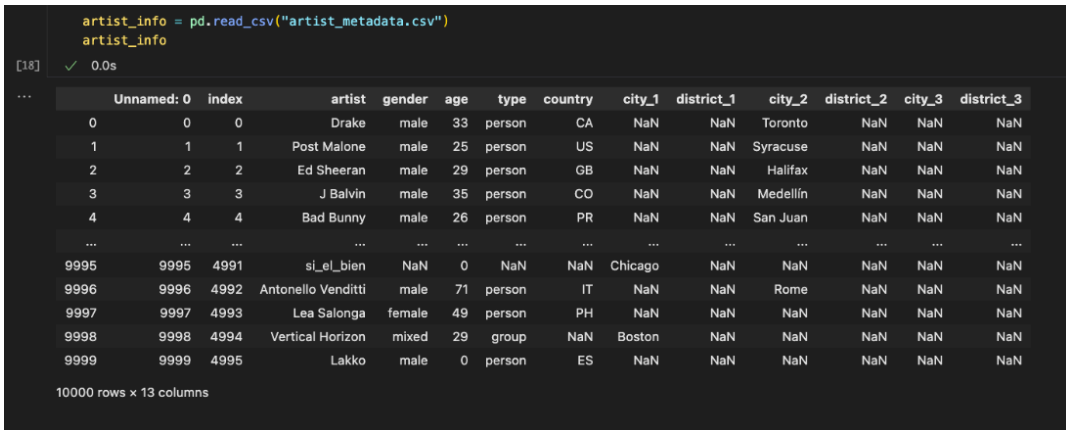
DSCI 311
Project 2 Data
Rayna Patel

The data I am working with is from Kaggle and is an extraction of data from Spotify: *Spotify Top Hit Playlist (2010-2023)* and *Spotify Artist Metadata Top 10K*. The Top Hit data is a list of the top songs by year with data about the track's audio features while the Artist Metadata includes artist gender, genres, and age. I want to use the artist metadata to assess questions about how an artist's demographic affects how popular the song is. For example, have female artists been gaining more popularity in recent years? Or, what audio features are the most popular by year? My main objective is to, based on audio features and artist demographics, predict what elements will make a song popular in future years.

Loading Data into Jupyter Notebook:



```
artist_info = pd.read_csv("artist_metadata.csv")
artist_info
```

[18] ✓ 0.0s

| | Unnamed: 0 | index | artist | gender | age | type | country | city_1 | district_1 | city_2 | district_2 | city_3 | district_3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | Drake | male | 33 | person | CA | NaN | NaN | Toronto | NaN | NaN | NaN |
| 1 | 1 | 1 | Post Malone | male | 25 | person | US | NaN | NaN | Syracuse | NaN | NaN | NaN |
| 2 | 2 | 2 | Ed Sheeran | male | 29 | person | GB | NaN | NaN | Halifax | NaN | NaN | NaN |
| 3 | 3 | 3 | J Balvin | male | 35 | person | CO | NaN | NaN | Medellín | NaN | NaN | NaN |
| 4 | 4 | 4 | Bad Bunny | male | 26 | person | PR | NaN | NaN | San Juan | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 9995 | 9995 | 4991 | si_el_bien | NaN | 0 | NaN | NaN | Chicago | NaN | NaN | NaN | NaN | NaN |
| 9996 | 9996 | 4992 | Antonello Venditti | male | 71 | person | IT | NaN | NaN | Rome | NaN | NaN | NaN |
| 9997 | 9997 | 4993 | Lea Salonga | female | 49 | person | PH | NaN | NaN | NaN | NaN | NaN | NaN |
| 9998 | 9998 | 4994 | Vertical Horizon | mixed | 29 | group | NaN | Boston | NaN | NaN | NaN | NaN | NaN |
| 9999 | 9999 | 4995 | Lakko | male | 0 | person | ES | NaN | NaN | NaN | NaN | NaN | NaN |

10000 rows × 13 columns

```
tracks = pd.read_csv("playlist_data.csv", encoding="latin1")
tracks
```

✓ 0.0s

| | playlist_url | year | track_id | track_name | track_popularity | album | ar |
|---|---|---|---|---|---|---|---|
| 0 | https://open.spotify.com/playlist/37i9dQZF1DWU... | 2000 | 6naxalmIoLFWR0siv8dnQQ | Oops!...I Did It Again | 81 | Oops!... I Did It Again | 26dSoYclwsYLMAKD |
| 1 | https://open.spotify.com/playlist/37i9dQZF1DWU... | 2000 | 2m1hi0nfMR9vdGC8UcrnwU | All The Small Things | 83 | Enema Of The State | 6FBDaR13swtiWwGhX |
| 2 | https://open.spotify.com/playlist/37i9dQZF1DWU... | 2000 | 3y4LxiYMgDl4RethdzpmNe | Breathe | 66 | Breathe | 25NQNriVT2YbSW80 |
| 3 | https://open.spotify.com/playlist/37i9dQZF1DWU... | 2000 | 0v1XpBHnsbkCn7iJ9Ucr1l | It's My Life | 81 | Crush | 58lV9VcRSjABbAbfW |
| 4 | https://open.spotify.com/playlist/37i9dQZF1DWU... | 2000 | 62bOmKYxYg7dhrC6gH9vFn | Bye Bye Bye | 75 | No Strings Attached | 6Ff53KvcvAj5U7Z1 |
| ... | ... | ... | ... | ... | ... | ... | |
| 2395 | https://open.spotify.com/playlist/6unJBM7ZGitZ... | 2023 | 0HD8mbiPjp3o94X3EaZp0o | exes | 56 | exes | 45dkTj5sMRSjrmBSE |
| 2396 | https://open.spotify.com/playlist/6unJBM7ZGitZ... | 2023 | 3XB5uhhlYSnkxpSihkNQwh | QLONA | 79 | MAÑANA SERÁ BONITO (BICHOTA SEASON) | 790FomKkXshlbRY |
| 2397 | https://open.spotify.com/playlist/6unJBM7ZGitZ... | 2023 | 4sx6NRwL6Ol3V6m9exwGlQ | LOVE AGAIN | 69 | LOVE AGAIN | 2tlP7SsRs7vjlcLrU |
| 2398 | https://open.spotify.com/playlist/6unJBM7ZGitZ... | 2023 | 2Zo1PcszsT9WQ0ANntJbID | Feather | 91 | emails i can t send fwd: | 74KM79TiuVKeVCqs8 |
| 2399 | https://open.spotify.com/playlist/6unJBM7ZGitZ... | 2023 | 11xC6P3iKYpFThT6Ce1KdG | Attention | 67 | Attention | 5cj0lLjcoR7YOSnhn |

2400 rows × 23 columns