

Mohammed Imtiyaz

## Assignment 4: Webscraping using PBI

Fau ID: Z23749265

1. I have logged into Power BI desktop using my FAU credentials and chosen <https://www.imdb.com/chart/boxoffice/> this website to proceed.
2. I have followed the exact steps as mentioned in the assignment and that led me to this navigator page. Where I see that a lot of transformation is needed on this. I clicked on transform data which took me to power query editor. There are no titles, a lot of data was to be split and corrected.

The top screenshot shows the Power BI Desktop interface with the Navigator window open. The 'Table 1' is selected in the 'Suggested Tables' list. The 'Table View' tab is active, displaying the following data:

Column1	Column2	Column3	Column4
1. Kung Fu Panda 4	6.5 (8.7K)	(8.7K)	\$30M
2. Dune: Part Two	8.9 (242K)	(242K)	\$29M
3. Arthur the King	6.9 (984)	(984)	\$7.5M
4. Imaginary	4.8 (3.3K)	(3.3K)	\$5.6M
5. Cabrin	8.0 (2.2K)	(2.2K)	\$2.8M
6. Love Lies Bleeding	7.3 (2.7K)	(2.7K)	\$2.5M
7. Bob Marley: One Love	6.5 (14K)	(14K)	\$2.3M
8. One Life	7.6 (13K)	(13K)	\$1.7M
9. The American Society of Magical Negroes	2.6 (2.7K)	(2.7K)	\$1.3M
10. Ordinary Angels	7.8 (1.6K)	(1.6K)	\$1M

The bottom screenshot shows the Power Query Editor with the same data table. The 'Table 1' is selected in the 'Queries' list. The 'Table.TransformColumnTypes' formula is applied to the table. The 'Properties' pane on the right shows the 'Table 1' query settings.

Column1	Column2	Column3	Column4	Column5
1. Kung Fu Panda 4	6.5 (8.6K)	(8.6K)	\$30M	\$108M
2. Dune: Part Two	8.9 (241K)	(241K)	\$29M	\$205M
3. Arthur the King	6.9 (970)	(970)	\$7.5M	\$7.5M
4. Imaginary	4.8 (3.3K)	(3.3K)	\$5.6M	\$19M
5. Cabrin	8.0 (2.2K)	(2.2K)	\$2.8M	\$13M
6. Love Lies Bleeding	7.3 (2.7K)	(2.7K)	\$2.5M	\$2.8M
7. Bob Marley: One Love	6.5 (14K)	(14K)	\$2.3M	\$93M
8. One Life	7.6 (13K)	(13K)	\$1.7M	\$1.7M
9. The American Society of Magical Negroes	2.6 (2.6K)	(2.6K)	\$1.3M	\$1.3M
10. Ordinary Angels	7.8 (1.6K)	(1.6K)	\$1M	\$18M

### 3. The final table after all the transformations looked like this:

Table 1

	MovieTitle	AvgRating	1.2 NumberofRatings(K)	WeekendGross(M)	TotalGross(M)	WeeksReleased
1	Kung Fu Panda 4	6.5	8.6	30.00	108.00	2
2	Dune: Part Two	8.9	241	29.00	205.00	3
3	Arthur the King	6.9	970	7.50	7.50	1
4	Imaginary	4.8	3.3	5.60	19.00	2
5	Cabrini	8	2.2	2.80	23.00	2
6	Love Lies Bleeding	7.3	2.7	2.50	2.80	2
7	Bob Marley: One Love	6.5	14	2.30	93.00	2
8	One Life	7.6	11	1.70	1.70	1
9	The American Society of Magical Negroes	2.6	2.8	1.30	1.30	1
10	Ordinary Angels	7.8	1.6	1.00	18.00	4

