THE WEEKND

BRIEF DATA EXPLORATORY STORIES

VOL. 1/1

Table of contents

Brief Data Exploratory Stories,

The Weeknd,

by Matthew Balogh,

Volume 1, Issue 1

Updated at 11, June, 2024.

About Brief Data Exploratory Stories

Brief Data Exploratory Stories encloses my practical experiments in the field of *Data Analysis* and *Data Science*, as its name implies, in a short written form.

It aims to cover exploring a given dataset, analysing it, and making easily interpretable insights, mainly visualizations from it.

It does not include code chunks or detailed descriptions of the processes.

With keeping it simple and straightforward, my goal is to get an understanding of the given dataset and context, and to extract insights while raising meaningful questions.

I expect that the topics, objectives, and methods used in these notebooks will transform overtime as I progress in the *Data Science* learning process. By considering these work products as essential pillars of my current studies, I decided to document these exercises, and to make more memorable, I designed them around a theme that gives the appearance of a magazine.

Github: https://github.com/matthew-balogh/bdxps

LinkedIn: https://linkedin.com/in/mate-balogh-a137771b7

About the Spotify Dataset

The dataset [1] contains the most popular songs on Spotify as of the end of 2023.

The toplist contains tracks regardless to their release year, and popularity of a song is measured in the total number of streams it has accumulated by the time of the data extraction.

A single observation includes information about the title of the song, the count and names of the involved artists, release date, and streams. Whether the given song takes place in a curated playlist or chart of Spotify, Apple, Deezer, or Shazam (only chart info), and if so how many of them, is also indicated for each observation.

Besides these, there are some technical details provided for each song like bpm, danceability, or liveness to name a few.

[1] Source: Kaggle, https://www.kaggle.com/datasets/zeesolver/spotfy

Dataset Characteristics

The original dataset consists of the following variables:

```
 [1] "track_name"            "artist(s)_name"
 [3] "artist_count"          "released_year"
 [5] "released_month"        "released_day"
 [7] "in_spotify_playlists" "in_spotify_charts"
 [9] "streams"               "in_apple_playlists"
[11] "in_apple_charts"       "in_deezer_playlists"
[13] "in_deezer_charts"      "in_shazam_charts"
[15] "bpm"                   "key"
[17] "mode"                  "danceability_%"
[19] "valence_%"             "energy_%"
[21] "acousticness_%"        "instrumentalness_%"
[23] "liveness_%"            "speechiness_%"
```

Since this report focuses on overall insights and not on specific technical details nor playlists or charts, most of the variables could be ignored with only the following ones [2] kept for the analysis:

```
[1] "track_name"     "artist_s_name"  "artist_count"
[4] "released_year"  "released_month" "released_day"
[7] "streams"
```

[2] note that the variable "artist(s)_name" had been renamed to "artist_s_name"

Besides the variable names, the dataset has the following structure:

```
Rows: 953
Columns: 7
$ track_name     <chr> "Seven (feat. Latto) (Explicit Ver.~
$ artist_s_name  <chr> "Latto, Jung Kook", "Myke Towers", ~
$ artist_count   <int> 2, 1, 1, 1, 1, 2, 2, 1, 1, 2, 2, 1,~
$ released_year  <int> 2023, 2023, 2023, 2019, 2023, 2023,~
$ released_month <int> 7, 3, 6, 8, 5, 6, 3, 7, 5, 3, 4, 7,~
$ released_day   <int> 14, 23, 30, 23, 18, 1, 16, 7, 15, 1~
$ streams        <dbl> 141381703, 133716286, 140003974, 80~
```

```
 track_name        artist_s_name       artist_count
Length:953         Length:953         Min.   :1.000
Class :character   Class :character   1st Qu.:1.000
Mode  :character   Mode  :character   Median :1.000
                                      Mean   :1.556
                                      3rd Qu.:2.000
                                      Max.   :8.000
 released_year  released_month    released_day
 Min.   :1930   Min.   : 1.000   Min.   : 1.00
 1st Qu.:2020   1st Qu.: 3.000   1st Qu.: 6.00
 Median :2022   Median : 6.000   Median :13.00
 Mean   :2018   Mean   : 6.034   Mean   :13.93
 3rd Qu.:2022   3rd Qu.: 9.000   3rd Qu.:22.00
 Max.   :2023   Max.   :12.000   Max.   :31.00
    streams
 Min.   :2.762e+03
 1st Qu.:1.417e+08
 Median :2.902e+08
 Mean   :5.139e+08
 3rd Qu.:6.738e+08
 Max.   :3.704e+09
```

The dataset contains 953 observations, that is 953 songs.

