**Project 2 – Data-Fu Pandas**

**Brief Description of the Final Database:** To create a consolidated movies database with streaming platforms information (For this project we are focusing on Netflix and Disney+)

**Why is our final database useful to a hypothetical organization?**

* Our database will give all the information of movies and their related streaming platforms to all the users in one place. This will save their time in searching for movies on separate platforms.
* Using our database, an organization can run an analysis on which streaming platform is more popular based on the movies it streams
* Users will also be able to filter out movies based on language, country, genres, directors, actors, etc.

**List of Data sources:**

* OMDB API (<http://www.omdbapi.com/>)
* Netflix streaming data (<https://data.world/chasewillden/netflix-shows/workspace/file?filename=netflix.xlsx>)
* Disney+ streaming data (<https://www.kaggle.com/unanimad/disney-plus-shows?select=disney_plus_shows.csv>)

**ETL STEPS:**

* Extract required movie titles from Netflix and Disney+ csv files
* Extract data from OMDB API based on movie titles present in Netflix and Disney+ csv files
* Transform JSON response into data frames with required columns. We will have two clean data frames each for Netflix and Disney+ movies
* Create valid ERD and define junction table
* Merge all the collected data into postgres sql database
* Create a comprehensive ReadMe File
* Create sample queries

**Team Members:**

* Dasa Simova ([github.com/DasaSimo](https://github.com/DasaSimo))
* Jay Dhruv ([github.com/jaybdhruv](https://github.com/jaybdhruv))
* Meet Sahni (<github.com/meetk5>)
* Shuchi Khandelwal (<github.com/shuchiK38>)