

## Imports

```
In [ ]: # %pip install spotipy
# %pip install python-Levenshtein
# %pip install fuzzywuzzy
import pandas as pd
import numpy as np
import spotipy.oauth2 as oauth2
import spotipy
from spotipy.oauth2 import SpotifyOAuth
from spotipy.oauth2 import SpotifyClientCredentials
from sklearn.preprocessing import MultiLabelBinarizer
import time
import fuzzywuzzy
```

## Spotipy Credentials

```
In [ ]: auth_manager = SpotifyClientCredentials(client_id = '7606efb22a6848bdbf48ecd67484b9d5',
client_secret = '5febce52b67b499db9f7295c5f6839'
sp = spotipy.Spotify(auth_manager=auth_manager)
```

The following cell loads in the grammy award winning and selects certain columns. It also drops duplicates and removes hyphens.

```
In [ ]: # Loadin and group grammy awards
grammyAwards = pd.read_csv("data/grammySongs_1999-2019.csv")
grammyAwards = grammyAwards[['Name', 'Artist', 'GrammyAward', 'GrammyYear', 'Genre']]
grammyAwards = grammyAwards.groupby(['Name', 'Artist', 'GrammyYear', 'Genre']).count().
grammyAwards = grammyAwards.drop_duplicates(subset=['Name', 'Artist'], keep='last').res
grammyAwards['Artist'] = grammyAwards['Artist'].apply(lambda s: s.replace("-", " "))
grammyAwards = grammyAwards.applymap(lambda s: s.upper() if type(s) == str else s)

grammyAwards.head(5)
```

```
Out[ ]:
```

	Name	Artist	GrammyYear	Genre	GrammyAward
0	'TIL SUMMER COMES AROUND	KEITH URBAN	2010	COUNTRY	1
1	24 FRAMES	ASON ISBELL, SONGWRITER	2015	AMERICAN ROOTS MUSIC	1
2	24K MAGIC	BRUNO MARS	2017	GENERAL	1
3	500 MILES HIGH	CHICK COREA, SOLOIST	2011	JAZZ	1
4	99 PROBLEMS	JAY Z	2004	RAP	1

The following code fixes the artist names by querying the spotipy API, or by performing some string splitting.

```
In [ ]: for index, row in grammyAwards.iterrows():
artistName = row['Artist']
trackName = row['Name']
year = row['GrammyYear']
```

```

try:
    artistList = artistName.split(",")
    artistName = artistList[0]
    artistName = artistName.split("&")[0]
    artistName = artistName.split(":")[0]
    artistName = artistName.split(";")[0]
    artistName = artistName.split(":")[0]
    artistName = artistName.replace("-", " ")
    if 'FEATURING' in artistName:
        grammyAwards.at[index, 'Artist'] = artistName.split('FEATURING')[0]
        continue
    if 'SOLOIST' in row['Artist']:
        grammyAwards.at[index, 'Artist'] = artistName.split(",")[0]
        continue
    if (artistName != row['Artist']):
        song = sp.search(q=f"track:{trackName}, year:{year}", type="track", limit=2)
        if len(song['tracks']['items']) == 0:
            song = sp.search(q=f"track:{trackName}", type="track", limit=2)
        artist = song['tracks']['items'][0]['artists'][0]['name'].upper()
        grammyAwards.at[index, 'Artist'] = artist
except:
    print("track not found", trackName)
grammyAwards.head(5)

```

track not found A FEATHER'S NOT A BIRD

```
Out[ ]:
```

	Name	Artist	GrammyYear	Genre	GrammyAward
0	'TIL SUMMER COMES AROUND	KEITH URBAN	2010	COUNTRY	1
1	24 FRAMES	JASON ISBELL	2015	AMERICAN ROOTS MUSIC	1
2	24K MAGIC	BRUNO MARS	2017	GENERAL	1
3	500 MILES HIGH	CHICK COREA	2011	JAZZ	1
4	99 PROBLEMS	JAY Z	2004	RAP	1

The following code loads in and cleans the billboard data, fills in null values, finds the maximum weeks on chart, top position, average weekly rank.

```

In [ ]: # Loading and cleaning billboard
billboardHot100 = pd.read_csv('data/billboardHot100_1999-2019.csv')
billboardHot100 = billboardHot100[['Artists', 'Name', 'Genre', 'Peak.position', 'Weeks.
billboardHot100['Weeks.on.chart'] = billboardHot100['Weeks.on.chart'].fillna(1)
billboardHot100['Peak.position'] = billboardHot100['Peak.position'].fillna(100)
billboardHot100['Weekly.rank'] = billboardHot100['Weekly.rank'].fillna(100)
billboardHot100['Genre'] = billboardHot100['Genre'].fillna('None')
billboardHot100['Date'] = billboardHot100['Date'].fillna('1999')
billboardHot100 = billboardHot100.groupby(['Artists', 'Name', 'Genre']).agg({'Peak.posit
billboardHot100.rename(columns = {'Weekly.rank': 'Avg.rank', 'Artists': 'Artist'}, inplace=
billboardHot100['Date'] = pd.to_datetime(billboardHot100['Date'])
billboardHot100['Date'] = billboardHot100['Date'].apply(lambda x : x.year)
billboardHot100 = billboardHot100.applymap(lambda s: s.upper() if type(s) == str else s
billboardHot100.head()

```

The following code gets song statistics for billboard songs through the spotipy api, saves it in a CSV

```
In [ ]: for index, row in billboardHot100.iterrows():
        artistName = row['Artist']
        trackName = row['Name']
        year = row['Date']
        try:
            song = sp.search(q=f"track:{trackName}, year:{year}", type="track", limit=2)
            if len(song['tracks']['items']) == 0:
                song = sp.search(q=f"track:{trackName}", type="track", limit=2)
            track_id = song['tracks']['items'][0]['uri']
            song_info = sp.audio_features(track_id)
            billboardHot100.at[index, 'Energy'] = song_info[0].get('energy')
            billboardHot100.at[index, 'Danceability'] = song_info[0].get('danceability')
            billboardHot100.at[index, 'Tempo'] = song_info[0].get('tempo')
            billboardHot100.at[index, 'Valence'] = song_info[0].get('valence')
        except:
            print("track not found", trackName)
        if (index % 500 == 0):
            print(index)
    # billboardHot100.to_csv("billboard_features.csv")
```