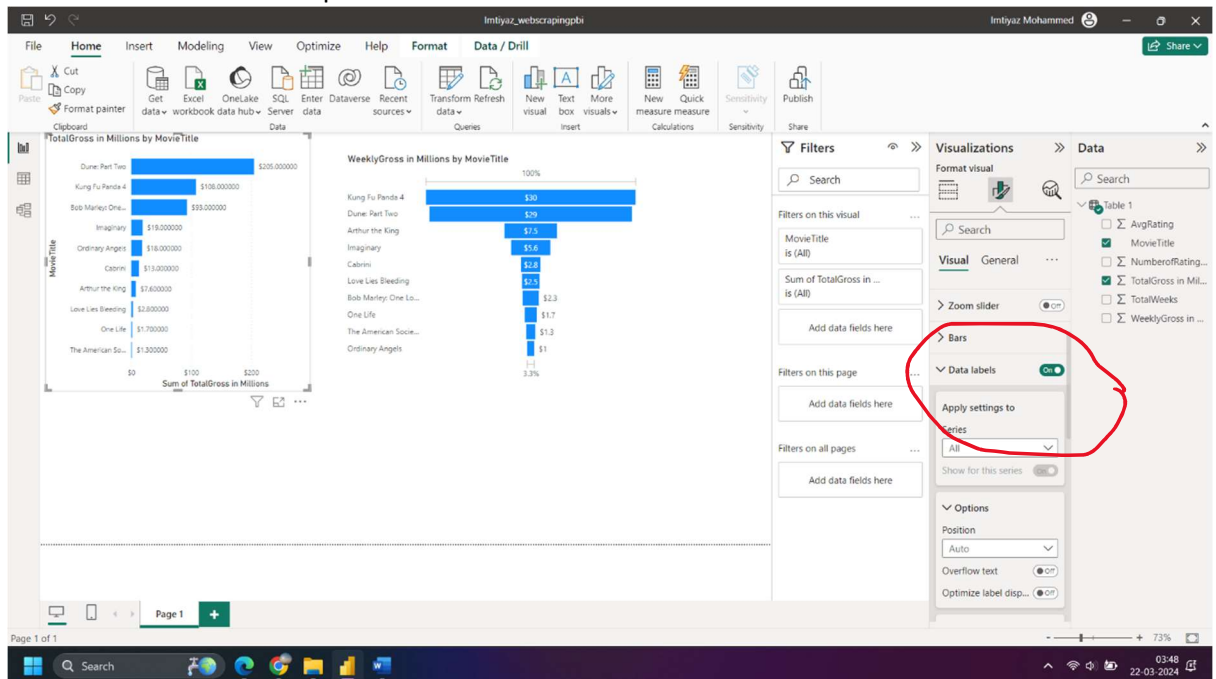
I have changed the column names for all of the columns. Changed the appropriate data type that is, Currency \$ for the revenues, number for ratings and total weeks. I had to use the split technique mentioned which made it easier for me to remove the unnecessary symbols and letters along with the numbers. I also had to manually change a few values which failed to convert correct decimal values.

### 4. After making necessary transformations. I clicked on close and apply. and this is how the table view looks after applying the changes.

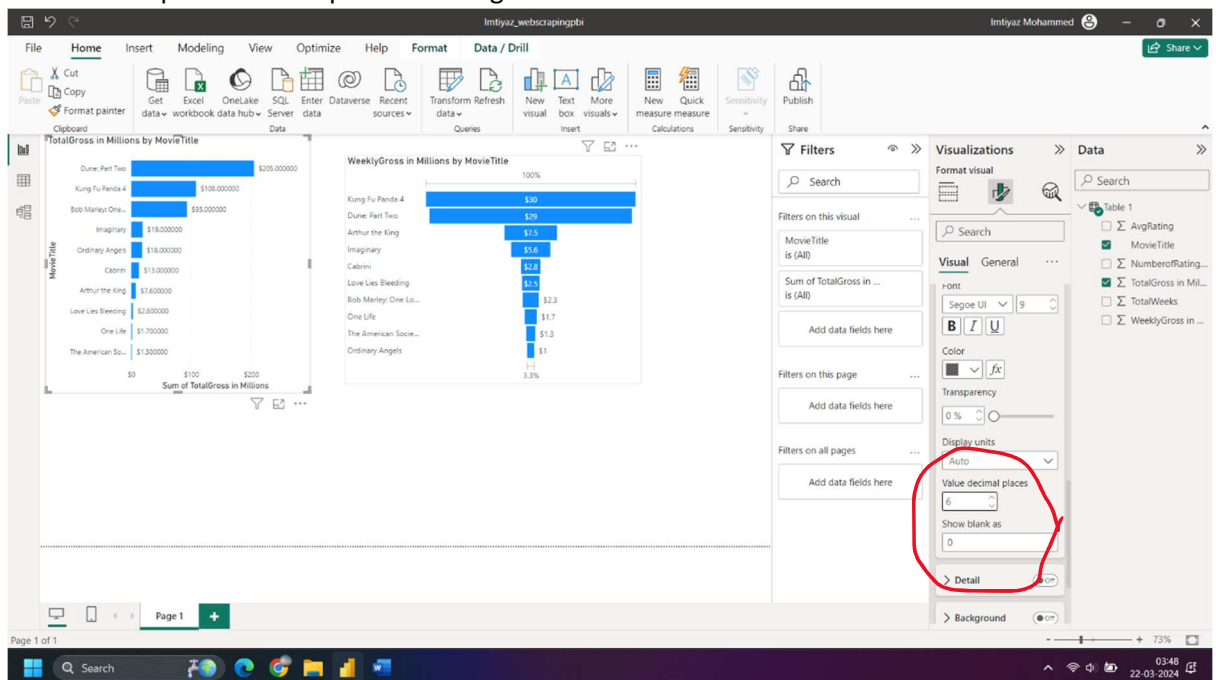
Table 1 (2) (10 rows) Column: NumberofRatings in thousands (9 distinct values)

