

ISM 6404 - Intro to Business Analytics and Big Data

Instructor: Dr. Tres Bishop

Assignment 4

(100 pts)

Task: Performing Web Scraping with Power BI

Provide 'clear' screen shots for each of the following. Submit your assignment in a separate Word/PDF document.

Madabhushi Sai Sravani(Z23572172)

1. Select a website page/URL of your own choice (look for table format) which has some numeric/measurable fields and values (10 points). Provide screenshot of the website page you want to scrape.

This is the website I have selected: <https://www.imdb.com/chart/boxoffice/>

The screenshot shows the IMDb Top Box Office (US) chart for the Weekend of March 8-10, 2024. The chart lists the top 10 movies with their box office earnings. The movies and their earnings are:

Rank	Title	Weekend Gross	Total Gross	Weeks Released	Avg. Per Ticket
1.	Kung Fu Panda 4	\$58M	\$75M	1	\$8.13M
2.	Dune: Part Two	\$44M	\$270M	2	\$8.13M
3.	Inspiry	\$39M	\$23M	1	\$8.13M
4.	Cabiri	\$27M	\$1.8M	1	\$8.13M
5.	Bob Marley: One Love	\$24M	\$1.8M	4	\$8.13M
6.	Ordinary Angels	\$22M	\$1.8M	2	\$8.13M
7.	Madame Web	\$11M	\$4.5M	4	\$7.5M
8.	Migration	\$11M	\$1.8M	12	\$8.13M
9.	Trolls	\$9.2M	\$1.8M	1	\$8.13M
10.	Episode 44.7	\$7.4K	\$3.4K	2	\$8.13M

The page also features a sidebar with links to 'More to explore' including 'Most Popular Movies', 'Top Rated English Movies', 'Most Popular TV Shows', 'Top 250 TV Shows', 'Lowest Rated Movies', 'Most Popular Celebs', and 'Movie News'. A promotional banner for 'INVINCIBLE' is visible at the top of the page.

2. Scrape it using Power BI desktop (see Tutorial below for steps). Provide screenshot of the final cleaned up table in the Power Query Editor before applying it to the report (30 points).

The screenshot shows the Microsoft Power Query Editor interface. The main area displays a table titled "Table 1" with the following columns and data:

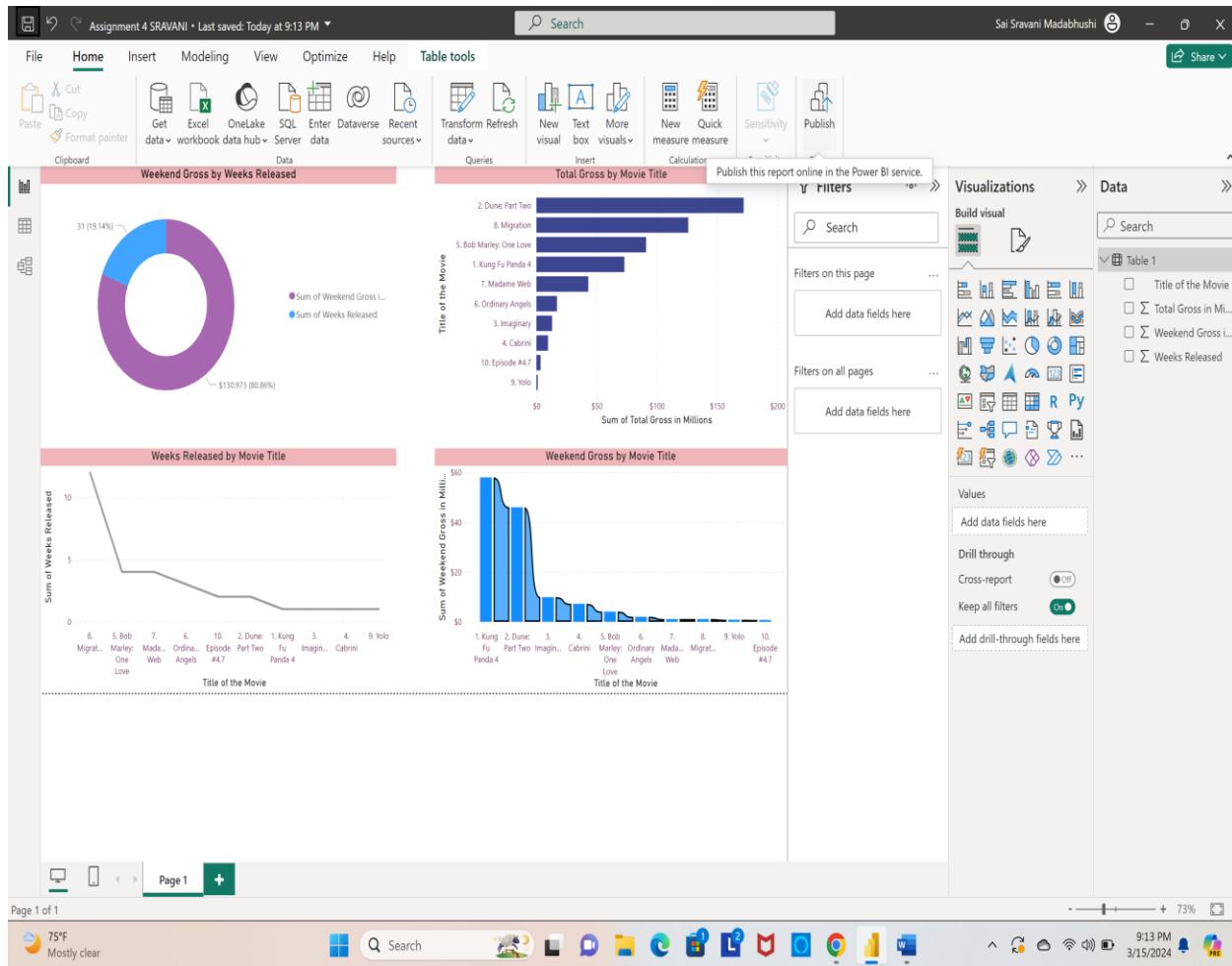
	Title of the Movie	Weekend Gross in Millions	Total Gross in Millions	Weeks Released
1.	Kung Fu Panda 4	58.00	73.00	1
2.	Dune: Part Two	46.00	176.00	2
3.	Imaginary	9.90	13.00	1
4.	Cabirini	7.20	9.60	1
5.	Bob Marley: One Love	4.10	91.00	4
6.	Ordinary Angels	2.00	17.00	3
7.	Madame Web	1.10	43.00	4
8.	Migration	1.10	126.00	12
9.	Yolo	0.83	1.20	1
10.	Episode #4.7	0.75	3.50	2

The "APPLIED STEPS" pane on the right lists the following transformations:

- Source
- Extracted Table From Html
- Changed Type
- Removed Columns
- Replaced Value
- Replaced Value1
- Replaced Value2
- Replaced Value3
- Renamed Column
- Changed Type1

At the bottom, the status bar indicates "4 COLUMNS, 10 ROWS" and "Column profiling based on top 1000 rows". The system tray shows the date and time as "3/15/2024 9:00 PM".

3. Create at least 4 visualizations and then Publish your report to Power BI cloud service. Give 'Titles' to each of the visualizations, they should have appropriate font size for easy readability,not too busy (use filters as applicable) (40 points).



4. In Power BI cloud service create a Dashboard showing the 4 visualizations. (20 points). Provide a screenshot of the full Dashboard.

