Nicholas De Santos

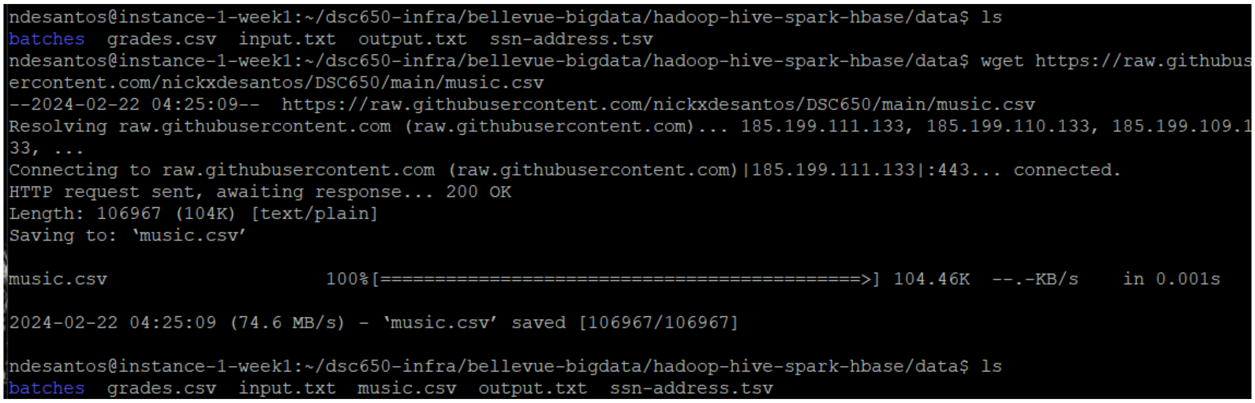
DSC 650

Winter 2023/2024

# Final Big Data Project

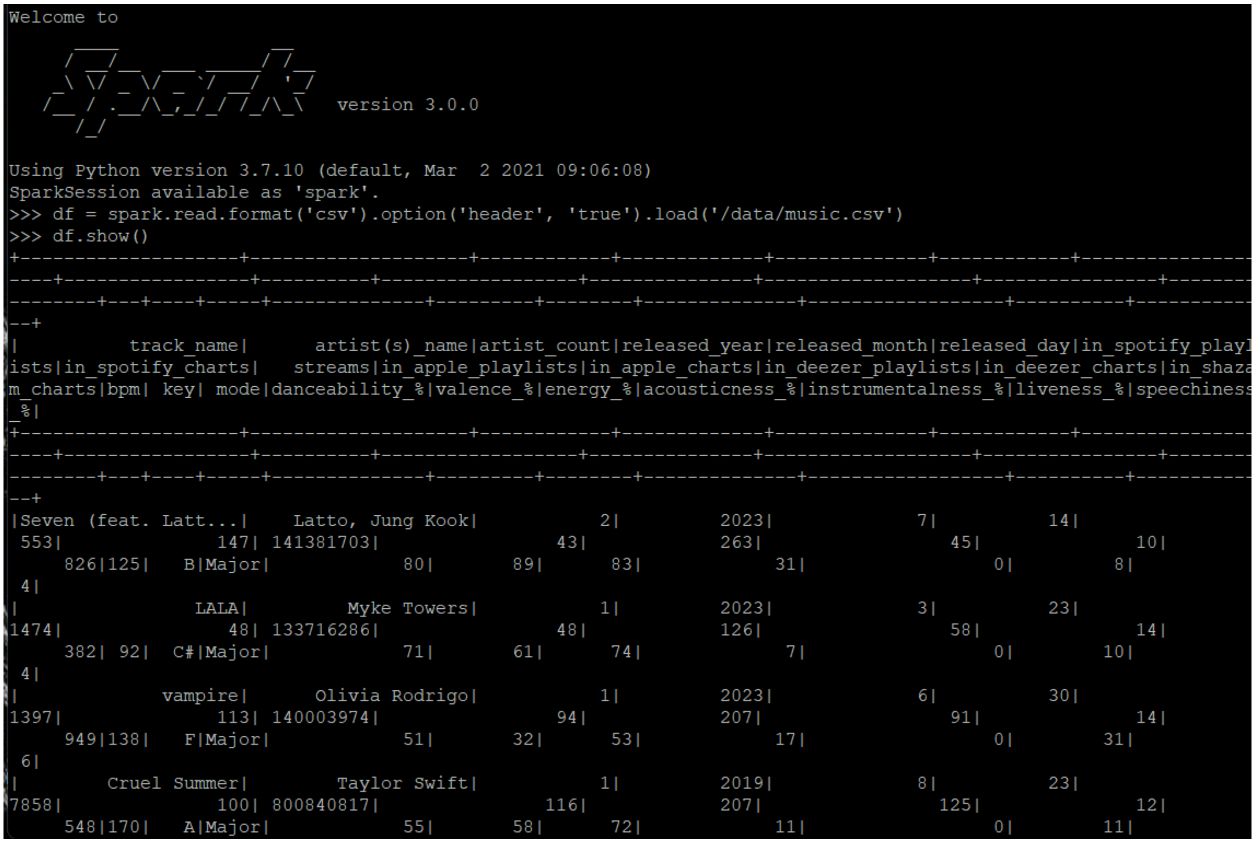
Music will always be part of the human experience. You can learn a lot about a person, about a culture, about a society from what type of music they listen to or what's popular during that time. Music of a time period can help people understand the main ideas of society of that time. Music has been a part of human culture for thousands of years and although it is constantly changing, it doesn't show signs of going away. The main objective of this report is to analyze recent music data to possibly help identify what features of songs and artists can help identify what type of music will be popular among the public. The data I chose to use for this analysis is data collected of the most streamed Spotify songs of 2023 (data can be found at https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023?resource=download ). The data contains nearly 1000 of the most streams songs of Spotify this last year. Ideally we would be working with more data for more years and larger datasets as well but the goal at the moment is to ultimately get an idea of what type of music and artists were most influential in 2023 and see the direction that music was moving and could be moving in the future.

