ETL Project – Week 13

Detailing the process of the extraction, transformation, and loading steps

* Extraction – Loaded and Extracted *Bojack\_Ratings.cvs* and *NFLX\_Stock.csv.*
* Transformation-
  + Bojack –
    1. Copied the "SEASON","EPISODE","RATING (IMDB)","YEAR","Release Date" columns into a new dataFrame, dropping the unnecessary columns from the original dataset.
    2. Cleaned the new headers up
  + Netflix
    1. Parsed out the date from a string to create month/date/year columns that can easily be queried.
    2. Copied the date", "Month", "Day", "year", "open", "high", "low", "close” columns into a new dataFrame.
    3. Cleaned up column headers
* Load
  + Created a new database and loaded my Netflix stock data into a new table called clean\_netflix\_db.
    1. I used this database to query specific months and dates to create my plots for stock high prices

What data sources you chose, and why?

* Bojack Horeseman Release Date information
  + Sourced from [Kaggle](https://www.kaggle.com/jaredhelton/bojack-horseman-season-1-to-5)
* Netflix Historical Stock Data
  + Sourced from [MacroTrends](https://www.macrotrends.net/stocks/charts/NFLX/netflix/stock-price-history)

Explain why you have performed the types of transformation you did

* Parsed out the date from a string to create month/date/year columns that can easily be queried from Netflix Stock Data.
* See above under ‘Transformation’ for specific information

Why you chose the type of final database

* I chose a relational database, because my two datasets need to be able to be queried together, using joins, to ensure all data is accurate.

Schema of the tables/collections in the final database

* Netflix
  + Index
  + Date
  + Month
  + Day
  + Year
  + Opening Price
  + High Price
  + Low Price
  + Closing Price
* Bojack
  + Index
  + Season
  + Episode
  + Rating (IMDB)
  + Year
  + Release Date

use case(s) for your database

* While this was a specific case use, Bojack Horseman could be expanded to all Netflix originals, to show potential investor trends.
* For further research, I would like to explore Netflix consumption and subscription trends following the addition/removal of popular shows.
  + My inspiration came from Netflix removing ‘Friends’ at the beginning of 2020 as well as ‘The Office’ leaving the streaming service later in the year.