**Data exploration.**

Starting Data Sources

* <http://screencritix.com/the-influence-of-anime-on-video-games/>
* <https://en.wikipedia.org/wiki/List_of_anime_based_on_video_games>
* <https://cartoonresearch.com/index.php/the-game-influence/>
* <https://www.anime-planet.com/anime/all>
* <https://myvideogamelist.com/allplatforms> (Specifically PS4, Switch, Wii/Wii U )
* <https://animanga.fandom.com/wiki/Special:Ask?offset=0&limit=500&q=%5B%5BCategory%3ASeries%5D%5D+%5B%5BMedia%3A%3AAnime%5D%5D&p=mainlabel%3D%2Fformat%3Dtable&po=%3FMedia%0A%3FDemographic%0A%3FGenre%0A%3FTag%3DTheme%0A>
* rawgpy API - <https://pypi.org/project/rawgpy/>

The Anime planet website was eventually dropped as a source because after attempting to scrape data it was found to be un-needed and there were technical challenges involved with the hover text where most of the data was stored. Other sources were found and we proceeded with the project.

Similarly, myvideogamelist.com was dropped as there were too many duplicates, the data was considered redundant.

Two tables were created, one based on anime games taken from the wiki list shown above. The other table was taken from csv files downloaded from the animanga.fandom.com website.

The tables were first cleaned in pandas, modifying column names, dropping duplicate data, adding id columns, concatenating different tables, etc.

The anime list (table 1) had to be filtered for only anime (manga was excluded). This was done by separating the media type and filtering for ‘Anime’ and ‘Games’. After the filtering was complete the dataframes were combined into one and output to a postgres database. There are 3 columns in this table: the primary key, the title of the original anime series and its genre.

The anime / game list (table 2) was filtered through the rawgpy API to find game genre info and then insert it into the data set. Duplicate names were dropped, and the genre data was edited to remove list brackets. Next the table was loaded into the postgres database where table 1 was stored. There are 7 columns in this table: the primary key, the title of the anime series, the game title of the series from the API, the game title of the series from the Wikipedia list, the genre of the game, the release year of the game and the rating of the game.