	AvgRating	NumberofRatings in thousands	WeeklyGross in Millions	TotalGross in Millions	TotalWeeks	MovieTitle
1	6.5	8.7	\$30	\$108	2	Kung Fu Panda 4
2	8.9	242	\$29	\$205	3	Dune: Part Two
3	6.9	0.984	\$7.5	\$7.5	1	Arthur the King
4	4.8	3.3	\$5.6	\$19	2	Imaginary
5	8	2.2	\$2.8	\$13	2	Cabrini
6	7.3	2.7	\$2.5	\$2.6	2	Love Lies Bleeding
7	6.5	14	\$2.3	\$93	5	Bob Marley: One Love
8	7.6	11	\$1.7	\$1.7	1	One Life
9	2.6	2.7	\$1.3	\$1.3	1	The American Society of Magical Negroes
10	7.8	1.6	\$1	\$18	4	Ordinary Angels

5. Now I went to the report view section and added a few visuals and appropriate rows and columns. To the charts. Kept data label on as mentioned.



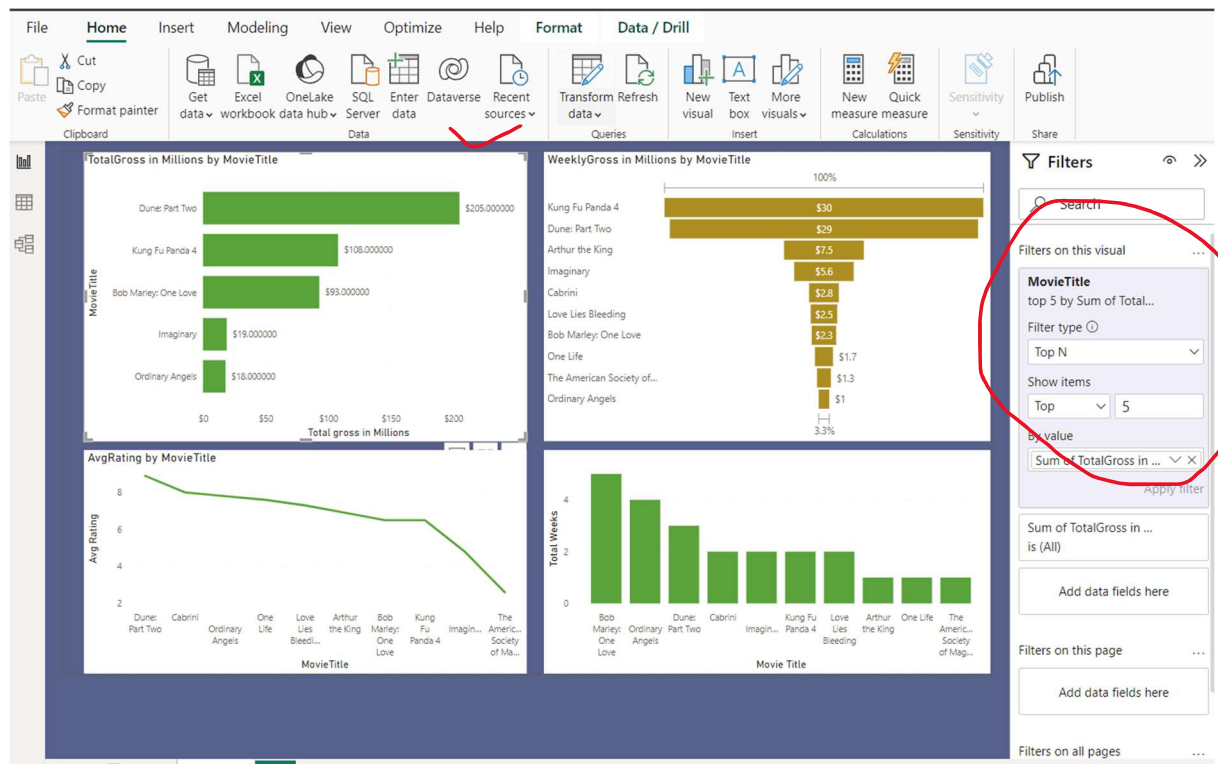
6. And values upto 6 decimal places is also given



