

upGrad
Campus 

Course : Python for
Analysis

Lecture On : IMDb

Movie Assignment

Instructor : Shambhavi
Gupta

About the Assignment

Where do I get the data from?

You can download the movie dataset from the IMDb Movie Assignment session of the Python for Analysis course.

Where do I write the code?

Download the 'Movie Assignment.ipynb' file from the IMDb Movie Assignment session of the Python for Analysis course. It is a commented Jupyter IPython Notebook file in which all the instructions and tasks to be performed are mentioned.

How do I submit the assignment?

Upload your 'ipynb' solution file in the 'submission' section on the session.

Some tips before starting the assignment

1. Go through the data dictionary thoroughly before starting with the assignment.
2. Read through each of the instructions carefully, identify the task to be performed, and only then proceed to write the required code.
3. In some cases, the variable names have already been assigned, and you just need to write code against them.
4. Always keep inspecting your data frame after you have performed a particular set of operations.
5. There might be some subtasks that involve the use of functions that you may not have used before. In such cases, you just need to do a simple search on Google/Stack Overflow to gain an understanding of the function
6. Always run the cells of the notebook sequentially/restart the kernel and run all the cells to avoid runtime errors.
7. For plot-related questions, please make sure you are using the appropriate chart size for better readability.
8. Optional subtasks are non-evaluative.

*Refer to the Problem Statement segment in the IMDb session for a detailed explanation

Problem Statement

You have the data for the 100 top-rated movies from the past decade along with various pieces of information about the *movie, its actors, and the voters who have rated these movies online*. In this assignment, you will try to find some interesting insights into these movies and their voters, using Python.



Example of what the data might look like:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Poster_Link	Series_Title	Released_Year	Certificate	Runtime	Genre	IMDB_Rating	Overview	Meta_scor	Director	Star1	Star2	Star3	Star4	No_of_Votes	Gross
2	https://m.media-amazon.com/images/M/MV5BMDFkYTc0MGEtZmNhMC00ZDZlZWYwOTY0ODM1ZmRlYWMwMTMxODk2OTU@_V1_UX67_CR0,0,67,98_AL_.jpg	The Shawshank Redemption	1994	A	142 min	Drama	9.3	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.	80	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110	2,83,41,469
3	https://m.media-amazon.com/images/M/MV5BM2MyNjYxNmUyYTAwNi00MTYxLWJmNWYyZjI0ODY3ZTk3OTF0XkEyXkFqcGdeQXVyNzkzMjQ5NzU@_V1_UY98_CR1,0,67,98_AL_.jpg	The Godfather	1972	A	175 min	Crime, Drama	9.2	An organized crime dynasty's aging patriarch transfers control of his clandestine empire to his reluctant son.	100	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367	13,49,66,411
4	https://m.media-amazon.com/images/M/MV5BMTMxNTMwODM0NF5BMjE5MjYxLWJmNWYyZjI0ODY3ZTk3OTF0XkEyXkFqcGdeQXVyNzkzMjQ5NzU@_V1_UX67_CR0,0,67,98_AL_.jpg	The Dark Knight	2008	UA	152 min	Action, Crime, Drama	9	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, Batman must accept one of the greatest psychological and physical tests of his ability to fight injustice.	84	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232	53,48,58,444
5	https://m.media-amazon.com/images/M/MV5BMWwMGQzZTItY2JlNC00OWZlLWlyMDctNDk2ZDQyYjRlZWYwOTY0ODM1ZmRlYWMwMTMxODk2OTU@_V1_UY98_CR1,0,67,98_AL_.jpg	The Godfather: Part II	1974	A	202 min	Crime, Drama	9	The early life and career of Vito Corleone in 1920s New York City is portrayed, while his son, Michael, expands and tightens his grip on the family crime syndicate.	90	Francis Ford Coppola	Al Pacino	Robert De Niro	Robert Duval	Diane Keaton	1129952	5,73,00,000
6	https://m.media-amazon.com/images/M/MV5BMWU4N2FjZmYxNTYwNi00NDQ0LTg0MjAtYTJlMjFhNGUxZDZlZWYwOTY0ODM1ZmRlYWMwMTMxODk2OTU@_V1_UX67_CR0,0,67,98_AL_.jpg	12 Angry Men	1957	U	96 min	Crime, Drama	9	A jury holdout attempts to prevent a miscarriage of justice by forcing his colleagues to reconsider the evidence.	96	Sidney Lumet	Henry Fonda	Lee J. Cobb	Martin Balsam	John Fiedler	689845	43,60,000
7	https://m.media-amazon.com/images/M/MV5BNzA5ZDNiZWMtMTMhNS00NDM1LTk4NDI0YTRmYzEwMzU@_V1_UX67_CR0,0,67,98_AL_.jpg	Lord of the Rings: The Return of the King	2003	U	201 min	Action, Adventure, Drama	8.9	Gandalf and Aragorn lead the World of Men against Sauron's army to draw his gaze from Frodo and Sam as they approach Mount Doom with the One Ring.	94	Peter Jackson	Elijah Wood	Viggo Mortensen	Ian McKellen	Orlando Bloom	1642758	37,78,45,905

*Please note that this is just to give you an idea and not the real dataset.