Variables of `track_name` and `artist_s_name` are character strings, while the other ones represent numbers. The minimum of artist

count (`artist_count`) is 1 representing the main artist, and goes up to a maximum of 8 across all the observations. Release date (`released_year`, `released_month`, `released_day`) goes from 1930 to 2023, streams (`streams`) from around 2700 up to 3.7 billion for a given song.

The following is a few-line sample from the dataset.

```
       track_name   streams released_year released_month
1 Seven (feat.... 141381703          2023              7
2            LALA 133716286          2023              3
3         vampire 140003974          2023              6
  released_day    artist_s_name artist_count
1           14 Latto, Jung Kook            2
2           23      Myke Towers            1
3           30   Olivia Rodrigo            1
```

Songs by The Weeknd

As for the songs by *The Weeknd*, there are 37 of them and can be summarized as follows:

```
 track_name         artist_s_name       artist_count
Length:37          Length:37          Min.   :1.000
Class :character   Class :character   1st Qu.:1.000
Mode  :character   Mode  :character   Median :1.000
                                      Mean   :1.486
                                      3rd Qu.:2.000
                                      Max.   :3.000
 released_year  released_month    released_day
 Min.   :2015   Min.   : 1.000   Min.   : 2.00
 1st Qu.:2020   1st Qu.: 1.000   1st Qu.: 7.00
 Median :2022   Median : 3.000   Median : 7.00
 Mean   :2021   Mean   : 4.486   Mean   :13.68
 3rd Qu.:2022   3rd Qu.: 8.000   3rd Qu.:22.00
 Max.   :2023   Max.   :12.000   Max.   :29.00
    streams
 Min.   :3.196e+07
 1st Qu.:1.011e+08
 Median :3.913e+08
 Mean   :6.468e+08
 3rd Qu.:6.981e+08
 Max.   :3.704e+09
```

The maximum artist count (`artist_count`) is 3 in this case, indicating that there are songs with 2 additional artists involved. Streams (`streams`) are from around 32 million up to 3.7 billion for a given song.

As for the release date, `released_year` spans the years from 2015 to 2023, `released_month` the months from January to December, `released_day` the days from the 2nd to the 29th day of the month. However, it is important to note, that these are only the two edges of the distributions of the variables, therefore the above summarization

does not indicate whether, for example, the artist has a popular song on the list with a release year of 2016, with a release month of February, or with a release day of 15.

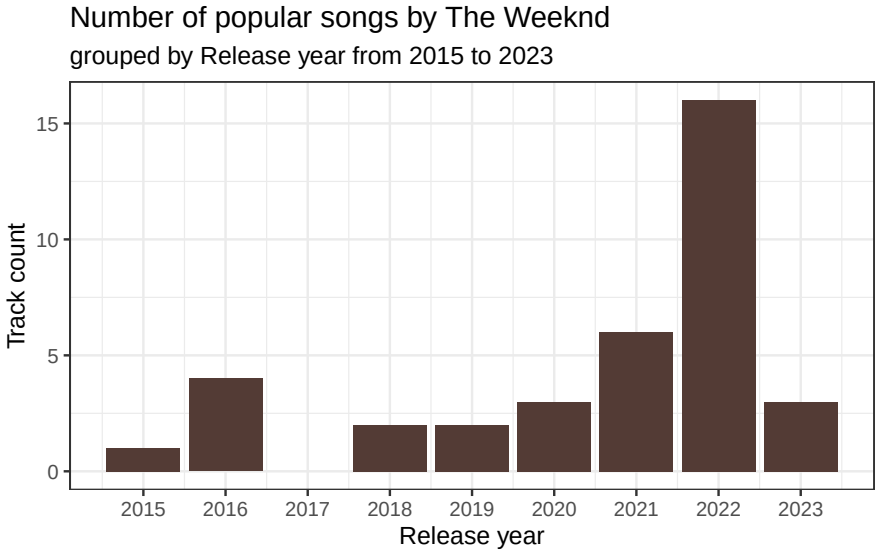
The following is a few-line sample from this subset of the dataset.

```
       track_name    streams released_year released_month
1 Popular (wit...  115364561          2023              6
2        Creepin'  843957510          2022             12
3     Die For You 1647990401          2016             11
4         Starboy 2565529693          2016              9
5 Die For You ...  518745108          2023              2
  released_day                       artist_s_name
1            2  The Weeknd, Madonna, Playboi Carti
2            2 The Weeknd, 21 Savage, Metro Boomin
3           24                          The Weeknd
4           21               The Weeknd, Daft Punk
5           24           Ariana Grande, The Weeknd
  artist_count
1            3
2            3
3            1
4            2
5            2
```