Getting into our analysis, we conducted an investigation within a hadoop envirnment first by using Spark to work with and analyze the data. First we had to make sure we could import and access the data freely in our environment. Here is the code where we imported the data.



```
ndesantos@instance-1-week1:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase/data$ ls
batches  grades.csv  input.txt  output.txt  ssn-address.tsv
ndesantos@instance-1-week1:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase/data$ wget https://raw.githubus
ercontent.com/nickxdesantos/DSC650/main/music.csv
--2024-02-22 04:25:09--  https://raw.githubusercontent.com/nickxdesantos/DSC650/main/music.csv
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.111.133, 185.199.110.133, 185.199.109.1
33, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.111.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 106967 (104K) [text/plain]
Saving to: 'music.csv'

music.csv                    100%[============================================>] 104.46K  --.-KB/s    in 0.001s

2024-02-22 04:25:09 (74.6 MB/s) - 'music.csv' saved [106967/106967]

ndesantos@instance-1-week1:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase/data$ ls
batches  grades.csv  input.txt  music.csv  output.txt  ssn-address.tsv
```

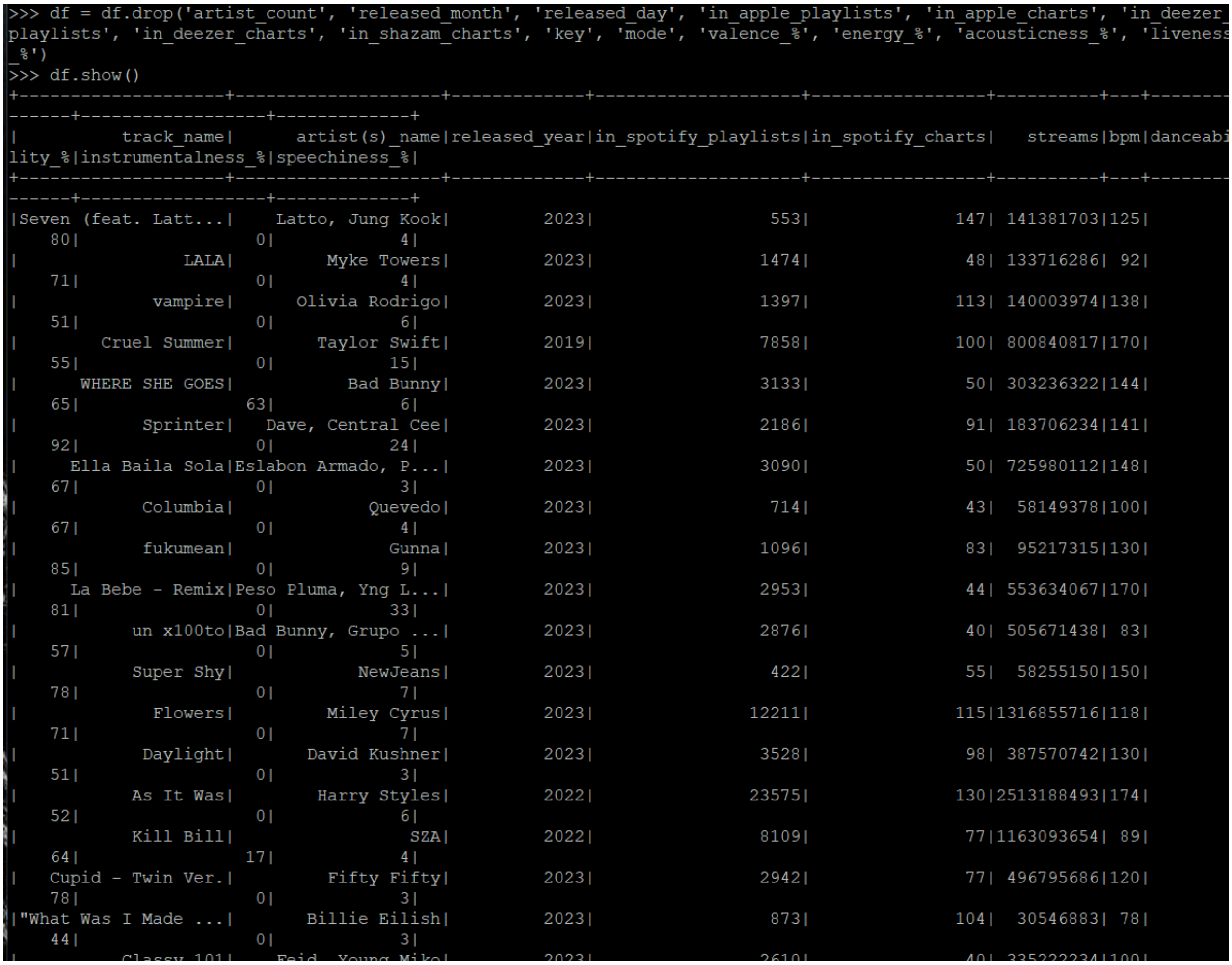
And here we can then see and verify that the data was loaded successfully and we're able to see and access that data that we're going to work with during this analysis.



```
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 3.0.0
      /_/

Using Python version 3.7.10 (default, Mar  2 2021 09:06:08)
SparkSession available as 'spark'.
>>> df = spark.read.format('csv').option('header', 'true').load('/data/music.csv')
>>> df.show()
+--------------------+--------------------+------------+-------------+--------------+------------+----------------
----+-----------------+----------+------------------+---------------+-------------------+----------------+--------
--------+---+----+-----+--------------+---------+--------+--------------+------------------+----------+-----------
--+
|          track_name|      artist(s)_name|artist_count|released_year|released_month|released_day|in_spotify_playl
ists|in_spotify_charts|   streams|in_apple_playlists|in_apple_charts|in_deezer_playlists|in_deezer_charts|in_shaza
m_charts|bpm| key| mode|danceability_%|valence_%|energy_%|acousticness_%|instrumentalness_%|liveness_%|speechiness
_%|
+--------------------+--------------------+------------+-------------+--------------+------------+----------------
----+-----------------+----------+------------------+---------------+-------------------+----------------+--------
--------+---+----+-----+--------------+---------+--------+--------------+------------------+----------+-----------
--+
|Seven (feat. Latt...|    Latto, Jung Kook|           2|         2023|             7|          14|
 553|              147| 141381703|                43|            263|                 45|              10|
     826|125|   B|Major|            80|       89|      83|            31|                 0|         8|
 4|
|                LALA|         Myke Towers|           1|         2023|             3|          23|
1474|               48| 133716286|                48|            126|                 58|              14|
     382| 92|  C#|Major|            71|       61|      74|             7|                 0|        10|
 4|
|             vampire|      Olivia Rodrigo|           1|         2023|             6|          30|
1397|              113| 140003974|                94|            207|                 91|              14|
     949|138|   F|Major|            51|       32|      53|            17|                 0|        31|
 6|
|        Cruel Summer|        Taylor Swift|           1|         2019|             8|          23|
7858|              100| 800840817|               116|            207|                125|              12|
     548|170|   A|Major|            55|       58|      72|            11|                 0|        11|
```

