Ex 6.1 Project Brief

Summary and Source: Anime, and Asian media in general, has only grown in popularity in the Western world, especially in the last decade. The younger generations have taken a liking to this ever-growing form of media, and because of this, I wanted to look into what is popular and how the production of these shows has changed in a way.
The dataset is taken from here on Kaggle, but only the anime set is going to be used in this project. It takes information from one of the more popular sites, myanimelist.com, where fans can keep track of which series or movies they've seen, want to see, or are currently watching.

Collection: The site is made by fans for fans. People can submit series/movies/music videos, and it gets approved as long as there's proof it exists (usually a link leading to a video or article).

Limitations: Again, this information is submitted by fans. Dozens of people submit something almost daily on the site, but many submissions are not as thorough as others. Looking through the data set, it's easy to see how many cells are blank because much of the information is unknown. It's gotten better for more recent series because of its popularity, but some submissions are of things that were never released outside of Asia, so information on them is limited. This also isn't entirely up to date

Ethics: There isn't any PII, but some of the series in this data set may not be suitable for all audiences. There may also be bias since all of these scores are averaged by the thousands of users who submitted a rating on the site.

Hypothesis: Anime has become more accessible because of streaming sites like Netflix and Crunchyroll. If this is the case, then series of the last decade are more popular than those from before 2010.

Key Questions:
1) How many series do we get per year compared to a decade ago? Two decades ago?
2) Does time have an effect on popularity? (Classic 1980/90s vs 2010/20s Older = better for some fans)
3) Are shorter series or movies more popular than longer series/movies?
4) Which source is most common?
5) Does the source affect popularity?
6) How many users scored the series as opposed to just adding it?

Column Name	Description	Missing/Action Taken/Reason	Time-Invariant /Variant	Structured/ Unstructured	Quantitative/ Qualitative	Nominal/Ordinal /Discrete/Continuous
anime_id	Numeric number of anime series/movie		Invariant	Structured	Quantitative	Nominal/Discrete
title	Title of anime series/movie (usually transliterated from JPN to ENG)		Invariant	Unstructured	Qualitative	Nominal
type	Type of media (TV series, Special, Movie)	Missing 58 values	Invariant	Structured	Qualitative	Nominal
score	Average score out of 10	Missing 8935 values	Invariant	Structured	Quantitative	Continuous
scored_by	Number of users who scored the media		Variant	Structured	Quantitative	Discrete
status	Current status of media (Airing, finished)		Variant	Structured	Qualitative	Nominal
episodes	Total number of episodes	Missing 547 values	Variant	Structured	Quantitative	Discrete
start_date	Beginning airdate (MM/DD/YY)	Dropped, unnecessary	Variant	Structured	Qualitative	Nominal
end_date	Ending airdate (MM/DD/YY)	Dropped, unnecessary	Variant	Structured	Qualitative	Nominal
source	Where the series/movie was adapted from (Manga, Webcomic, original)	Missing 3561 values	Invariant	Structured	Qualitative	Nominal
members	Total number of members who have added the anime to their list		Variant	Structured	Qualitative	Discrete

favorites	Total number of members who have added the anime to their list of favorites		Variant	Structured	Qualitative	Discrete
episode_duration	Total run time of episode/movie (23-25 minutes, or 60-120 minutes for films)	Missing 598 values	Variant	Structured	Quantitative	Continuous
total_duration	Total episode of series/movie			Structured	Quantitative	Continuous
rating	MPA rating (G, PG, PG-13, R, etc.)	Made letters (G, PG, R) uppercase		Structured	Qualitative	Nominal/Ordinal
sfw	Is the series safe for work (True/False)	Dropped, unnecessary	Invariant	Structured	Qualitative	Nominal
approved	When a series is added, there is a wait period to see if it is approved to be added to the database.	Dropped, unnecessary	Invariant	Structured	Qualitative	Nominal
created_at	When the anime was added	Dropped, unnecessary	Invariant	Structured	Qualitative	Nominal
updated_at	When the entry (or its info) was updated	Dropped, unnecessary	Variant	Structured	Qualitative	Nominal
start_year	Year it premiered	Missing 4684 values, took the year listed in the start date	Variant	Structured	Quantitative	Ordinal
start_season	Season it premiered (Winter, spring, summer, fall)	Capitalized the first letter of each season for uniform	Variant	Structured	Qualitative	Nominal
real_start_date	Same as start_date	Dropped, unnecessary	Variant	Structured	Qualitative	Nominal

real_end_date	Same as end_date	Dropped, unnecessary	Variant	Structured	Qualitative	Nominal
broadcast_day	Day of the week it airs/aired	Dropped, missing 21535 values	Variant	Structured	Qualitative	Nominal
broadcast_time	Time of day it airs/aired (24-hour)	Dropped, missing 21672 values	Variant	Structured	Qualitative	Nominal
genres	Genres listed for the series/movie	Replaced 4988 **[]** with Unknown	Variant	Unstructured	Qualitative	Nominal
themes	Themes shown or covered in series/movie	Dropped, missing 11020, unnecessary	Variant	Unstructured	Qualitative	Nominal
demographics	Who it is aimed towards	Changed ['Kids', Shounen'] and ['Kids', 'Shoujo'] to simply Kids, and removed [' '] from each value	Variant	Unstructured	Qualitative	Nominal
studios	Animation studio(s)	Dropped, missing 10418 values	Variant	Unstructured	Qualitative	Nominal
producers	Studios/Companies who produced the series/movies	Dropped, missing 13301 values	Variant	Unstructured	Qualitative	Nominal
licensors	Companies who licensed and distributed the series/movie outside of Japan	Missing 20234 values, dropped	Variant	Unstructured	Qualitative	Nominal
synopsis	Synopsis of series/movie	Dropped, not necessary	Variant	Unstructured	Qualitative	*Text-N/A*
background	A bit of trivia for the series/movies	Dropped, not necessary	Variant	Unstructured	Qualitative	*Text-N/A*

main_picture	The photo shown on the website for the series.	Dropped, not necessary	Invariant	Unstructured	Qualitative	*Link-N/A*
url	Link to the website for the series/movie	Dropped, not necessary	Invariant	Unstructured	Qualitative	*Link-N/A*
trailer_url	Trailer for the series/movie	Dropped, not necessary	Invariant	Unstructured	Qualitative	*Link-N/A*
title_english	Title translated into English	Dropped, not necessary	Invariant	Unstructured	Qualitative	Nominal
title_japanese	The title in its original Japanese	Dropped, not necessary	Invariant	Unstructured	Qualitative	Nominal
title_synonyms	Alternative titles	Dropped, not necessary	Invariant	Unstructured	Qualitative	Nominal
country_origin	Created to show the country that created the anime	I created this myself cross-checking a few sites and sources	Invariant	Unstructured	Qualitative	Nominal
story_origin	For any series that was based on a book or existing story. Created to show the country that the story originated.	I created this myself cross-checking a few sites and sources	Invariant	Unstructured	Qualitative	Nominal