

Spotify Streaming Data Analysis Project

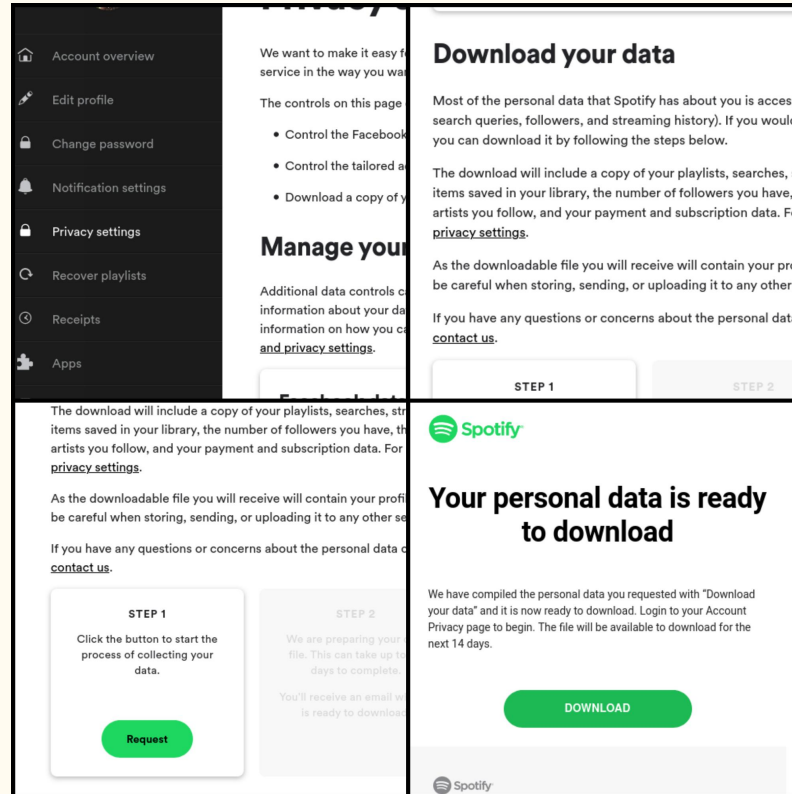
Bhupesh K. Chandrawanshi

Outline:

- Introduction to Spotify Dataset
- Dataset Statistics and User Details
- Exploration and Visualisation
- Insight from Data
- Summary and Future Work

Introduction to Spotify Dataset

Downloading Dataset : [Link](#)



<input type="checkbox"/> 0 ▾		📁 / Downloads / MyData (2)	
		📁 ..	
<input type="checkbox"/>	📄	Follow.json	
<input type="checkbox"/>	📄	Identity.json	
<input type="checkbox"/>	📄	Inferences.json	
<input type="checkbox"/>	📄	Payments.json	
<input type="checkbox"/>	📄	Playlist1.json	
<input type="checkbox"/>	📄	Read_Me_First.pdf	
<input type="checkbox"/>	📄	SearchQueries.json	
<input type="checkbox"/>	📄	StreamingHistory0.json	
<input type="checkbox"/>	📄	StreamingHistory1.json	
<input type="checkbox"/>	📄	Userdata.json	
<input type="checkbox"/>	📄	YourLibrary.json	

Dataset Statistics & Spotify User Details

Stream History Dataset Content

```
: df = pd.read_json("StreamingHistory0.json") # Load the JSON File into a dataframe  
df.head() # Read first few lines of the dataframe
```

	endTime	artistName	trackName	msPlayed
0	2020-07-26 05:32	Fotty Seven	Chaud	192311
1	2020-07-27 07:34	DaBaby	ROCKSTAR (feat. Roddy Ricch)	181733
2	2020-07-27 07:38	THE SCOTTS	THE SCOTTS	165977
3	2020-07-27 07:39	Logic	Perfect	100291
4	2020-07-27 07:43	Roddy Ricch	The Box	196652

