

Taylor Swift Analysis Initial Data Report

PROJECT OVERVIEW:

Motivation: Taylor Swift has taken the world by storm with her Era's Tour and record-breaking *Midnights* album. Taylor has gained unique fame for her unprecedented re-recording process to own all of her work. While other analysts have analyzed Taylor's Spotify song attributes, no other analyst has publicly examined variables that are unique to this project. The original steps of the analysis are a part of a CareerFoundry program. However, this is a launch to a larger independent project, in which I seek to build a relational database about Taylor Swift's music and develop data insights in a way that is accessible to music fans and industry experts. As a "Swiftie" myself, I have a particular interest in what variables to investigate. This also means I have bias. However, being a fan means that I have a unique insight into data cleaning, accuracy, and patterns due to domain-specific knowledge developed over years of research.

Objective: The goal of this stage of the project is to conduct an exploratory visual analysis to build an interactive dashboard featuring data about Taylor Swift albums and songs across success measures. It will showcase regression, geospatial, time-series, and clustering analysis gleaned from her musical success. There will be a particular emphasis on Billboard charts, album sales, regional patterns in the United States, song characteristics, and "Taylor's Versions". This project will culminate in an interactive Tableau dashboard featuring technical aspects of the analysis and will launch a larger analysis that is accessible to wider audiences.

Scope: The focus of this analysis is on Taylor Swift's studio albums and songs from 2006 through 2023. The three categories of albums include original versions not owned by Taylor that either have been or will be re-released as Taylor's Versions, albums that have always been owned by Taylor, and "(Taylor's Versions)," which are re-records of originals as Taylor takes back ownership of her work. While Taylor's Version albums have the same name and feature the same tracks as originals, with slight variations, re-records also include additional "vault" tracks. The analysis includes the following albums:

Taylor Swift Debut (October 2006), *Fearless* (November 2008), *Speak Now* (October 2010), *Red* (October 2012), *1989* (October 2014), *Reputation* (November 2017), *Lover* (August 2019), *folklore* (July 2020), *evermore* (December 2020), *Fearless*(Taylor's Version)(April 2021), *Red* (Taylor's Version)(November 2021), *Midnights*(October 2022), *Speak Now* (Taylor's Version)(July 2023), *1989*(Taylor's Version)(October 2023).

The exploratory analysis will set the stage for potential predictive analysis. The predictive aspect of this analysis will focus on her upcoming release of reputation (Taylor's Version) and rumored new album. While her Eras tour is a major aspect of her success, the tour itself is not the focus of this analysis.

KEY QUESTIONS TO EXPLORE

After cleaning and reviewing all dataframes, here are suggested questions to guide further analysis:

1. How does Taylor's music perform across different success measures and are these measures related? **Week 1 Album Sales | Billboard Charts | Spotify Streams | Google Search Trends**

2. Do song characteristics relate to any success metrics?

Spotify's [audio features song measures](#)

3. Are there regional album differences across the United States?

Google Search Trends ([How Does Search Trends Work?](#))

4. How do Taylor's original albums compare to her re-releases (Taylor's Versions)?

DATA SOURCES

Taylor Swift Google Search Popularity (Primary geographical dataset): The main variable of this data frame is the "Google Trend's" search score popularity for the keyword "Taylor Swift." during periods near each album release. Each period has the search trend score for 51 U.S. regions (50 states and D.C.), relevant album name, distance from release date, as well as average United States and worldwide google average search scores derived from a separate file that includes a daily U.S. and Global score from Google Trends. This dataset includes time "periods" (each are 13 days) for each studio album from October 2006-November 21st, 2023. Each album release has 5 "periods" of data (-2, -1, 0, 1, 2), with the first day of period 0 being the album release date. [Source: <https://trends.google.com>, manually collected and compiled into cohesive csv file]

Taylor Swift Spotify Dataset: all albums listed on Spotify for Taylor Swift and is updated monthly. The variables are song characteristics, all of which were set by Spotify. Here is a complete list of what each song measure means [Source: Jarred Priester's dataset from [Kaggle](#), collected using Spotify's API]

Taylor Swift Song List and Spotify Streams : list of every Taylor Swift song from her studio albums, including track name, track length, album, and Spotify streams as of November 9th, 2023. [source: Spotify, with manual collection on 11/9/23]]

Taylor Swift Total Album Sales: A compilation of Taylor Swift's studio album (2006-2023) sales for first-week performance in traditional album sales, album units, streams, vinyl, and CD sales. [Compile from a variety of sources listed in references, manually imputed 11/11/23]

Billboard Hot 100 Songs & Top 200 Albums (Primary time-series data set): Hot 100 data set includes the chart week, chart position, song title, song artist, prior position, peak position, and weeks on a chart for the Billboard Hot 100 songs. There are 100 entries listed for each chart week. The Top 200 Albums data set includes the chart week, rank, song title, song artist, prior rank, peak position, and weeks on a chart for the Top 200 Albums. Both data sets span from 1958 through 12/31/23. The data is from United States listeners only. [source: [GitHub](#) user using the Billboard API to create updated charts from [Billboard.com](#)]

LIMITATIONS

Personal Bias:

As a "Swiftie" myself, I have a particular interest in what variables to investigate and have a positive opinion on Taylor Swift's music and persona. This also means I have bias. However, being a fan means that I have a unique insight into data cleaning, accuracy, and patterns due to domain specific knowledge, which would take years for a person new to this content to build.

Google Search Trends:

- This data was time-consuming to download and unrealistic to maintain long-term.
- Since I manually compiled much of this data, it is prone to human error in input and therefore needs to be checked thoroughly for accuracy, which was noted during my data cleaning and wrangling steps.
- There is some exclusion bias across data sets, since I decided which "periods" of time to include from the Google search trends and at times left out variables based on EDA.
- Another key limitation is that I am assuming that a state's Google search trends score can help demonstrate a state's popularity and preference for Taylor Swift. Another limitation is that the search trends scores are only derived from searches conducted on Google, not all search providers, and therefore have skewed data.

Spotify Audio Features and Streaming Data:

- Spotify was the organization that produced the song data. Certain attributes, such as "track length", are objective, while others, such as "popularity index" are based on Spotify algorithms that are not always explicit. Furthermore, measures such as "danceability" and "valence" are not objective and at times contrast greatly depending on who is scoring each song.

