Purposes of the analysis :

Britta needs help to create an automated pipeline that takes in new data, performs the appropriate transformations, and loads the data into existing tables.

I was using my skills in refactoring the code from this module to create one function that takes in the three files—Wikipedia data, Kaggle metadata, and the MovieLens rating data—and performs the ETL process by adding the data to a PostgreSQL database.

Outcome :

To help Britta , I created four technical analysis deliverables

* Deliverable 1: Write an ETL Function to Read Three Data Files
* Deliverable 2: Extract and Transform the Wikipedia Data
* Deliverable 3: Extract and Transform the Kaggle data
* Deliverable 4: Create the Movie Database

Supported with screen shots and uploading the files to my Github

For deliverable number oneI refactored the given code to generate clean Wikimovies data frame , kaggle metadata frame and rating data frame

For deliverable number two I did refactoring, extract and transform the Wikipedia data so you can merge it with the Kaggle metadata. While extracting the IMDb IDs using a regular expression string and dropping duplicates.

Deliverable number three I did the coding for refactoring, extract and transform the Kaggle metadata and MovieLens rating data, then converted the transformed data into separate DataFrames. Then, merged the Kaggle metadata DataFrame with the Wikipedia movies DataFrame to create the movies Finally, I merged the MovieLens rating data DataFrame with the , movies data frame to create the movies with rating data frame .