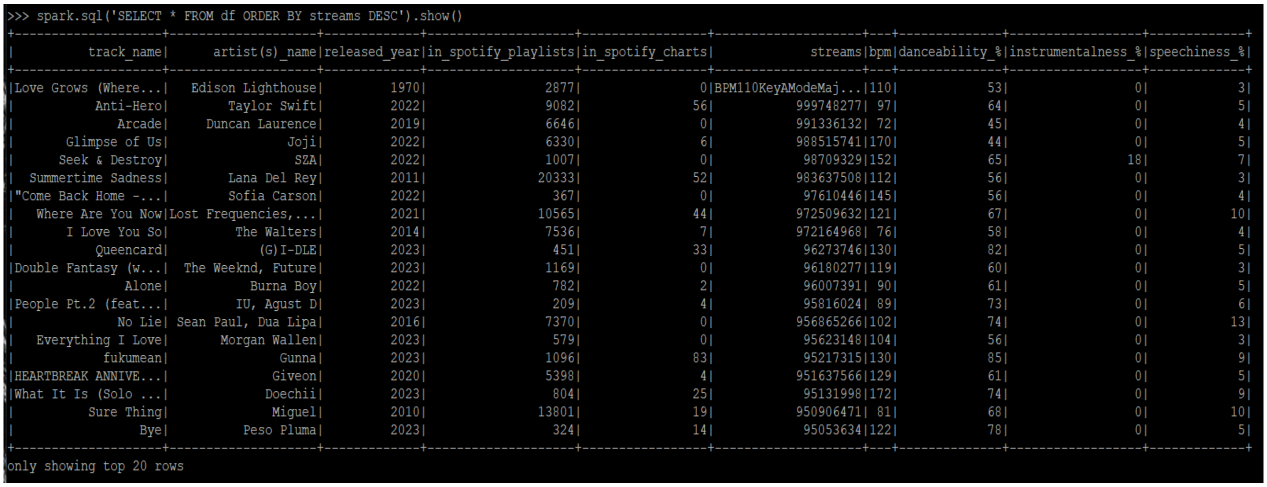
Once we had our data imported and verified we can now work on some cleaning and rearranging for our analysis. As far as cleaning went, we only had to get rid of one observation due to a missing value. One observation as missing values for some of the columns so we removed that observation.

Next since we're concerned with what music is popular, we identify our target variable as the number of streams a song has to tell us how popular it was. Once we had our target variable we had to choose what variables we will keep as our "predicting" variables. The following image shows the data ad column after only keeping the columns we believe to be important to the investigation.



```
>>> df = df.drop('artist_count', 'released_month', 'released_day', 'in_apple_playlists', 'in_apple_charts', 'in_deezer_
playlists', 'in_deezer_charts', 'in_shazam_charts', 'key', 'mode', 'valence_%', 'energy_%', 'acousticness_%', 'liveness
_%')
>>> df.show()
+--------------------+--------------------+-------------+--------------------+-----------------+----------+---+--------
------+------------------+-------------+
|          track_name|      artist(s)_name|released_year|in_spotify_playlists|in_spotify_charts|   streams|bpm|danceabi
lity_%|instrumentalness_%|speechiness_%|
+--------------------+--------------------+-------------+--------------------+-----------------+----------+---+--------
------+------------------+-------------+
|Seven (feat. Latt...|    Latto, Jung Kook|         2023|                 553|              147| 141381703|125|
    80|                 0|            4|
|                LALA|         Myke Towers|         2023|                1474|               48| 133716286| 92|
    71|                 0|            4|
|             vampire|      Olivia Rodrigo|         2023|                1397|              113| 140003974|138|
    51|                 0|            6|
|        Cruel Summer|        Taylor Swift|         2019|                7858|              100| 800840817|170|
    55|                 0|           15|
|      WHERE SHE GOES|           Bad Bunny|         2023|                3133|               50| 303236322|144|
    65|                63|            6|
|            Sprinter|   Dave, Central Cee|         2023|                2186|               91| 183706234|141|
    92|                 0|           24|
|     Ella Baila Sola|Eslabon Armado, P...|         2023|                3090|               50| 725980112|148|
    67|                 0|            3|
|            Columbia|             Quevedo|         2023|                 714|               43|  58149378|100|
    67|                 0|            4|
|            fukumean|               Gunna|         2023|                1096|               83|  95217315|130|
    85|                 0|            9|
|     La Bebe - Remix|Peso Pluma, Yng L...|         2023|                2953|               44| 553634067|170|
    81|                 0|           33|
|           un x100to|Bad Bunny, Grupo ...|         2023|                2876|               40| 505671438| 83|
    57|                 0|            5|
|           Super Shy|            NewJeans|         2023|                 422|               55|  58255150|150|
    78|                 0|            7|
|             Flowers|         Miley Cyrus|         2023|               12211|              115|1316855716|118|
    71|                 0|            7|
|            Daylight|       David Kushner|         2023|                3528|               98| 387570742|130|
    51|                 0|            3|
|           As It Was|        Harry Styles|         2022|               23575|              130|2513188493|174|
    52|                 0|            6|
|           Kill Bill|                 SZA|         2022|                8109|               77|1163093654| 89|
    64|                17|            4|
|   Cupid - Twin Ver.|         Fifty Fifty|         2023|                2942|               77| 496795686|120|
    78|                 0|            3|
|"What Was I Made ...|       Billie Eilish|         2023|                 873|              104|  30546883| 78|
    44|                 0|            3|
|          Classy 101|    Feid, Young Miko|         2023|                2610|               40| 335222234|100|
```

Now, our data is ready for investigation. We first dive into an investigation of the top streamed songs of 2023. The following image shows those results.

