Time Series Analysis - Forecasting Deaths in Game of Thrones

Sireesh Limbu

May 2019

1 Introduction

Based on George R.R. Martin's best-selling book series "A Song of Ice and Fire" - Game of Thrones is an American fantasy-fiction television drama series created by David Benioff and D.B. Weiss for the popular satellite television and cable network HBO. The title has been taken from the first novel from the series - "A Game of Thrones".

Set in the fictional continents of Westeros and Essos, Game of Thrones revolves around nine noble families and their fight to gain control over Westeros and survive against the ancient undead enemies that were hitherto dormant.

Filmed in Belfast, US,Canada and scenic parts of Europe based out of countries like Croatia, Morocco, Malta,Scotland, Northern Ireland, Iceland and Spain, the whole series has been concluded in eight seasons with a total of 72 episodes with the last episode to be aired on Monday - 20th May in Sweden.

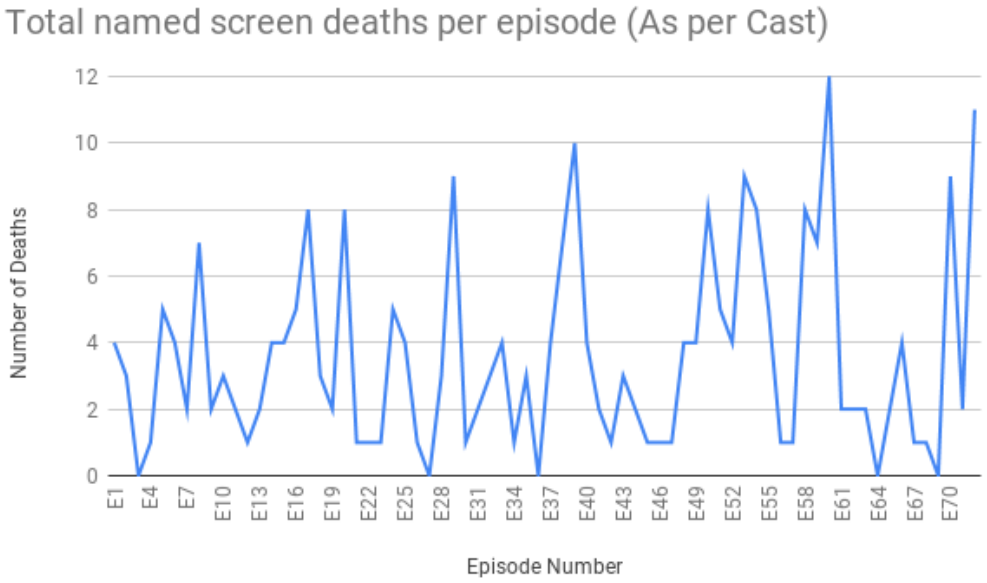
2 Goal

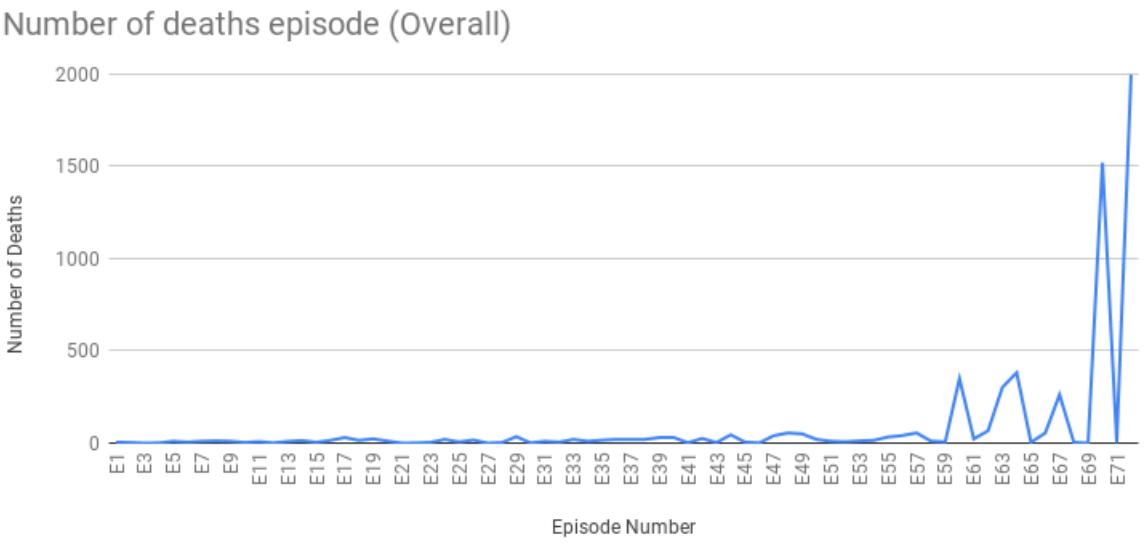
Game of Thrones, apart from being famous for various ingenious factors such as story-plot, CGI and well-drawn characters is also known for it's violence in the form of battles, wars or just random killings. This has amounted to an average of 81 killings per episode and a total of 5863 on-screen killings spanning over 8 seasons until the second-last episode. The goal of this project is

