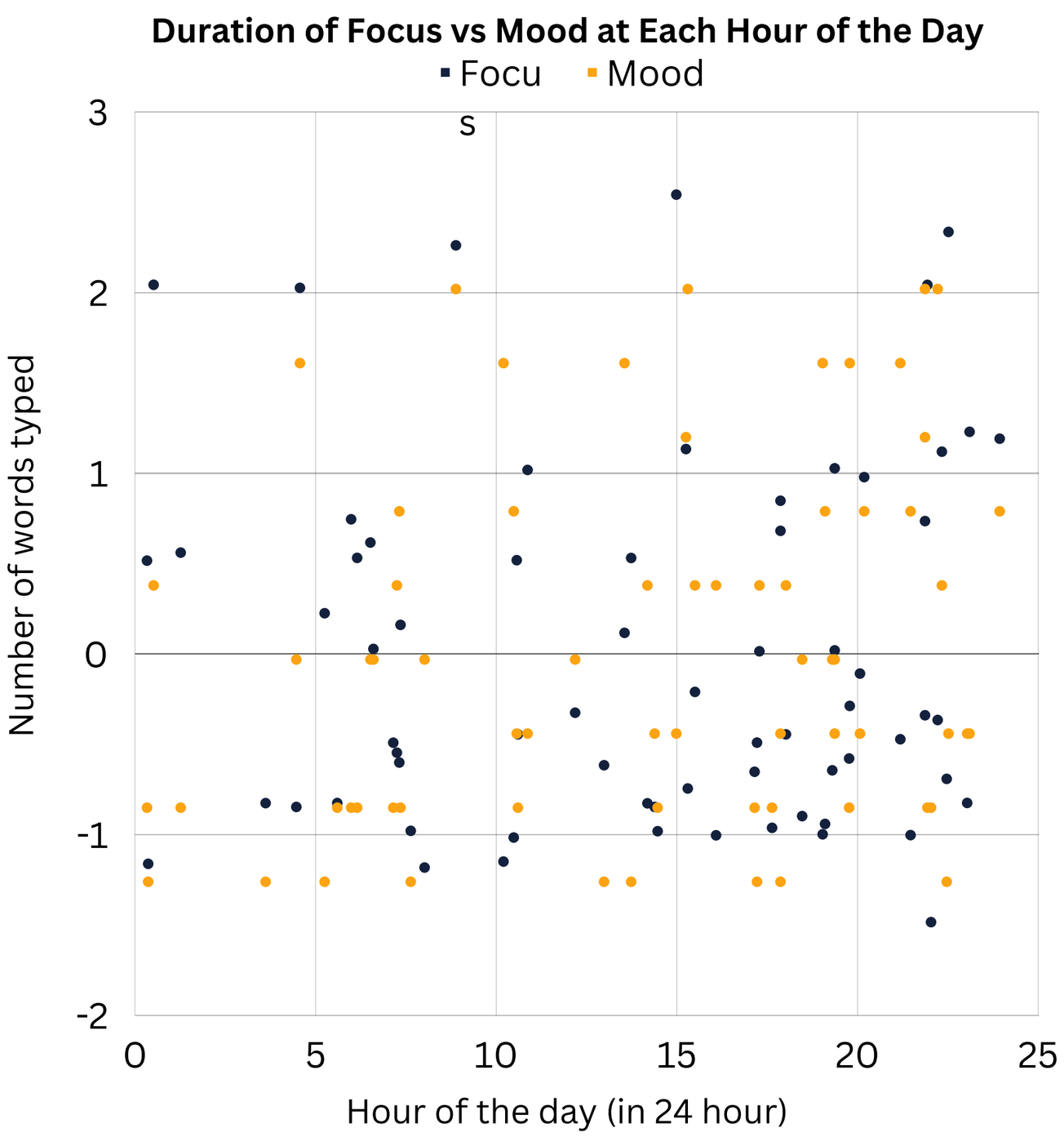


I graphed the number of words as a scaled representation of their type typed with the time of day I started working hoping to see a relationship and if there was a specific time that I worked better in, and although it is very slight, I tend to have higher word counts during the afternoon and night than compared to the morning. Something that I think is interesting to note though, is that on days that I have a low fatigue score (low means not tired), I don't tend to write much amounts compared to when I have a high fatigue score, which leads me to believe that fatigue is a **non-determining** factor to my productivity, even at different times of the day.