Formatted Dataset

```
: spotify_stream_df.head() # To check the newly formed dataset with additional columns
```

	endTime	artistName	trackName	msPlayed	Play-Time	year	month	day	weekday	time	hours	day-name	Count	Time-Played (hh-mm-ss)	Listening Time(Hours)
0	2020-07-26 05:32	Fotty Seven	Chaud	192311	2020-07-26 05:32:00	2020	7	26	6	05:32:00	5	Sunday	1	0 days 00:03:12.311000	0.053
1	2020-07-27 07:34	DaBaby	ROCKSTAR (feat. Roddy Ricch)	181733	2020-07-27 07:34:00	2020	7	27	0	07:34:00	7	Monday	1	0 days 00:03:01.733000	0.050
2	2020-07-27 07:38	THE SCOTTS	THE SCOTTS	165977	2020-07-27 07:38:00	2020	7	27	0	07:38:00	7	Monday	1	0 days 00:02:45.977000	0.046
3	2020-07-27 07:39	Logic	Perfect	100291	2020-07-27 07:39:00	2020	7	27	0	07:39:00	7	Monday	1	0 days 00:01:40.291000	0.028
4	2020-07-27 07:43	Roddy Ricch	The Box	196652	2020-07-27 07:43:00	2020	7	27	0	07:43:00	7	Monday	1	0 days 00:03:16.652000	0.054

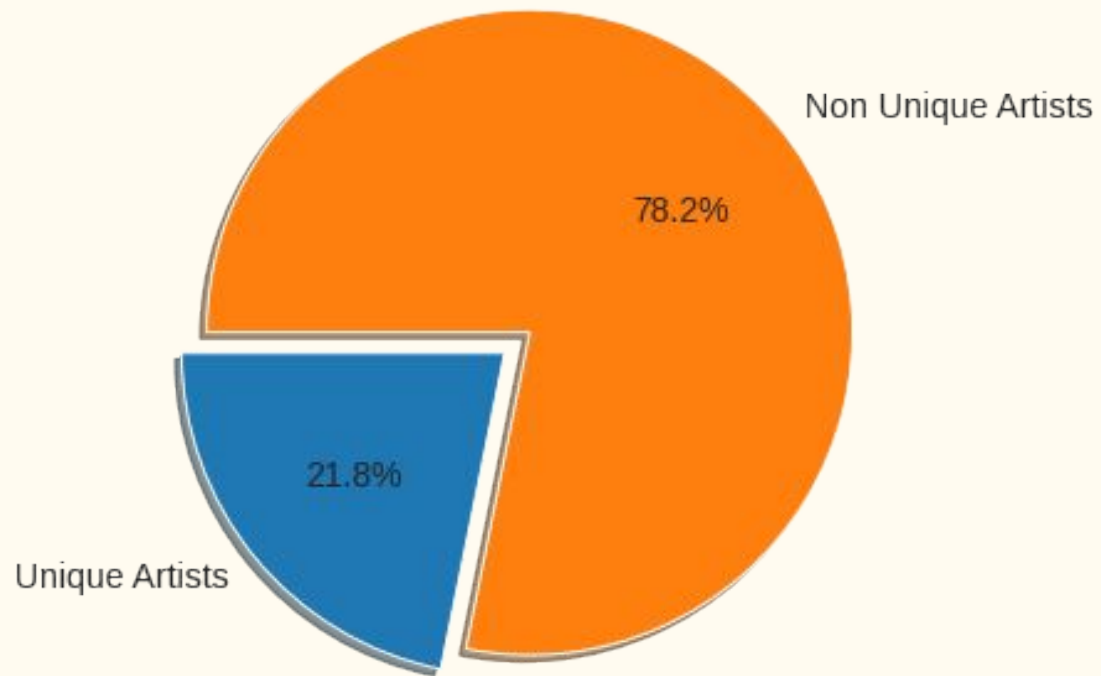
Quick Statistics

```
: spotify_stream_df.describe() # Final check for any abnormality
```