to analyze the data of total on-screen deaths per episode for both 'Named' and 'Overall' characters and eventually forecast the total deaths in the last episode.

3 Data Collection and Initial Analysis

The total deaths statistics survey from Washington post and fandom wiki solely dedicated to Game of thrones has been the two major online sources of data for this project. The collection of total deaths for both named and un-named characters together per episode due to lack of reliable sources has been modified to add up to total deaths per season which in turn has been verified by the Washington Post survey. But, the fandom website provides reliable source of the number of on-screen deaths of named characters per episode. Due to this, the project has been divided into two parts: first to analyze, fit the model and forecast deaths for named characters and finally do the same for all characters. Following are the Number of Deaths Vs Episode Number plots for overall and 'Named' and 'Overall' deaths.

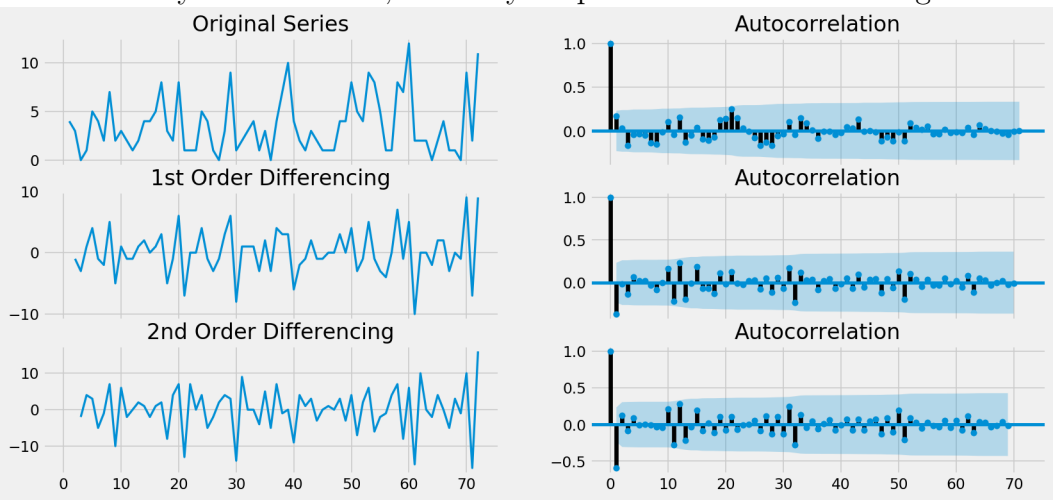




As we can see there are no clear patterns. We further analyze with their respective autocorrelation plots.

4 Analysis (Named character)

Plotting the original series along with it's correlation we get a first impression of stationarity. Furthermore, we analyse upto 2nd order differencing.

