1. Movie & TV Show on Netflix

A list of movies and TV shows on Netflix, including information about show\_id, type, title, director, cast, country, date\_added, release\_year, rating, duration, listed\_in, description

<https://www.kaggle.com/shivamb/netflix-shows>

Something we like about the dataset:

* It is supportive of multiple search methods. Users can search from different aspects, such as title, genre, release year, or ratings. The dataset contains various information about movies and TV shows on Netflix.
* We can explore into network analysis of Actors / Directors.
* It is interesting to understand what content is available in different countries.

Something we don’t like about the dataset:

* It only contains one table.
* It has some missing data that “actors contained in this show” didn’t give a full list of actors.

2. Formula 1 Race Data

This dataset contains data from 1950 all the way through the 2017 season, and consists of tables listing information about constructors, race drivers, lap times, pit stops and more.

<https://www.kaggle.com/cjgdev/formula-1-race-data-19502017?select=circuits.csv>

Something we like about the dataset:

* The dataset consists of several csv tables that can have connections through driverID, so we are able to return different kinds of information with respect to users’ requests, such as lap times and pit stops.
* The data contains information from 1950 to 2017, allowing users and us to explore formula 1 race cars’ changes on a large scale of time.
* The dataset contains links of wikipedia pages for drivers and seasons, which we think can be linked from our web page.

Something we don’t like about the dataset:

* The dataset contains too much information which may extend the loading time for users to obtain search results after entering keywords.

3. Covid-19 Data

COVID Daily Updates of Confirmed number of cases, number of recovered patients, number of deaths, and date of report from January 22nd to October 3rd. It has Covid-19 data of different countries, including data for different states of the US and different provinces of China, with data of their latitudes and longitudes.

<https://www.kaggle.com/gpreda/coronavirus-2019ncov>

Something we like about the dataset:

* It contains data of latitude and longitude so we can compare Covid-19 information from the geographical perspective.
* We can easily sort by date and get the number of increased cases in each region
* There are many topics related to Covid-19 data. We are able to explore our interests to the search of information in other fields, such as connection of Covid cases with economic status of the countries.

Something we don’t like about the dataset:

* There may not be any correlation between Covid-19 cases and latitude and longitude
* There are uneven numbers of data for each country. For example, eight regions in Australia got reported everyday, but more number of states in the US got reported
* The dataset is not perfectly written. For example Chicago is written as “Chicago” or “Chicago, IL” in the region column.