	year	month	day	weekday	hours	Count	Listening Time(Hours)	Listening Time(Minutes)
count	10804.000000	10804.000000	10804.000000	10804.000000	10804.000000	10804.0	10804.000000	10804.000000
mean	2020.561829	5.819326	16.268327	2.943817	13.245927	1.0	0.034158	2.049620
std	0.496185	3.755626	8.981046	1.960852	5.085855	0.0	0.028722	1.723019
min	2020.000000	1.000000	1.000000	0.000000	0.000000	1.0	0.000000	0.000000
25%	2020.000000	2.000000	9.000000	1.000000	9.000000	1.0	0.004000	0.250000
50%	2021.000000	4.000000	16.000000	3.000000	13.000000	1.0	0.039000	2.333000
75%	2021.000000	9.000000	24.000000	4.000000	17.000000	1.0	0.056000	3.350000
max	2021.000000	12.000000	31.000000	6.000000	23.000000	1.0	0.388000	23.300000

Exploration & Visualisation

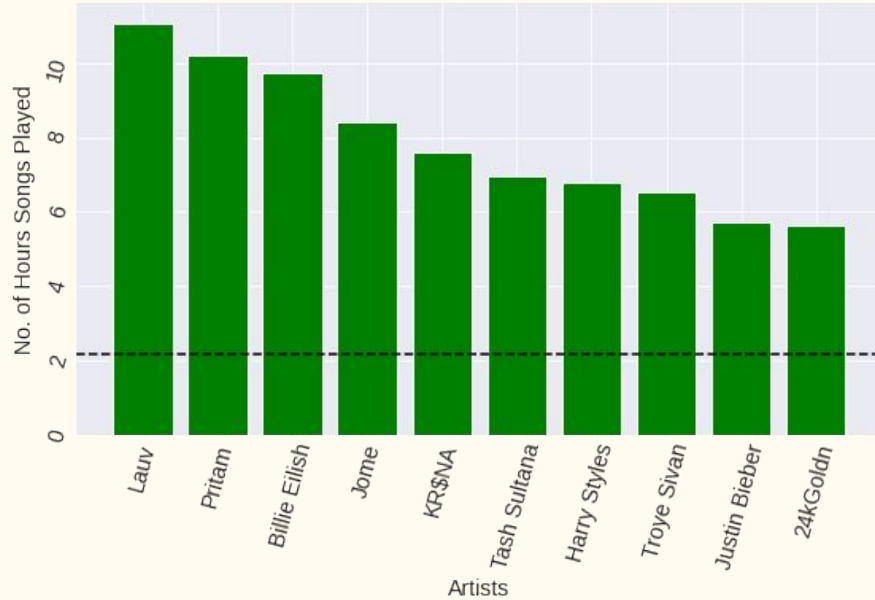
Unique Artist Percentage



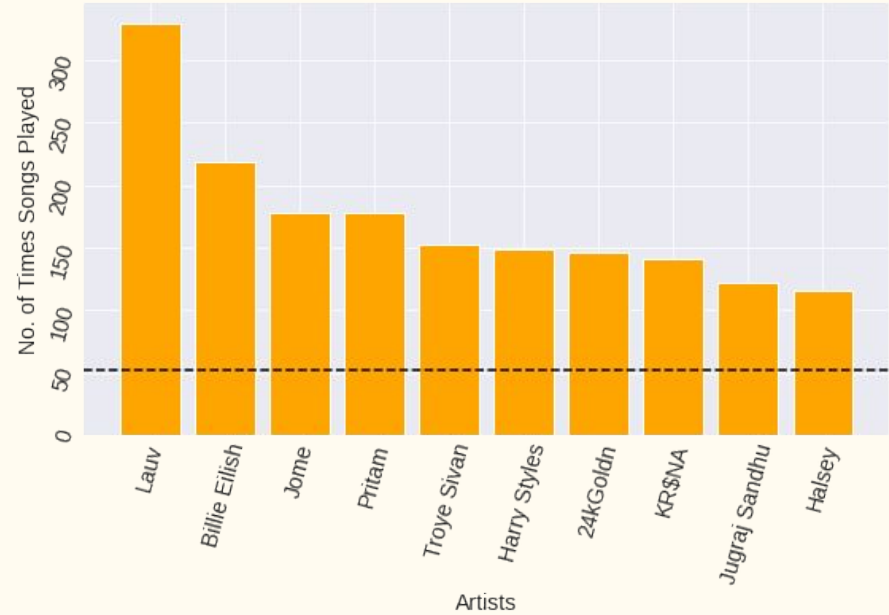
*The unique artist percentage comes around 21.83 %.
I usually listen to the 21% of my favourite artist most
of the time.*