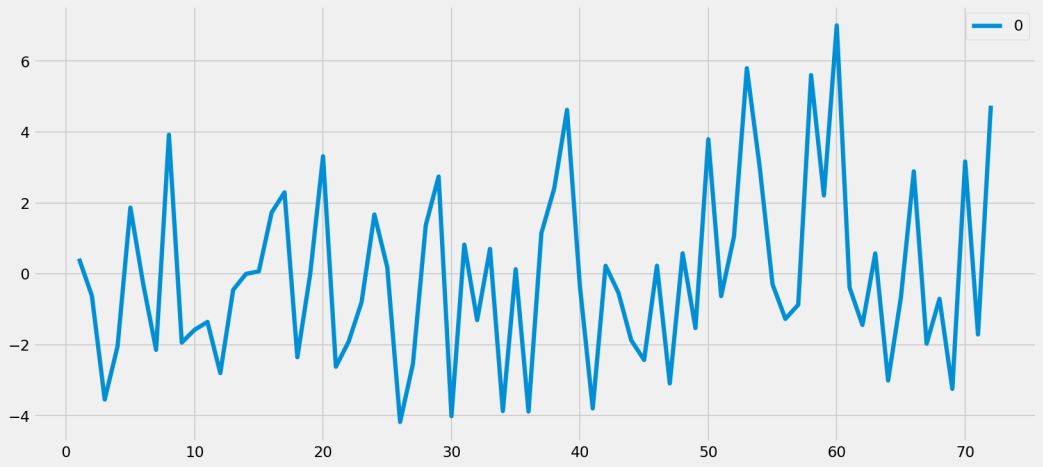


We notice that there is no proper visible trend or seasonality. We further perform the Dickey-Fuller test for checking the same.

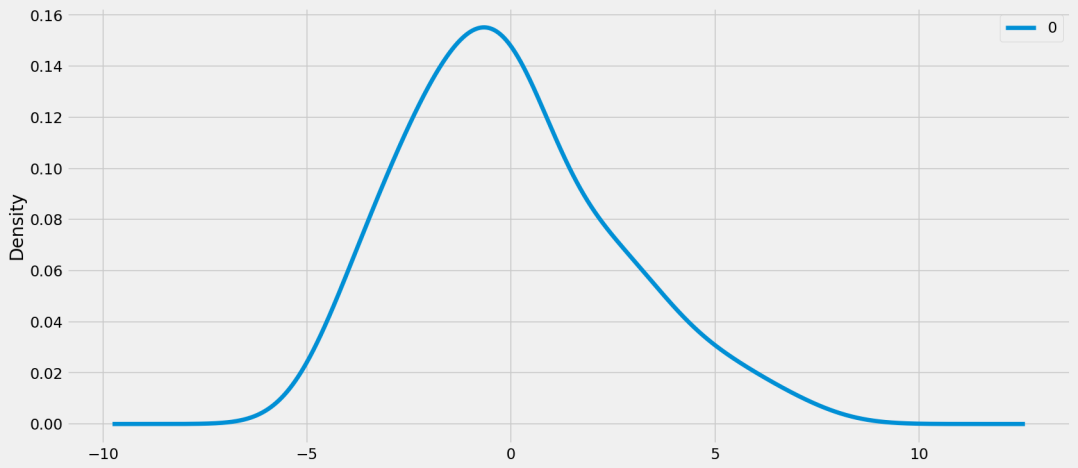
We get the following output:-
ADF = -6.518858048374907
p-value = 1.0544759906810726e-08
The small p-value indicates that the series is stationary hence we use the inital data as it is, without differencing for model fitting and forecasting.

5 Model fitting and Forecasting (Named Characters)

As the available data for this forecasting is limited we use a simple AR model by calling the ARMA model with zero differencing, zero moving-average window and 14 lag order for auto-regression for analyzing. A simple model fitting and forecast gives us the following output:-
array([1.57700512]) around 2
Following are the residual error and density plots for the model fit:-



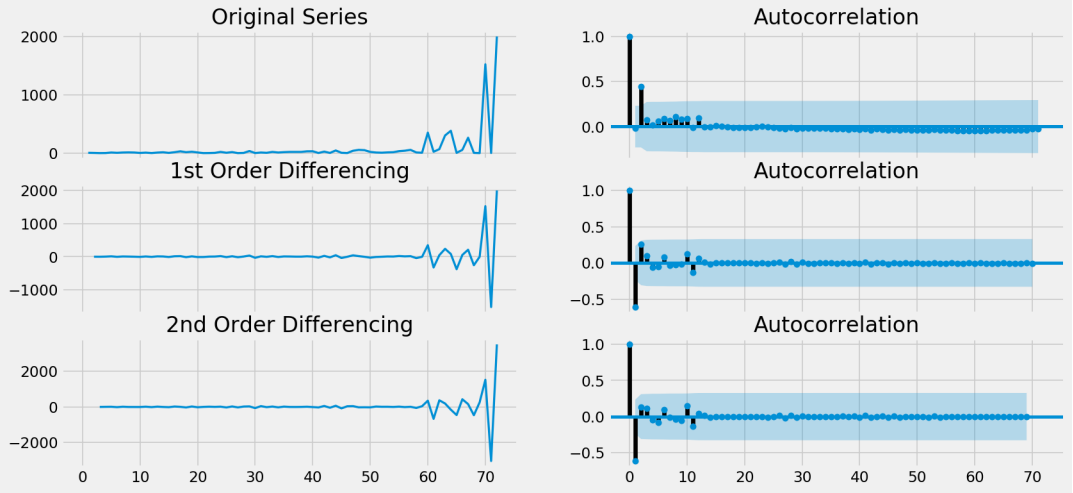
The line plot for errors suggests there are still trend information that has not been successfully captured by the model.



