Conclusion

There are 2550 rows and 17 columns in our data.

There is no missing value and the data is already cleaned as it is imported from TED website directly. Only the speaker_occupation was missing, this clearly can be seen in msno.matrix graph, but it can be ignored as well as it won't contribute much to the analysis. We decided to drop them to clear our data.

Some insights:

'Parrots, the universe and everything' by Douglas Adams has the maximum duration with 1.46hr(5256 seconds).

Observations on views:

Ken Robinson's talk on 'Do Schools Kill Creativity?' is the most popular TED Talk of all time with 47.2 million views. Also, it is also one of the oldest talks on the TED website. Robinson's talk is closely followed by Amy Cuddy's talk on "Your Body Language May Shape Who You Are." There are only 2 talks that have surpassed the 40 million mark and 4 talks that have crossed the 30 million mark. The average number of views on TED Talks in 1.7 million and the median number of views is 1.12 million. This suggests a very high average level of popularity of TED Talks.

When we checked the describe method on languages column we found that:

On average, there are 191.6 comments on every TED Talk. We can conclude that the TED Online Community is highly involved in discussions. There is a huge standard deviation associated with the comments. In fact, it is even larger than the mean suggesting that the measures may be sensitive to outliers. The minimum number of comments on a talk is 2 and the maximum is 6404. The minimum number, could be as a result of the talk is newly posted or not just not many people have something to say, also, the talk was not popular among viewers.

The correlation matrix on 'vews' & 'comments' columns showed that the pearson coefficient is slightly more than 0.5. This suggests a medium to strong correlation between the two quantities.

When we grouped by 'comments' & 'duration' and sorted values, we realized that:
TED Talks with duration less than 2000(seconds) have more comments. Maximum value of comments is 6404. The TED Talk with maximum comments is Militant atheism by Richard Dawkins.

The correlation between comments and views is 0.53 which suggests a medium correlation between the two quantities. Also, The correlation between duration and views is 0.048 which is quite less and suggests very low correlation between the two quantities.

As we saw from 'comments' graph, Richard Dawkins' talk on Militant Atheism generated the greatest amount of discussion and opinions despite having significantly lesser views than Ken Robinson's talk, which is second in the list.

This raised some interesting questions of:

Which talks tend to attract the largest amount of discussion?
To answer this question, we defined a new feature discussion quotient which was simply the ratio of the number of comments to the number of views. Then, we checked which talks have the largest discussion quotient and found that extremely interesting insights. Half of the talks in the top 10 are on the lines of Faith and Religion. Science and religion is still a very hotly debated topic even in the 21st century.

The most discusses talk, though, is The Case for Same Sex Marriage (which has religious undertones). This is not that surprising considering the amount of debate the topic caused back in 2009.

The objective of this work was to understand a little about the TED Talks popularity. They are interesting metrics to evaluate a TED Talk and this work explored some results of correlation between duration & comments, which topics were popular among viewers, talks which are shorter than 30 minutes had more discussion rather than long ones, how people evaluate them and what may influence these results.

There are more analyses that could be performed with these data. Many of the correlations mentioned in this work could be calculated and tested. And insights and observations could be further explored and different models can be built based on them.