

Spotify Personal Streaming Analysis

*insert dashboard photo with link to live BI report

Project Background

Introduction

This project uses my personal streaming history from Spotify to identify my music taste and listening habits from 2015-2024 / year-to-year change. This dataset is really cool, as it is completely unique to me. Use this data to learn more about myself.

About the Data

My full personal streaming history was delivered from Spotify via email, stored in 23 JSON files containing 327,123 total records. Each song or podcast episode streamed throughout the lifespan of my account is summarized in a row. ...information regarding when and for how long the record played. The unique identifier is the date-time-stamp (stored as **ts**) of when the song or podcast was streamed.

Records detailing a song include the artist, album name, and song title

Podcasts include the show name and episode name.

All media types in the dataset use *ms_played* representing the duration that the media was played for in milliseconds.

Executive Summary

Data Cleaning & Preparation

Data Aggregation

- Loaded all 23 files into Power Query and used *Append Queries as New* to insert all records into one usable dataset titled ***full_streaming_history***
 - Since the column names and data structure in each original JSON file matched, no additional manipulation was required to standardize the data.
- After combining the datasets, I would typically disable loading for each individual source query to optimize performance. In this case, since each file contains sensitive personal information, I replaced the combined query with a static, hard-coded copy to eliminate dependencies and secure the model for public sharing.

Data Cleaning

Deleted Columns

- ***ip_addr***, ***conn_country***, and ***platform*** to protect sensitive personal information
- ***audiobook_title***, ***audiobook_uri***, ***audiobook_chapter_uri***, and ***audiobook_chapter_title***, as audiobooks are not a significant part of my listening profile
- ***incognito_mode***, as this feature has never been used on my account

Renamed Columns

- *spotify_track_uri* to **track_id**
- *spotify_episode_uri* to **episode_id**
- *master_metadata_track_name* to **track_name**
- *master_metadata_album_artist_name* to **artist_name**
- *master_metadata_album_album_name* to **album_name**
- *episode_show_name* to **show_name**

Created Columns

- **sec_played**, **min_played**, and **hrs_played** as alternative measures of record duration
- **month_streamed**, **hr_of_day_streamed**, and **day_of_week_streamed** to identify listening habits
 - also created **day_of_week_id** to order days of week Mon-Sun in visuals
- **song** to indicate if a record represents a song (1) or podcast (0)

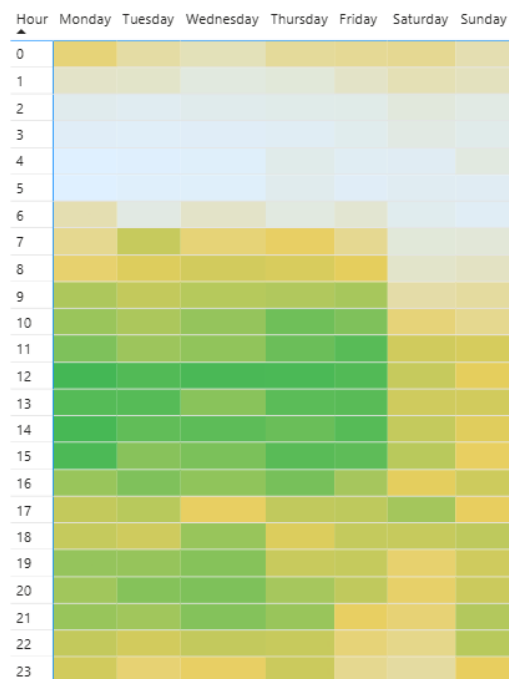
Changed Data Types

- changed datatypes of all indicator variables from TRUE/FALSE to Whole Number

Summary of Insights

Listening Patterns

- I streamed the most media on weekdays, with spikes during the middle of the day and the evening.



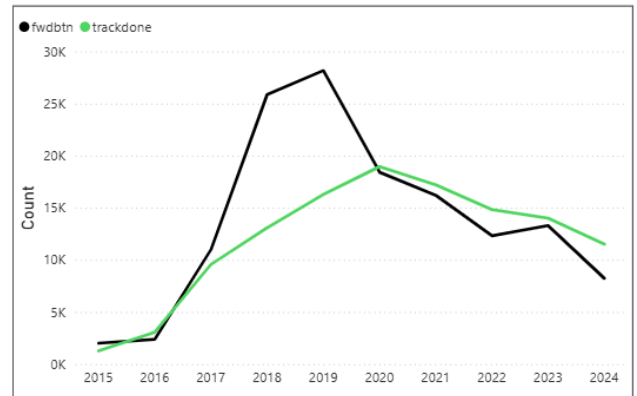
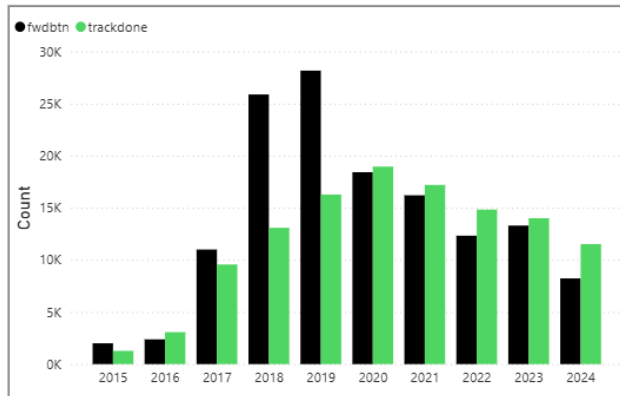
Seasonality

