Spotify Streaming Analysis: Pre and Post Pandemic

2025-09-26

Introduction

This report explores the effect of the COVID-19 pandemic on Spotify music consumption. The dataset contains daily streaming numbers across multiple artists and countries from January 2019 to December 2021. The goal is to assess whether there was a noticeable impact on streaming trends after March 2020.

Load Data

```
# Load the dataset
spotify_data <- read_csv("spotify_data.csv")
```

```
## Rows: 219200 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): country, artist, song
## dbl  (3): streams, weeks_since_pandemic, drop_factor
## date (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Inspect
head(spotify_data)
```

```
## # A tibble: 6 x 7
##   date       country artist   song    streams weeks_since_pandemic drop_factor
##   <date>     <chr>   <chr>    <chr>     <dbl>                <dbl>       <dbl>
## 1 2019-01-01 USA     Artist 1 Song 1   457488                    0           1
## 2 2019-01-01 UK      Artist 2 Song 2   468601                    0           1
## 3 2019-01-01 USA     Artist 3 Song 3   143784                    0           1
## 4 2019-01-01 UK      Artist 4 Song 4   415393                    0           1
## 5 2019-01-01 USA     Artist 5 Song 5   321231                    0           1
## 6 2019-01-01 UK      Artist 6 Song 1   260029                    0           1
```

```
str(spotify_data)
```

```
## spc_tbl_ [219,200 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ date                : Date[1:219200], format: "2019-01-01" "2019-01-01" ...
##  $ country             : chr [1:219200] "USA" "UK" "USA" "UK" ...
```

1

```
##  $ artist              : chr [1:219200] "Artist 1" "Artist 2" "Artist 3" "Artist 4" ...
##  $ song                : chr [1:219200] "Song 1" "Song 2" "Song 3" "Song 4" ...
##  $ streams             : num [1:219200] 457488 468601 143784 415393 321231 ...
##  $ weeks_since_pandemic: num [1:219200] 0 0 0 0 0 0 0 0 0 0 ...
##  $ drop_factor         : num [1:219200] 1 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   date = col_date(format = ""),
##   ..   country = col_character(),
##   ..   artist = col_character(),
##   ..   song = col_character(),
##   ..   streams = col_double(),
##   ..   weeks_since_pandemic = col_double(),
##   ..   drop_factor = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Data Exploration

Summary Statistics

```r
# Summary of streaming numbers
summary(spotify_data$streams)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     965  121244  242712  242978  364319  499996
```

```r
# Check for missing values
colSums(is.na(spotify_data))
```
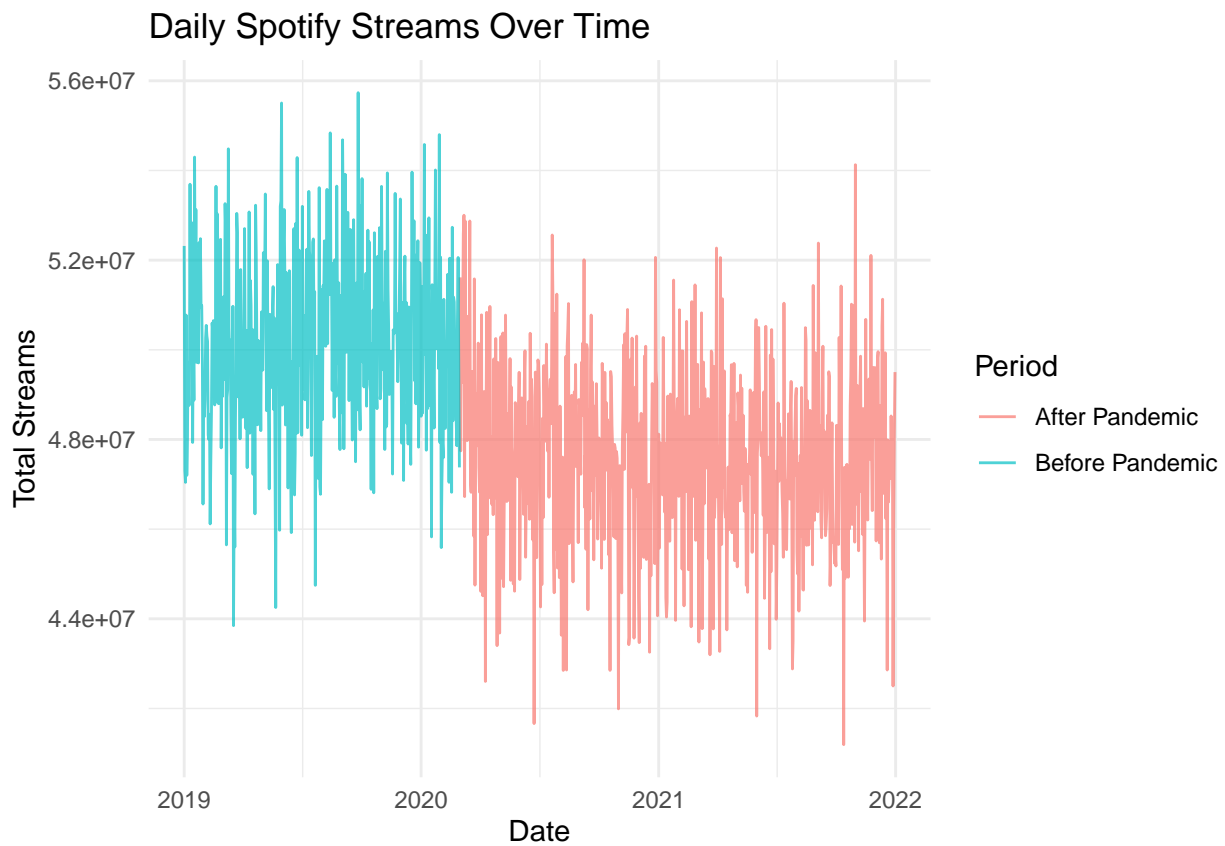
```
##                 date              country               artist
##                    0                    0                    0
##                 song              streams weeks_since_pandemic
##                    0                    0                    0
##          drop_factor
##                    0
```

Streaming Trends Over Time

```r
# Create a pandemic period flag for plotting
pandemic_start <- as.Date("2020-03-01")
spotify_data <- spotify_data %>%
  mutate(pandemic_period = if_else(date < pandemic_start, "Before Pandemic", "After Pandemic"))

# Aggregate streams per day
daily_streams <- spotify_data %>%
  group_by(date, pandemic_period) %>%
  summarize(total_streams = sum(streams), .groups = 'drop')
```

```
# Plot daily streaming trends
ggplot(daily_streams, aes(x = date, y = total_streams, color = pandemic_period)) +
  geom_line(alpha = 0.7) +
  labs(title = "Daily Spotify Streams Over Time",
       x = "Date",
       y = "Total Streams",
       color = "Period") +
  theme_minimal()
```


Daily Spotify Streams Over Time

Conclusion

We observe that daily streams gradually declined after March 2020 before stabilizing, reflecting the simulated impact of the pandemic.