Results



I graphed both the average duration of the session, and the average duration of focus over the course of the data collection, and this tells me that I am really not spending my time well in comparison to how much time I am dedicating to working, most of it ends up being distracted. Even now, leading up into the weeks before finals you can find that my average duration of focus is starting to reach around the same level it was at during the time of midterms, which is just a cool sort of idea of how I am naturally forced to focus as soon as projects start coming up.

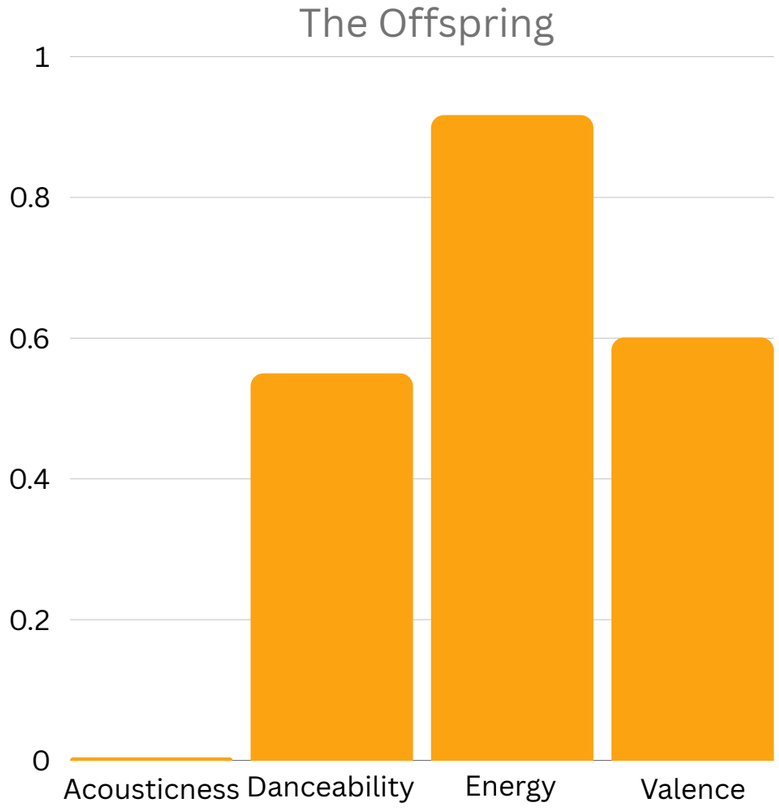


I graphed a scaled number representing the minutes of focus I have per session, compared to the time of day that I started working, and found that there really isn't much of a relationship between the time of day and how well I can focus. However, interestingly when also deciding to include my mood as a comparison to a duration of focus, I tend to focus better when my mood is also better, and vise versa. There is almost a 1:1 relationship between focus and mood which can say that mood **IS** a determining factors on how well I focus.