Top 10 Artists

My Top 10 Favourite Artist (based on Hours)

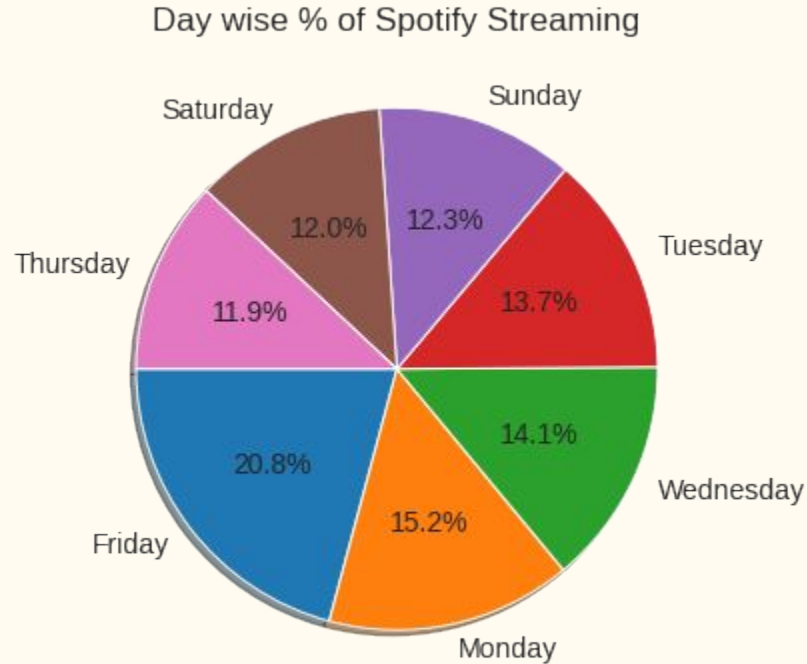


My Top 10 Favourite Artist (based on Counts)



Top 10 Artists list is divided into two sections and they are quite different.

