Final Video

Behind hit songs.

Background on Spotify and Streaming

Spotify Background

- Spotify is a digital music, podcast, and video streaming service

- The company was founded in 2006 in Stockholm, Sweden, by Daniel Ek.

Spotify for Artists

- Spotify maintains a separate section of their application/ website solely for artists

- They have a six step process to help artists post their art successfully on the application

Spotify Impact

- Spotify has transformed how users consume music through machine learning algorithms that are able to curate desirable playlists for each individual user.

Why we are interested in this topic

Interest in Music

- We found the most in common within the subject of music.
- All four of us use music as a form of relaxation

Interest in our artists

- Our group was most interested in finding commonalities with our different music tastes.
- To find a decipherable formula artists follow to create popular music.

Interest in success

- Our team was curious about why certain songs were listened to.
- Social media trends have a strong influence on music's popularity

Our Question

What criteria should be met in order for an up and coming artist to make it in the music industry?

More specifically what musical style elements result in the most Spotify streams for artists?

Hypothesis

We predict that there will be a strong correlation between song characteristics and total streams of a song. This correlation would suggest there are key elements that would predict the popularity of songs.

Approach to Testing Our Hypothesis

- analyze variables and evaluate if there are correlations between the total number of streams.
- seek correlations between variables.
- complete a linear regression for popularity of songs based on song characteristics.

The Datasets

hundf = 2000song.csv
# of observations: 2000

Index	Title	Artist	Top Genre	Year	Per Minute (BPM)	Energy	Danceability	Loudness (dB)	Liveness	Valence	Length (Duration)	Acoustic
1	Sunrise	Norah Jones	adult standards	2004	157	30	53	-14	11	68	201	
2	Black Night	Deep Purple	album rock	2000	135	79	50	-11	17	81	207	
3	Clint Eastwood	Gorillaz	alternative hip hop	2001	168	69	66	-9	7	52	341	

moredf = 116ksong.csv
# of observations: 130663

	artist_name	track_id	track_name	acousticness	danceability	duration_ms	energy	instrumentalness
0	YG	2RM4jf1Xa9zPgMGRDiht8O	Big Bank feat. 2 Chainz, Big Sean, Nicki Minaj	0.00582	0.743	238373	0.339	0.0
1	YG	1tHDG53xJNGsItRA3vfVgs	BAND DRUM (feat. A$AP Rocky)	0.02440	0.846	214800	0.557	0.0
2	R3HAB	6Wosx2euFPMT14UXiWudMy	Radio Silence	0.02500	0.603	138913	0.723	0.0

songs = charts.csv
# of observations: 1556

title	rank	date	artist	url	region	chart	trend	streams
Despacito (Featuring Daddy Yankee)	1	2017-03-01	Luis Fonsi	https://open.spotify.com/track/4aWmUDTfIPGksMN...	Argentina	top200	SAME_POSITION	365941.0
El Amante	2	2017-03-01	Nicky Jam	https://open.spotify.com/track/3umS4y3uQDkqekN...	Argentina	top200	SAME_POSITION	179697.0
Reggaetón Lento (Bailemos)	3	2017-03-01	CNCO	https://open.spotify.com/track/3AEZUABDXNtecAO...	Argentina	top200	SAME_POSITION	169647.0

The Song Characteristics (Variables)

song_name - The name of a song as listed on Spotify

song_artist - The main artist of a song

BPM - Beats per minute or tempo of a song

energy - perceptual measure of intensity and activity

loudness - overall loudness of a track in decibels (dB)

liveness - the presence of an audience in the recording

valence - describing the musical positiveness conveyed by a track.

song_duration - Duration of a song in seconds.