Insights

Q1: In which year were these hits released?

As the bar chart shows below, these songs were released between 2015 and 2023. The artist's hits were released in all years except 2017 of this period. From all the songs that he produced in the year 2022, 16 was popular in 2023. With its tall bar in the chart, these songs - released in that single year - alone contributed to his result with an approximate of 43%.

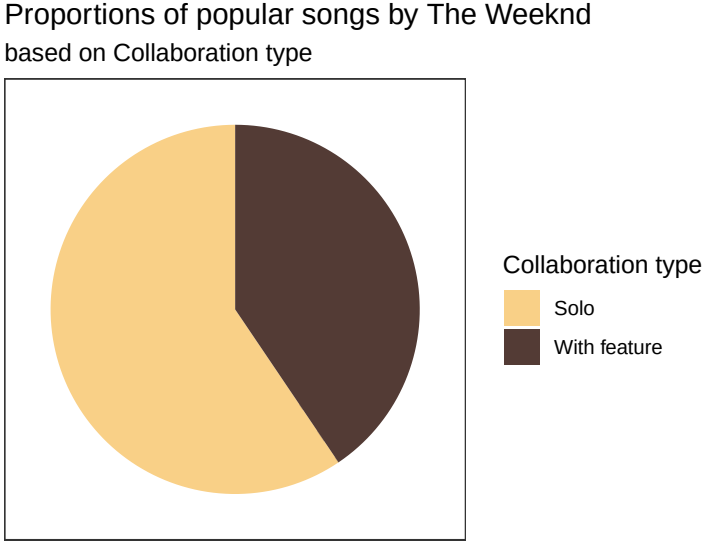


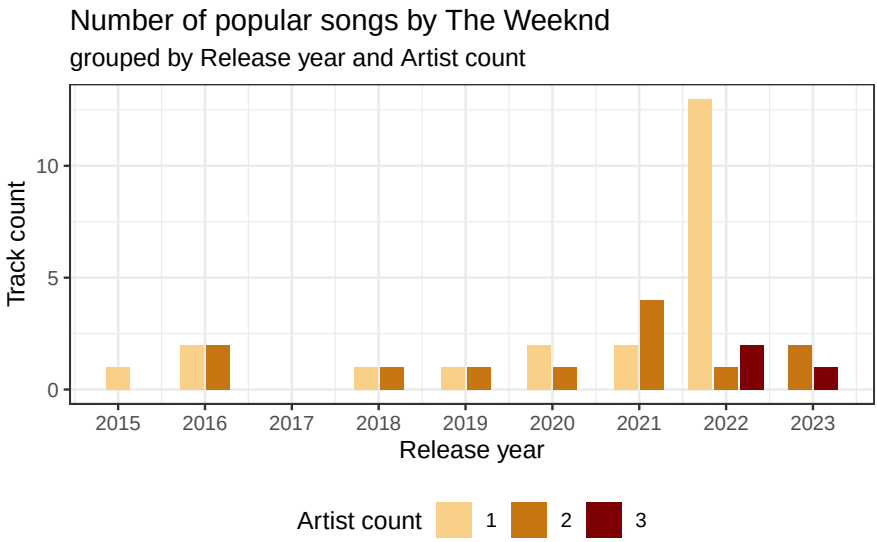
Q2: Were these hits solo tracks or included other artists as features?