- Spotify does not provide historical data on daily streams for each song across time, but rather only shows the total from a specific point in time across time. The data only captures the total streams on November 9th, therefore each day makes this data less accurate. I did normalize some figures based on dividing total streams by the days since a song was released, which is not fully accurate due to the impact of time for the most recent albums.

Sales Data

- Data on first-week sales was also manually compiled. Taylor Nation, taylorsswift.com, nor Taylor Swift herself release album sales directly to the public. I cross-checked album sales totals whenever possible and left information blank when no reliable source could provide information.
- Traditional album sales differ from equivalent album units, which also takes into account streams. Future analysis could examine albums that have week 1 equivalent album units in order to compare overall first-week success (that includes streams) into account while looking for relationships.

Billboard Charts Data

- Billboard is recognized as a reputable source of data in the music industry. However, they are not explicit about the exact algorithms they use to calculate their charts other than a list of factors they consider.

TERMINOLOGY

Google Trends Search Score: Google Trends score for Interest By Subregion works by displaying numbers between 0 to 100 based on the relative search interest of a topic in a particular location. However, values are calculated based on the fraction of total searches in that geographic area. This means that the subregion with the highest value (100) had a higher proportion of searches for that query; not a higher absolute query count. Scores of 0 mean that an insufficient amount of data was available for the query based on the sample dataset. It doesn't necessarily correlate with 0 search volume in Google's search engines.

Eras Tour: The name of Taylor Swift's most recent (and current) tour, which features songs from all 10 of her studio albums. The tour has been named the highest grossing tour of all time and has created unprecedented fan behavior, popularity, press, and musical success.

Studio Albums: Every album that is recorded in a studio and released to the public. For this analysis, I will use the “Deluxe” version, or equivalent, which is the album version with the most number of songs for each album. The following studio albums are included:

The list of the 14 albums included in this analysis are as follows:

Red (Taylor's Version)	30 tracks, album_id = REDTV
Fearless (Taylor's Version)	26 tracks, album_id = FEARTV
Midnights (The Til Dawn Edition)	23 tracks, album_id = MIDN
1989 (Taylor's Version) [Deluxe]	22 tracks, album_id = 1989TV
Speak Now (Taylor's Version)	22 tracks, album_id = SPEAKTV
Red (Deluxe Edition)	22 tracks, album_id = RED
Speak Now (Deluxe Edition)	20 tracks, album_id = SPEAK
1989 (Deluxe Edition)	19 tracks, album_id = 1989
Fearless Platinum Edition	19 tracks, album_id = FEAR
Lover	18 tracks, album_id = LOVE
evermore (deluxe version)	17 tracks, album_id = EVER
folklore (deluxe version)	17 tracks, album_id = FOLK
reputation	15 tracks, album_id = REP
Debut	15 tracks, album_id = DEBUT

album_id	album_title	release_date	release_year
DEBUT	Debut	10/24/2006	2006
FEAR	Fearless	11/11/2008	2008
SPEAK	Speak Now	10/25/2010	2010
REDTV	Red	10/22/2012	2012
1989	1989	11/15/2014	2014
REP	Reputation	11/10/2017	2017
LOVE	Lover	8/23/2019	2019
FOLK	folklore	7/24/2020	2020
EVER	evermore	12/11/2020	2020
FEARTV	Fearless (Taylor's Version)	4/9/2021	2021
REDTV	Red (Taylor's Version)	11/12/2021	2021
MIDN	Midnights	10/21/2022	2022
SPEAKTV	Speak Now (Taylor's Version)	7/7/2023	2023
1989TV	1989 (Taylor's Version)	10/27/2023	2023

Swiftie: This is the term used to refer to Taylor Swift's fans, who are also called “Swifties”. She is famous for having a very passionate fan base and known for interacting with her fans through her song choices, social media, events, and “Easter Eggs.” There is an endless amount of Taylor “lore” that her fanbase has developed in a way that is unique to Taylor Swift as an artist and person.

reputation Stadium Tour Surprise Song Playlist
 folklore: the long pond studio sessions (from the Disney+ special) [\[deluxe edition\]](#)
 1989 (Taylor's Version)
 Midnights (3am Edition)
 folklore,
 Red,
 Speak Now World Tour Live,
 evermore,
 Speak Now,
 Midnights,
 1989,
 Fearless,
 Live From Clear Channel Stripped 2008

These are albums with either song repeats from long studio albums or non-studio albums and are not part of this analysis.]

(Taylor's Version): Taylor Swift has been recording music since 2006. [In 2019, her record company at the time, Big Machine, sold her music without consent.](#) Swift has since outlined instances of bullying from this record label. She decided to re-record all of the studio albums that were a part of this deal, as a stance for artists to own their own work and to prevent a harmful label from benefiting from her sales. The label was also connected to Kanye West, who infamously publicly shamed Taylor Swift at the 2009 MTV Music Awards. This is the first time a major artist has taken on this endeavor. The rerecords mean that Taylor will own the masters of each new album and therefore will have complete control over the albums. Taylor has been “re-releasing” the rerecords as “(Taylor's Version)” and musical platforms have jumped on board, playing only the Taylor's Version of songs once they are re-released.