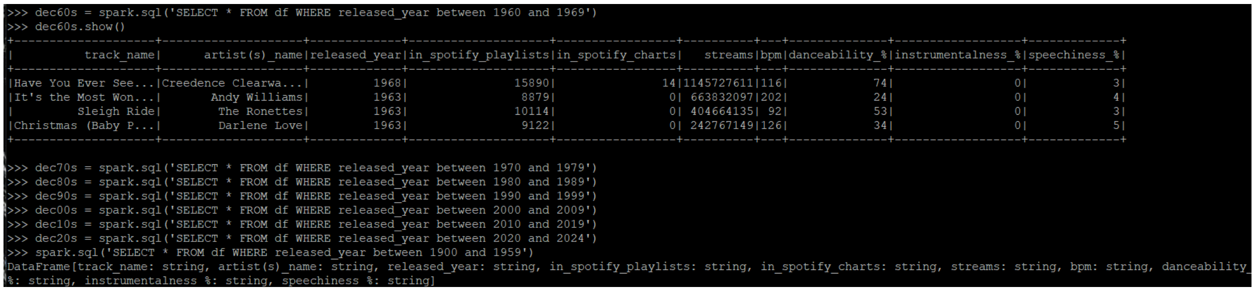


```
>>> spark.sql('SELECT * FROM df ORDER BY streams DESC').show()
+--------------------+--------------------+-------------+--------------------+-----------------+--------------------+---+--------------+------------------+-------------+
|          track_name|      artist(s)_name|released_year|in_spotify_playlists|in_spotify_charts|             streams|bpm|danceability_%|instrumentalness_%|speechiness_%|
+--------------------+--------------------+-------------+--------------------+-----------------+--------------------+---+--------------+------------------+-------------+
|Love Grows (Where...|   Edison Lighthouse|         1970|                2877|                0|BPM110KeyAModeMaj...|110|            53|                 0|            3|
|           Anti-Hero|        Taylor Swift|         2022|                9082|               56|           999748277| 97|            64|                 0|            5|
|              Arcade|     Duncan Laurence|         2019|                6646|                0|           991336132| 72|            45|                 0|            4|
|       Glimpse of Us|                Joji|         2022|                6330|                6|           988515741|170|            44|                 0|            5|
|      Seek & Destroy|                 SZA|         2022|                1007|                0|            98709329|152|            65|                18|            7|
|  Summertime Sadness|        Lana Del Rey|         2011|               20333|               52|           983637508|112|            56|                 0|            3|
|"Come Back Home -...|        Sofia Carson|         2022|                 367|                0|            97610446|145|            56|                 0|            4|
|   Where Are You Now|Lost Frequencies,...|         2021|               10565|               44|           972509632|121|            67|                 0|           10|
|       I Love You So|         The Walters|         2014|                7536|                7|           972164968| 76|            58|                 0|            4|
|           Queencard|            (G)I-DLE|         2023|                 451|               33|            96273746|130|            82|                 0|            5|
|Double Fantasy (w...|  The Weeknd, Future|         2023|                1169|                0|            96180277|119|            60|                 0|            3|
|               Alone|           Burna Boy|         2022|                 782|                2|            96007391| 90|            61|                 0|            5|
|People Pt.2 (feat...|         IU, Agust D|         2023|                 209|                4|            95816024| 89|            73|                 0|            6|
|              No Lie| Sean Paul, Dua Lipa|         2016|                7370|                0|           956865266|102|            74|                 0|           13|
|   Everything I Love|       Morgan Wallen|         2023|                 579|                0|            95623148|104|            56|                 0|            3|
|            fukumean|               Gunna|         2023|                1096|               83|            95217315|130|            85|                 0|            9|
|HEARTBREAK ANNIVE...|              Giveon|         2020|                5398|                4|           951637566|129|            61|                 0|            5|
|What It Is (Solo ...|             Doechii|         2023|                 804|               25|            95131998|172|            74|                 0|            9|
|          Sure Thing|              Miguel|         2010|               13801|               19|           950906471| 81|            68|                 0|           10|
|                 Bye|          Peso Pluma|         2023|                 324|               14|            95053634|122|            78|                 0|            5|
+--------------------+--------------------+-------------+--------------------+-----------------+--------------------+---+--------------+------------------+-------------+
only showing top 20 rows
```

As you can see, our results show that Anti-Hero by Taylor swift is the most streamed song of 2023, followed by Arcade by Duncan Laurence, Glimpse of Us by Joji, Seek and Destroy by SZA, and Summertime Sadness by Lana Del Rey as the top 5 steamed songs of 2023. Love Grows by Edison Lighthouse shows up at the top because we have missing values for the amount of streams as well as other solumns for that observation, ultimately why we ended up removing this observation from the dataset.

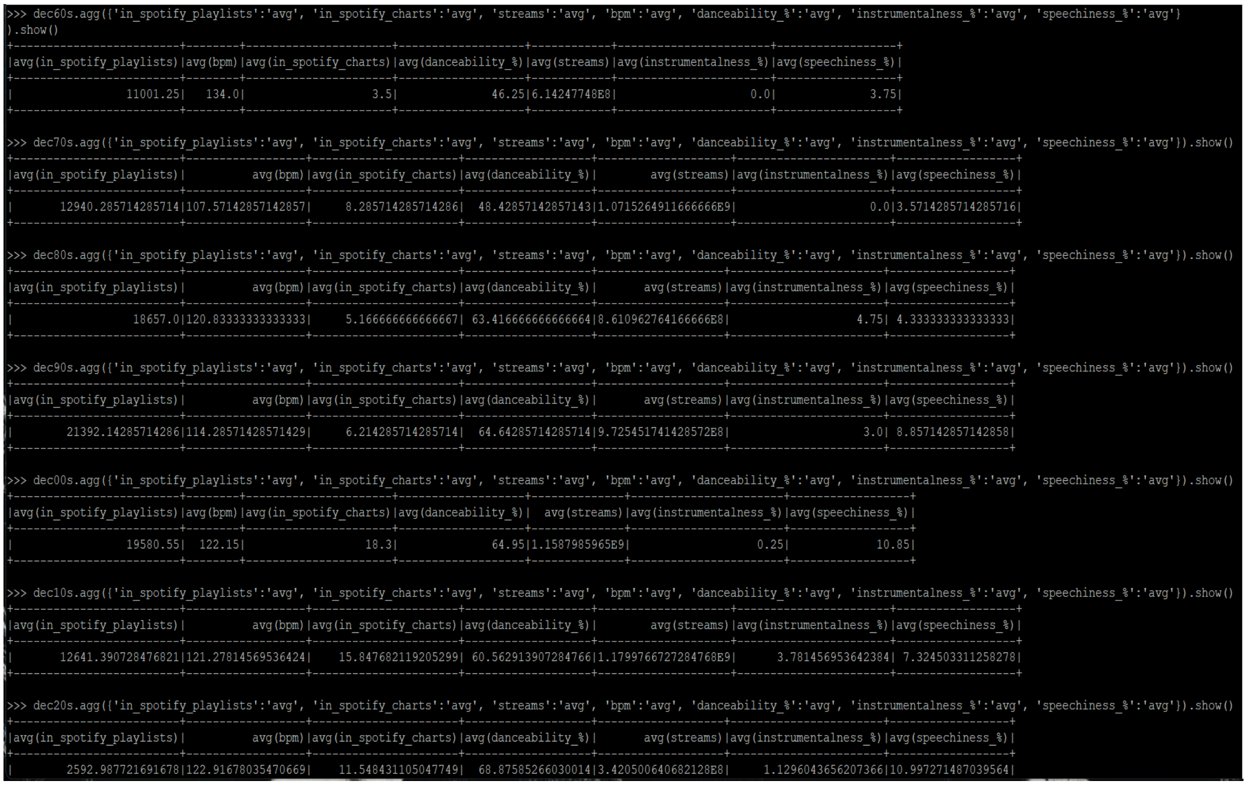
While there is no genre column within the data to let us know what type of music is played, for one, that information can always be investigated after our analysis is done to help further analyze our results but also we have other columns to tell us information about the most streamed songs. For example, we know that the dancability percentage of those top songs was around the 40%-60% range, the instrumentalness percentage was typically 0% except for SZA's song and speechiness was always below 7% for these most streamed songs of 2023.