acousticness - The confidence level of a song being acousitc

speechiness - The relative rate of spoken words in a song

Data Wrangling

```
#This reordered our variables and eliminated characteristics we are not interested in.
hundf = hundf[['Title','Artist','Beats Per Minute (BPM)','Energy','Loudness (dB)','Liveness','Valence','Length (Duration)','A
moredf = moredf[['track_name','artist_name','tempo','energy','loudness','liveness','valence','duration_ms','acousticness','sp
songs = songs[['title','artist','streams']]

hundf.rename(columns={'Title': 'song_name','Artist': 'song_artist','Beats Per Minute (BPM)':'BPM','Energy':'energy','Loudness
moredf.rename(columns={'track_name': 'song_name','artist_name': 'song_artist','tempo':'BPM','duration_ms':'song_duration'}, i
songs.rename(columns={'title':'song_name','artist':'artist_name'},inplace=True)
```

```
hundf.head(3)
```

	song_name	song_artist	BPM	energy	loudness	liveness	valence	song_duration	acousticness	speechiness
0	Sunrise	Norah Jones	157	30	-14	11	68	201	94	3
1	Black Night	Deep Purple	135	79	-11	17	81	207	17	7
2	Clint Eastwood	Gorillaz	168	69	-9	7	52	341	2	17

```
moredf.head(3)
```

	song_name	song_artist	BPM	energy	loudness	liveness	valence	song_duration	acousticness	speechiness
0	Big Bank feat. 2 Chainz, Big Sean, Nicki Minaj	YG	203.927	0.339	-7.678	0.0812	0.118	238373	0.00582	0.4090
1	BAND DRUM (feat. A$AP Rocky)	YG	159.009	0.557	-7.259	0.2860	0.371	214800	0.02440	0.4570
2	Radio Silence	R3HAB	114.966	0.723	-5.890	0.0824	0.382	138913	0.02500	0.0454

```
songs.head(3)
```

	song_name	artist_name	streams
0	Despacito (Featuring Daddy Yankee)	Luis Fonsi	365941.0
1	El Amante	Nicky Jam	179697.0
2	Reggaetón Lento (Bailemos)	CNCO	169647.0

Data wrangling

```
for i in range(len(hundf.song_name)):
    if 'bad guy' == hundf.song_name[i]:
        indx = i
hundf.iloc[indx]
```

```
song_name                bad guy
song_artist         Billie Eilish
BPM                          135
energy                        43
loudness                     -11
liveness                      10
valence                       56
song_duration                194
acousticness                  33
speechiness                   38
Name: 786, dtype: object
```

```
for i in range(len(moredf.song_name)):
    if 'bad guy' == moredf.song_name[i]:
        indx = i
moredf.iloc[indx]
```

```
song_name                bad guy
song_artist         Billie Eilish
BPM                      131.926
energy                     0.418
loudness                 -10.998
liveness                     0.1
valence                    0.578
song_duration             194088
acousticness               0.308
speechiness                0.368
Name: 99368, dtype: object
```

```
def standardize_BPM(val):
    try:
        output = val * .97722963

    except:
        output = np.nan

    return output

def standardize_loud(val):
    try:
        output = val / 1.0001851

    except:
        output = np.nan
```

```
hundf['BPM'] = hundf['BPM'].apply(standardize_BPM)
hundf['loudness'] = hundf['loudness'].apply(standardize_loud)
moredf['energy'] = moredf['energy'].apply(standardize_energy)
moredf['liveness'] = moredf['liveness'].apply(standardize_live)
moredf['valence'] = moredf['valence'].apply(standardize_valence)
moredf['song_duration'] = moredf['song_duration'].apply(standardize_dur)
moredf['acousticness'] = moredf['acousticness'].apply(standardize_acou)
moredf['speechiness'] = moredf['speechiness'].apply(standardize_speech)
```

```
for i in range(len(moredf.song_name)):
    if 'bad guy' == moredf.song_name[i]:
        indx = i
Billie2 = moredf.iloc[indx]
for i in range(len(hundf.song_name)):
    if 'bad guy' == hundf.song_name[i]:
        indx = i
Billie1 = hundf.iloc[indx]
print(Billie1)
print(Billie2)
```