DATA PROFILES

Google Search Trends Dataframe: gsearch_st

The original data frame is 3672x14

column	variable description	example	Structured/ unstructured	qualitative or quantitative	ordinal/nomi nal/binary	discrete/ continuous	time variant
period_state	Used as the unique identifier and for sorting the records in accurate order by state	2006/09/28, Alabama	structured	qualitative	ordinal	NA	yes
period	13 day date range used to group time trends from album release dates.	2006/09/28- 2006/10/10	structured	qualitative	ordinal	NA	yes
period_start_date	The first day of the period	2006-09-28	structured	qualitative	ordinal	NA	yes
state	U.S. region (all 50 states plus Washington D.C.)	Colorado	structured	qualitative	ordinal	NA	no
g_search_score	Each state's Google Trends search score for the keyword "Taylor Swift" from the corresponding time period. See Google Trends terminology outlining score methodology.	98	structured	quantitative	NA	discrete	yes
album_event	Name of album (all included albums) or album announcements (for Midnights only) from the corresponding time period.	Debut	structured	qualitative	nominal	NA	no
album_id	Album identifier, which will be used to connect other dataframes.	DEBUT, 1989, LOVE, REP	structured	qualitative	nominal	NA	no
album_release_date	Date of each album's release from the corresponding time period.	2006-10-24	structured	qualitative	ordinal	NA	no
album_release_day	The day of the week the album from the period was released	Tuesday	structured	qualitative	ordinal	NA	no
periods_from_release	How many periods from the album release date, with 0 being the period of album release.	-2, -1, 0, 1, 2, 3	structured	quantitative	NA	discrete	yes
us_search_avg	The average U.S. search score for "Taylor Swift" using Google Trends for the corresponding time period. See Google Trends terminology outlining score methodology. The Google Trends score average was created by aggregating the daily score for that period.	65	structured	quantitative	NA	discrete	yes
us_search_peak	The day corresponding to the specific period when the United States had the highest search score for "Taylor Swift". Currently formatted as mm/dd/yyyy	2006-10-23	structured	qualitative	ordinal	NA	yes
global_search_avg	The average worldwide search score for "Taylor Swift" using Google Trends for the corresponding time period. See Google Trends terminology outlining score methodology. The Google Trends score average was created by aggregating the daily score for that period.	78	structured	quantitative	NA	discrete	yes
global_search_peak	The day corresponding to the specific period when the world had the highest search score for "Taylor Swift". Currently formatted as mm/dd/yyyy	2006-10-23	structured	qualitative	ordinal	NA	yes

Initial Descriptive Analysis: gsearch_st

NUMERICAL VARIABLES

```
In [23]: df_gsearch_st.describe()
```

Out[23]:

	g_search_score	periods_from_release	us_search_avg	global_search_avg
count	3672.000000	3672.000000	3672.000000	3672.000000
mean	54.990741	-0.194444	62.222222	65.222222
std	20.353841	1.595687	16.467038	15.760056
min	0.000000	-5.000000	20.000000	27.000000
25%	44.000000	-1.000000	51.750000	56.000000
50%	56.000000	0.000000	66.000000	69.500000
75%	67.000000	1.000000	75.000000	77.250000
max	100.000000	2.000000	85.000000	86.000000

QUALITATIVE VARIABLES

```
: df_gsearch_st['state'].value_counts()
[12]: state
Alabama      72
Pennsylvania 72
Nevada        72
New Hampshire 72
New Jersey   72
New Mexico   72
New York     72
North Carolina 72
North Dakota 72
Ohio         72
Oklahoma     72
Oregon       72
Rhode Island 72
..          ..
Name: count, dtype: int64

album_event
Debut      255
Speak Now  255
1989       255
Reputation 255
Midnights  255
Speak Now (Taylor's Version) 255
evermore   255
Fearless (Taylor's Version) 255
folklore   255
Lover      255
Red (Taylor's Version) 255
Red        255
Fearless  255
1989 (Taylor's Version) 204
Midnights_announce 153
Name: count, dtype: int64

[14]: df_gsearch_st['album_id'].value_counts()
Out[14]: album_id
MIDN      408
DEBUT     255
SPEAK     255
1989      255
REP       255
SPEAKTV   255
EVER      255
FEARTV    255
FOLK      255
LOVE      255
REDTV     255
RED       255
FEAR      255
1989TV    204
Name: count, dtype: int64

[15]: df_gsearch_st['album_release_date'].value_counts()
Out[15]: album_release_date
2022-10-27    408
2006-10-24    255
2010-10-25    255
2014-10-27    255
2017-11-10    255
2023-07-07    255
2020-12-11    255
2021-04-09    255
2020-07-24    255
2019-08-23    255
2021-11-12    255
2012-10-22    255
2008-11-11    255
2023-10-27    204
Name: count, dtype: int64

In [16]: df_gsearch_st['album_release_day'].value_counts()
Out[16]: album_release_day
Friday      1989
Monday      765
Tuesday     510
Thursday    408
Name: count, dtype: int64

[17]: df_gsearch_st['periods_from_release'].value_counts()
Out[17]: periods_from_release
-2      714
-1      714
0       714
1       714
2      663
-5       51
-4       51
-3       51
Name: count, dtype: int64
```

```
df_gsearch_st.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3672 entries, 0 to 3671
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  --
0   period_state           3672 non-null   object
1   period                 3672 non-null   object
2   period_start_date      3672 non-null   object
3   state                  3672 non-null   object
4   g_search_score         3672 non-null   int64
5   album_event            3672 non-null   object
6   album_id               3672 non-null   object
7   album_release_date     3672 non-null   object
8   album_release_day      3672 non-null   object
9   periods_from_release   3672 non-null   int64
10  us_search_avg          3672 non-null   int64
11  us_peak_search         3672 non-null   object
12  global_search_avg      3672 non-null   int64
13  global_peak_search     3672 non-null   object
dtypes: int64(4), object(10)
memory usage: 401.8+ KB
```

Cleaned Dataframe Variables

Spotify Song Sound Characteristics Dataframe: song_info

All variable descriptions provided by the [Kaggle](#) user, defined by Spotify.

The original data frame is 530 x 17.