After this investigation, we also want to determine if time period has an impact in popularity of a song. Of course the more recent a song was released the more likely it is to be heard in that moment ratherr than music that come out even just 10 years ago let alone something like 40 or 50 years ago. But some music is timeless and if some of those old songs are popular an investigation can also be done on what type of song characteristics were popular in that time period and that can possible shed some light as to which of those vintage song features still hold today. The image below shows how the data was individually divided by decades and what those subsetted datasets might look like. It also shows that we had no data prior to the 1960s in this dataset.



```
>>> dec60s = spark.sql('SELECT * FROM df WHERE released_year between 1960 and 1969')
>>> dec60s.show()
+--------------------+--------------------+-------------+--------------------+-----------------+----------+---+--------------+------------------+-------------+
|          track_name|      artist(s)_name|released_year|in_spotify_playlists|in_spotify_charts|   streams|bpm|danceability_%|instrumentalness_%|speechiness_%|
+--------------------+--------------------+-------------+--------------------+-----------------+----------+---+--------------+------------------+-------------+
|Have You Ever See...|Creedence Clearwa...|         1968|               15890|               14|1145727611|116|            74|                 0|            3|
|It's the Most Won...|       Andy Williams|         1963|                8879|                0| 663832097|202|            24|                 0|            4|
|         Sleigh Ride|        The Ronettes|         1963|               10114|                0| 404664135| 92|            53|                 0|            3|
|Christmas (Baby P...|        Darlene Love|         1963|                9122|                0| 242767149|126|            34|                 0|            5|
+--------------------+--------------------+-------------+--------------------+-----------------+----------+---+--------------+------------------+-------------+

>>> dec70s = spark.sql('SELECT * FROM df WHERE released_year between 1970 and 1979')
>>> dec80s = spark.sql('SELECT * FROM df WHERE released_year between 1980 and 1989')
>>> dec90s = spark.sql('SELECT * FROM df WHERE released_year between 1990 and 1999')
>>> dec00s = spark.sql('SELECT * FROM df WHERE released_year between 2000 and 2009')
>>> dec10s = spark.sql('SELECT * FROM df WHERE released_year between 2010 and 2019')
>>> dec20s = spark.sql('SELECT * FROM df WHERE released_year between 2020 and 2024')
>>> spark.sql('SELECT * FROM df WHERE released_year between 1900 and 1959')
DataFrame[track_name: string, artist(s)_name: string, released_year: string, in_spotify_playlists: string, in_spotify_charts: string, streams: string, bpm: string, danceability_
%: string, instrumentalness_%: string, speechiness_%: string]
```

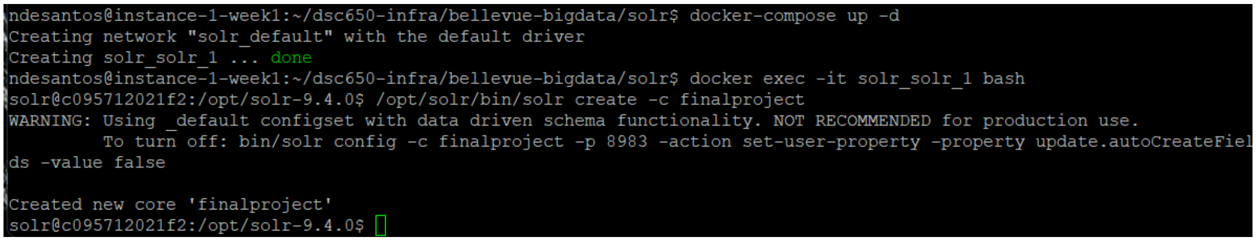
Once we separated the data, a small statistical summary was conducted for each decade including the averages for all the columns in order to be able to see if there are any major changes between decades as well as any patterns we might be able to find in the results. The following image shows the results of this investigation.