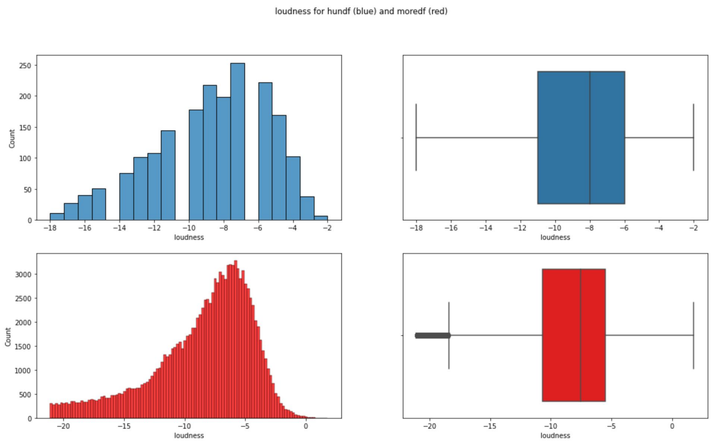
```
song_name                bad guy
song_artist         Billie Eilish
BPM                      131.926
energy                        43
loudness                 -10.998
liveness                      10
valence                       56
song_duration                194
acousticness                  33
speechiness                   38
Name: 786, dtype: object
song_name                bad guy
song_artist         Billie Eilish
BPM                      131.926
energy                        43
loudness                 -10.998
liveness                      10
valence                       56
song_duration                194
acousticness                  33
speechiness                   38
Name: 99368, dtype: object
```

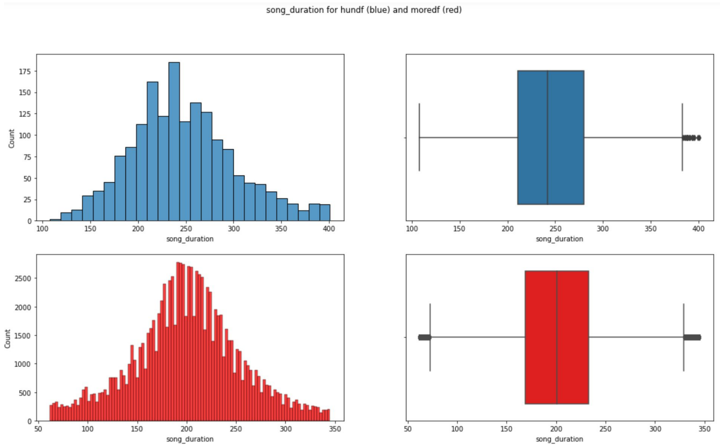
Analysis

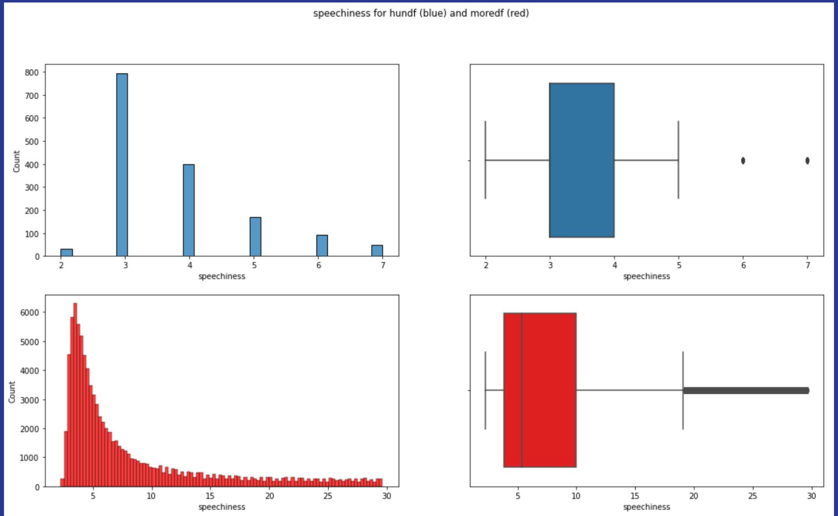
Below is an image of the Pearson Correlation heatmap of variable correlations computed with pandas and seaborn

- Acousticness & Energy Strong negative correlation
- Acousticness & Loudness strong negative correlation
- Lack of large correlation with Total Streams