column	variable description	example	Structured/ unstructured	qualitative or quantitative	ordinal/nomi nal/binary	discrete/ continuous	time variant
index	Numbered 0 to 529	3	structured	qualitative	ordinal	NA	yes
name	The name of the song	York (Taylor's Version)	structured	qualitative	nominal	NA	no
album	The name of the album	1989 (Taylor's Version) [Deluxe]	structured	qualitative	nominal	NA	no
release_date	The day, month, and year the album was released	45226	structured	qualitative	ordinal	NA	yes
track_number	The order the song appears in the album	1	structured	qualitative	ordinal	NA	no
id	Spotify id for the song	4WUepByoeqcedHoY hSNHRt	structured	qualitative	ordinal	NA	no
uri	Spotify uri for the song	spotify:track:4WUep ByoeqcedHoYhSNHRt	structured	qualitative	ordinal	NA	no
acousticness	A confidence measure from 0.0 to 1.0 whether a track is acoustic. 1.0 represents high confidence that a track is acoustic.	0.00942	structured	quantitative	NA	continuous	no
danceability	A measure from 0.0 to 1.0 for how suitable a track is for dancing based on the combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is the least danceable and 1.0 is the most danceable.	0.757	structured	quantitative	NA	continuous	no
energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.	0.61	structured	quantitative	NA	continuous	no
instrumentalness	Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.	3.66	structured	quantitative	NA	continuous	no
liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.	0.367	structured	quantitative	NA	continuous	yes
loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.	-4.84	structured	quantitative	NA	continuous	no
speechiness	Detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.	0.0327	structured	quantitative	NA	continuous	no
tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.	116.998	structured	quantitative	NA	continuous	no
valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).	0.685	structured	quantitative	NA	continuous	no
popularity	The popularity of the song from 0 to 100, which is influenced by first week listens, total streams of a song, how recently a song has been played, and the frequency the track has been played.	80	structured	quantitative	NA	discrete	yes
duration	The duration of the track in milliseconds.	212600	structured	quantitative	NA	discrete	no

Initial Descriptive Analysis: song_info

NUMERICAL VARIABLES

	track_number	acousticness	danceability	energy	instrumentalness	liveness	loudness	speechiness	tempo	valence	song_popularity	duration_ms
count	262.000000	262.000000	262.000000	262.000000	262.000000	262.000000	262.000000	262.000000	262.000000	262.000000	262.000000	262.000000
mean	11.000000	0.268904	0.585863	0.598149	0.001714	0.145617	-6.938294	0.057705	124.939370	0.416386	68.725191	242937.889313
std	6.582224	0.309119	0.109234	0.180482	0.015231	0.085980	2.728841	0.087605	31.142591	0.194393	13.012846	47622.857455
min	1.000000	0.000191	0.292000	0.118000	0.000000	0.035700	-17.932000	0.023100	68.534000	0.049900	39.000000	107133.000000
25%	6.000000	0.021525	0.515250	0.480250	0.000000	0.094125	-8.463000	0.029325	99.979500	0.265500	59.000000	216023.250000
50%	11.000000	0.111000	0.592500	0.616000	0.000000	0.114500	-6.488500	0.035800	121.967000	0.404500	72.000000	237166.000000
75%	16.000000	0.508500	0.648500	0.735250	0.000022	0.168750	-4.998750	0.053500	146.102750	0.542500	78.000000	263494.000000
max	30.000000	0.971000	0.897000	0.950000	0.179000	0.657000	-1.909000	0.912000	208.918000	0.942000	100.000000	613026.000000

QUALITATIVE VARIABLES

```
df_song_info_stud_alb['album'].value_counts()
Red (Taylor's Version)      30
Fearless (Taylor's Version) 26
Midnights (The Til Dawn Edition) 23
1989 (Taylor's Version) [Deluxe] 22
Speak Now (Taylor's Version) 22
Red (Deluxe Edition)         22
Speak Now (Deluxe Edition)   20
1989 (Deluxe Edition)        19
Fearless Platinum Edition    19
Lover                        18
evermore (deluxe version)     17
folklore (deluxe version)     17
reputation                   15
Debut                        15
Name: album, dtype: int64

df_song_info_stud_alb['album_release_date'].value_counts()
11/12/2021      30
4/9/2021        26
5/26/2023       23
10/27/2023      22
7/7/2023        22
10/22/2012      22
10/25/2010      20
10/27/2014      19
11/11/2008      19
8/23/2019       18
1/7/2021        17
8/18/2020       17
11/10/2017      15
10/24/2006      15

df_song_info_stud_alb['album_id'].value_counts()
REDTV      30
FEARTV     26
MIDN       23
1989TV     22
SPEAKTV    22
RED         22
SPEAK       20
1989        19
FEAR        19
LOVE        18
EVER        17
FOLK        17
REP         15
DEBUT       15
Name: album_id, dtype: int64

df_song_info_stud_alb['track_number_str'].value_counts()
01      14
03      14
04      14
05      14
06      14
07      14
08      14
09      14
10      14
11      14
12      14
13      14
14      14
15      14
02      14
17      12
16      12
18      10
19      9
20      7
21      6
22      6
23      3
24      2
25      2
26      2
27      1
28      1
29      1
30      1
Name: track_number_str, dtype: int64

df_song_info_stud_alb.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 285 entries, 0 to 529
Data columns (total 18 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   song_title          285 non-null    object
 1   album_title         285 non-null    object
 2   album_release_date  285 non-null    object
 3   track_number        285 non-null    int64
 4   acousticness        285 non-null    float64
 5   danceability        285 non-null    float64
 6   energy              285 non-null    float64
 7   instrumentalness    285 non-null    float64
 8   liveness            285 non-null    float64
 9   loudness            285 non-null    float64
10   speechiness         285 non-null    float64
11   tempo               285 non-null    float64
12   valence             285 non-null    float64
13   song_popularity     285 non-null    int64
14   duration_ms         285 non-null    int64
15   track_number_str    285 non-null    object
16   album_id            285 non-null    object
17   song_id             285 non-null    object
dtypes: float64(9), int64(3), object(6)
memory usage: 42.3+ KB
```

Cleaned Dataframe Variables

DATAFRAMES CLEANING: Google and Song Data

Columns dropped, renamed, changed (data type)

Consistency Checks: missing values, duplicates, inaccuracies, outliers

Other

Wrangling Steps		Columns dropped, renamed, changed (data type)	
		Consistency Checks: missing values, duplicates, inaccuracies, other	
dataframe	variable	issue/action	notes
gsearch_st	period_state	none	
	period	none	
	period_start_date	none	
	state	none	
	g_search	none	
	album_event	none	
	album_id	none	
	album_release_date	none	
	album_release_day	none	
	periods_from_release	none	
	us_search_avg	none	
	us_search_peak	none	
	global_search_avg	none	
	global_search_peak	none	
song_info	index	drop	not needed
	name	column renamed	song_name
	album	column renamed , rows removed, value updated	Renamed to album_title, removed all irrelevant albums and save to another dataframe and kept relevant ones as df_song_info_stud_alb, replaced "Taylor Swift" with "Debut" for clarity
	release_date	column renamed	album_release_date
	track_number	data type, accuracy	added additional column called "track_number_str" in case manipulation with string type is handy in future. Also checked track
	id	drop	not needed
	uri	drop	not needed
	acousticness	none	
	danceability	none	
	energy	none	
	instrumentalness	none	
	liveness	none	
	loudness	accuracy	investigate min, max, and means based on positive/negative values
	speechiness	none	
	tempo	none	
	valence	none	
	popularity	none	
	duration	column change	song_duration
	track_number_str	added column, format stri	see above, same steps as with track_number, formatted the str to be two digit with 0X for prior
	album_id	added variable	added based on album_name, to match other
	song_id	added variable	added based on album_id and track_number_str to match individual songs with other dataframes

