

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 9-10. The chart lists 10 titles with their weekend gross, total gross, and number of weeks in release. The top title is 'Kung Fu Panda 4' with a weekend gross of \$58M. The page also features a 'You have rated' section, 'More to explore' links, and a 'Movie News' section.

Rank	Title	Weekend Gross	Total Gross	Weeks Released
1	Kung Fu Panda 4	\$58M	\$58M	1
2	Dune: Part Two	\$48M	\$173M	2
3	Imaginary	\$8.9M	\$18M	1
4	Cabrini	\$7.2M	\$8.4M	1
5	Bob Marley: One Love	\$4.1M	\$8.1M	4
6	Ordinary Angels	\$3.9M	\$17M	3
7	Madame T. T.	\$1.1M	\$1.1M	4
8	Miguel	\$1.1M	\$1.1M	12
9	Yolo	\$52K	\$52K	1
10	Episode 4.7	\$14K	\$14K	2

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).

FileHomeTransformAdd ColumnViewToolsHelp

Close & Apply

New Source

Recent Sources

Enter Data

Data source settings

Manage Parameters

Refresh Preview

Advanced Editor

Choose Columns

Remove Columns

Keep Rows

Remove Rows

Sort

Split Column

Group By

Replace Values

Data Type: Whole Number

Use First Row as Headers

Replace Values

Merge Queries

Append Queries

Combine Files

Text Analytics

Vision

Azure Machine Learning

AI Insights

Table 1

Table.RenameColumns(#"Changed Type1",{"Column1","Title of the Movie"},{"Column4","Weekend Gross in Millions"},{"Column5","Total Gross in Millions"},{"Column6","Weeks Released"})

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

Query Settings

PROPERTIES

Name

Table 1

All Properties

APPLIED STEPS

Source

Extracted Table From Html

Changed Type

Removed Columns

Replaced Value

Replaced Value1

Replaced Value2

Replaced Value3

Renamed Columns

Changed Type1

Renamed Columns1

4 COLUMNS, 10 ROWS

Column profiling based on top 1000 rows

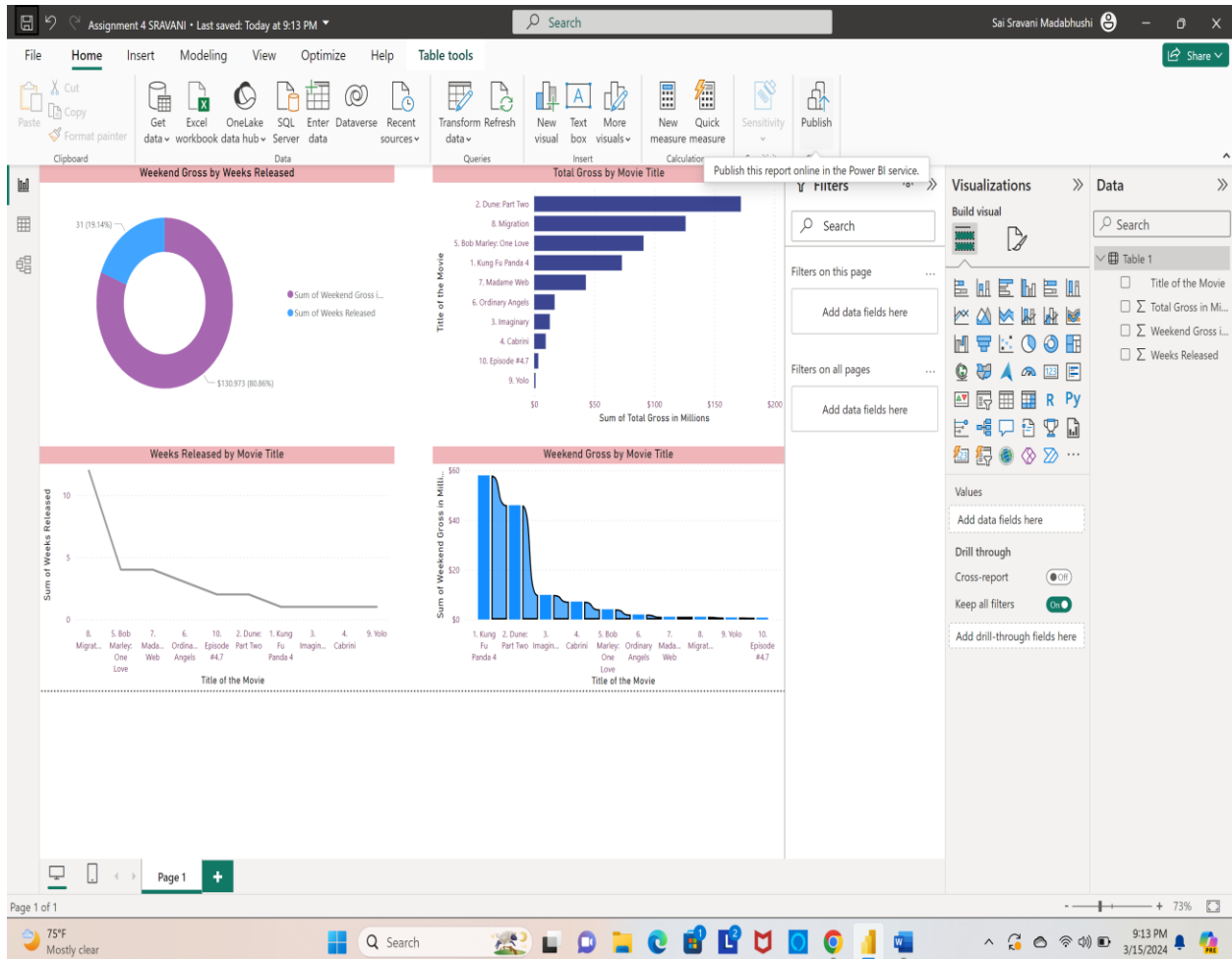
PREVIEW DOWNLOADED AT 8:57 PM

76°F Clear

Search

9:00 PM 3/15/2024

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.

