Predicting Track Popularity Based on Spotify Song Features

Jacob Silber

Capstone Proposal

Spotify is a streaming service that contains millions of songs, podcasts, and videos from artists worldwide. Listeners subscribe to Spotify and pay a monthly fee for access to all of the content that they offer. Users can have a free account that requires the user to listen to ads dispersed throughout their listening experience, or they can pay $9.99 for ad free streams. Spotify has disrupted the way that people listen to their music and artists put their content out into the world.

I recently discovered that Spotify allows people to access their API to explore audio features and perform in-depth analysis on music tracks that are uploaded to Spotify. I received Spotify API credentials to access the API and begin my exploration of the data that is offered by the streaming service.

Using the json formatted data returned by the Spotify API, which includes quantifiable features about tracks such as "danceability", "energy", "valence", etc., I will use my knowledge of data scraping, parsing, exploratory analysis, python, and multiple linear regression to create a model that predicts any given tracks popularity, another track feature, based on Spotify song features.

1. Question to answer:
   Is it possible to accurately predict any given song's popularity?

2. Data:
   I have access to every album, artist, and track that is on Spotify thanks to Spotify's API. For each track that I query, I will return the name of the track, artist, album name, and the list of attributes for each song: popularity, time, release date, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, livenss, valence, tempo, and time signature.

3. Minimum Viable Product

   - MVP:
     - Script that creates a class called "SpotifyPlaylist"
     - Script that creates a "SpotifyPlaylist" of 10,000 tracks (the maximum number of tracks allowed for a playlist) and exports the playlist and all of its attributes to a CSV
     - A jupyter notebook in which I will read in the CSV that I exported from the previous script where I will display my data exploration and multiple linear regression analysis

