This Project will examine data from Global shark attacks (https://www.kaggle.com/teajay/global-shark-attacks). This data source contains 5992 instances of human shark interactions.From 1900 to 2016. This information of shark attacks comes from all across the world. There is no problem statement or business problem as such.The initial analysis is done to each feature and provides descriptive statistics on each country,Area,Age etc.Then further analysis is done by cleaning the data and creating month column from the given case numbers.

Further Hypothesis tests are carried out namely the chi-square test which tests for dependencies between two categories.

The final portion of this project deals with the predictive modelling which classifies fatal and non-fatal attacks. The ultimate goal of this project is to establish a proper difference between classifying fatalities with proper features like Age,Country and Month or just classifying fatalities with just a single injury text column.

Deliverables for this project will be 1. The code used for statistical tests and visual analysis 2. A PPT which will represent story and methods used.