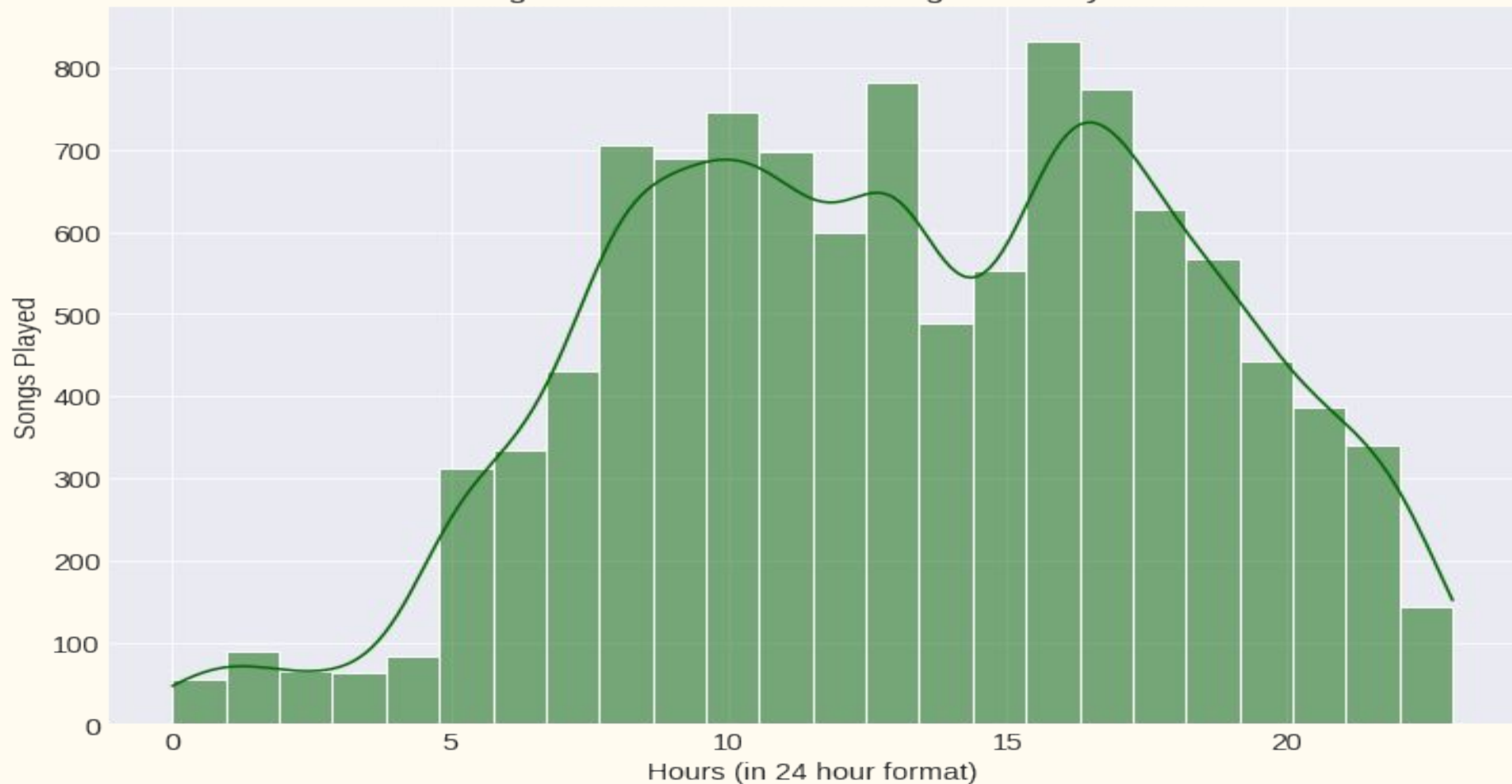
Weekly Spotify Streaming Statistics



*Unlike others, I usually stream less on Weekends
because I travel and go out.*

*Friday has the largest chunk because I work on my
projects on Fridays and I like to listen my playlist
while working.*

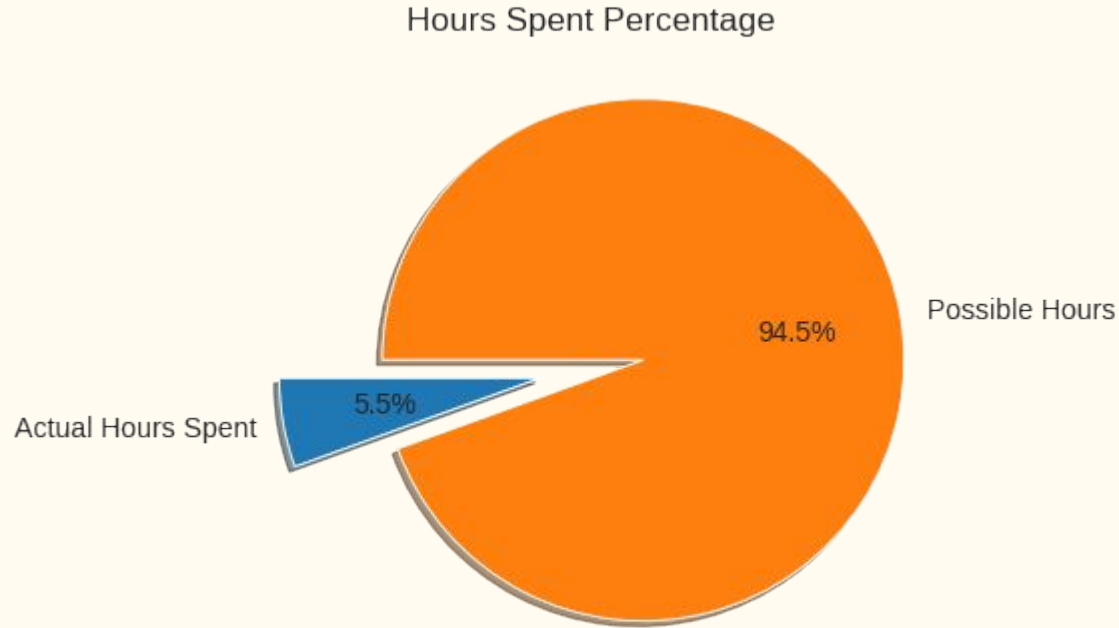
Average Distribution of Streaming Over Day Hours