Load in the billboard csv, because the above cell does a large number of queries that we don't want to repeat. It also creates a column notating if the billboard song won a grammy or not.

```
In [ ]: billboardHot100 = pd.read_csv('billboard_features.csv')
        billboardHot100['GrammyAward'] = np.zeros((len(billboardHot100)), dtype=int)
        billboardHot100.head(5)
```

Out [ ]:

	Unnamed: 0	Artist	Name	Genre	Peak.position	Weeks.on.chart	A
0	0	"WEIRD AL" YANKOVIC	CANADIAN IDIOT	COMEDY,PARODY,ROCK	82.0	3.0	91
1	1	"WEIRD AL" YANKOVIC	WORD CRIMES	PARODY,COMEDY,POP	39.0	2.0	65
2	2	'N SYNC	BYE BYE BYE	ADULT CONTEMPORARY,DANCE- POP,TEEN POP,BOY BAND...	4.0	23.0	14
3	3	'N SYNC	GONE	ADULT CONTEMPORARY,ACOUSTIC,SOUL POP,R&B,BOY ...	11.0	24.0	27
4	4	'N SYNC	IT'S GONNA BE ME	MEMES,DANCE-POP,TEEN POP,BOY BAND,POP	1.0	25.0	21

The following code merges the grammyAwards table and the billboardHot100 tables with a fuzzy merge. The naming convention for the artists is slightly different between the two tables, so we are forced to use levenstein distance to compute similar rows. This is a slow process and requires a lot

of computation, but we could not figure out a better way. Saves results in a csv. This function also updates the artists of the grammy csv to match the billboard. Also, this sets grammy winners as winners in the billboard csv.

```
In [ ]: # code for merging

'''
for each one, search in the billboard for the artistName, true if satisfies fuzzy and
'''

from fuzzywuzzy import fuzz
for index, row in grammyAwards.iterrows():
    if index % 50 == 0:
        print(index)
    artist1 = row['Artist']
    track = row['Name']
    found = False
    for id2, r2 in billboardHot100.iterrows():
        artist2 = r2['Artist']
        t2 = r2['Name']
        artistRatio = fuzz.partial_ratio(artist1, artist2)
        trackRatio = fuzz.partial_ratio(track, t2)
        if trackRatio > 90 and artistRatio > 90:
            print(artist1, artist2)
            print(track, t2)
            found = True
            grammyAwards.at[index, 'Artist'] = artist2
            grammyAwards.at[index, 'Energy'] = r2['Energy']
            grammyAwards.at[index, 'Danceability'] = r2['Danceability']
            grammyAwards.at[index, 'Tempo'] = r2['Tempo']
            grammyAwards.at[index, 'Valence'] = r2['Valence']
            grammyAwards.at[index, 'Peak.position'] = r2['Peak.position']
            grammyAwards.at[index, 'Weeks.on.chart'] = r2['Weeks.on.chart']
            grammyAwards.at[index, 'Avg.rank'] = r2['Avg.rank']
            billboardHot100.at[id2, 'GrammyAward'] = 1
            break
    if not found:
        grammyAwards.at[index, 'Energy'] = -1
        grammyAwards.at[index, 'Danceability'] = -1
        grammyAwards.at[index, 'Tempo'] = -1
        grammyAwards.at[index, 'Valence'] = -1
        grammyAwards.at[index, 'Peak.position'] = -1
        grammyAwards.at[index, 'Weeks.on.chart'] = -1
        grammyAwards.at[index, 'Avg.rank'] = -1
        print('no match in top 100', track, artist1)
grammyAwards.head(10)
grammyAwards.to_csv("grammy_award_data2.csv")
```

```
0
no match in top 100 'TIL SUMMER COMES AROUND KEITH URBAN
no match in top 100 24 FRAMES JASON ISBELL
BRUNO MARS BRUNO MARS
24K MAGIC 24K MAGIC
no match in top 100 500 MILES HIGH CHICK COREA
JAY Z JAY
99 PROBLEMS 99 PROBLEMS
MIGUEL MIGUEL
ADORN ADORN
LENNY KRAVITZ LENNY KRAVITZ
```

AGAIN AGAIN  
CHRISTINA AGUILERA CHRISTINA AGUILERA  
AIN'T NO OTHER MAN AIN'T NO OTHER MAN  
PARAMORE PARAMORE  
AIN'T IT FUN AIN'T IT FUN  
KENDRICK LAMAR KENDRICK LAMAR  
ALRIGHT ALRIGHT  
KENDRICK LAMAR KENDRICK LAMAR  
ALRIGHT ALRIGHT  
LENNY KRAVITZ LENNY KRAVITZ  
AMERICAN WOMAN AMERICAN WOMAN  
no match in top 100 ANGEL LALAH HATHAWAY  
no match in top 100 ANGRY WORLD NEIL YOUNG  
SYSTEM OF A DOWN SYSTEM OF A DOWN  
B.Y.O.B. B.Y.O.B.  
LADY GAGA LADY GAGA  
BAD ROMANCE BAD ROMANCE  
SKRILLEX SKRILLEX  
BANGARANG BANGARANG  
no match in top 100 BARTON HOLLOW THE CIVIL WARS  
no match in top 100 BE OK WILLI.AM  
MARY J BLIGE MARY J. BLIGE  
BE WITHOUT YOU BE WITHOUT YOU  
CHRISTINA AGUILERA CHRISTINA AGUILERA  
BEAUTIFUL BEAUTIFUL  
U2 U2  
BEAUTIFUL DAY BEAUTIFUL DAY  
U2 U2  
BEAUTIFUL DAY BEAUTIFUL DAY  
CARRIE UNDERWOOD CARRIE UNDERWOOD  
BEFORE HE CHEATS BEFORE HE CHEATS  
no match in top 100 BEFORE I FORGET SLIPKNOT  
CHER CHER  
BELIEVE BELIEVE  
no match in top 100 BEST PART HE.R.  
LITTLE BIG TOWN LITTLE BIG TOWN  
BETTER MAN BETTER MAN  
FANTASIA FANTASIA  
BITTERSWEET BITTERSWEET  
DAVID BOWIE DAVID BOWIE  
BLACKSTAR BLACKSTAR  
no match in top 100 BLAME IT CHRISSETTE MICHELE  
RASCAL FLATTS RASCAL FLATTS  
BLESS THE BROKEN ROAD BLESS THE BROKEN ROAD  
CARRIE UNDERWOOD CARRIE UNDERWOOD  
BLOWN AWAY BLOWN AWAY  
TONY BENNETT TONY BENNETT, AMY WINEHOUSE  
BODY AND SOUL BODY AND SOUL  
ELLA MAI ELLA MAI  
BOO'D UP BOO'D UP  
GREEN DAY GREEN DAY  
BOULEVARD OF BROKEN DREAMS BOULEVARD OF BROKEN DREAMS  
no match in top 100 BRAND NEW DAY STING  
FAITH HILL FAITH HILL  
BREATHE BREATHE  
EVANESCENCE EVANESCENCE  
BRING ME TO LIFE BRING ME TO LIFE  
CHRIS STAPLETON CHRIS STAPLETON  
BROKEN HALOS BROKEN HALOS  
no match in top 100 BUBBLIN ANDERSON PAAK