Analysis Continued





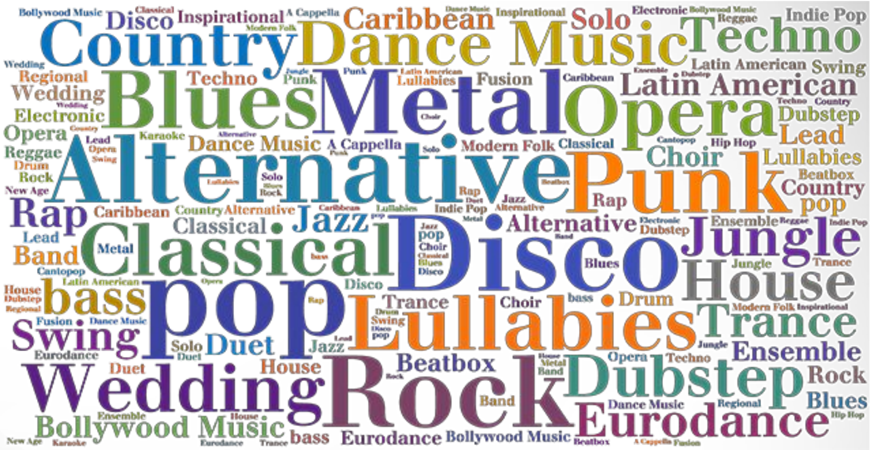


Analysis Continued

P-Value Chart For Top 2000 Vs Non Top 2000			
BPM	**Energy**	**Loudness**	**Liveness**
1.57144615875282e-11	4.472378677849976e-12	6.987422139267793e-12	0.010084174387O4752
Valence	**Song Duration**	**Acousticness**	**Speechiness**
2.2406855730029122e-12	2.722374677863O816e-243	0.008277785598854276	1.3583278171736077e-157

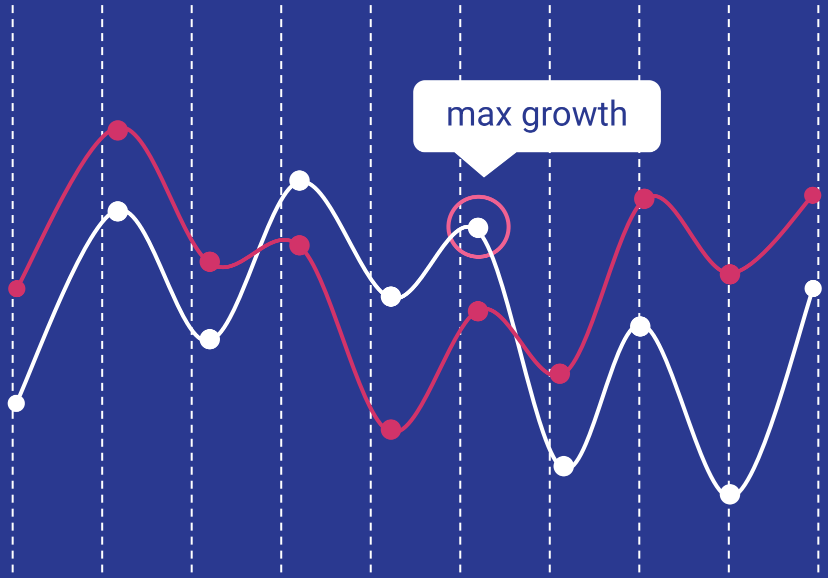
P-Value Chart For More Vs Less Than 600,000 Streams			
BPM	**Energy**	**Loudness**	**Liveness**
0.9262109435798871	0.603835767667495	6.416562893048198e-07	0.08453091360089497
Valence	**Song Duration**	**Acousticness**	**Speechiness**
0.025800604421793064	0.0001634850839733911	0.4052842253766348	8.204908901299235e-06

Limitations & Future Direct



Conclusion

- Acousticness had a negative correlation with energy and loudness variables
- In comparative analysis…
  - All variables in the top 2000 songs were statistically different compared to songs that were not
  - Some variables in songs with more than 600,000 streams were statistically different compared to songs that had less than 600,000 streams
- This may be helpful for up and coming artist who may want to make a song that can get 600,000 streams or more, or even, make it to the top 2000 streamed songs!



Thank you!