**Project Description: Extraction of IMDB data through web scraping**

|  |
| --- |
| **Project Name:** Extraction of IMDB data through web scraping |
| **Website Name:** <www.imdb.com> |
| **Technology:** Python |
| **Team Members:** |
| Giorgi BENASHVILI |
| Madeena Shaw SHAIK |
| Sahana Devaraje GOWDA |
| Ganesh Kanann IYYANAR |
| Venkat Kumari NATARAJAN |

|  |
| --- |
| **Project Abstract:** |
| Initially we divided the project into phases. Each phase is executed based on the sequence and its importance. The phases and its details are below:  **Phase 1:**  The project is about extraction of IMDB data via web scraping. The Technologies used for extracting the data is below:  1) Methods with one argument;  2) Basic FOR loops;  3) Try/except rules for handling errors/exceptions;  4) Urllib.request library;  5) Bs4 library;  6) Urllib.error library;  7) Read/find/find all functions  **Phase 2:**  The data is extracted by the sequence below:   1. The highest-rated movie names 2. Years of the highest-rated movies 3. Posters of the highest-rated movies 4. Ratings of the highest-rated movies 5. Names of the most popular celebrities 6. Short descriptions of the most popular celebrities 7. Images of the most popular celebrities   **Phase 3:**  Structure involved in creating & storing data in Json  1) methods to read extracted data;  2) storing in argument;  3) generating output file/ parent directory;  4) pandas as pd;  5) pathlib;  6) IPython.core.display  7) urllib.request library;  8) bs4 library;  9) urllib.error library;  **Phase 4:**  System involved in web application using Flask.   1. SQL Alchemy in conjunction with SQLite to store information about movies 2. Installing packages locally 3. Installing web app dependencies 4. Creating SQLite database 5. Creating movie model manager 6. Create table movie database 7. Listing movies table in the database 8. Updating movies table in the database 9. Deleting movies table in the database   **Phase 5:**  Completion of the project with expected results. |