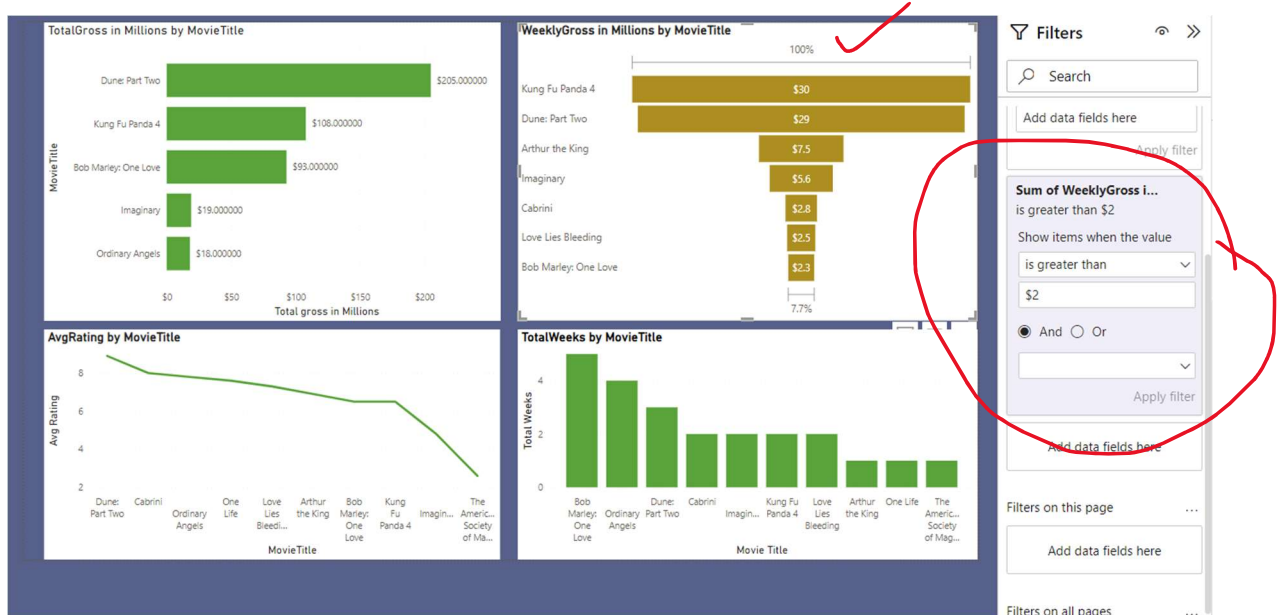
7. Added some more visuals



8. For adding filter. I chose to apply filter to the first visual. To show only top 5 movies based on total revenue.



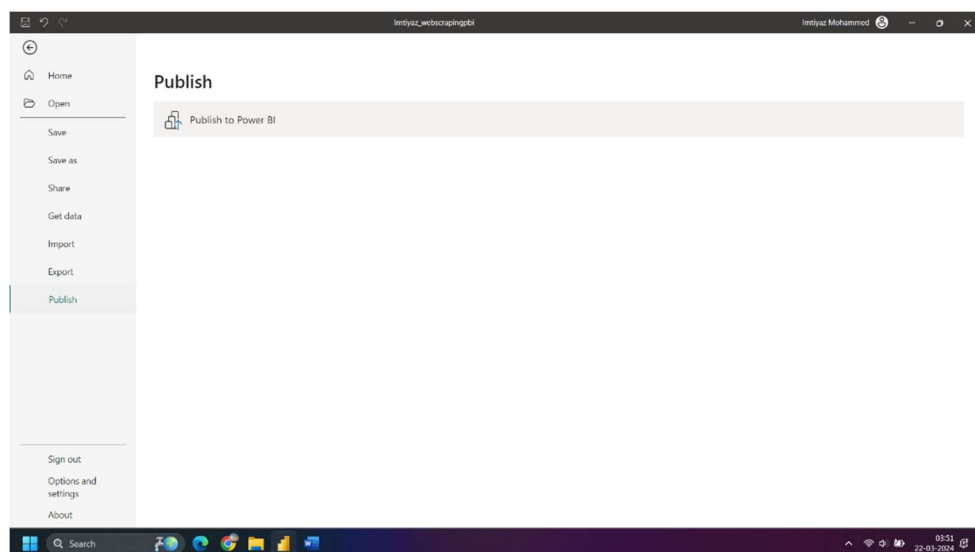
9. For the next visual too I decided to use filter to see weekly gross of movies above 2 million dollars