no match in top 100 BUTTERFLIES KACEY MUSGRAVES  
PRINCE PRINCE  
CALL MY NAME CALL MY NAME  
no match in top 100 CAN'T STOP THE FEELING! JAMMY JAMS  
ADELE ADELE  
CHASING PAVEMENTS CHASING PAVEMENTS  
no match in top 100 CHEROKEE CHRISTIAN MCBRIDE  
no match in top 100 CHOICES GEORGE JONES  
no match in top 100 CIRICE GHOST  
no match in top 100 CITY OF BLINDING LIGHTS U2  
ZEDD ZEDD  
CLARITY CLARITY  
50  
USHER USHER  
CLIMAX CLIMAX  
COLDPLAY COLDPLAY  
CLOCKS CLOCKS  
KYLIE MINOGUE KYLIE MINOGUE  
COME INTO MY WORLD COME INTO MY WORLD  
SHANIA TWAIN SHANIA TWAIN  
COME ON OVER COME ON OVER  
no match in top 100 COUSIN DUPREE STEELY DAN  
EMINEM EMINEM, DR. DRE, 50 CENT  
CRACK A BOTTLE CRACK A BOTTLE  
SOLANGE SOLANGE  
CRANES IN THE SKY CRANES IN THE SKY  
LINKIN PARK LINKIN PARK  
CRAWLING CRAWLING  
GNARLS BARKLEY GNARLS BARKLEY  
CRAZY CRAZY  
no match in top 100 CROSS MY MIND JILL SCOTT  
FAITH HILL FAITH HILL  
CRY CRY  
JUSTIN TIMBERLAKE JUSTIN TIMBERLAKE  
CRY ME A RIVER CRY ME A RIVER  
no match in top 100 D.O.A. (DEATH OF AUTO-TUNE) JAY Z  
LUTHER VANDROSS LUTHER VANDROSS  
DANCE WITH MY FATHER DANCE WITH MY FATHER  
no match in top 100 DANCIN' 4 CHICKEN TERENCE BLANCHARD  
no match in top 100 DANGEROUSLY IN LOVE 2 BEYONCÉ  
RED HOT CHILI PEPPERS RED HOT CHILI PEPPERS  
DANI CALIFORNIA DANI CALIFORNIA  
JOHN MAYER JOHN MAYER  
DAUGHTERS DAUGHTERS  
no match in top 100 DAYDREAMIN' LUPE FIASCO  
DIRTY VEGAS DIRTY VEGAS  
DAYS GO BY DAYS GO BY  
KANYE WEST KANYE WEST  
DIAMONDS FROM SIERRA LEONE DIAMONDS FROM SIERRA LEONE  
LENNY KRAVITZ LENNY KRAVITZ  
DIG IN DIG IN  
no match in top 100 DISORDER IN THE HOUSE BRUCE SPRINGSTEEN  
no match in top 100 DISRESPECTFUL MARY J BLIGE  
no match in top 100 DISSIDENT AGGRESSOR JUDAS PRIEST  
no match in top 100 DON'T FENCE ME IN JOHN DAVERSA  
NORAH JONES NORAH JONES  
DON'T KNOW WHY DON'T KNOW WHY  
THE CHAINSMOKERS THE CHAINSMOKERS  
DON'T LET ME DOWN DON'T LET ME DOWN  
no match in top 100 DON'T LET ME BE LONELY TONIGHT JAMES TAYLOR

