1. **Problem:**

Game of Thrones is one of the most popular shows of the modern era. It was created based on the book “A song of ice and fire”. It is famous that every character in this book including the main characters is subject to death regardless of their popularity. The question arises, are these deaths random, or are they related to features and characteristics that every character from the book has. This is the problem that we are going to solve.

2. **Exploratory data analysis:**

The dataset consisted of 1946 characters and 26 features of which 13 had missing values. The data related to age had two observations that had an abnormal age value that was treated according to external research

The categorical data related to “culture” and “house” was also treated by grouping similar culture under one category

The missing values of age were later imputed with the mean. The categorical data was also filled with a categorical variable called “Unknown”

3. **Models:**

For the models, we have used KNN, Logistic Regression, Decision Tree, and Random Forest.

Logistic regression seemed to return better results and a better accuracy which made it our final model after cross validation which resulted in AUC = 0.781

4. Insights:

In order to be able to predict what character was going die or live, we have built a model that predicts those outcomes based on the most important features for every character, after running an algorithm that selects the most important features, the following features were found to be the most contributing were:

- Popularity was the most important feature
- Appearing in book 4 was the most important feature, followed by book 1, book 5 and book3
- Also, Gender, age and order of appearance were features to consider predicting the outcome for each character .