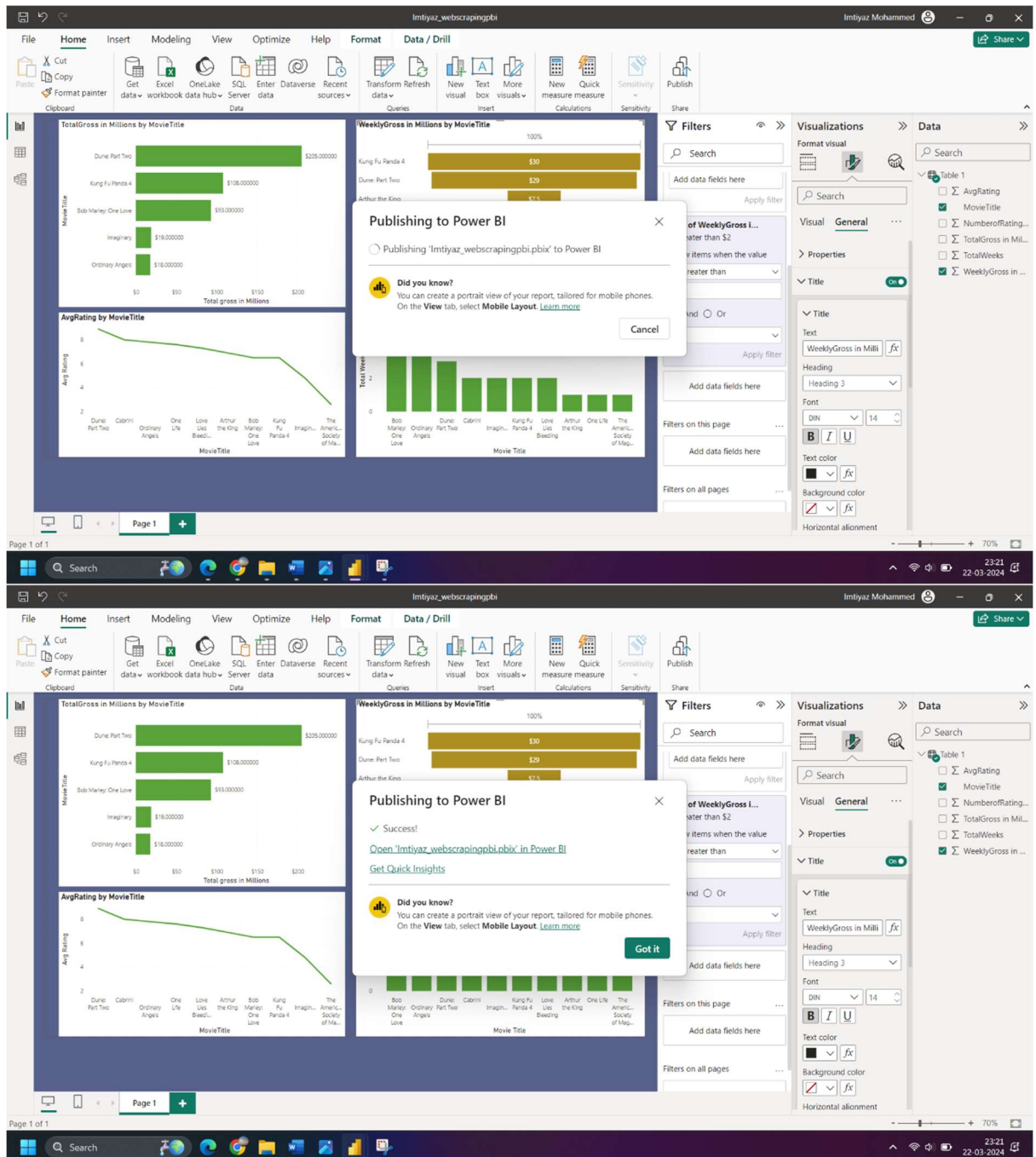


The first 2 visuals show the weekly gross , total gross of the movies.

And the third one is about the average ratings of the movies. We can see that the Dune is the highly rate movie. The American society is the least rated movie. The final chart speaks