THE BLACK EYED PEAS THE BLACK EYED PEAS  
DON'T PHUNK WITH MY HEART DON'T PHUNK WITH MY HEART  
no match in top 100 DON'T WANNA FIGHT ALABAMA SHAKES  
no match in top 100 DON'T WANNA FIGHT JOHN HEINRICH  
TRAIN TRAIN  
DROPS OF JUPITER DROPS OF JUPITER  
no match in top 100 DRUNK IN LOVE BEYONCÉ  
no match in top 100 DYSTOPIA MEGADETH  
no match in top 100 EARNED IT (FIFTY SHADES OF GREY) THE WEEKND  
CHRIS STAPLETON CHRIS STAPLETON  
EITHER WAY EITHER WAY  
no match in top 100 EL DORADO IRON MAIDEN  
no match in top 100 ELECTRIC MESSIAH HIGH ON FIRE  
SILK CITY SILK CITY, DUA LIPA  
ELECTRICITY ELECTRICITY  
no match in top 100 ELEVATION U2  
no match in top 100 ELITE DEFTONES  
JAY-Z JAY  
EMPIRE STATE OF MIND EMPIRE STATE OF MIND  
no match in top 100 EYES OF THE INSANE SLAYER  
ALICIA KEYS ALICIA KEYS  
FALLIN' FALLIN'  
no match in top 100 FAMILY AFFAIR HANS ZIMMER  
GORILLAZ GORILLAZ  
FEEL GOOD INC. FEEL GOOD INC  
PORTUGAL THE MAN PORTUGAL. THE MAN  
FEEL IT STILL FEEL IT STILL  
no match in top 100 FINAL SIX SLAYER  
no match in top 100 FINGERPRINTS CHICK COREA  
100  
no match in top 100 FOOL FOR YOU CELO GREEN  
no match in top 100 FOR ONCE IN MY LIFE TONY BENNETT  
no match in top 100 FORGOT ABOUT DRE DR DRE  
no match in top 100 FROM THE BOTTOM OF MY HEART STEVIE WONDER  
no match in top 100 FROM THIS VALLEY THE CIVIL WARS  
no match in top 100 FUCK YOU CEE LO GREEN  
no match in top 100 FUTURE BABY MAMA PRINCE  
no match in top 100 GALVANIZE THE CHEMICAL BROTHERS  
no match in top 100 GENTLE ON MY MIND JOHN MAILANDER  
DAFT PUNK DAFT PUNK  
GET LUCKY GET LUCKY  
no match in top 100 GET RIGHT WITH GOD LUCINDA WILLIAMS  
MISSY "MISDEMEANOR" ELLIOTT MISSY "MISDEMEANOR" ELLIOTT  
GET UR FREAK ON GET UR FREAK ON  
no match in top 100 GIRL CRUSH NEW POLITICS  
LITTLE BIG TOWN LITTLE BIG TOWN  
GIRL CRUSH GIRL CRUSH  
BRUCE SPRINGSTEEN BRUCE SPRINGSTEEN  
GIRLS IN THEIR SUMMER CLOTHES GIRLS IN THEIR SUMMER CLOTHES  
no match in top 100 GIVE MY LOVE TO ROSE JOHNNY CASH  
no match in top 100 GLORY LIL WAYNE  
no match in top 100 GOD PROVIDES TAMELA MANN  
no match in top 100 GOD IS DEAD? BLACK SABBATH  
DRAKE DRAKE  
GOD'S PLAN GOD'S PLAN  
KANYE WEST KANYE WEST  
GOLD DIGGER GOLD DIGGER  
no match in top 100 GONE GONE GONE (DONE MOVED ON) ALISON KRAUSS  
KANYE WEST KANYE WEST  
GOOD LIFE GOOD LIFE

no match in top 100 GRAVEDIGGER DAVE MATTHEWS  
RAGE AGAINST THE MACHINE RAGE AGAINST THE MACHINE  
GUERRILLA RADIO GUERRILLA RADIO  
no match in top 100 HALO BEYONCÉ  
PHARRELL WILLIAMS PHARRELL WILLIAMS  
HAPPY (LIVE) HAPPY  
no match in top 100 HARDER BETTER FASTER STRONGER DAFT PUNK  
no match in top 100 HE THINK I DON'T KNOW MARY J BLIGE  
TONI BRAXTON TONI BRAXTON  
HE WASN'T MAN ENOUGH HE WASN'T MAN ENOUGH  
no match in top 100 HEAVEN JOHN LEGEND  
LOS LONELY BOYS LOS LONELY BOYS  
HEAVEN HEAVEN  
ADELE ADELE  
HELLO HELLO  
no match in top 100 HERE WE GO AGAIN NORAH JONES  
KORN KORN  
HERE TO STAY HERE TO STAY  
NO DOUBT NO DOUBT  
HEY BABY HEY BABY  
OUTKAST OUTKAST  
HEY YA! HEY YA!  
TRAIN TRAIN  
HEY, SOUL SISTER (LIVE) HEY, SOUL SISTER  
no match in top 100 HOLY SPIRIT RYAN ELLIS  
no match in top 100 HOT HOUSE CHICK COREA  
NELLY NELLY  
HOT IN HERRE HOT IN HERRE  
DRAKE DRAKE  
HOTLINE BLING HOTLINE BLING  
no match in top 100 HOUSE OF MERCY SARAH JAROSZ  
no match in top 100 HOW DEEP IS YOUR LOVE PJ MORTON  
AULI'I CRAVALHO AULI'I CRAVALHO  
HOW FAR I'LL GO HOW FAR I'LL GO  
TIM MCGRAW TIM MCGRAW  
HUMBLE AND KIND HUMBLE AND KIND  
KENDRICK LAMAR KENDRICK LAMAR  
HUMBLE. HUMBLE.  
NELLY FURTADO NELLY FURTADO  
I'M LIKE A BIRD I'M LIKE A BIRD  
GLEN CAMPBELL GLEN CAMPBELL  
I'M NOT GONNA MISS YOU I'M NOT GONNA MISS YOU  
no match in top 100 IF WE WERE VAMPIRES JASON ISBELL AND THE 400 UNIT  
150  
ALICIA KEYS ALICIA KEYS  
IF I AIN'T GOT YOU IF I AIN'T GOT YOU  
no match in top 100 IMAGINE MARCOS YAROIDE  
no match in top 100 IN MY PLACE COLDPLAY  
no match in top 100 IRON MAN BLACK SABBATH  
no match in top 100 IS THIS LOVE CORINNE BAILEY RAE  
no match in top 100 IS YOU IS, OR IS YOU AIN'T (MY BABY) STEVE LALLIER'S TUXEDO JUNCTION  
ALAN JACKSON ALAN JACKSON, JIMMY BUFFETT  
IT'S FIVE O'CLOCK SOMEWHERE IT'S FIVE O'CLOCK SOMEWHERE  
no match in top 100 IT'S NOT RIGHT BUT IT'S OKAY WHITNEY HOUSTON  
no match in top 100 I'M SO LONESOME I COULD CRY JOHN SCOFIELD  
no match in top 100 JESUS CHILDREN TIMMY SINGERS  
KANYE WEST KANYE WEST  
JESUS WALKS JESUS WALKS  
CARRIE UNDERWOOD CARRIE UNDERWOOD  
JESUS, TAKE THE WHEEL JESUS, TAKE THE WHEEL