*My usual peak listening/streaming hours lies in
between 8:00 AM and 5:00 PM*

Insights from our Data Analysis

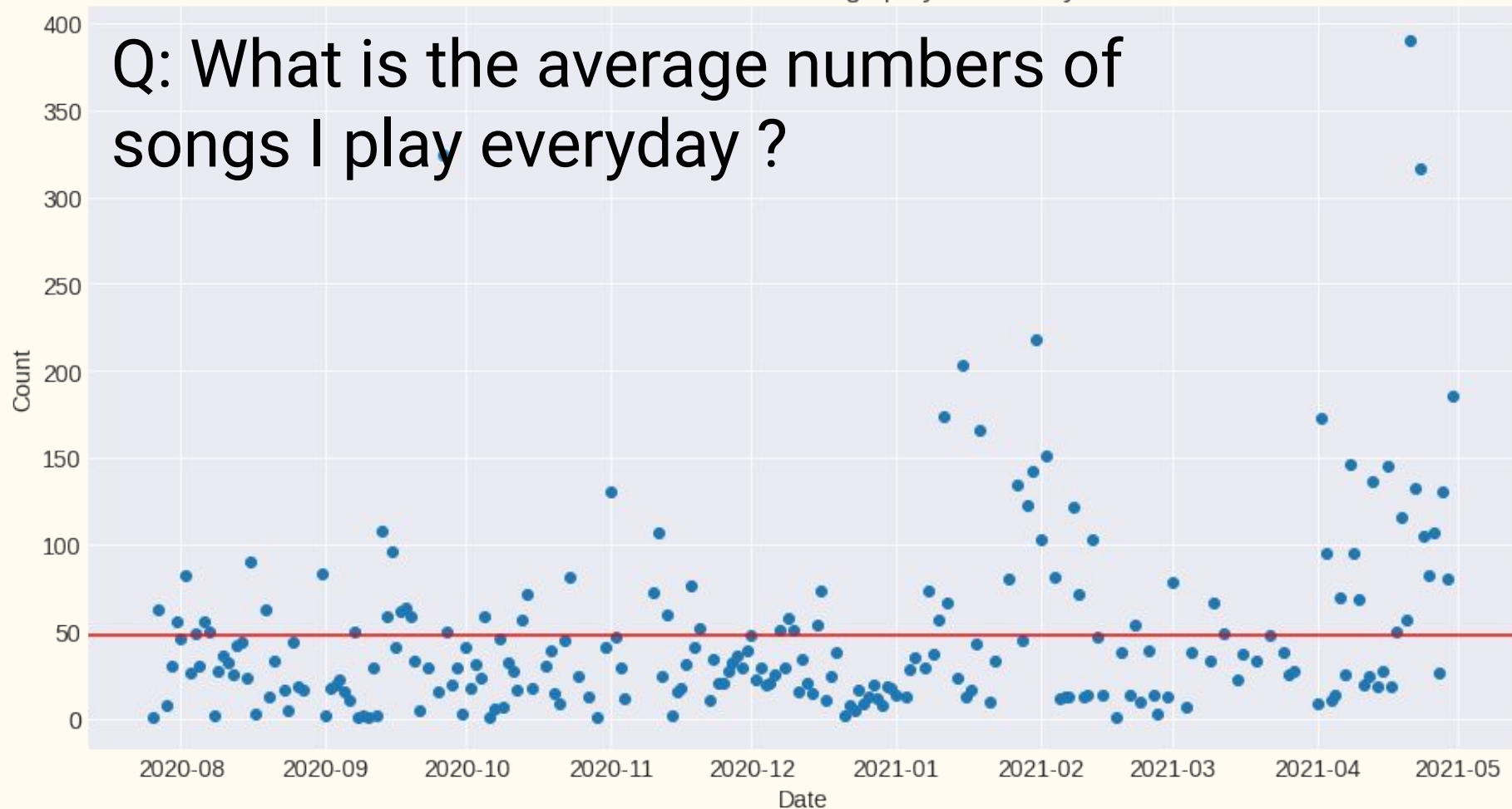
Q: What is my actual usage percentage ?