Flow of Analysis

1. **Reading the data**
 1. Read the Movies Data.
 2. Inspect the Dataframe
2. **Data Analysis**
 1. Reduce the Digits
 2. Calculate and analyze the profit
 3. Analysis on the general audience and the critics (*find out the highest rated movies liked by both*)
 4. Analysis on the Most Popular Trios (*check the popularity is the Facebook likes of each of these actors*)
 5. Runtime Analysis
 6. R-Rated Movies

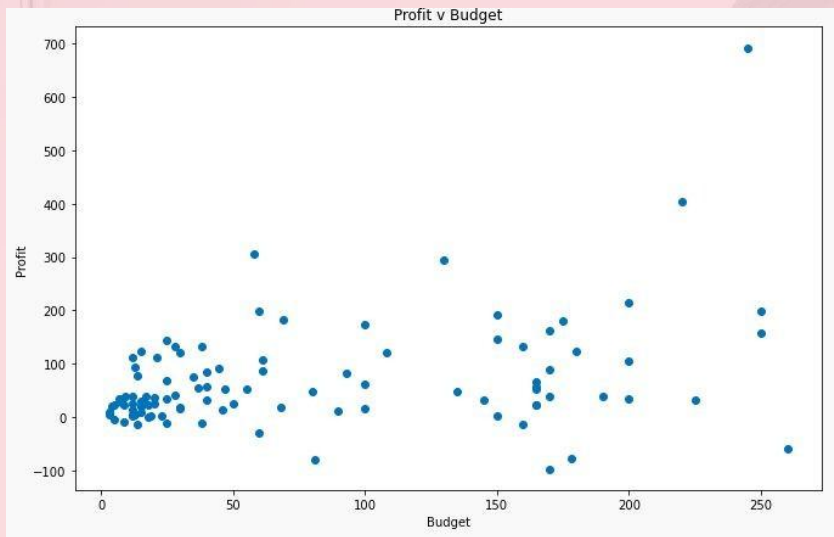
3. Demographic analysis

1. Combine the Dataframe by Genres
2. Count the Genre (plot graphs for the same)
3. Gender and Genre
4. US vs non-US Cross Analysis
5. Top 1000 Voters Vs Genres

NOTE:

- Plot graphs wherever necessary for the analysis
- Feel free to run sub-analysis to get to your answer
- Refer to the jupyter notebook for the detailed task explanation

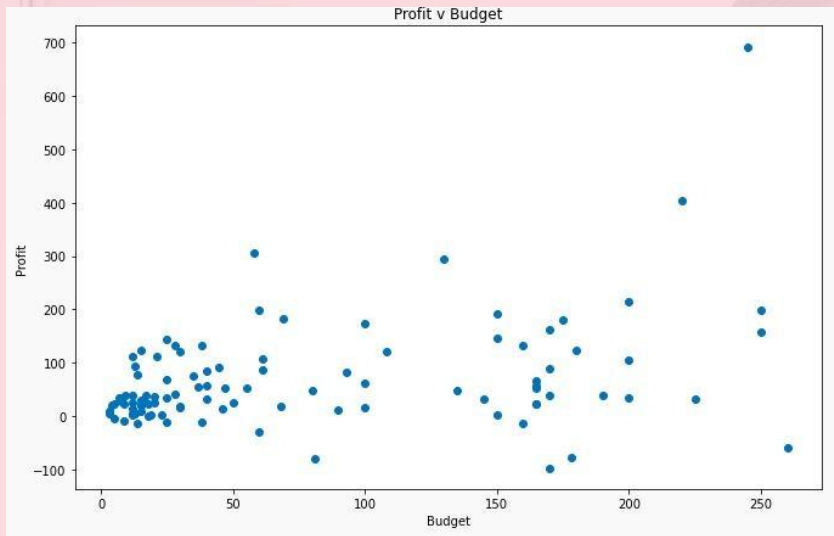
Example of charts



Example of inferences/observations

1. The average rating across all genres shows a trend with increasing age group
2. is the LEAST popular genre among U18M audience and is the LEAST popular genre among U18F audience
3. What is the trend among *movies* with *higher budgets*?

Example of charts



Example of inferences/observations

1. The average rating across all genres shows a downward trend with increasing age group
2. Romance is the LEAST popular genre among U18M audience and Crime is the LEAST popular genre among U18F audience
3. Movies with higher budgets are not necessarily profitable



All The Best!