23F Document Automation Python

Project Report

Hari Ravinkumar Patel – 200552675

Bhagyesh Kachiya – 200556134

Kushal Patel – 200528706

# **Movies Data Explorer**

**Problem Statement:**

Create a Python Application that scraps and organizes data of highest grossing movies from the Wikipedia’s page of Highest Grossing Movies. The user inputs a specific year and gets a well-organized output of the films released in the same year and stores the data to the excel spreadsheet.

**Overview:**

Movie Data Explorer is a Python application that enables users to explore and organize information about the highest grossing movies. This application extracts data from Wikipedia’s list of highest grossing movies, allowing users to specify a desired year and receive a detailed dataset of movies released in that year.

**Working of the Program:**

1. User Input Validation:
   * The application ensures user input is valid and falls within the acceptable range (1900 to 2023).
2. Web Scraping:
   * Utilizes web scraping techniques to fetch film data from Wikipedia.
   * Incorporates a variety of user agents for a seamless data retrieval experience.
3. Data Organization:
   * Organizes film data based on the user-specified year, including rank, peak position, title, gross revenue, and release year.
4. Excel Export:
   * Stores the organized data according to the input year into the excel spreadsheet.
5. Exception Handling:
   * Implements exception handling to manage unexpected errors during data extraction.

**Result:**

* This project provides a solution to organize and retrieve the data related to one of the years highest grossing films and if no movies are found for a specified year this application ensures the transparency by communicating to the user.