- *Time Spent : 369 Hours*
- *Total Possible : 6688.5 Hours*
- *Time Spent Percentage : 5.52%*

Maximum number of songs played in a day

Q: What is the average numbers of songs I play everyday ?



Average Songs Player per Day : 39

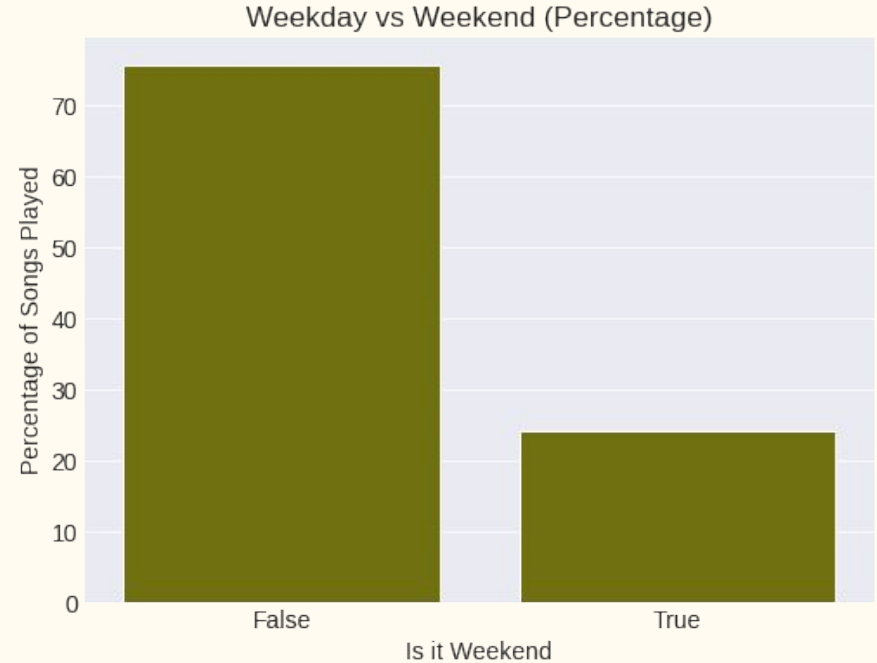
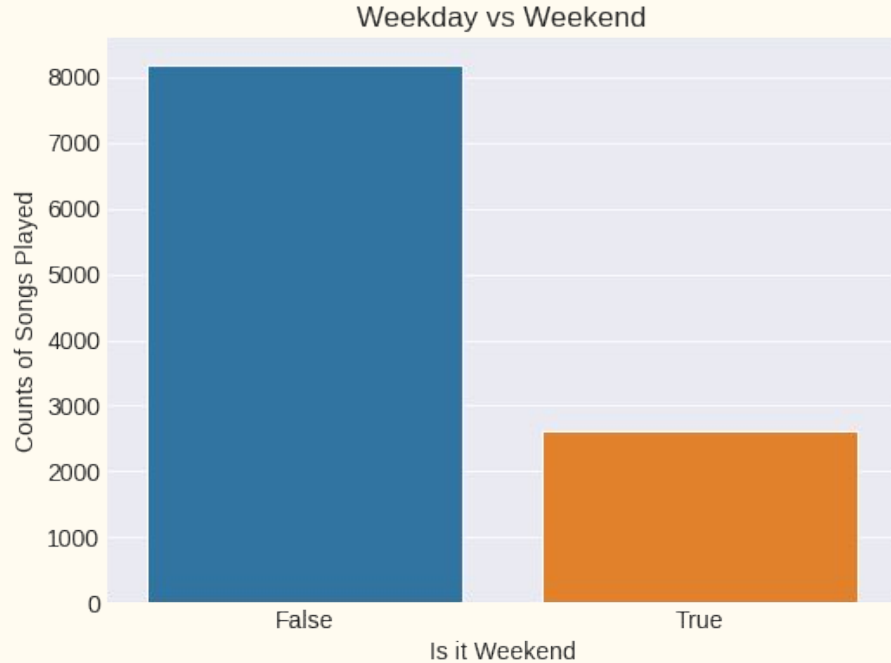
*Maximum number of Songs Played in a Day: 390
(2021-04-21)*