- Overall seasonality effect
- I streamed more podcast episodes and fewer songs during the summer and fall months due to NFL topics being discussed on my regularly streamed podcast channels.

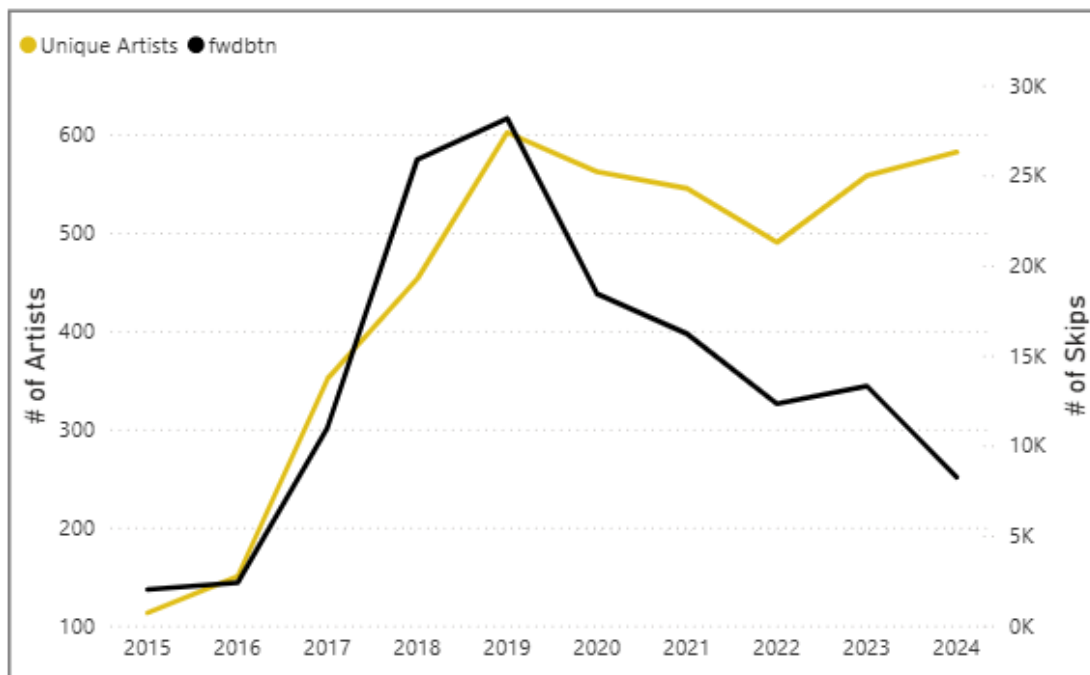
Behaviors/Habits (year-to-year)

- Podcast vs. Music Split

- The ratio of skipped songs to finished songs visualized on the left chart indicates that I have become a more patient listener year over year. The right chart illustrates the peak of my impatience



- But what caused the spike? The visual below charts the number of skipped songs vs. the number of unique artists streamed by year.



- As I ventured out, streaming an increased number of unique artists each year, I had to figure out what music I liked. At the same time, Spotify's algorithm took my feedback. So, we both were learning what kind of music I enjoy.
- my increased patience, but also Spotify's powerful algorithm. As I have logged more streaming, Spotify's ability to suggest songs I enjoy, and therefore am less likely to skip, is enhanced.

Evolution of Music Taste

Taste by Year

Top artists and songs by year

CURRENT TASTE

Recommendations

- I listen to a healthy mix of rock, alternative and bluegrass. This is a major shift from my high school and early college years, where I streamed mostly rap.
- My favorite artist is Green Day, streaming the most hours across multiple hours over many years. Walters / Olivia Rodrigo, but exclusively one album from each.
- I have grown into an inquisitive listener, as I skip fewer songs and branch out with more new artists. Part of this is due to the expert algos from spot.
- Since I enjoy spending time outside, I can expect my streaming frequency to drop significantly.