DATA PROFILES CONTINUED

Song Spotify Plays: song_streams

column	variable description	example	Structured/ unstructured	qualitative or quantitative	ordinal/nomi nal/binary	discrete/ continuous	time variant
song_id	Unique identifier built using the album id and song track number	MIDN05	structured	qualitative	ordinal	NA	no
track_number	The track number of the song in order of the album	05	structured	qualitative	ordinal	NA	no
song_title	The title of the song	Time McGraw	structured	qualitative	nominal	NA	no
spotify_plays	The number of spotify plays as of the most recent update, taken directly from spotify counts on the day of update as a total number of streams ever.	112,816, 077	structured	quantitative	NA	discrete	yes
track_length	The length of the song in MM:SS	3:51	structured	quantitative	NA	discrete	no
album_title	The name of the album, only including albums from the scope of this analysis (new albums August 2022 onward and respective originals if re-released)	Lover	structured	qualitative	nominal	NA	no
album_id	The album identifier, created by Leo to be a shortened title of the album and key to other dataframes	LOVE	structured	qualitative	nominal	NA	no
release_date	The date the album was released, written as mm/dd/yyyy	06/28/2008	structured	qualitative	ordinal	NA	no
updated	Last date that data was sourced from Spotify (streams)	45239	structured	qualitative	ordinal	NA	yes

Initial Descriptive Analysis: song_streams

NUMERICAL VARIABLES

```
stud_alb_song_streams.describe()
```

	tot_spot_plays_song	track_length_secs
count	2.850000e+02	285.000000
mean	2.094616e+08	240.687719
std	2.522123e+08	48.118370
min	4.443209e+06	107.000000
25%	5.012600e+07	212.000000
50%	1.172463e+08	235.000000
75%	2.516694e+08	261.000000
max	1.591542e+09	613.000000

```
stud_alb_song_streams['track_number'].value_counts()
track_number
1    14
3    14
4    14
5    14
6    14
7    14
8    14
9    14
10   14
11   14
12   14
13   14
14   14
15   14
2    14
17   12
16   12
18   10
19    9
20    7
22    6
21    6
23    3
24    2
25    2
26    2
27    1
28    1
29    1
30    1
Name: count, dtype: int64
```

QUALITATIVE VARIABLES

```
stud_alb_song_streams['album_release_date']
```

```
album_release_date
11/12/2021    30
4/9/2021      26
5/26/2023     23
10/22/2012    22
10/27/2023    22
7/7/2023      22
10/25/2010    20
10/27/2014    19
11/11/2008    19
8/23/2019     18
8/18/2020     17
1/7/2021      17
10/24/2006    15
11/10/2017    15
Name: count, dtype: int64
```

```
stud_alb_song_streams['album_id']
```

```
album_id
REDTV    30
FEARTV   26
MIDN     23
RED      22
1989TV   22
SPEAKTV  22
SPEAK    20
1989     19
FEAR     19
LOVE     18
FOLK     17
EVER     17
DEBUT    15
REP      15
Name: count, dtype: int64
```

```
stud_alb_song_streams['album_title'].value_counts()
```

```
album_title
Red (Taylor's Version)    30
Fearless (Taylor's Version) 26
Midnights (The Til Dawn Edition) 23
Red (Deluxe Edition)      22
1989 (Taylor's Version) [Deluxe] 22
Speak Now (Taylor's Version) 22
Speak Now (Deluxe Edition)  20
1989 (Deluxe Edition)       19
Fearless Platinum Edition  19
Lover                       18
folklore (deluxe version)   17
evermore (deluxe version)   17
Debut                       15
reputation                  15
Name: count, dtype: int64
```

Cleaned Dataframe Variables: : song_streams

```
stud_alb_song_streams.info()

<class 'pandas.core.frame.DataFrame'>
Index: 285 entries, 0 to 529
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   song_id             285 non-null    object
 1   track_number        285 non-null    object
 2   song_title          285 non-null    object
 3   tot_spot_plays_song 285 non-null    int32
 4   track_length        285 non-null    object
 5   album_title         285 non-null    object
 6   album_id            285 non-null    object
 7   album_release_date  285 non-null    object
 8   last_update         285 non-null    object
 9   track_length_secs   285 non-null    int64
dtypes: int32(1), int64(1), object(8)
memory usage: 23.4+ KB
```