about the number of weeks a movie was in theatre. So bob marley was there for 5 weeks. And again American society is for just a week

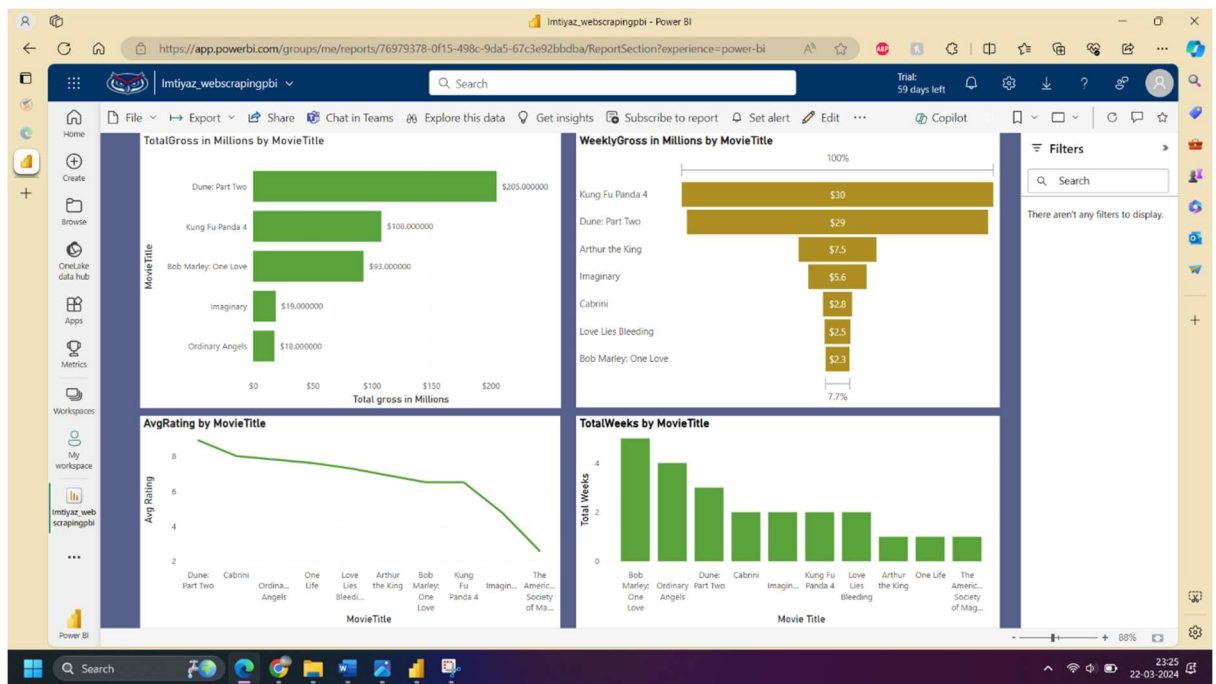
10. Then I had to get to the final step to publish the report.





We can see it here





The final dashboard looks like this