no match in top 100 JOANNE (WHERE DO YOU THINK YOU'RE GOIN'?) LADY GAGA  
no match in top 100 JOLENE PENTATONIX  
BRUNO MARS BRUNO MARS  
JUST THE WAY YOU ARE JUST THE WAY YOU ARE  
no match in top 100 KEEP ON THE SUNNY SIDE JUNE CARTER CASH  
no match in top 100 KID SISTER THE TIME JUMPERS  
no match in top 100 KILLER DILLER BLUES ALABAMA SHAKES  
CHRISTINA AGUILERA CHRISTINA AGUILERA, LIL' KIM, MYA, P!NK  
LADY MARMALADE LADY MARMALADE  
no match in top 100 LAKE BY THE OCEAN MAXWELL  
CARRIE UNDERWOOD CARRIE UNDERWOOD  
LAST NAME LAST NAME  
no match in top 100 LAZARETTO JACK WHITE  
IDINA MENZEL IDINA MENZEL  
LET IT GO LET IT GO  
THE BLACK EYED PEAS THE BLACK EYED PEAS  
LET'S GET IT STARTED LET'S GET IT STARTED  
BRAD PAISLEY BRAD PAISLEY  
LETTER TO ME LETTER TO ME  
no match in top 100 LITTLE GHETTO BOY LALAH HATHAWAY  
no match in top 100 LITTLE THINGS INDIAARIE  
TIM MCGRAW TIM MCGRAW  
LIVE LIKE YOU WERE DYING LIVE LIKE YOU WERE DYING  
LIL WAYNE LIL WAYNE  
LOLLIPOP LOLLIPOP  
THE BLACK KEYS THE BLACK KEYS  
LONELY BOY LONELY BOY  
EMINEM EMINEM  
LOSE YOURSELF LOSE YOURSELF  
no match in top 100 LOVE OF MY LIFE (AN ODE TO HIP HOP) ERYKAH BADU  
no match in top 100 LOVE ON TOP BEYONCÉ  
no match in top 100 LOVESTONED/I THINK SHE KNOWS JUSTIN TIMBERLAKE  
KENDRICK LAMAR KENDRICK LAMAR  
LOYALTY. LOYALTY.  
no match in top 100 LUCKY COURTNEY JOHN  
no match in top 100 MAKE IT MINE JASON MRAZ  
MAROON 5 MAROON 5  
MAKES ME WONDER MAKES ME WONDER  
SHANIA TWAIN SHANIA TWAIN  
MAN! I FEEL LIKE A WOMAN! MAN! I FEEL LIKE A WOMAN!  
SANTANA SANTANA  
MARIA MARIA MARIA MARIA  
no match in top 100 MASSEDUCTION ST VINCENT  
TAYLOR SWIFT TAYLOR SWIFT  
MEAN MEAN  
no match in top 100 MESSENGERS LECRAE  
no match in top 100 MILES BEYOND JOHN MCLAUGHLIN  
NE YO NE  
MISS INDEPENDENT MISS INDEPENDENT  
LUDACRIS LUDACRIS  
MONEY MAKER MONEY MAKER  
OUTKAST OUTKAST  
MS. JACKSON MS. JACKSON  
METALLICA METALLICA  
MY APOCALYPSE MY APOCALYPSE  
USHER USHER AND ALICIA KEYS  
MY BOO MY BOO  
MAREN MORRIS MAREN MORRIS  
MY CHURCH MY CHURCH  
200

THE BLACK EYED PEAS THE BLACK EYED PEAS  
MY HUMPS MY HUMPS  
no match in top 100 MY NAME IS EMINEM  
no match in top 100 N\*\*\*\*S IN PARIS LIANNE LA HAVAS  
no match in top 100 N\*\*\*\*S IN PARIS LIANNE LA HAVAS  
LADY ANTEBELLUM LADY ANTEBELLUM  
NEED YOU NOW NEED YOU NOW  
no match in top 100 NEVER ALONE TORI KELLY  
no match in top 100 NEVER HAVE TO BE ALONE CECE WINANS  
no match in top 100 NEW FANG THEM CROOKED VULTURES  
no match in top 100 NEXT BIG THING VINCE GILL  
no match in top 100 NO GREATER LOVE SMOKIE NORFUL  
ALICIA KEYS ALICIA KEYS  
NO ONE NO ONE  
CHANCE THE RAPPER CHANCE THE RAPPER  
NO PROBLEM NO PROBLEM  
EMINEM EMINEM  
NOT AFRAID NOT AFRAID  
DIXIE CHICKS DIXIE CHICKS  
NOT READY TO MAKE NICE NOT READY TO MAKE NICE  
no match in top 100 ON TO THE NEXT ONE CASSIDY  
RIHANNA RIHANNA  
ONLY GIRL (IN THE WORLD) ONLY GIRL  
no match in top 100 ORBITS WAYNE SHORTER  
JOHN LEGEND JOHN LEGEND  
ORDINARY PEOPLE ORDINARY PEOPLE  
no match in top 100 OTIS JAY-Z  
no match in top 100 PEARLS BAND OF SKULLS  
no match in top 100 PLEASE COME HOME GARY CLARK JR  
no match in top 100 PLEASE READ THE LETTER ALISON KRAUSS  
LADY GAGA LADY GAGA  
POKER FACE POKER FACE  
LITTLE BIG TOWN LITTLE BIG TOWN  
PONTOON PONTOON  
MAXWELL MAXWELL  
PRETTY WINGS PRETTY WINGS  
no match in top 100 PUT YOUR LIGHTS ON SANTANA  
no match in top 100 RADIO NOWHERE BRUCE SPRINGSTEEN  
IMAGINE DRAGONS IMAGINE DRAGONS  
RADIOACTIVE RADIOACTIVE  
CLEAN BANDIT CLEAN BANDIT  
RATHER BE RATHER BE  
no match in top 100 REALLY LOVE D'ANGELO  
CHILDISH GAMBINO CHILDISH GAMBINO  
REDBONE REDBONE  
GRETCHEN WILSON GRETCHEN WILSON  
REDNECK WOMAN REDNECK WOMAN  
AMY WINEHOUSE AMY WINEHOUSE  
REHAB REHAB  
no match in top 100 RICH WOMAN ALISON KRAUSS  
CHAMILLIONAIRE CHAMILLIONAIRE  
RIDIN RIDIN'  
ADELE ADELE  
ROLLING IN THE DEEP ROLLING IN THE DEEP  
LORDE LORDE  
ROYALS ROYALS  
BRYSON TILLER BRYSON TILLER  
RUN RUN ME DRY  
JAY-Z JAY  
RUN THIS TOWN RUN THIS TOWN

