Bethzaida Guerra

Alisha Hill-McElroy

Thi Q

ETL-Project

**Introduction**

**Inspiration**

We began the project with the idea of analyzing the airline industry. However, the information didn’t pan out to our satisfaction. After additional brainstorming, we remembered the Week 6, Day 1, Activity 6 OMDB Request class assignment from the SMU-DAL-DATA-PT-06-2021 Boot Camp. We then decided to tackle that topic and adding our creative twist to it. It is very common for analyst to analyze the best of the best, so we decided to analyze the worst of the worst. The Bottom 100 Movies!!!

**Sources of Data**

Our first thought was to check a familiar source of Kaggle.com, however we didn’t find a csv to meet our needs. Our next stop was IMDB.com, a very creditable movie critic website that would provide all sorts if data for any movie you could think of. Lastly, we scrolled to OMDB.com to retrieve additional data to complete our data set.

**Extract**

The primary extraction came from IMDB.com in which web scraping was performed and bottom 100 movies along with the rank and the movie id was collected. This scrape provide the necessary information to pull from OMDB. While discussing the project with the professor, it was said that we would have to manually request all 100 movies. We then broke it down and assigned 1/3 each. Well, that didn’t sit well with Thi so she wrote code to loop through the request and it returned all 100 movies’ metadata. This process saved us a lot of time. During this process we noticed several values were Naan. This is because we used the movie titles instead of the ID to perform the request. The discrepancies in

**Transform**

**Load**

**References**

Website

https://www.imdb.com/chart/bottom

OMDB API

"http://www.omdbapi.com/?t="