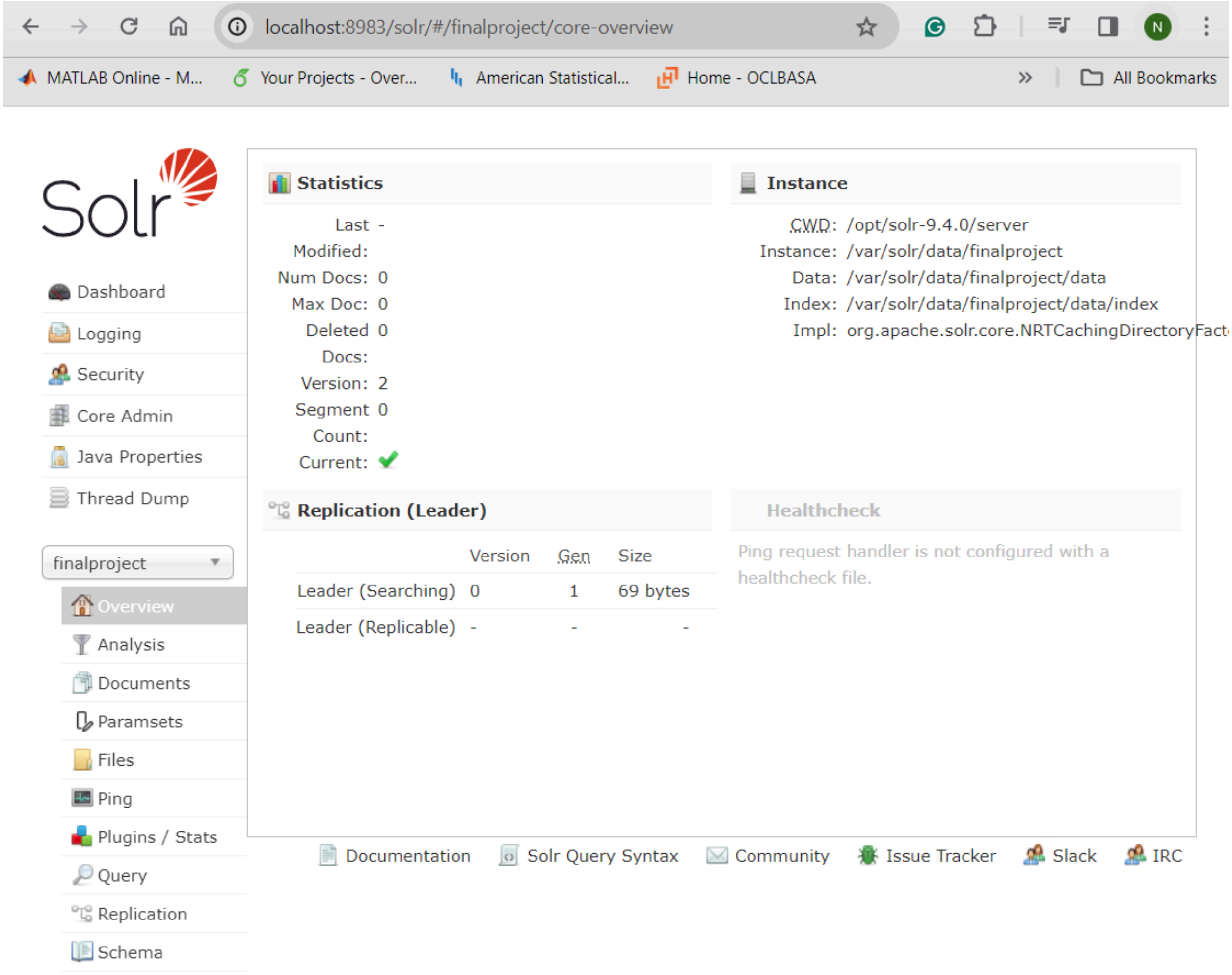
```
>>> dec60s.agg({'in_spotify_playlists':'avg', 'in_spotify_charts':'avg', 'streams':'avg', 'bpm':'avg', 'danceability_%':'avg', 'instrumentalness_%':'avg', 'speechiness_%':'avg'}
).show()
+-------------------------+--------+----------------------+-------------------+------------+-----------------------+------------------+
|avg(in_spotify_playlists)|avg(bpm)|avg(in_spotify_charts)|avg(danceability_%)|avg(streams)|avg(instrumentalness_%)|avg(speechiness_%)|
+-------------------------+--------+----------------------+-------------------+------------+-----------------------+------------------+
|                 11001.25|   134.0|                   3.5|              46.25|6.14247748E8|                    0.0|              3.75|
+-------------------------+--------+----------------------+-------------------+------------+-----------------------+------------------+

>>> dec70s.agg({'in_spotify_playlists':'avg', 'in_spotify_charts':'avg', 'streams':'avg', 'bpm':'avg', 'danceability_%':'avg', 'instrumentalness_%':'avg', 'speechiness_%':'avg'}).show()
+-------------------------+------------------+----------------------+-------------------+--------------------+-----------------------+------------------+
|avg(in_spotify_playlists)|          avg(bpm)|avg(in_spotify_charts)|avg(danceability_%)|        avg(streams)|avg(instrumentalness_%)|avg(speechiness_%)|
+-------------------------+------------------+----------------------+-------------------+--------------------+-----------------------+------------------+
|       12940.285714285714|107.57142857142857|     8.285714285714286|  48.42857142857143|1.0715264911666666E9|                    0.0|3.5714285714285716|
+-------------------------+------------------+----------------------+-------------------+--------------------+-----------------------+------------------+

>>> dec80s.agg({'in_spotify_playlists':'avg', 'in_spotify_charts':'avg', 'streams':'avg', 'bpm':'avg', 'danceability_%':'avg', 'instrumentalness_%':'avg', 'speechiness_%':'avg'}).show()
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+
|avg(in_spotify_playlists)|          avg(bpm)|avg(in_spotify_charts)|avg(danceability_%)|       avg(streams)|avg(instrumentalness_%)|avg(speechiness_%)|
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+
|                  18657.0|120.83333333333333|     5.166666666666667| 63.416666666666664|8.610962764166666E8|                   4.75| 4.333333333333333|
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+

>>> dec90s.agg({'in_spotify_playlists':'avg', 'in_spotify_charts':'avg', 'streams':'avg', 'bpm':'avg', 'danceability_%':'avg', 'instrumentalness_%':'avg', 'speechiness_%':'avg'}).show()
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+
|avg(in_spotify_playlists)|          avg(bpm)|avg(in_spotify_charts)|avg(danceability_%)|       avg(streams)|avg(instrumentalness_%)|avg(speechiness_%)|
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+
|        21392.14285714286|114.28571428571429|     6.214285714285714|  64.64285714285714|9.725451741428572E8|                    3.0| 8.857142857142858|
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+

>>> dec00s.agg({'in_spotify_playlists':'avg', 'in_spotify_charts':'avg', 'streams':'avg', 'bpm':'avg', 'danceability_%':'avg', 'instrumentalness_%':'avg', 'speechiness_%':'avg'}).show()
+-------------------------+--------+----------------------+-------------------+--------------+-----------------------+------------------+
|avg(in_spotify_playlists)|avg(bpm)|avg(in_spotify_charts)|avg(danceability_%)|  avg(streams)|avg(instrumentalness_%)|avg(speechiness_%)|
+-------------------------+--------+----------------------+-------------------+--------------+-----------------------+------------------+
|                 19580.55|  122.15|                  18.3|              64.95|1.1587985965E9|                   0.25|             10.85|
+-------------------------+--------+----------------------+-------------------+--------------+-----------------------+------------------+

>>> dec10s.agg({'in_spotify_playlists':'avg', 'in_spotify_charts':'avg', 'streams':'avg', 'bpm':'avg', 'danceability_%':'avg', 'instrumentalness_%':'avg', 'speechiness_%':'avg'}).show()
+-------------------------+------------------+----------------------+-------------------+--------------------+-----------------------+------------------+
|avg(in_spotify_playlists)|          avg(bpm)|avg(in_spotify_charts)|avg(danceability_%)|        avg(streams)|avg(instrumentalness_%)|avg(speechiness_%)|
+-------------------------+------------------+----------------------+-------------------+--------------------+-----------------------+------------------+
|       12641.390728476821|121.27814569536424|    15.847682119205299| 60.562913907284766|1.1799766727284768E9|      3.781456953642384| 7.324503311258278|
+-------------------------+------------------+----------------------+-------------------+--------------------+-----------------------+------------------+

>>> dec20s.agg({'in_spotify_playlists':'avg', 'in_spotify_charts':'avg', 'streams':'avg', 'bpm':'avg', 'danceability_%':'avg', 'instrumentalness_%':'avg', 'speechiness_%':'avg'}).show()
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+
|avg(in_spotify_playlists)|          avg(bpm)|avg(in_spotify_charts)|avg(danceability_%)|       avg(streams)|avg(instrumentalness_%)|avg(speechiness_%)|
+-------------------------+------------------+----------------------+-------------------+-------------------+-----------------------+------------------+
|        2592.987721691678|122.91678035470669|    11.548431105047749|  68.87585266030014|3.420500640682128E8|     1.1296043656207366|10.997271487039564|
```

From these results we know that, based on streams, the 90s were the most popular decade. Songs from the 90s were also within the most spotify playlists as well. The 90s also had a smaller average BPM for their music to we might consider that characteristic of a slower tempo as something that might be a characteristic of more popular music. This observation can also be confirmed by looking back at the top 20 streamed songs (excluding Love Grows), the majority of the top 5 streamed songs have a bpm of 114 of lower. Another trend we see is that around the 80s we see a huge increase in the dancability of music which stayed pretty consistent in following decades. We see the same spike with the instrumentalness percentage of the song. There is basically no sintrumentalness percentage in the 60s and 70s. And as for speechiness percentage, the precent of speech in music appears to increase as time progresses. With each decade the average speechiness percentage went up every time with the exception of the 2010s. But then it quickly went back up in the following decade in the 2020s.