JOHN MAYER JOHN MAYER  
SAY SAY  
DESTINY'S CHILD DESTINY'S CHILD  
SAY MY NAME SAY MY NAME  
A GREAT BIG WORLD A GREAT BIG WORLD, CHRISTINA AGUILERA  
SAY SOMETHING SAY SOMETHING  
RED HOT CHILI PEPPERS RED HOT CHILI PEPPERS  
SCAR TISSUE SCAR TISSUE  
SKRILLEX SKRILLEX  
SCARY MONSTERS AND NICE SPRITES SCARY MONSTERS AND NICE SPRITES  
TOOL TOOL  
SCHISM SCHISM  
no match in top 100 SCREAM A.K.A. ITCHIN' MISSY ELLIOTT  
no match in top 100 SEE THAT MY GRAVE IS KEPT CLEAN MAVIS STAPLES  
ADELE ADELE  
SET FIRE TO THE RAIN [LIVE] SET FIRE TO THE RAIN  
THE WHITE STRIPES THE WHITE STRIPES  
SEVEN NATION ARMY SEVEN NATION ARMY  
no match in top 100 SEXY BACK DJ NASTY  
250  
MURPHY LEE NELLY, P. DIDDY, MURPHY LEE  
SHAKE YA TAILFEATHER SHAKE YA TAILFEATHER  
LADY GAGA LADY GAGA, BRADLEY COOPER  
SHALLOW SHALLOW  
ED SHEERAN ED SHEERAN  
SHAPE OF YOU SHAPE OF YOU  
no match in top 100 SHE WALKS THIS EARTH (SOBERANA ROSA) STING  
no match in top 100 SHINE DOLLY PARTON  
JOHN LENNON LEN  
SHINE STEAL MY SUNSHINE  
KELLY CLARKSON KELLY CLARKSON  
SINCE U BEEN GONE SINCE U BEEN GONE  
no match in top 100 SINGLE LADIES (PUT A RING ON IT) BEYONCÉ  
VELVET REVOLVER VELVET REVOLVER  
SLITHER SLITHER  
SANTANA SANTANA  
SMOOTH SMOOTH  
SANTANA SANTANA  
SMOOTH SMOOTH  
no match in top 100 SO AMAZING 50 CENT  
SADE SADE  
SOLDIER OF LOVE SOLDIER OF LOVE  
no match in top 100 SOLITARY MAN JOHNNY CASH  
GOTYE GOTYE  
SOMEBODY THAT I USED TO KNOW SOMEBODY THAT I USED TO KNOW  
ADELE ADELE  
SOMEONE LIKE YOU SOMEONE LIKE YOU  
no match in top 100 SOMETHING SNARKY PUPPY WITH LALAH HATHAWAY  
CARRIE UNDERWOOD CARRIE UNDERWOOD  
SOMETHING IN THE WATER SOMETHING IN THE WATER  
U2 U2  
SOMETIMES YOU CAN'T MAKE IT ON YOUR OWN SOMETIMES YOU CAN'T MAKE IT ON YOUR OWN  
U2 U2  
SOMETIMES YOU CAN'T MAKE IT ON YOUR OWN SOMETIMES YOU CAN'T MAKE IT ON YOUR OWN  
no match in top 100 SOUTHSIDE COMMON  
no match in top 100 SPACE COWBOY KACEY MUSGRAVES  
no match in top 100 ST. ANGER METALLICA  
SUGARLAND SUGARLAND  
STAY STAY  
SAM SMITH SAM SMITH

STAY WITH ME (DARKCHILD VERSION) STAY WITH ME  
no match in top 100 STAY WITH ME (BY THE SEA) JOHN LEGEND  
no match in top 100 STAYING POWER BARRY WHITE  
SHERYL CROW SHERYL CROW  
STEVE MCQUEEN STEVE MCQUEEN  
TWENTY ONE PILOTS TWENTY ONE PILOTS  
STRESSED OUT STRESSED OUT  
KANYE WEST KANYE WEST  
STRONGER STRONGER  
U2 U2  
STUCK IN A MOMENT YOU CAN'T GET OUT OF STUCK IN A MOMENT YOU CAN'T GET OUT OF  
KEITH URBAN KEITH URBAN  
STUPID BOY STUPID BOY  
no match in top 100 SULTAN'S CURSE MASTODON  
no match in top 100 SUNRISE NORAH JONES  
ALICIA KEYSS ALICIA KEYS  
SUPERWOMAN SUPERWOMAN  
no match in top 100 SWAGGA LIKE US T.I.  
no match in top 100 SWEET CHILD O' MINE SHERYL CROW  
KEITH URBAN KEITH URBAN  
SWEET THING SWEET THING  
no match in top 100 TEQUILA DAN SHAY  
BRUNO MARS BRUNO MARS  
THAT'S WHAT I LIKE THAT'S WHAT I LIKE  
no match in top 100 THE CLOSER I GET TO YOU LUTHER VANDROSS  
no match in top 100 THE CONNECTION EMMYLOU HARRIS  
SANTANA SANTANA  
THE GAME OF LOVE THE GAME OF LOVE  
MIRANDA LAMBERT MIRANDA LAMBERT  
THE HOUSE THAT BUILT ME THE HOUSE THAT BUILT ME  
no match in top 100 THE JOKE BRANDI CARLILE  
no match in top 100 THE LAST IN LINE TENACIOUS D  
no match in top 100 THE LUCKY ONE ALISON KRAUSS & UNION STATION  
FOO FIGHTERS FOO FIGHTERS  
THE PRETENDER THE PRETENDER  
EMINEM EMINEM  
THE REAL SLIM SHADY THE REAL SLIM SHADY  
no match in top 100 THE REASON WHY VINCE GILL  
300  
BRUCE SPRINGSTEEN BRUCE SPRINGSTEEN  
THE RISING THE RISING  
OUTKAST OUTKAST  
THE WHOLE WORLD THE WHOLE WORLD  
USHER USHER  
THERE GOES MY BABY THERE GOES MY BABY  
no match in top 100 THERE GOES THE NEIGHBORHOOD SHERYL CROW  
ED SHEERAN ED SHEERAN  
THINKING OUT LOUD THINKING OUT LOUD  
no match in top 100 THINKING OUT LOUD BOYCE AVENUE  
no match in top 100 THRIFT SHOP KIDZ BOP KIDS  
no match in top 100 THY WILL HILLARY SCOTT & THE SCOTT FAMILY  
no match in top 100 TONITE LCD SOUNDSYSTEM  
BRITNEY SPEARS BRITNEY SPEARS  
TOXIC TOXIC  
CHRIS STAPLETON CHRIS STAPLETON  
TRAVELLER TRAVELLER  
no match in top 100 TROUBLE PINK  
NO DOUBT NO DOUBT  
UNDERNEATH IT ALL UNDERNEATH IT ALL  
D'ANGELO D'ANGELO