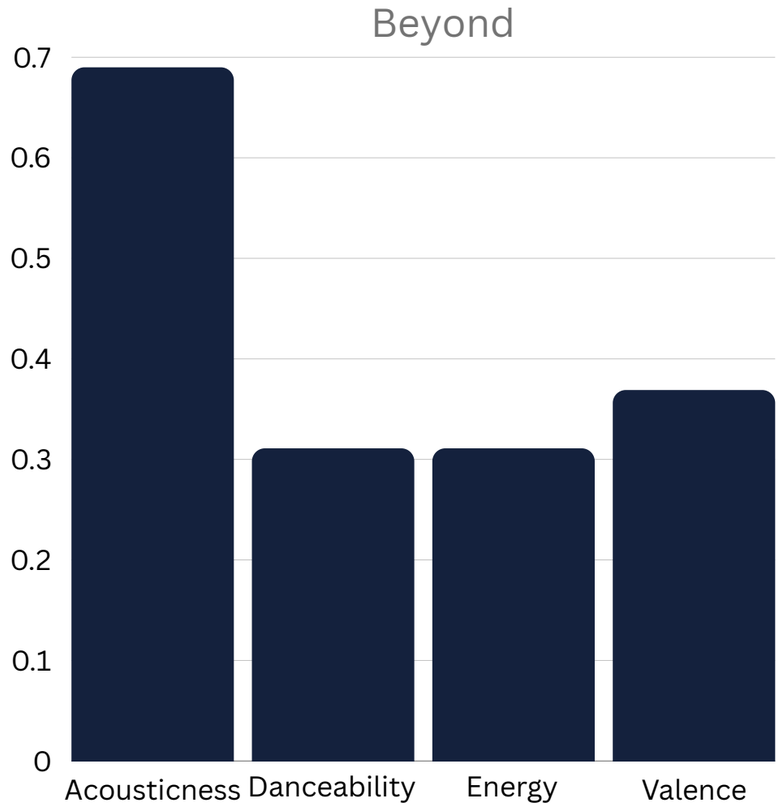
Results

TOP 9 MOST LISTENED SONGS + AUDIO FEATURES

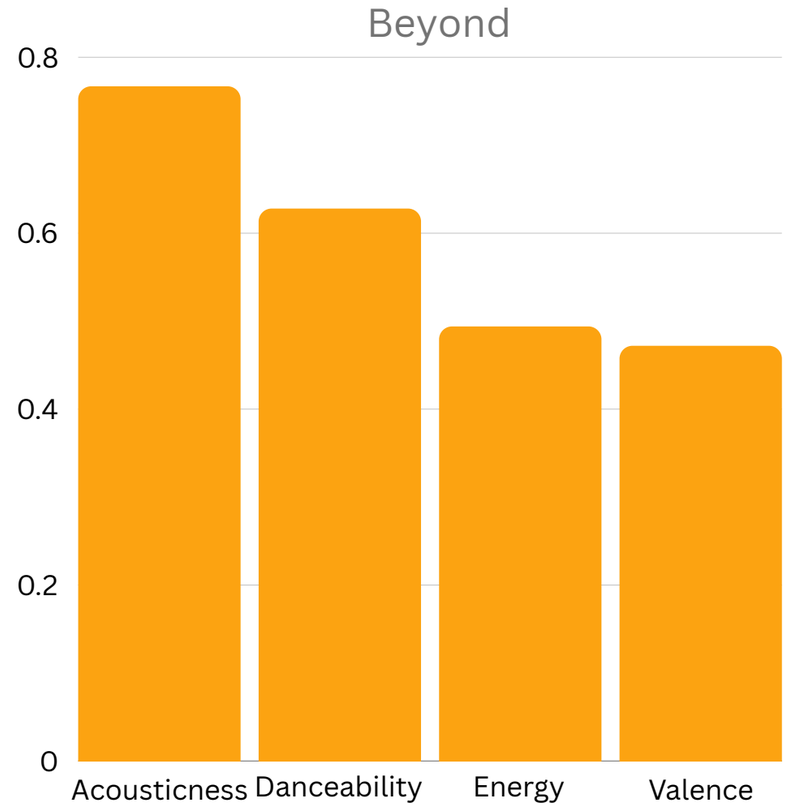
You're Gonna Go Far, Kid
The Offspring



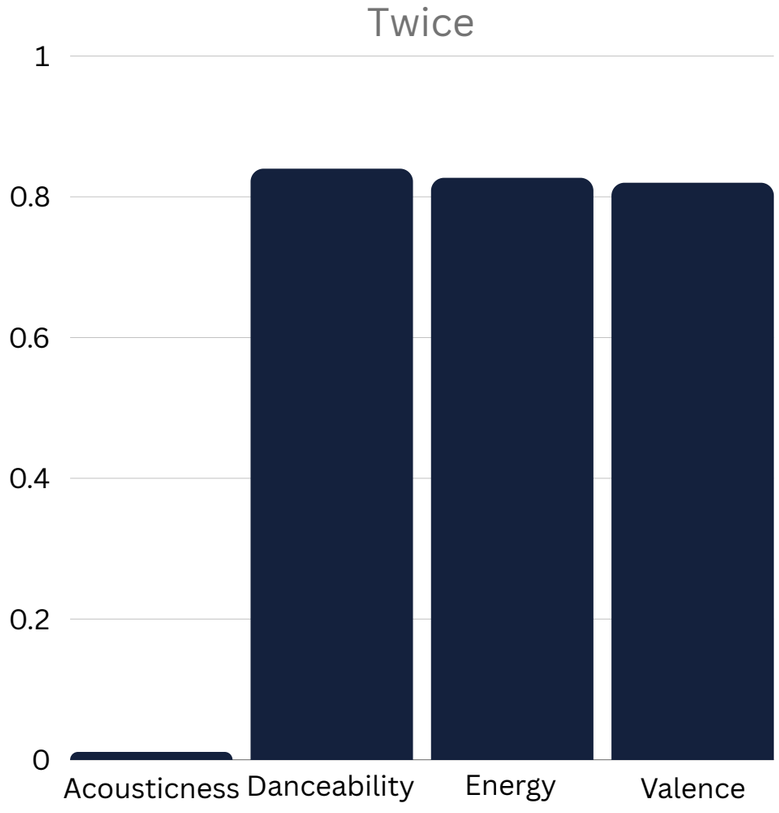
海闊天空
Beyond



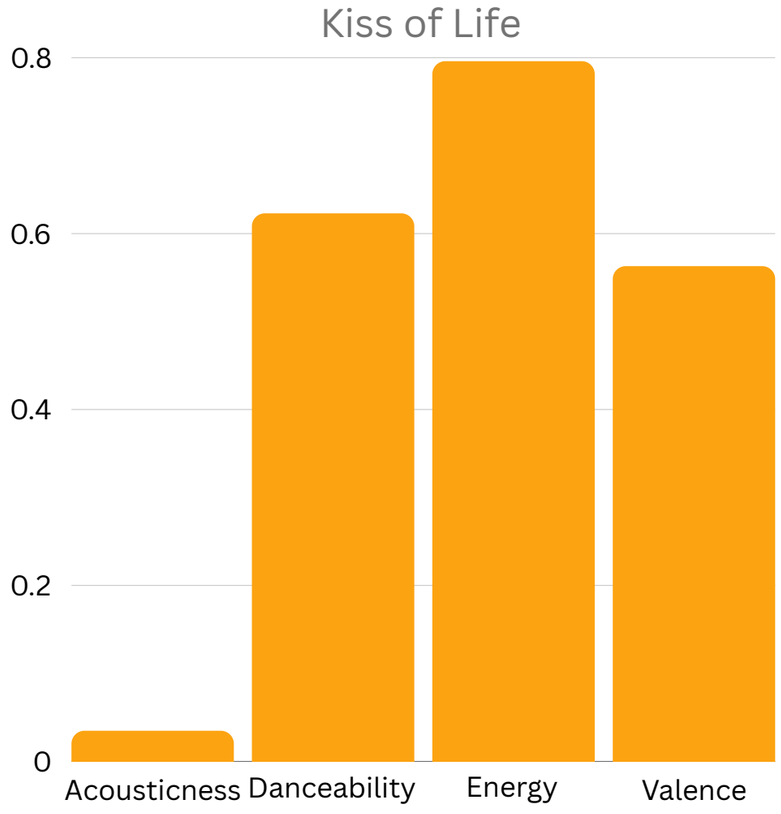
光輝歲月
Beyond



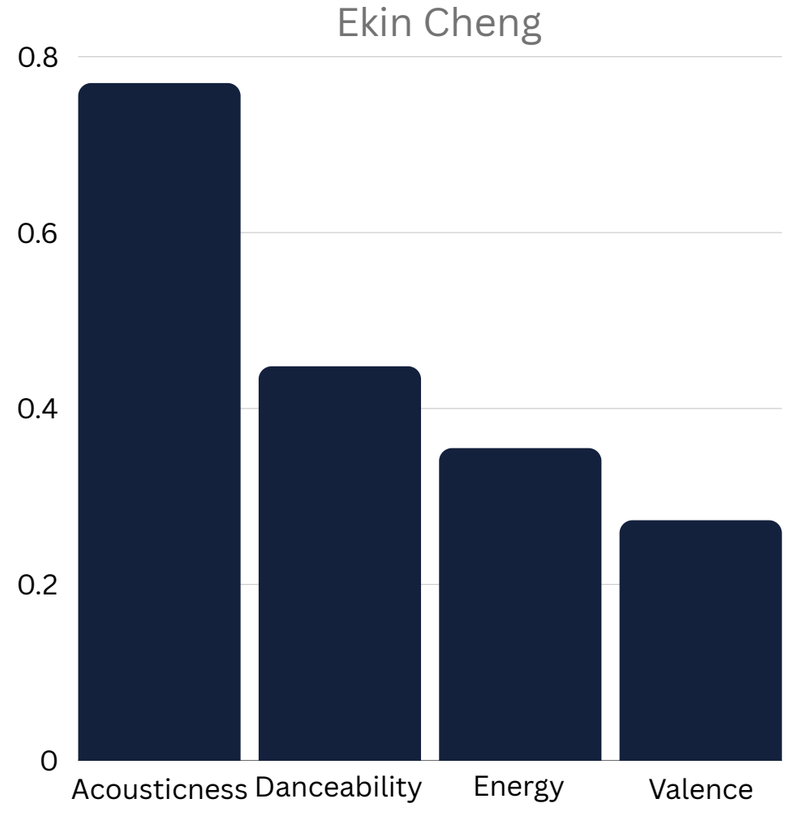
CRUEL
Twice



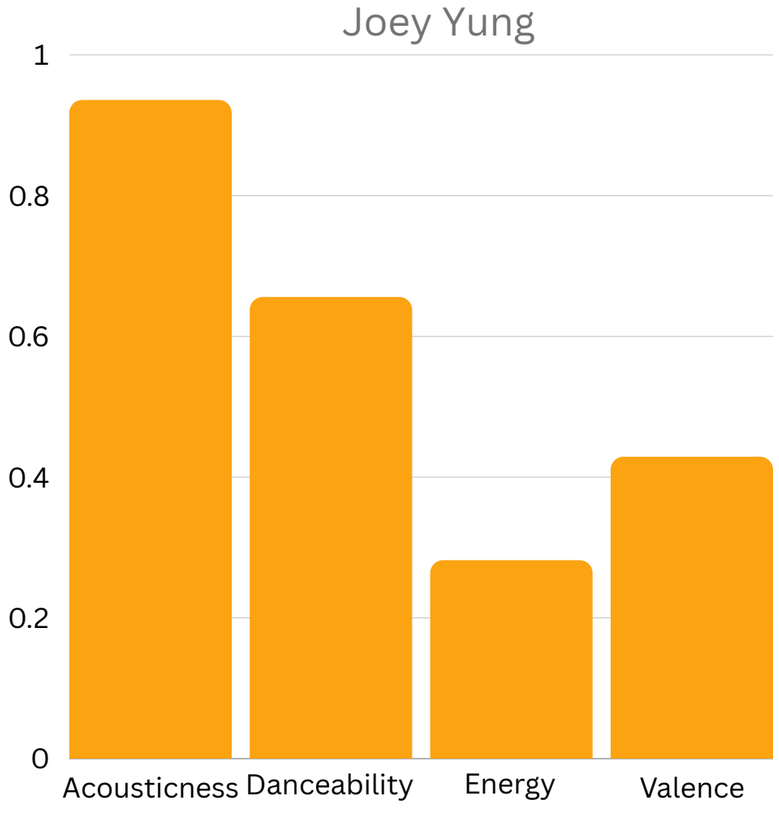
Shhh
Kiss of Life



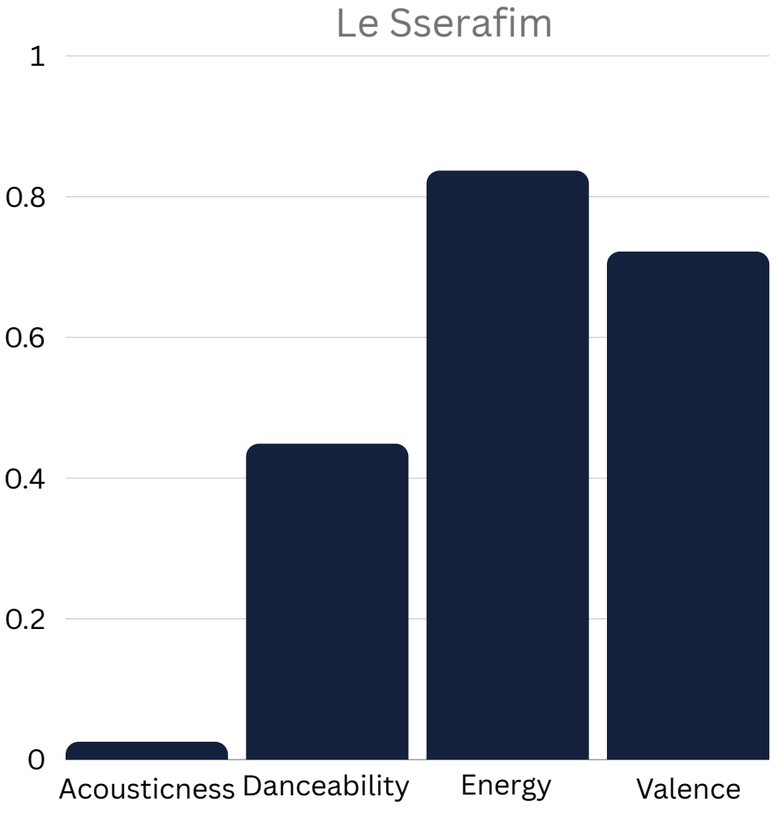
甘心替代你
Ekin Cheng



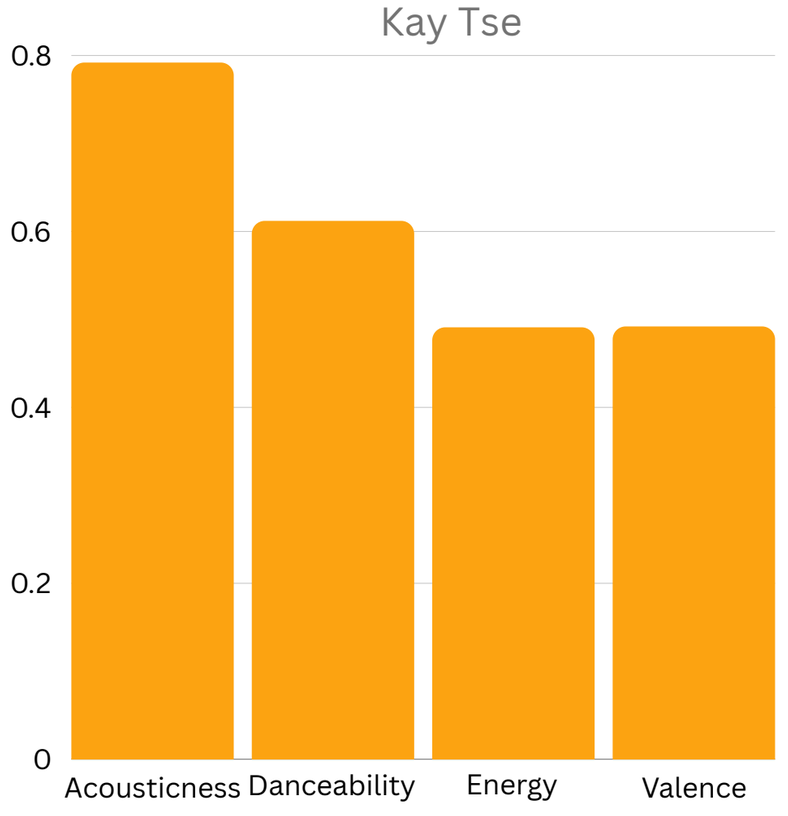
我的驕傲
Joey Yung



CHOICES
Le Sserafim



囍帖街
Kay Tse



Results

ANALYSIS OF MY MOST LISTENED SONGS

A track’s audio features can be described as the list of attributes that can help generalize songs of similar nature or genres of music. Some of these attributes include: acousticness, danceability, energy or valence.