Total Album Sales Dataframe: album_sales

Table: The original data frame is 14x17

column	variable description	example	Structured/ unstructured	qualitative or quantitative	ordinal/nominal/ binary	discrete/ continuous	time variant
album_id	shortened album identifier	FEAR	structured	qualitative	nominal	NA	no
album_title	Name of album	Debut	structured	qualitative	nominal	NA	no
album_announce_mo	Month the album was announced	Aug	structured	qualitative	ordinal	NA	no
album_announce_day	Day of month the album was announced	13	structured	qualitative	ordinal	NA	no
album_announce_date	Date the album was announced, as yyyy-mo-dd	2021-12-10	structured	qualitative	ordinal	NA	no
album_announce_year	Year the album was announced	2019	structured	qualitative	ordinal	NA	no
album_release_mo	Month the album was released	Oct	structured	qualitative	ordinal	NA	no
album_release_day	Day of month the album was released	22	structured	qualitative	ordinal	NA	no
album_release_year	Year album was released	2006	structured	qualitative	ordinal	NA	no
album_release_date	Date album was released, as yyyy-mm-dd	2006-10-24	structured	qualitative	ordinal	NA	no
chart_date	Chart date that captured week 1 sales	2017-12-05	structured	qualitative	ordinal	NA	no
wk1_equiv_album_units	Number of equivalent album units. This is an industry standard to normalize sales figures to represent one unit as one album sale, which is derived from downloads, streams, and traditional sales.	592,300	structured	quantitative	NA	discrete	no
wk1_trad_album_sales	Physical and digital item sales (CDs, vinyls, album downloads)	1208000	structured	quantitative	NA	discrete	no
wk1_sea_units	Sales of digital downloads based on individual songs or entire albums, which are purchased.	227000	structured	quantitative	NA	discrete	no
wk1_on_demand_streams	Number of streams of an artist songs across streaming platforms (Spotify, Apple Music, YouTube, etc.)	226000000	structured	quantitative	NA	discrete	no
wk1_vinyl_sales	Total physical vinyl sales	67000	structured	quantitative	NA	discrete	no
wk1_cd_sales	Total CD sales.	554000	structured	quantitative	NA	discrete	no

Initial Descriptive Analysis: album_sales

NUMERICAL VARIABLES

	album_announce_day	album_announce_yr	album_release_day	album_release_year	w1_trad_album_sales
count	14.000000	14.000000	14.000000	14.000000	1.400000e+01
mean	15.142857	2016.857143	17.214286	2016.857143	7.429857e+05
std	7.220240	5.789342	7.083877	5.789342	4.620321e+05
min	4.000000	2006.000000	7.000000	2006.000000	3.900000e+04
25%	10.250000	2012.500000	11.000000	2012.500000	4.035000e+05
50%	14.000000	2019.500000	18.000000	2019.500000	6.470000e+05
75%	19.000000	2021.000000	23.750000	2021.000000	1.191000e+06
max	28.000000	2023.000000	27.000000	2023.000000	1.359000e+06

	wk1_equiv_album_units	w1_sea_units	wk1_on_demand_streams		wk1_vinyl_sales
count	1.200000e+01	8.000000	8.000000e+00	count	5.000000
mean	8.751917e+05	226125.000000	2.969831e+08	mean	343500.000000
std	4.620988e+05	93601.644522	1.226376e+08	std	278608.147763
min	2.910000e+05	109000.000000	1.429800e+08	min	67000.000000
25%	5.709750e+05	173000.000000	2.246225e+08	25%	114000.000000
50%	7.810000e+05	212000.000000	2.792075e+08	50%	268500.000000
75%	1.248500e+06	242250.000000	3.212950e+08	75%	575000.000000
max	1.653000e+06	419000.000000	5.492600e+08	max	693000.000000

QUALITATIVE VARIABLES

```
179]: album_id
      DEBUT      1
      FEAR       1
      SPEAK      1
      RED        1
      1989       1
      REP        1
      LOVE       1
      FOLK       1
      EVER       1
      FEARTV     1
      REDTV      1
      MIDN       1
      SPEAKTV    1
      1989TV     1
      Name: count, dtype: int64

      album_announce_mo
      August      5
      June        3
      September   1
      August      1
      July        1
      December    1
      February    1
      May         1
      Name: count, dtype: int64

      album_release_day
      19      2
      13      2
      24      2
      15      1
      4        1
      10       1
      11       1
      18       1
      28       1
      5        1
      9         1
      Name: count, dtype: int64

      album_release_year
      2020      2
      2021      2
      2023      2
      2006      1
      2008      1
      2010      1
      2012      1
      2014      1
      2017      1
      2019      1
      2022      1
      Name: count, dtype: int64

      df_album_sales['album_anno']
      album_announce_date
      2006-06-19      1
      2008-09-15      1
      2010-08-04      1
      2012-08-13      1
      2014-08-19      1
      2017-08-24      1
      2019-06-13      1
      2020-07-24      1
      2020-12-10      1
      2021-02-11      1
      2021-06-18      1
      2022-08-28      1
      2023-05-05      1
      2023-08-09      1
      Name: count, dtype: int64

      df_album_sales['album_rele']
      album_release_mo
      October      5
      November     4
      July         2
      August        1
      December     1
      April         1
      Name: count, dtype: int64

      album_release_date
      2006-10-24      1
      2008-11-11      1
      2010-10-25      1
      2012-10-22      1
      2014-11-15      1
      2017-11-10      1
      2019-08-23      1
      2020-07-24      1
      2020-12-11      1
      2021-04-09      1
      2021-11-12      1
      2022-10-21      1
      2023-07-07      1
      2023-10-27      1
      Name: count, dtype: int64
```

Album Sale Subsets:

Note: These small dataframes will be incorporated into other larger dataframes based on next steps of the analysis. The isolation of variables allows for numerical statistics and pattern finding.

```
Will have the following subsets in order to compute descriptive stats:

alb_sales_trad_all = all albums are included, wk1_trad_album_sales is only sales figure ('album_id
== 'DEBUT', 'FEAR', 'SPEAK', 'REDTV', '1989', 'REP', 'LOVE', 'FOLK', 'EVER', 'FEARTV', 'REDTV',
'MIDN', 'SPEAKTV', '1989TV')

alb_sales_unit = 13/14 albums (DEBUT not included), wk1_equiv_album_units and wk1_trad_album_sales
included ('album_id == 'FEAR', 'SPEAK', 'REDTV', '1989', 'REP', 'LOVE', 'FOLK', 'EVER', 'FEARTV',
'REDTV', 'MIDN', 'SPEAKTV', '1989TV')

alb_sales_streams = All albums Lover onward (8/14 albums), wk1_equiv_album_units and
wk1_trad_album_sales, w1_sea_units, and wk1_on_demand_streams included ('album_id == 'LOVE', 'FOLK',
'EVER', 'FEARTV', 'REDTV', 'MIDN', 'SPEAKTV', '1989TV')

alb_sales_vinyl = All albums FEARTV onward (5/14 albums), wk1_equiv_album_units and
wk1_trad_album_sales, w1_sea_units, wk1_on_demand_streams, and wk1_vinyl_sales included ('album_id ==
'FEARTV', 'REDTV', 'MIDN', 'SPEAKTV', '1989TV')

Will also export entire clean df for reference, with all 14 albums.
```