Evidence of Spotify Data

track	artist	album	popularity	time	release_date	track_number	danceability	energy	key
Southern Cross	Crosby, Stills & Nash	Greatest Hits	47	280960	2005-03-14	6	0.668	0.437	2
Teach Your Children	Crosby, Stills & Nash	Greatest Hits	40	179040	2005-03-14	10	0.564	0.28	2
Love the One You're With	Crosby, Stills & Nash	Replay	55	184653	1980-12-08	7	0.609	0.489	0
The Weight	The Band	To Kingdom Come (The Definitive Collection)	20	278533	1989	6	0.626	0.379	9
I Shall Be Released	The Band	To Kingdom Come (The Definitive Collection)	17	196533	1989	7	0.474	0.239	4
Sugar Mountain	Neil Young	Decade	0	341493	1977-10-28	6	0.479	0.0882	10
Cowgirl In the Sand	Neil Young	Decade	8	602493	1977-10-28	12	0.266	0.569	5
Heart of Gold	Neil Young	Decade	11	187280	1977-10-28	22	0.538	0.48	7
Comes A Time - Live	Neil Young	Live Rust	0	195466	1979-11-19	3	0.403	0.373	0
Quinn The Eskimo (The Mighty Quinn)	Bob Dylan	Self Portrait	0	165306	1970-06-08	17	0.381	0.537	9
Closer to Fine	Indigo Girls	Indigo Girls (Expanded Edition)	56	241600	1989	1	0.54	0.677	9
Wild World	Yusuf / Cat Stevens	The Best Of Cat Stevens 20th Century Masters The Millennium Collection	43	199240	2007-01-01	2	0.476	0.48	0
Peace Train	Yusuf / Cat Stevens	The Best Of Cat Stevens 20th Century Masters The Millennium Collection	31	249493	2007-01-01	7	0.608	0.585	0
Moonshadow	Yusuf / Cat Stevens	The Best Of Cat Stevens 20th Century Masters The Millennium Collection	32	169360	2007-01-01	6	0.645	0.278	2
Father And Son	Yusuf / Cat Stevens	The Best Of Cat Stevens 20th Century Masters The Millennium Collection	33	220946	2007-01-01	8	0.497	0.335	7
Cat's in the Cradle	Harry Chapin	The Gold Medal Collection	21	222951	1988	6	0.489	0.577	10
Brown Eyed Girl	Van Morrison	Ultimate Van Morrison Collection	0	185626	2012-11-11	1	0.484	0.519	7
Leader of the Band	Dan Fogelberg	The Very Best Of Dan Fogelberg	3	258626	2001	6	0.63	0.208	8
Part of the Plan	Dan Fogelberg	The Very Best Of Dan Fogelberg	0	198040	2001	2	0.643	0.932	2
American Pie	Don McLean	The Best Of Don McLean	71	515866	1988-01-01	1	0.527	0.509	7
Friend of the Devil - 2001 Remaster	Grateful Dead	The Very Best Of Grateful Dead	47	201960	2003-09-16	6	0.647	0.745	7
Uncle John's Band - 2001 Remaster	Grateful Dead	The Very Best Of Grateful Dead	50	282773	2003-09-16	5	0.729	0.425	7
Ripple - 2001 Remaster	Grateful Dead	The Very Best Of Grateful Dead	43	250400	2003-09-16	17	0.608	0.483	7
Fire on the Mountain - 2003 Remaster	Grateful Dead	The Very Best Of Grateful Dead	41	226400	2003-09-16	14	0.713	0.644	1
I Know You Rider - Live in Paris 1972 Remastered Version	Grateful Dead	Europe '72 [Live] [Expanded]	32	303440	2004-10-05	7	0.355	0.762	7
Big Yellow Taxi	Joni Mitchell	Ladies Of The Canyon	68	134800	1970-03-01	10	0.611	0.47	4
You Can Get It If You Really Want	Jimmy Cliff	Many Rivers To Cross : The Best Of Jimmy Cliff	0	161026	2003-01-01	7	0.935	0.544	11
You Can't Always Get What You Want	The Rolling Stones	Let It Bleed	65	448720	1969-12-05	9	0.32	0.62	0
No Woman, No Cry - Live At The Lyceum, London/1975	Bob Marley & The Wailers	Legend (Deluxe Edition)	33	429200	2002-01-01	2	0.642	0.65	1
Redemption Song	Bob Marley & The Wailers	Legend (Deluxe Edition)	38	227333	2002-01-01	12	0.668	0.254	0
Ariel	Dean Friedman	Dean Friedman	35	263639	1977-02-04	2	0.675	0.564	2
Company	Dean Friedman	Dean Friedman	15	200611	1977-02-04	1	0.583	0.245	7
Big Yellow Taxi	Counting Crows	Films About Ghosts: The Best Of...	59	225426	2003-01-01	10	0.685	0.878	1
Me and Julio Down by the Schoolyard	Paul Simon	Paul Simon	70	164813	1972-01-14	6	0.626	0.816	9
Mother and Child Reunion	Paul Simon	Paul Simon	54	187400	1972-01-14	1	0.712	0.632	9
Mrs. Robinson - Single Mix	Simon & Garfunkel	Greatest Hits	53	234346	1972-06-14	1	0.602	0.456	6
The 59th Street Bridge Song (Feelin' Groovy)	Simon & Garfunkel	The Essential Simon & Garfunkel	35	113640	2003-10-14	12	0.428	0.366	3
With A Little Help From My Friends	Joe Cocker	The Anthology	28	313466	1999-01-01	8	0.312	0.434	2
Imagine	Claudio Medeiros	Jazz Do It	0	347946	2010-11-01	13	0.449	0.103	10
The General	Dispatch	Bang Bang [Remastered]	1	245413	2004-10-26	3	0.599	0.673	10
Save Tonight	Eagle-Eye Cherry	Desireless	65	239226	1998-07-13	1	0.547	0.693	0
Here's to the Night	Eve 6	Horrorscope	56	249840	2000	5	0.583	0.652	7
Closing Time	Semisonic	20th Century Masters: The Millennium Collection: Best Of Semisonic	69	274160	2003-01-01	1	0.484	0.841	7
Better Together	Jack Johnson	In Between Dreams	72	207680	2005-01-01	1	0.866	0.347	7
Swing Life Away	Rise Against	Siren Song Of The Counter-Culture	58	200066	2004-01-01	11	0.477	0.426	6
Home	Edward Sharpe & The Magnetic Zeros	Up From Below	26	303200	2009-07-14	6	0.542	0.589	2
Best I Ever Had (Grey Sky Morning)	Vertical Horizon	Everything You Want	57	270800	1999-06-15	4	0.441	0.563	6