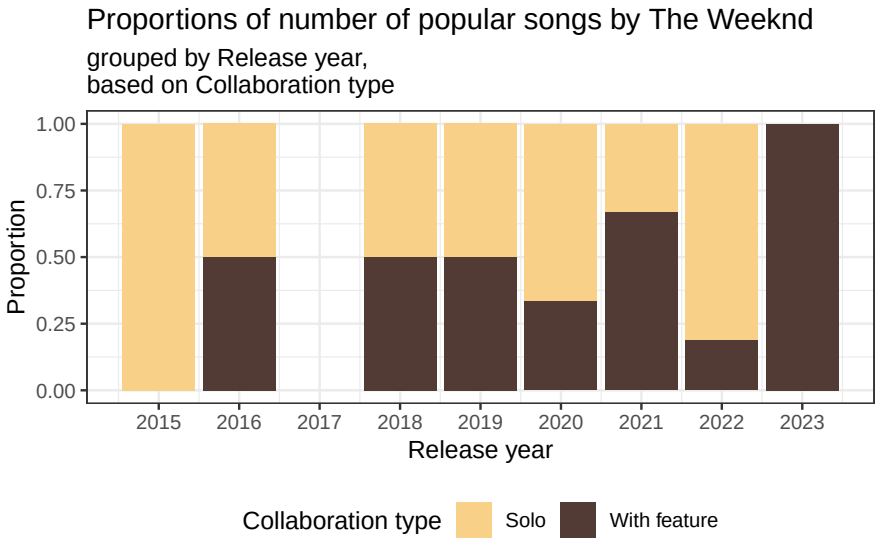
A major part of these songs were solo productions. In 2015, only *Solo* songs were released, that is, without collaboration. However, between 2016 and 2021, tracks include collaborations with another single artist in a somewhat balanced manner.

Some of the tracks from 2022 and 2023 have also been released in collaboration with another artist and some of them even have multiple artists involved in the collaboration (*Multiple features*).

Interestingly, his tracks from 2022 that made to the toplist were 80% *Solo* productions, while no *Solo* track of the artist from 2023 landed a position in the toplist.




Number of popular songs by The Weeknd
grouped by Release year and Artist count
Track count
10
5
0
2015
2016
2017
2018
2019
2020
2021
2022
2023
Release year
Artist count
1
2
3


Proportions of number of popular songs by The Weeknd
grouped by Release year,
based on Collaboration type
Proportion
1.00
0.75
0.50
0.25
0.00
2015
2016
2017
2018
2019
2020
2021
2022
2023
Release year
Collaboration type
Solo
With feature

Q3: From these songs of the toplist, which are the most successful ones in terms of streams on the platform?

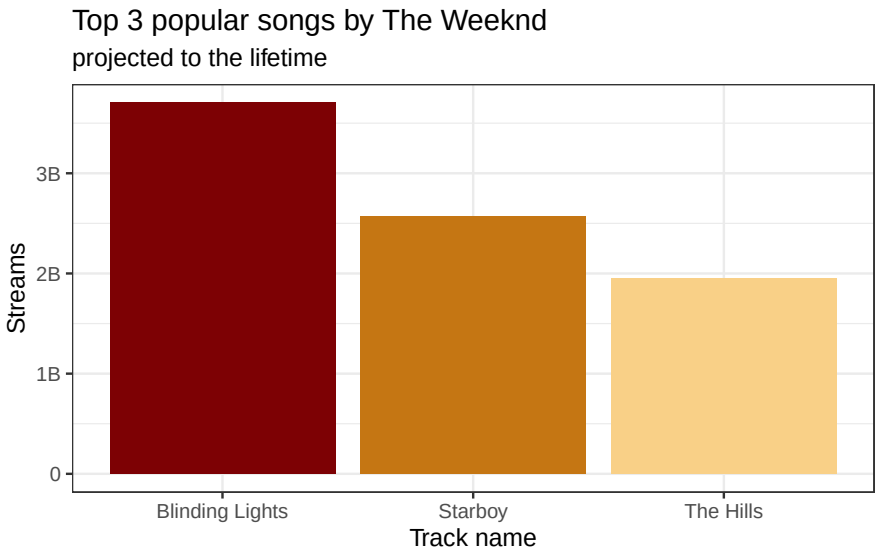
The most successful ones in terms of total streams are:

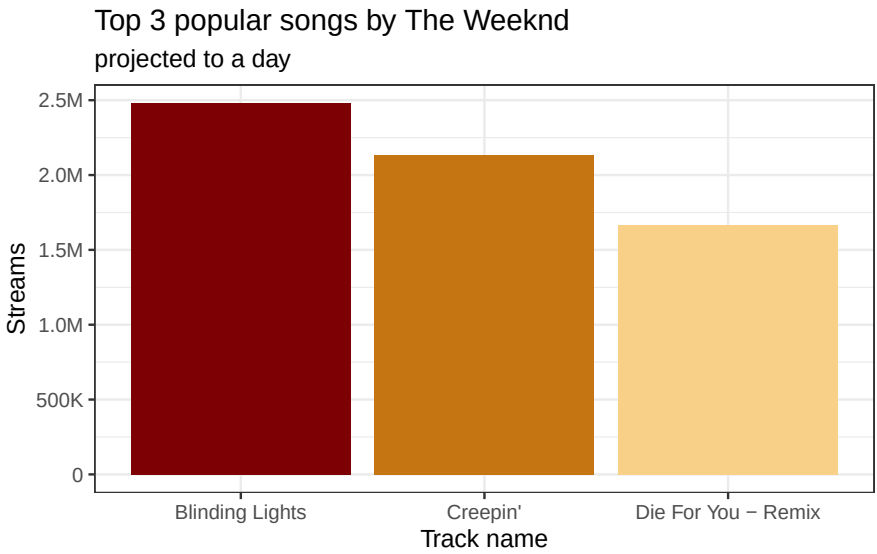
```
       track_name    streams
1 Blinding Lights 3703895074
2         Starboy 2565529693
3       The Hills 1947371785
```

Given that songs differ in their release dates, a different list is acquired if the streams are projected to a day as follows:

```
           track_name streams_day
1     Blinding Lights     2479180
2             Creepin'     2136601
3 Die For You - Remix     1667991
```

Either way, the track titled as *Blinding Lights*, is at the first place with its 3.7 billion total streams and an almost 2.5 million streams projected to a day.


Top 3 popular songs by The Weeknd
projected to the lifetime
Streams
3B
2B
1B
0
Blinding Lights
Starboy
The Hills
Track name


Top 3 popular songs by The Weeknd
projected to a day
Streams
2.5M
2.0M
1.5M
1.0M
500K
0
Blinding Lights
Creepin'
Die For You − Remix
Track name

Q4: Have these top 3 songs of the artist been out for a while or are there fresh hits in it?

The artist's top 3 songs, considering lifetime total streams, all have been released back in 2019, 2016, and 2015 respectively. Even though older songs have better chance to accumulate more total streams compared to fresh ones, his top 2 songs released in 2019 and 2016 separately have more streams than the one released in 2015.

```
       track_name    streams released_year
1 Blinding Lights 3703895074          2019
2         Starboy 2565529693          2016
3       The Hills 1947371785          2015
```

If streams are projected to a day of the lifetime of a song, then the result includes two relatively fresh songs from 2022, and 2023, yet the first place is still occupied by the same song from 2019, as it has been already indicated before.

```
           track_name streams_day released_year
1     Blinding Lights     2479180          2019
2            Creepin'     2136601          2022
3 Die For You - Remix     1667991          2023
```

Q4.1: What about the top 10?