UNTITLED (HOW DOES IT FEEL) UNTITLED  
no match in top 100 UPTOWN FUNK MARK RONSON  
KINGS OF LEON KINGS OF LEON  
USE SOMEBODY USE SOMEBODY  
U2 U2  
VERTIGO VERTIGO  
U2 U2  
VERTIGO VERTIGO  
COLDPLAY COLDPLAY  
VIVA LA VIDA VIVA LA VIDA  
DARIUS RUCKER DARIUS RUCKER  
WAGON WHEEL WAGON WHEEL  
JOHN MAYER JOHN MAYER  
WAITING ON THE WORLD TO CHANGE WAITING ON THE WORLD TO CHANGE  
FOO FIGHTERS FOO FIGHTERS  
WALK WALK  
no match in top 100 WALK ON U2  
no match in top 100 WANNA BE HAPPY? KIRK FRANKLIN  
no match in top 100 WAR MACHINE AC/DC  
no match in top 100 WAX SIMULACRA THE MARS VOLTA  
MARIAH CAREY MARIAH CAREY  
WE BELONG TOGETHER WE BELONG TOGETHER  
no match in top 100 WE ARE YOUNG FUN  
no match in top 100 WELCOME TO JAMROCK DAMIAN MARLEY  
no match in top 100 WHAT GOES AROUND...COMES AROUND JUSTIN TIMBERLAKE  
no match in top 100 WHAT A BEAUTIFUL NAME HILLSONG WORSHIP  
no match in top 100 WHAT YOU KNOW TI.  
no match in top 100 WHENEVER I SAY YOUR NAME STING  
no match in top 100 WHIPLASH MOTÖRHEAD  
no match in top 100 WHISKEY IN THE JAR METALLICA  
TAYLOR SWIFT TAYLOR SWIFT  
WHITE HORSE WHITE HORSE  
BAHA MEN BAHÁ MEN  
WHO LET THE DOGS OUT WHO LET THE DOGS OUT  
CREED CREED  
WITH ARMS WIDE OPEN WITH ARMS WIDE OPEN  
no match in top 100 WOMAN WOLF MOTHER  
no match in top 100 WORK IT MISSY ELLIOTT  
KEITH URBAN KEITH URBAN  
YOU'LL THINK OF ME YOU'LL THINK OF ME  
no match in top 100 A CHANGE IS GONNA COME HERBIE HANCOCK  
no match in top 100 A FEATHER'S NOT A BIRD ROSANNE CASH, ARTIST  
LIL WAYNE LIL WAYNE  
A MILLI A MILLI  
FOO FIGHTERS FOO FIGHTERS  
ALL MY LIFE ALL MY LIFE  
JANET JACKSON JANET  
ALL FOR YOU ALL FOR YOU  
no match in top 100 ALL OF THE LIGHTS EVIL EMPIRE  
KENDRICK LAMAR C  
I BOYFRIEND/GIRLFRIEND  
THE BLACK EYED PEAS THE BLACK EYED PEAS  
I GOTTA FEELING I GOTTA FEELING  
LEE ANN WOMACK LEE ANN WOMACK  
I HOPE YOU DANCE I HOPE YOU DANCE  
350  
no match in top 100 I SEE THE LIGHT (FROM TANGLED) MANDY MOORE  
MACY GRAY MACY GRAY  
I TRY I TRY  
SARAH MCLACHLAN SARAH MCLACHLAN

I WILL REMEMBER YOU I WILL REMEMBER YOU  
 no match in top 100 0 DEATH RALPH STANLEY  
 MAROON 5 MAROON 5  
 THIS LOVE THIS LOVE  
 CHILDISH GAMBINO CHILDISH GAMBINO  
 THIS IS AMERICA THIS IS AMERICA  
 USHER USHER  
 U DON'T HAVE TO CALL U DON'T HAVE TO CALL  
 USHER USHER  
 U REMIND ME U REMIND ME  
 no match in top 100 WHEN BAD DOES GOOD CHRIS CORNELL  
 no match in top 100 WHERE WERE YOU (WHEN THE WORLD STOPPED TURNING) ALAN JACKSON  
 no match in top 100 WHERE ARE Ü NOW SKRILLEX AND DIPLO WITH JUSTIN BIEBER  
 ALICIA KEYS ALICIA KEYS  
 YOU DON'T KNOW MY NAME YOU DON'T KNOW MY NAME  
 no match in top 100 YOU GOT ME THE ROOTS  
 LAUREN DAIGLE LAUREN DAIGLE  
 YOU SAY YOU SAY  
 no match in top 100 YOU WANT IT DARKER LEONARD COHEN  
 JOHN MAYER JOHN MAYER  
 YOUR BODY IS A WONDERLAND YOUR BODY IS A WONDERLAND

Load in the CSV containing grammy award winning songs with billboard data.

```
In [ ]: grammyAwards = pd.read_csv('grammy_award_data.csv')
```