A word cloud featuring the names of various Indian celebrities and musicians. The names are arranged in a circular pattern, with 'Lauv' being the largest and most prominent name in the center. Other large names include 'Justin Bieber', 'Billie Eilish', 'Pritam', 'Jome', 'Troye', 'Sivan', 'Harry Styles', 'KR\$NA', '24kGoldn', and 'The Score'. Smaller names like 'Jas Gawronski', 'Arijit Singh', 'Neha Kakkar', 'Shreya Ghoshal', 'Alia Bhatt', 'Vishal Mishra', 'Kunal Khemu', 'Ranveer Singh', 'Katrina Kaif', 'Ayushmann Khurrana', 'Vicky Kaushal', 'Nikhil D'Souza', 'Rohit Shetty', 'Ajay Devgn', 'Anil Kapoor', 'Shah Rukh Khan', 'Deepika Padukone', 'Priyanka Chopra', 'Kareena Kapoor', 'Sonam Kapoor', 'Alia Bhatt', 'Vishal Mishra', 'Kunal Khemu', 'Ranveer Singh', 'Katrina Kaif', 'Ayushmann Khurrana', 'Vicky Kaushal', 'Nikhil D'Souza', 'Rohit Shetty', 'Ajay Devgn', 'Anil Kapoor', 'Shah Rukh Khan', 'Deepika Padukone', 'Priyanka Chopra', 'Kareena Kapoor', 'Sonam Kapoor' are also visible. The names are in various colors and fonts, creating a vibrant and dynamic visual.

[illegible]

Both list can be further converted into Top 100 Songs Playlist and Top 100 Artist Playlist.

Q: Usage distribution of Weekday vs Weekend ?



	Counts	Percentage
Weekend	8187	75.77749
Weekday	2617	24.22251

Summary & Future Work

- *2359 Unique Artist which comprises 21.83 % of total artist in my streaming history.*
- *4068 Unique Songs which comprises 37.65 % of total songs in my streaming history.*

- *Most Active Day: Fridays*
- *Most Active Hours: 8 AM to 7 PM.*
- *Maximum usage around 4PM-5PM and Minimum at 12AM-1PM hour period.*

- *39 songs played on average per day.*
- *On 21st April 2021 I played 390 Songs which is my personal maximum.*
- *279 Days Spotify Usage (Total Days)*
- *369 hours of streaming on spotify till now.*
- *5.5% of the possible 6688.5 hours*

- *My favourite artist whom I've played most is Lauv.*
- *While my favourite song is bitches broken hearts by Billie Eilish.*

Future Work:

- *Create random and custom playlist out of the User Data*
- *Look out for ways to increase the usage time to make more profit.*
- *Make pipelines to automate these process.*

Thank You!

Bhupesh K. Chandrawanshi