Cleaned Dataframe Variables:

```
df_album_sales.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 17 columns):
#   Column              Non-Null Count  Dtype
---  -
0   album_id             14 non-null    object
1   album_title          14 non-null    object
2   album_announce_mo    14 non-null    object
3   album_announce_day   14 non-null    int64
4   album_announce_yr    14 non-null    int64
5   album_announce_date  14 non-null    object
6   album_release_mo     14 non-null    object
7   album_release_day    14 non-null    int64
8   album_release_year   14 non-null    int64
9   album_release_date   14 non-null    object
10  chart_date           9 non-null     object
11  wk1_equiv_album_units 14 non-null    object
12  w1_trad_album_sales   14 non-null    int64
13  w1_sea_units          14 non-null    object
14  wk1_on_demand_streams 14 non-null    object
15  wk1_vinyl_sales       14 non-null    object
16  wk1_cd_sales          14 non-null    object
dtypes: int64(5), object(12)
memory usage: 2.0+ KB
```

Billboard Hot 100 Dataframe: [bill_top_songs](#)

Table: The original data frame is 340,600 x 7

column	variable description	example	Structured/ unstructured	qualitative or quantitative	ordinal/nomi nal/binary	discrete/ continuous	time variant
chart_week	Week of the chart, each with 100 entries for top 100 songs according to Billboard.com rankings	11/11/2023	structured	qualitative	ordinal	NA	yes
current_position	Position on chart, ranging from 1 to 100. 1 is the best rank	2	structured	qualitative	ordinal	NA	yes
title	Song title	Is It Over Now? (Taylor's Version) [From The Vault]	structured	qualitative	nominal	NA	no
performer	Song artist	Taylor Swift	structured	qualitative	nominal	NA	no
last_week	Position on the chart from the prior week	0	structured	qualitative	ordinal	NA	yes
peak_position	Highest (lowest in number, highest in ranking) position a song has held	3	structured	qualitative	ordinal	NA	yes
wks_on_chart	Number of weeks a song has been on the Top 100 chart.	4	structured	quantitative	NA	discrete	yes

Initial Descriptive Analysis: bill_top_songs

NUMERICAL VARIABLES

```
bill_top_songs_ts.describe()
```

	current_song_position	last_week	peak_pos	wks_on_chart
count	1337.000000	1165.000000	1337.000000	1337.000000
mean	39.981301	32.902146	17.656694	12.462229
std	26.651001	24.921127	21.187738	11.594714
min	1.000000	0.000000	1.000000	1.000000
25%	17.000000	12.000000	2.000000	3.000000
50%	37.000000	29.000000	9.000000	9.000000
75%	61.000000	48.000000	23.000000	18.000000
max	100.000000	100.000000	94.000000	53.000000

QUALITATIVE VARIABLES

```
bill_top_songs_ts['chart_week'].value_counts(dropna=False)
: chart_week
11/11/2023    22
7/22/2023     22
11/27/2021    21
11/5/2022     19
11/12/2022    19
..
4/7/2018       1
3/31/2018       1
3/24/2018       1
1/27/2018       1
9/23/2006       1
Name: count, dtype: int64

bill_top_songs_ts['artist'].value_counts(dropna=False)
: artist
Taylor Swift    1337
Name: count, dtype: int64

bill_top_songs_ts['song_title'].value_counts(dropna=False)
: song_title
Anti-Hero                53
Shake It Off              50
You Belong With Me        50
Love Story                49
Teardrops On My Guitar    48
..
Come Back...Be Here (Taylor's Version) 1
Starlight (Taylor's Version)            1
Forever & Always (Taylor's Version)      1
Fearless (Taylor's Version)             1
Long Story Short                        1
Name: count, Length: 204, dtype: int64

bill_top_songs_ts['current_song_position'].value_counts(dropna=False)
: current_song_position
1      31
39     30
34     27
4      27
5      25
..
100     4
92      4
58      4
95      3
99      2
Name: count, Length: 1

bill_top_songs_ts['peak_pos'].value_counts(dropna=False)
: peak_pos
1      233
2     130
6      74
13     68
4      59
...
43      1
73      1
31      1
70      1
93      1
..
Name: count, Length: 94, dtype: int64
```

Cleaned Dataframe Variables:

```
bill_top_songs_ts.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1337 entries, 0 to 89485
Data columns (total 7 columns):
#   Column                      Non-Null Count  Dtype
---  -
0   chart_week                  1337 non-null   object
1   current_song_position       1337 non-null   int64
2   song_title                  1337 non-null   object
3   artist                      1337 non-null   object
4   last_week                   1165 non-null   float64
5   peak_pos                    1337 non-null   int64
6   wks_on_chart                1337 non-null   int64
dtypes: float64(1), int64(3), object(3)
```

Billboard Top 200 Albums Dataframe: [bill_top_albums](#)

Table: The original data frame is 590,692 x 7

column	variable description	example	Structured/ unstructured	qualitative or quantitative	ordinal/nomi nal/binary	discrete/ continuous	time variant
chart_week	Week of the chart, each with 100 entries for top 100 songs according to Billboard.com rankings	11/11/2023	structured	qualitative	ordinal	NA	yes
current_position	Position on chart, ranging from 1 to 100. 1 is the best rank	2	structured	qualitative	ordinal	NA	yes
title	Album title	Midnights	structured	qualitative	nominal	NA	no
performer	Album artist	Taylor Swift	structured	qualitative	nominal	NA	no
last_week	Position on the chart from the prior week	0	structured	qualitative	ordinal	NA	yes
peak_position	Highest (lowest in number, highest in ranking) position a song has held	3	structured	qualitative	ordinal	NA	yes
wks_on_chart	Number of weeks a song has been on the Top 100 chart.	4	structured	quantitative	NA	discrete	yes

Initial Descriptive Analysis: [bill_top_albums](#)