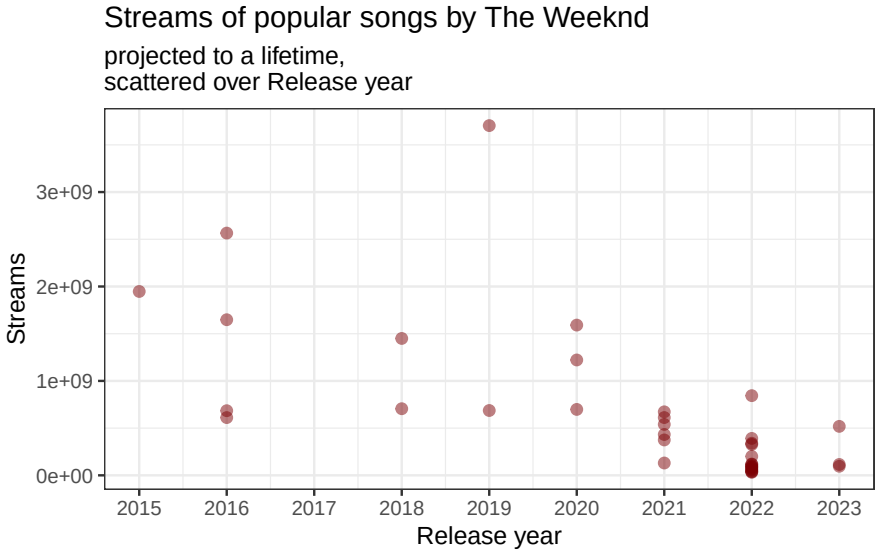
Taking a look at the top 10 songs of the artist in the lists below, it is hard to tell whether fresh or old songs have accumulated more streams.

```
                              track_name released_year
1                         Blinding Lights          2019
2                                 Starboy          2016
3                               The Hills          2015
4                            Die For You          2016
```

```
5                            Save Your Tears          2020
6                          Call Out My Name           2018
7  Save Your Tears (with Ariana Grande) ...           2020
8                                  Creepin'           2022
9                          I Was Never There          2018
10                               After Hours          2020
```

```
                                  track_name released_year
1                            Blinding Lights          2019
2                                  Creepin'           2022
3                        Die For You - Remix          2023
4                            Save Your Tears          2020
5                                    Starboy          2016
6  Save Your Tears (with Ariana Grande) ...           2020
7          Moth To A Flame (with The Weeknd)          2021
8                                  You Right          2021
9                            Call Out My Name          2018
10           One Right Now (with The Weeknd)          2021
```

A scatterplot of streams over the release year of songs could be a better way to answer this question.

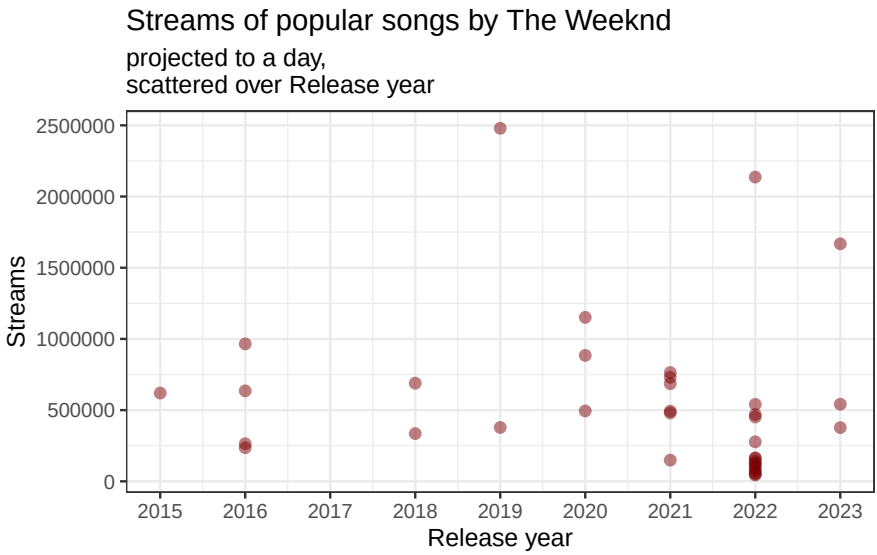


The scatterplot depicts that as release year increases, that is the more fresh a song is, streams count seems to decrease, that is the less streams a song has (except a few outliers - the top 3 songs). Based on the visualization, there might be a correlation between the two variable and if there is, it is probably negative. Visualization, however, can easily deceive one.

As it appears, there really is a negative correlation, however, as the top 3 list and the coefficient of determination show, it is not a very reliable indicator.

	Correlation with Streams
Release year	-0.66
	Coefficient of determination (%)
Release year	43.17

As expected, the scatterplot of streams projected to a day of a song's lifetime over the release year shows a more rectangular-ish positioning, that is, other than a few outliers (the top 3 songs), the streams count projected to a day does not change with the change of release year. This is verified by the table below, where the correlation coefficient is very close to 0 and where the coefficient of determination is extremely low.



	Correlation with Streams (projected to a day)
Release year	-0.14

	Coefficient of determination (%)
Release year	1.86