Team 5 – Ivan Aguirre, minu khosla, shweta Kalavar

November 02, 2019

# Overview

## Project Title

|  |  |
| --- | --- |
|  | The Movie Dataset |

## Project Description Outline

|  |  |
| --- | --- |
|  | The Movie Dataset consists of movies released on or before July 2017. Data points include cast, crew, genera, budget, revenue, release date, rating, content rating, language, production company, production country. |

## Research Question to Answer

|  |  |
| --- | --- |
|  | This dataset contains a list of csv files that we will be researching to extract, transform and load |

* movies\_metadata.csv: Contains Cast and Crew Information for all movies in the movies\_metadata.csv file.
* credits.csv: Contains Cast and Crew Information for all movies in the movies\_metadata.csv file.
* links.csv: Contains IMDB IDs of all movies featured in the ratings.csv file (About 45,000 movies).
* ratings.csv: Contains ratings from users.

## Data Sets to be used

|  |  |
| --- | --- |
|  | The Movie Dataset was from [www.kaggle.com](http://www.kaggle.com)  <https://www.kaggle.com/rounakbanik/the-movies-dataset> |

## Rough Breakdown of Tasks

|  |  |
| --- | --- |
|  |  |

* Extract, Transform & Load the Data Set
* Review & analyze the CSV files
* Extracted the data from the CSV files
* Transformed the data by cleaning the dataset using pandas
* Designed the table structure
* Created the Database tables using Prosgres
* Loaded the data from the data frame into the final tables within in Database using Prosgres