NUMERICAL VARIABLES

```
In [254]: bill_top_alb_ts.describe()
:
```

	current_week	last_week	peak_pos	wks_on_chart
count	2568.000000	2568.000000	2568.000000	2568.000000
mean	73.991433	66.815421	2.192368	124.664330
std	58.297993	56.059475	4.259821	99.622675
min	1.000000	0.000000	1.000000	1.000000
25%	24.000000	19.000000	1.000000	46.000000
50%	58.000000	51.000000	1.000000	104.000000
75%	122.000000	112.000000	1.000000	176.000000
max	200.000000	200.000000	88.000000	464.000000

```
In [254]: bill_top_alb_ts.shape
Out[254]: (2568, 7)
```

```
In [254]: bill_top_alb_ts['alb_title'].value_counts(dropna=False)
:
```

alb_title	count
1	64
2	43
9	35
21	34
10	33
..	..
154	3
184	3
73	3
198	2
118	2

Name: count, Length: 2568

QUALITATIVE VARIABLES

```
bill_top_alb_ts['alb_title'].value_counts(dropna=False)
alb_title
1989                                464
Taylor Swift                        284
reputation                         267
Fearless                          261
Lover                              219
Speak Now                          193
Red                                185
Folklore                           171
Evermore                           151
Fearless (Taylor's Version)        121
Red (Taylor's Version)              103
Midnights                          54
The Taylor Swift Holiday Collection (EP)  43
Beautiful Eyes (EP)                 21
Speak Now (Taylor's Version)         17
Speak Now: World Tour Live CD + DVD  11
Lover: Live From Paris                1
Folklore: The Long Pond Studio Sessions (Soundtrack)  1
1989 (Taylor's Version)              1
Name: count, dtype: int64
```

Cleaned Dataframe Variables:

```
Index: 2568 entries, 0 to 177618
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   chart_week  2568 non-null   object
1   current_week  2568 non-null   int64
2   alb_title    2568 non-null   object
3   artist       2568 non-null   object
4   last_week    2568 non-null   int64
5   peak_pos     2568 non-null   int64
6   wks_on_chart 2568 non-null   int64
```

DATAFRAME CLEANING: album sales, Billboard Hot Songs, Billboard Top Albums

Columns dropped, renamed, changed (data type)

Consistency Checks: missing values, duplicates, inaccuracies, outliers

Other

album_sales	album_id	duplicate/accuracy	REDTV to "RED"
	album_title	formatting	reputation
	album_announce_mo	none	
	album_announce_day	none	
	album_announce_date	none	
	album_announce_year	none	
	album_release_mo	none	
	album_release_day	none	
	album_release_year	none	
	album_release_date	none	
	chart_date	none for now	mixed type
	wk1_equiv_album_units	none	See image for subset info.
	wk1_trad_album_sales	missing info treatment/subset	See image for subset info.
bill_top_songs	wk1_sea_units	missing info treatment/subset	See image for subset info.
	wk1_on_demand_streams	missing info treatment/subset	See image for subset info.
	wk1_vinyl_sales	missing info treatment/subset	See image for subset info.
	wk1_cd_sales	missing values	subset since not clear pattern of missing values and don't have access to this
bill_top_songs	chart_week	none	Created subset TS only df
	current_position	name change	song_current_position
	title	name change	song_title
	performer	name change	artist
	last_week	none	
	peak_position	name change	song_peak_pos
	wks_on_chart	name change	song_wks_on_chart
bill_top_albums	chart_week	none	
	current_week	name change	alb_current_pos
	title	name change	alb_title
	performer	name change	artist
	last_week	none	
	peak_position	name change	alb_peak_pos
	wks_on_chart	name change	alb_wks_on_chart

Appendix

References and Further Reading:

[Taylor Swift's '1989 \(Taylor's Version\)' Google vault puzzles leave fans frustrated | CNN](#)
[Taylor Swift First With Entire Top 10 on Hot 100, 'Anti-Hero' No. 1 – Billboard:](#)
[Taylor Swift \(forbes.com\)](#)
[Taylor Swift's Record-Setting Eras Tour Ticket Sales: Who Gets Paid? – Billboard](#)
[Taylor Swift's 'Cruel Summer' Now Her Longest Leading No. 1 on Pop Airplay – Billboard](#)
[Taylor Swift's 'Is It Over Now?' Debuts at No. 1 on Hot 100 – Billboard](#)
[Taylor Swift '1989' Nearly Half of Albums Sold in the U.S. Last Week – Billboard](#)
[Taylor Swift '1989 \(Taylor's Version\)' Breaks Spotify Streaming Record – Billboard](#)
[The 10 Most-Streamed Artists of All Time on Spotify \(Sep 2023\) \(ledgernote.com\)](#)
[Taylor Swift's 15 Most Popular Songs, Ranked \(According To Spotify\) \(msn.com\)](#)
[The STAGGERING Numbers Behind Taylor Swift's Tour, How It's Boosting America \(msn.com\)](#)
[Spotify Popularity — A unique insight into the Spotify algorithm and how to influence it | by Oskar Eichler | The Songstats Lab](#)
[Taylor Swift Charts All 21 Songs From '1989 \(Taylor's Version\)' on the Hot 100 \(msn.com\)](#)

Additional Data Sources for Future Use:

Eras Tour Dates and Locations from taylorswift.com
[Taylor Swift's Tweets\(Kaggle.com\)](#)
[Most Streamed Spotify Songs 2023 \(Kaggle.com\)](#)
[Taylor Swift - Spotify Top Albums \(kworkb.net\)](#)
[Spotify Daily Chart - United States \(kworkb.net\)](#)
Album Images sourced from taylorswift.com

Taylor Swift Projects Completed By Other Analysts:

[The Evolution of Taylor Swift's Music as Told by Spotify Data | by Emily Wang | Medium](#)
[An analysis of Taylor Swift's Spotify data to see what makes her music popular](#)
[Taylor Swift Data Analysis : Is Taylor Swift's Song Making Your Mood? - SwiftUp \(saraswatisepti.com\)](#)
[Text analysis and data visualization with Taylor Swift songs \(github.com\)](#)
[A Data Scientist Breaks Down All 10 Taylor Swift Albums – A Dash of Data](#)
[An Exploratory Data Analysis of Taylor Swift's Music | by Aimi Wen | Medium](#)
[Taylor Swift Data from Spotify and YouTube \(kaggle.com\)](#)
[Midnights Single Prediction | Kaggle](#)