The next portion of our investigation was done using Solr's document searching capabilities. The investigation focuses on the text data within our data set. Before we get into the data, we first create a new collection within our Solr interface to be able to host our data and work with it. The following images show the creation of a new solr collection as well as the confirmation that Solr is connected to our environment by showing our collection on their web interface.



```
ndesantos@instance-1-week1:~/dsc650-infra/bellevue-bigdata/solr$ docker-compose up -d
Creating network "solr_default" with the default driver
Creating solr_solr_1 ... done
ndesantos@instance-1-week1:~/dsc650-infra/bellevue-bigdata/solr$ docker exec -it solr_solr_1 bash
solr@c095712021f2:/opt/solr-9.4.0$ /opt/solr/bin/solr create -c finalproject
WARNING: Using _default configset with data driven schema functionality. NOT RECOMMENDED for production use.
         To turn off: bin/solr config -c finalproject -p 8983 -action set-user-property -property update.autoCreateFiel
ds -value false

Created new core 'finalproject'
solr@c095712021f2:/opt/solr-9.4.0$ 
```



**Statistics**

Last Modified: -
Num Docs: 0
Max Doc: 0
Deleted Docs: 0
Version: 2
Segment Count: 0
Current: ✓

**Instance**

CWD: /opt/solr-9.4.0/server
Instance: /var/solr/data/finalproject
Data: /var/solr/data/finalproject/data
Index: /var/solr/data/finalproject/data/index
Impl: org.apache.solr.core.NRTCachingDirectoryFact

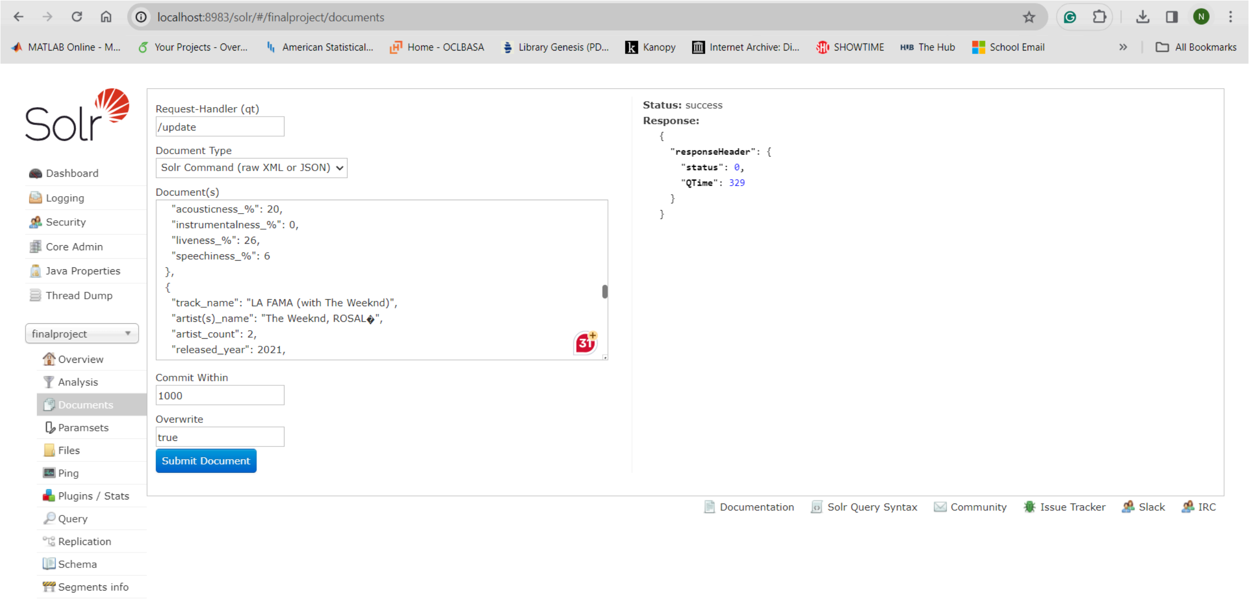
**Replication (Leader)**

| | Version | Gen | Size |
|---|---|---|---|
| Leader (Searching) | 0 | 1 | 69 bytes |
| Leader (Replicable) | - | - | - |

**Healthcheck**

Ping request handler is not configured with a healthcheck file.

Once we have that set up we can move on adding our data to this environment. The original data was in CSV format but we converted it to a JSON format for the sake of working with Solr's default settings.



Request-Handler (qt): /update

Document Type: Solr Command (raw XML or JSON)

Document(s)

```
  "acousticness_%": 20,
  "instrumentalness_%": 0,
  "liveness_%": 26,
  "speechiness_%": 6
 },
 {
  "track_name": "LA FAMA (with The Weeknd)",
  "artist(s)_name": "The Weeknd, ROSAL�",
  "artist_count": 2,
  "released_year": 2021,
```

