**2/16/19**

**University of MN Data Visualization and Analytics Boot Camp**

**Cohort 2**

**Team FuzzyWuzzy Project 2.**

**Members: Anupa, Nandu, Eyamba, Nisha, Abe**

**Abstract:** Team FuzzyWuzzy will be using movie/film data from multiple csv files. The original data from these files is from over 45,000 movies from the International Movie Database and The Movie Database. From this data we will be extracting and transforming data to create new tables that will then be joined to create a comprehensive database that can be used for query and analysis.

**Data Sources:** We are using the following csv files to create our database.

* Movie\_metadata.csv
* Keywords.csv
* Links.csv
* Credits.csv

The movie\_metadata.csv file has multiple columns that contain what look like lists of dictionaries. We will be parsing those columns to create new tables as follows:

* Base: Will store all movie titles
* Finance: Table will store financial data such as budget and revenue numbers for each film.
* Genre: Will describe the genre categories each film falls within.
* Language: Will display the original language in which the movie was filmed.
* Production Company: Will display the production company or companies for each film.
* Production Country/Location: will display location(s) for each movie.
* Cast: Will list actors associated with each film.
* Crew: Will display the crew members (e.g., director, producer) for each film
* Keywords: Will display keywords associated with each film.
  + For example, if one wanted to do a query to display movies associated with car chases.
* Links: houses the link id’s for both IMDB and TMDB.