ETL of Thrones

* Extract
  + Our data were taken from three csv files stashed in a Kaggle dataset called Game of Thrones (<https://www.kaggle.com/mylesoneill/game-of-thrones>).
  + One set called battles.csv contained the information used to populate Battles table in database, one called deaths.csv was used to generate Deaths and Characters tables, one called predictions.csv used for Discussion.
* Transform
  + The transform step in this ETL task consisted primarily of stripping the original dataframes created from our three CSV files to the columns needed for our four tables.
  + The key challenge came in generating two tables –– Characters and Deaths –– from the same dataset, deaths.csv.
    - This task also required filtering out all living characters in the original deaths.csv file from the deaths data frame before setting that data frame to sql as the Deaths table. The result was a Characters table with unique character IDs generated from all rows in the deaths.csv file, along with a Deaths table with only those characters who have died in the A Song of Ice and Fire books, with unique death IDs generated in both tables only for those dead characters.
  + Battles and Discussion tables were pretty straightforward in terms of the transformations required, although there was an issue matching the characters in predictions.csv to deaths.csv, stemming from the inclusion of hundreds of non-ASOIAF, but still canonical, characters in the former file.
* Load
  + Tables
    - Characters
      * CharacterName, Gender, Book1, Book2, Book3, Book4, Book5, Character\_ID, House, Death\_ID
    - Deaths
      * Character\_ID, CharacterName, Book of Death, Death\_ID, DeathYear
    - Battles
      * BattleName, Battle\_ID, Battle\_Year, Attacker, Defender, Attacker, Attacker\_Outcome
    - Predictions
      * Discussion\_ID Character\_ID, CharacterName, Popularity
  + We used MySQL to generate all IDs in the above tables using the primary key auto increment command.
    - Then we used the common IDs to join the Characters and Deaths and Characters and Discussion tables.