Each of these are described by Spotify as:

- Acousticness - a confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
- Energy - a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
- Danceability - describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
- Valence - a measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

Credit to Spotify for defining these features:
https://developer.spotify.com/documentation/web-api/reference/get-audio-features

Although I don’t really have any genre asides from Cantopop rock, or korean pop in my most listened songs, I feel that most of them follow a similar genre and style of music. I feel personally, that when I need to do work I’m never really looking for something to gain out of the music, and instead I am mostly looking for something to fill the background noise. I think that also listening to such high energetic music most of the time, helps to prevent fatigue from falling low, as well as having singable, high valence songs also prevents my mood from dropping, which in an ironic sort of way keeps me focused on the tasks at hand.

Reflection

WHAT HAVE I LEARNED

One of the biggest takeaways I have gotten from this project is designing an actual data model and then using that model to create inferences from the data. Although this data was relatively simple to understand, I learned a lot about understanding and creating visualizations from variables, or defining custom metrics from a given set of data. I also learned some very important data visualization skills such as scaling a value in order to keep everything within the same relative range to create an effective visualization (used in mood and fatigue scatter plots).

MY THOUGHTS

As some pointers to keep in mind for the next time I attempt another project like this, I would tell myself to really think about what variables are actually needed before just collecting a variable that wouldn't give any insight or relation to other values in my data. I would also probably try and ask for more records and variety of data to work with, as despite this being a relatively small dataset, it felt that there wasn't much room for me to represent data in anything other than scatter plots, bar graphs or line charts. I also think that storing numbers "as is" is actually not a bad approach when dealing with purely quantitative data. As such, I had a lot of freedom to create my own relationships between the variables as opposed to if I was locked into using a specific format of data for specific variables. Something else I really apreciated while working on this project was learning the BASH scripting language to help me automate the process of collecting the data, because I really don't think I would've been bothered to track everything manually, and if I were to do a similar data analysis project again, I would definately look for ways to automate it. One thing I regret doing, was not tracking which songs I listened to the most for each sesion, as I feel like that could've added more value to my last visualization where I speak about how I feel that song choice and selection is imoprtant. I also think that the volume of data was too much for certain visualizations, and maybe I should've broken some of these into graphs by month