Commit Within: 1000

Overwrite: true

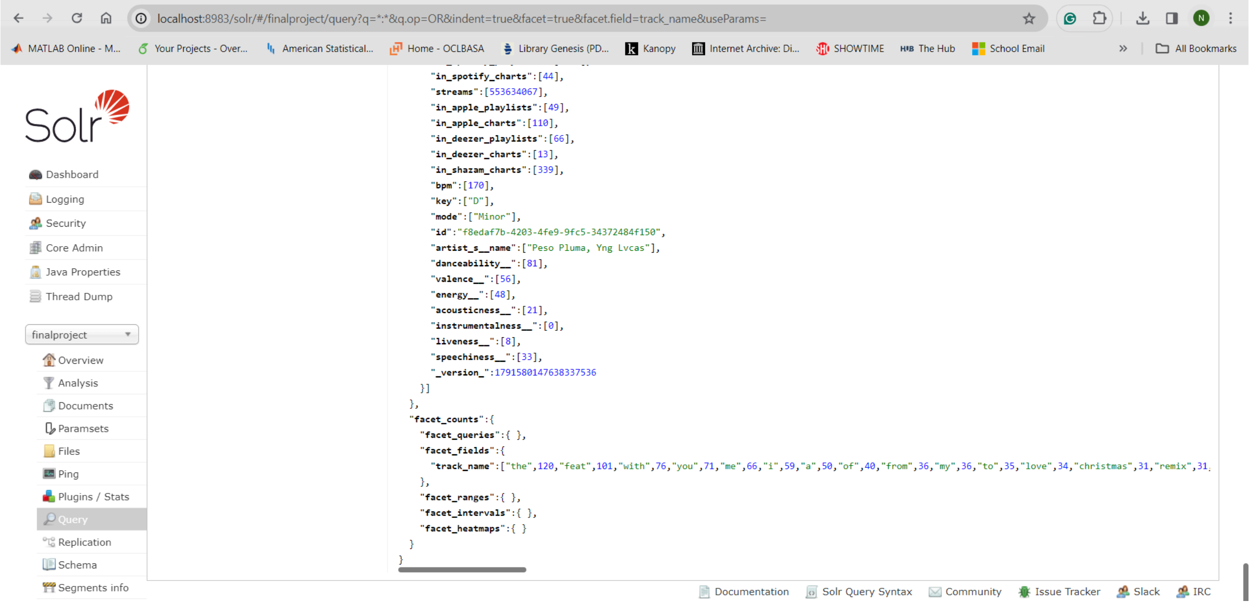
Submit Document

Status: success

Response:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 329
  }
}
```

Once we converted our data to the correct format, our goal was to first look at the song title data and look for what kind of key words are most popular in the title if any. This is an effort to try and see what kind of topics are most popular in songs.

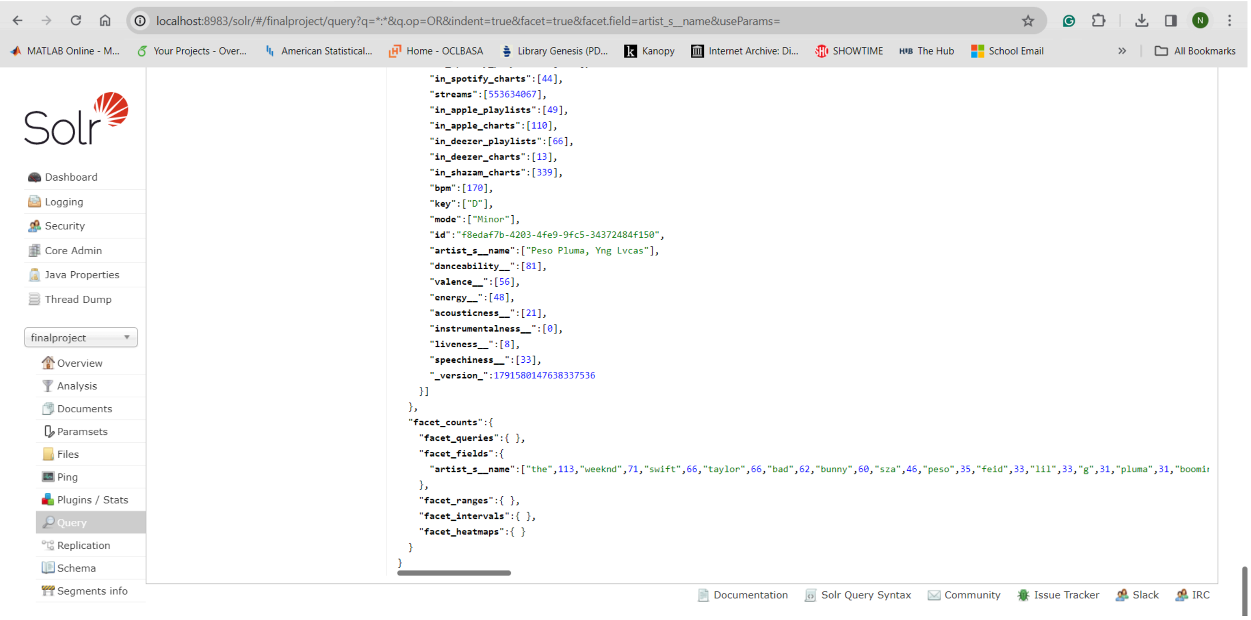


localhost:8983/solr/#/finalproject/query?q=*:*&q.op=OR&indent=true&facet=true&facet.field=track_name&useParams=

```
      "in_spotify_charts":[44],
      "streams":[553634067],
      "in_apple_playlists":[49],
      "in_apple_charts":[110],
      "in_deezer_playlists":[66],
      "in_deezer_charts":[13],
      "in_shazam_charts":[339],
      "bpm":[170],
      "key":["D"],
      "mode":["Minor"],
      "id":"f8edaf7b-4203-4fe9-9fc5-34372484f150",
      "artist_s__name":["Peso Pluma, Yng Lvcas"],
      "danceability__":[81],
      "valence__":[56],
      "energy__":[48],
      "acousticness__":[21],
      "instrumentalness__":[0],
      "liveness__":[8],
      "speechiness__":[33],
      "_version_":1791580147638337536
    }]
  },
  "facet_counts":{
    "facet_queries":{ },
    "facet_fields":{
      "track_name":["the",120,"feat",101,"with",76,"you",71,"me",66,"i",59,"a",50,"of",40,"from",36,"my",36,"to",35,"love",34,"christmas",31,"remix",31,
    },
    "facet_ranges":{ },
    "facet_intervals":{ },
    "facet_heatmaps":{ }
  }
}
```

After looking at those results of our faceted query, and after disregarding coming stop words we see that songs with titles that contain words like “With”, “You”, “me” “love” and “christmas” all appear in more than 30 songs within the dataset. Over a hundred songs contain the word “feat” meaning a lot of the most streams songs of 2023 were some kind of collaboration which might show some direction for aspiring music artists that want to become more popular.

We then conducted the same type of investigation for artists that were included in this dataset. The following image shows that query as well as the results.



```
localhost:8983/solr/#/finalproject/query?q=*:*&q.op=OR&indent=true&facet=true&facet.field=artist_s__name&useParams=
```

```
      "in_spotify_charts":[44],
      "streams":[553634067],
      "in_apple_playlists":[49],
      "in_apple_charts":[110],
      "in_deezer_playlists":[66],
      "in_deezer_charts":[13],
      "in_shazam_charts":[339],
      "bpm":[170],
      "key":["D"],
      "mode":["Minor"],
      "id":"f8edaf7b-4203-4fe9-9fc5-34372484f150",
      "artist_s__name":["Peso Pluma, Yng Lvcas"],
      "danceability__":[81],
      "valence__":[56],
      "energy__":[48],
      "acousticness__":[21],
      "instrumentalness__":[0],
      "liveness__":[8],
      "speechiness__":[33],
      "_version_":1791580147638337536
    }]
  },
  "facet_counts":{
    "facet_queries":{ },
    "facet_fields":{
      "artist_s__name":["the",113,"weeknd",71,"swift",66,"taylor",66,"bad",62,"bunny",60,"sza",46,"peso",35,"feid",33,"lil",33,"g",31,"pluma",31,"boomir
    },
    "facet_ranges":{ },
    "facet_intervals":{ },
    "facet_heatmaps":{ }
  }
}
```

The results for artists are a little more straight forward since artist names are a little more recognizable than sone titles. From the results it’s easy to see people like The Weeknd, Taylor Swift, Bad BUnny, SZA, Peso Pluma, Metro Boomin, Drake and Kendrick Lamar as the top streamed artists of 2023. These results can provide frame work on which of those artists’ song characteristics make them so popular.

Our investigation led us through different styles of music from different artists from different decades. Our goal was to investigate what characteristics of different song make them more popular among people. We looked at recent spotify data from 2023 in order to analyze music trends and found that the most popular songs tended to have a dancability percentage of around the 40%-60% range, the instrumentalness percentage was typically 0% except for SZA's song and speechiness was always below 7% for these most streamed songs of 2023. We also did an individual investigation of different decades of music and found that songs from the 90s were the most popular in 2023 and one major characteristics of those songs is a relatively lower tempo (BPM). Further investigation into text data showed us that these most streamed songs of 2023 had "most common" words such as "with", "you", "me" "love" and "christmas" telling us that the most popular songs as far as streaming goes tend to be about love or some kind of relationship or christmas. We also saw that "feat" was another really common word appearing in song titles showing the popularity of collaborations. Lastly, we also found that the most streamed artists of 2023 included The Weeknd, Taylor Swift, Bad BUnny, SZA, Peso Pluma, Metro Boomin, Drake and Kendrick Lamar.