

DS201

Lab 3 – IMDB Movie database

Using basic data exploration and visualization techniques in Microsoft Excel to answer these questions by Visualization

1. Use the provided link to download the dataset
2. Create an ID column for each movie
3. Match and get IMDB score for each genre.
(Hint as following image)

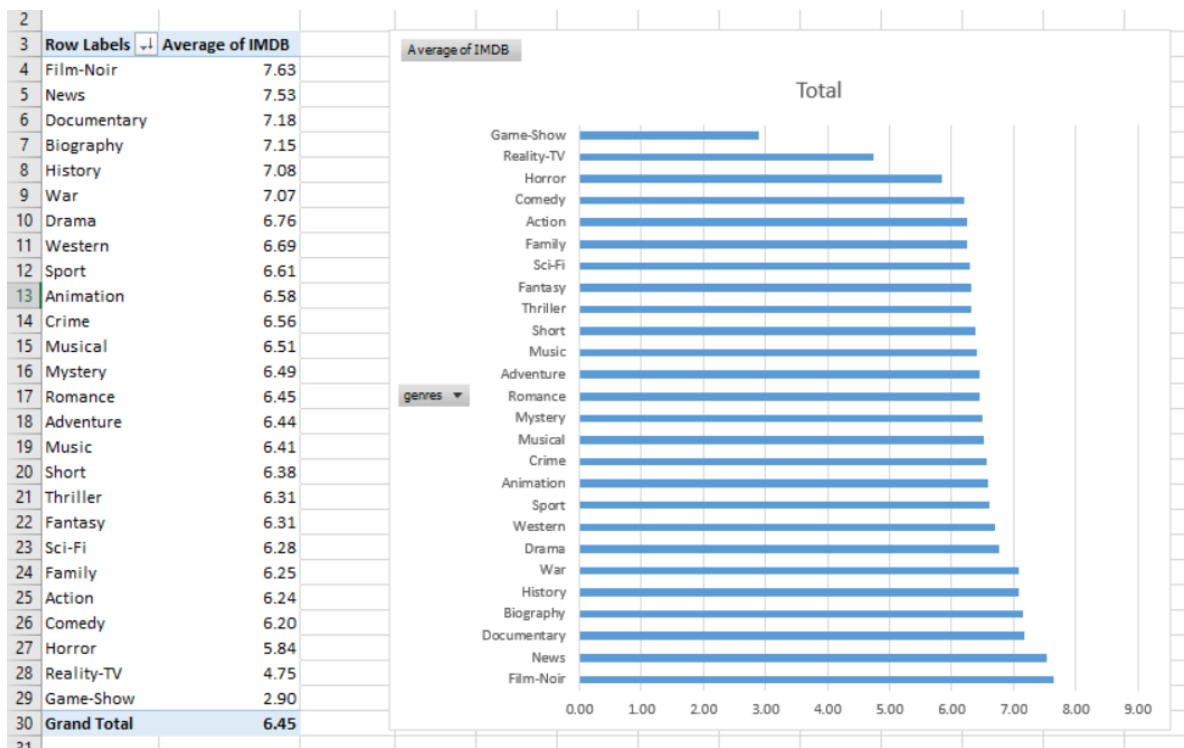
	A	B	C	D	E	F	G
1	ID	genres	IMDB			ID	imdb_score
2	1	Action	7.9			1	7.90
3	1	Adventure	=VLOOKUP(A3,\$F\$2:\$G\$5044,2,FALSE)				7.10
4	1	Fantasy	VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])				6.80
5	1	Sci-Fi	7.9			4	8.50
6	2	Action	7.1			5	7.10
7	2	Adventure	7.1			6	6.60
8	2	Fantasy	7.1			7	6.20
9	3	Action	6.8			8	7.80
10	3	Adventure	6.8			9	7.50
11	3	Thriller	6.8			10	7.50
12	4	Action	8.5			11	6.90
13	4	Thriller	8.5			12	6.10
14	5	Document	7.1			13	6.70
15	6	Action	6.6			14	7.30
16	6	Adventure	6.6			15	6.50
17	6	Sci-Fi	6.6			16	7.20
18	7	Action	6.2			17	6.60
19	7	Adventure	6.2			18	8.10

	A	B	C	D	E	F	G
1	ID	genres	IMDB			ID	imdb_score
5041	1480	Sport	6.3			5040	7.50
5042	1481	Comedy	3.8			5041	6.30
5043	1481	Family	3.8			5042	6.30
5044	1482	Action	5.2			5043	6.60
5045	1482	Horror	5.2				
5046	1482	Sci-Fi	5.2				
5047	1483	Action	6				
5048	1483	Adventure	6				

	A	B	C	D	E	F	G	H
1	ID	genres	IMDB			ID	imdb_score	
14492	5038	Drama	6.4					
14493	5039	Comedy	7.7					
14494	5039	Drama	7.7					
14495	5040	Crime	7.5					
14496	5040	Drama	7.5					
14497	5040	Mystery	7.5					
14498	5040	Thriller	7.5					
14499	5041	Drama	6.3					
14500	5041	Horror	6.3					
14501	5041	Thriller	6.3					
14502	5042	Comedy	6.3					
14503	5042	Drama	6.3					
14504	5042	Romance	6.3					
14505	5043	Document	6.6					
14506								

Answer these following questions:

- Who are the top 20 Director based on average IMDB score?
- Based on IMDB score, who should be the main actor that investor might prefer in the next movie? Please show the table.
- Based on IMDB score, which genre you should invest in the next movie? Show the list of genres in a table with a bar (or column) chart.



Note: images provided on this document is for a hint, your data workspace and result does not have to be the same.