One hot encodes the genres in the billboard

```
In [ ]: from sklearn.preprocessing import MultiLabelBinarizer

genres_as_list = [g.split(",") for g in billboardHot100['Genre']]
one_hot = MultiLabelBinarizer()
genre_encoding = one_hot.fit_transform(genres_as_list)
genres = one_hot.classes_
genres = [genre.replace(';', ' ') for genre in genres]

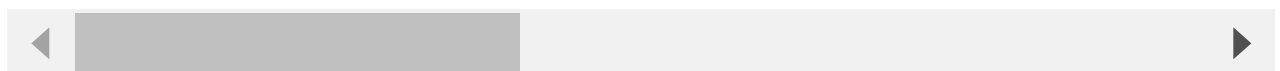
genre_df = pd.DataFrame(genre_encoding, columns = genres)
billboardHot100 = billboardHot100.join(genre_df)
billboardHot100 = billboardHot100.drop(columns=['Genre'])
billboardHot100
```

```
Out[ ]:
```

	Unnamed: 0	Artist	Name	Peak.position	Weeks.on.chart	Avg.rank	Date	Energy	Dance
0	0	"WEIRD AL" YANKOVIC	CANADIAN IDIOT	82.0	3.0	91.333333	2006	0.697	
1	1	"WEIRD AL" YANKOVIC	WORD CRIMES	39.0	2.0	65.000000	2014	0.430	
2	2	'N SYNC	BYE BYE BYE	4.0	23.0	14.260870	2000	0.926	
3	3	'N SYNC	GONE	11.0	24.0	27.125000	2001	0.409	

	Unnamed: 0	Artist	Name	Peak.position	Weeks.on.chart	Avg.rank	Date	Energy	Dance
4	4	'N SYNC	IT'S GONNA BE ME	1.0	25.0	21.680000	2000	0.873	
...	...	...	...	...	...	...	...	...	...
7209	7209	WILL.I.AM	FALL DOWN	100.0	1.0	58.000000	2013	0.902	
7210	7210	WILL.I.AM	I GOT IT FROM MY MAMA	31.0	9.0	56.666667	2007	0.777	
7211	7211	WILL.I.AM	IT'S A NEW DAY	78.0	2.0	81.000000	2008	0.644	
7212	7212	WILL.I.AM, BRITNEY SPEARS	SCREAM & SHOUT	3.0	24.0	14.583333	2012	0.686	
7213	7213	WILL.I.AM, NICKI MINAJ	CHECK IT OUT	24.0	15.0	54.133333	2010	0.694	

7214 rows × 346 columns



Finds the 50 most popular genres

```
In [ ]: genre_count = []
genre_dict= {}
for genre in genres:
    count = billboardHot100[genre].sum()
    genre_count.append(count)
    genre_dict[genre] = count
sorted_genres = dict(sorted(genre_dict.items(), key=lambda genre:genre[1]))
popular_genres_top_50 = list(sorted_genres.keys())[-50:]
```

Gets rid of all genres but the 50 most popular

```
In [ ]: billboardHot100 = billboardHot100[list(billboardHot100.columns[1:12]) + popular_genres_
```

Adds columns in the grammyAwards matrix for genres

```
In [ ]: df = pd.DataFrame(data=np.zeros((360, len(popular_genres_top_50)), dtype=int), columns=
df.head()
grammyGenres = pd.concat([grammyAwards, df], axis=1)
grammyGenres.head(5)
```

```
Out[ ]: Name Artist GrammyYear Genre GrammyAward Energy Danceability Tempo Valence
```

	Name	Artist	GrammyYear	Genre	GrammyAward	Energy	Danceability	Tempo	Valence
0	'TIL SUMMER COMES AROUND	KEITH URBAN	2010	COUNTRY	1	-1.000	-1.000	-1.000	-1.000
1	24 FRAMES	JASON ISBELL	2015	AMERICAN ROOTS MUSIC	1	-1.000	-1.000	-1.000	-1.000
2	24K MAGIC	BRUNO MARS	2017	GENERAL	1	0.803	0.818	106.970	0.632
3	500 MILES HIGH	CHICK COREA	2011	JAZZ	1	-1.000	-1.000	-1.000	-1.000
4	99 PROBLEMS	JAY	2004	RAP	1	0.887	0.494	89.554	0.548

5 rows × 62 columns



Uses fuzzy to match each song with its related genres

```
In [ ]: for index, row in grammyGenres.iterrows():
        givenGenre = row['Genre']
        for genre in popular_genres_top_50:
            genreRatio = fuzz.partial_ratio(givenGenre, genre)
            if (genreRatio > 90):
                grammyGenres.at[index, genre] = 1
        grammyGenres = grammyGenres.drop(columns="Genre")
        grammyGenres.head(5)
```

Out [ ]:

	Name	Artist	GrammyYear	GrammyAward	Energy	Danceability	Tempo	Valence	Peak.position
0	'TIL SUMMER COMES AROUND	KEITH URBAN	2010	1	-1.000	-1.000	-1.000	-1.000	-
1	24 FRAMES	JASON ISBELL	2015	1	-1.000	-1.000	-1.000	-1.000	-
2	24K MAGIC	BRUNO MARS	2017	1	0.803	0.818	106.970	0.632	
3	500 MILES HIGH	CHICK COREA	2011	1	-1.000	-1.000	-1.000	-1.000	-
4	99 PROBLEMS	JAY	2004	1	0.887	0.494	89.554	0.548	3

5 rows × 61 columns





Saves final CSVs

```
In [ ]: grammyGenres = grammyGenres[['Name', 'Artist', 'GrammyAward', 'Avg.rank', 'Peak.position',  
    'Danceability', 'Energy', 'Tempo', 'Valence']+popular_genres_top_50]  
grammyGenres.to_csv("Grammys.csv")  
billboardHot100 = billboardHot100[['Name', 'Artist', 'GrammyAward', 'Avg.rank', 'Peak.position',  
    'Danceability', 'Energy', 'Tempo', 'Valence']+popular_genres_top_50]  
billboardHot100.to_csv("Billboard.csv")
```