The errors although follow a Gaussian trend, it is clearly not centered at zero which hints a bias in the prediction.

6 Analysis(Overall Deaths)

We perform the same analysis, this time for all on-screen deaths for each episode. Following are the autocorrelation plots.

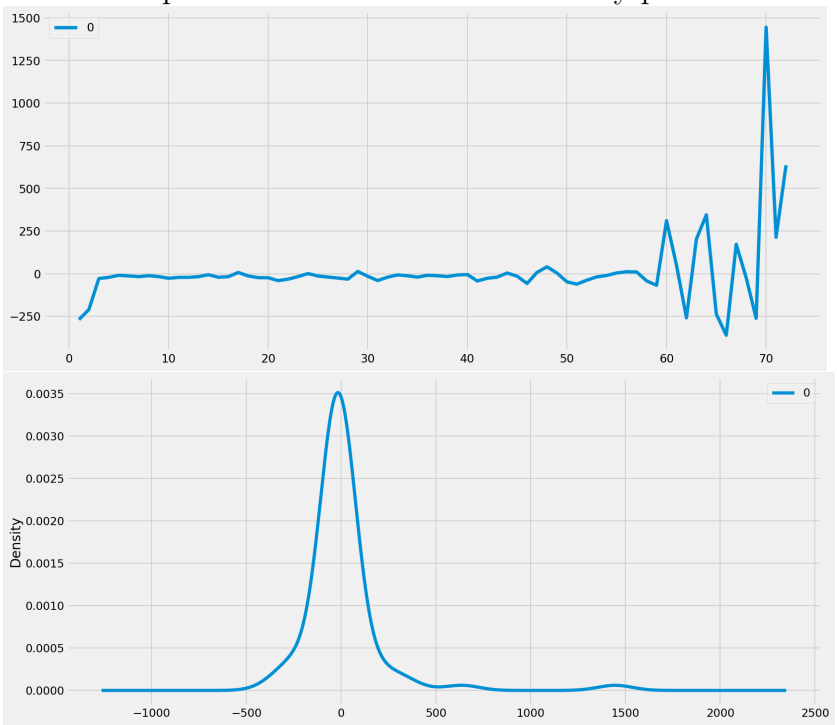


Clearly, there is neither a trend nor seasonality.

We perform the stationarity test further on. The Dickey-Fuller stationarity test gives the following output:-
ADF = 5.217848071894859
p-value = 1.0
The high p-value suggests, as expected non-stationarity of the data.

7 Model fitting and Forecasting (Overall Characters)

We again use the AR model by calling ARMA and plugging the Moving Average window to zero, to fit and forecast the non-stationary data. We get the following forecast for the last episode:-
20.615011 around 20
dtype: float64
We further plot the residual error and density plots for the model fit.



In this case the errors clearly follow the Gaussian trend and are centered almost at zero.

8 Results

After choosing simple AR models for both the data of on-screen deaths we conclude a prediction of around **2** 'Named' character deaths and **20** 'Overall' deaths in the final episode of Game of Thrones.

9 References

1.
Article title: Game of Thrones death count 2019 — Statistic
Website title: Statista
URL: https://www.statista.com/statistics/420077/game-of-thrones-deadliest-seasons/
2.
Article title: Game of Thrones Wiki
Website title: Gameofthrones.fandom.com
URL: https://gameofthrones.fandom.com/wiki/Game$_o f_T hrones_W iki$