

After importing the csv file in SQL Server, rounding the Imdb rating to 1 digit for the convenience.

```
Update [GameOfThrones].[dbo].[got_imdb]
SET imdb_rating = Round(imdb_rating, 1);
```

Questions and Queries

1. What are the top-rated episodes of Game of Thrones according to the IMDb ratings?

```
SELECT ep.title, ep.season, max(imdb.imdb_rating) as 'imdb_rating',
DENSE_RANK() OVER (ORDER BY max(imdb.imdb_rating) DESC) as 'rank'
FROM [GameOfThrones].[dbo].[got_episodes] ep JOIN [GameOfThrones].[dbo].[got_imdb]
imdb
ON ep.original_air_date = imdb.original_air_date
GROUP BY ep.title, ep.season;
```

The above query gives us the rating of all the episodes. We can put this query in CTE and from that can find the top-rated episodes like (top 5/top 10) as per our requirement. E.g., We want to find top 5 episodes by rating , the query is –

```
WITH temp_table as (
SELECT ep.title, ep.season, max(imdb.imdb_rating) as 'imdb_rating',
DENSE_RANK() OVER (ORDER BY max(imdb.imdb_rating) DESC) as 'rank'
FROM [GameOfThrones].[dbo].[got_episodes] ep JOIN [GameOfThrones].[dbo].[got_imdb]
imdb
ON ep.original_air_date = imdb.original_air_date
GROUP BY ep.title, ep.season
)

SELECT * FROM temp_table WHERE rank <= 5;
```

Result –

Results		Messages		
	title	season	imdb_rating	rank
1	The Rains of Castamere	3	9.9	1
2	Battle of the Bastards	6	9.9	1
3	The Winds of Winter	6	9.9	1
4	The Lion and the Rose	4	9.7	2
5	The Mountain and the Viper	4	9.7	2
6	The Spoils of War	7	9.7	2
7	The Watchers on the Wall	4	9.6	3
8	The Children	4	9.6	3
9	Baelor	1	9.6	3
10	And Now His Watch Is Ended	3	9.5	4
11	Fire and Blood	1	9.4	5
12	The Dance of Dragons	5	9.4	5
13	The Dragon and the Wolf	7	9.4	5

We see from result that 3 episodes received highest ratings. From above query we can also answer the question- Which episode received the highest IMDb rating?

2. Which episode received the lowest IMDb rating?

```
WITH temp_table as (  
SELECT ep.title, ep.season, min(imdb.imdb_rating) as 'imdb_rating',  
DENSE_RANK() OVER (ORDER BY min(imdb.imdb_rating)) as 'rank'  
FROM [GameOfThrones].[dbo].[got_episodes] ep JOIN [GameOfThrones].[dbo].[got_imdb]  
imdb  
ON ep.original_air_date = imdb.original_air_date  
GROUP BY ep.title, ep.season  
)  
  
SELECT * FROM temp_table WHERE rank = 1;
```

Result –

	title	season	imdb_rating	rank
1	The Iron Throne	8	4.1	1

From results, its seen that the last episode of the last season was lowest rated ever.

3. What is the average IMDb rating of all episodes?

```
SELECT ROUND(AVG(imdb_rating),1) as 'average_rating'  
FROM [GameOfThrones].[dbo].[got_imdb];
```

Result –

	average_rating
1	8.8

4. How does the IMDb rating vary across seasons?

```
SELECT season, ROUND(AVG(imdb_rating),1) AS 'average_rating'
FROM [GameOfThrones].[dbo].[got_imdb]
GROUP BY season;
```

Result –

	season	average_rating
1	1	9
2	2	8.8
3	3	9
4	4	9.2
5	5	8.7
6	6	9
7	7	9
8	8	6.4

5. What are the most-watched episodes of Game of Thrones based on the number of viewers?

```
WITH temp_table as(
SELECT season, title, MAX(us_viewers) as 'us_viewers',
RANK() OVER (ORDER BY MAX(us_viewers) DESC) as 'rank'
FROM [GameOfThrones].[dbo].[got_episodes]
GROUP BY season, title
)
SELECT * FROM temp_table WHERE rank = 1;
```

Result –

	season	title	us_viewers	rank
1	8	The Iron Throne	13610000	1

Indeed, the finale episode of the last season is the most watched episode.

6. Which episode had the lowest number of viewers?

Let's use a sub-query for this one.

```
SELECT season, episode_num_in_season, title, us_viewers
FROM [GameOfThrones].[dbo].[got_episodes]
WHERE us_viewers = (
SELECT MIN(us_viewers) FROM [GameOfThrones].[dbo].[got_episodes]);
```

Result –

Results		Messages		
	season	episode_num_in_season	title	us_viewers
1	1	2	The Kingsroad	2200000

7. How does the average number of viewers vary across seasons?

```
SELECT season, AVG(us_viewers) AS 'average_viewers'
FROM [GameOfThrones].[dbo].[got_episodes]
GROUP BY season;
```

Result –

Results		Messages	
	season	average_viewers	
1	1	2515000	
2	2	3795000	
3	3	4966000	
4	4	6846000	
5	5	6880000	
6	6	7688000	
7	7	10261428	
8	8	11993333	

We see that there is linear increase in the viewership as the seasons progress.

8. Are there any episodes that have a high IMDb rating but a relatively low number of viewers or total votes or vice versa?

```
SELECT imdb.season, imdb.episode_num, imdb.title, ep.us_viewers,
DENSE_RANK() OVER (ORDER BY ep.us_viewers DESC) as 'rank_by_viewers',
imdb.imdb_rating,
DENSE_RANK() OVER (ORDER BY imdb.imdb_rating DESC) as 'rank_by_rating',
imdb.total_votes,
DENSE_RANK() OVER (ORDER BY imdb.total_votes DESC) as 'rank_by_votes'
FROM [GameOfThrones].[dbo].[got_episodes] ep JOIN [GameOfThrones].[dbo].[got_imdb]
imdb
ON ep.original_air_date = imdb.original_air_date;
```

This query gives us the ranks based on imdb rating, number of viewers and number of votes.

Result –

	season	episode_num	title	us_viewers	rank_by_viewers	imdb_rating	rank_by_rating	total_votes	rank_by_votes
1	8	6	The Iron Throne	13610000	1	4.1	19	242548	1
2	8	3	The Long Night	12020000	3	7.5	18	212837	2
3	6	9	Battle of the Bastards	7660000	17	9.9	1	203879	3
4	6	10	The Winds of Winter	8890000	12	9.9	1	143780	4
5	8	1	Winterfell	11760000	4	7.6	17	130346	5
6	8	2	A Knight of the Seven Kingdoms	10290000	6	7.9	16	128712	6
7	3	9	The Rains of Castamere	5220000	30	9.9	1	101517	7
8	7	4	The Spoils of War	10170000	8	9.7	2	86600	8
9	7	6	Beyond the Wall	10240000	7	9	8	63062	9
10	7	7	The Dragon and the Wolf	12070000	2	9.4	5	62010	10

From the output one interesting conclusion is that, the finale episode of the series was most viewed and most voted but received lowest rating.