Executive Summary

# What makes a book successful?

Since the creation of the Gutenberg press the numbers of writers and readers have increased. Change of power, revolution and peace have been influenced by words. The internet eased this process and sharing a book has never been easier.

This study has the objective of trying to find what factors make a successful book and provide a recommendation engine to publishers.

Attributes of final data model are:

* Year Of Publication: which year the book being evaluated has been publish
* Series?: Is the book part of a Serie
* 1st Publication Era: when has the book have first been published?
* Awards?: Has the book received an award?
* Genre1 - First level of genre:
  + Non-Fiction
  + Fiction
* Genre2 - Second level of genre:
  + Classics
  + Historical
  + Thriller
  + Biography
  + Contemporary
  + Fantasy
  + Others
* User Country:
  + USA
  + Other COuntries
* Group Age:
  + Child/Teenager (Age <= 19)
  + Young Adult (19 < Age < 39)
  + Adult/Senior (Age > 39)
* Successful?
  + Successful (Rating >= 8)
  + Regular

Target: **Successful?**

## Data sources:

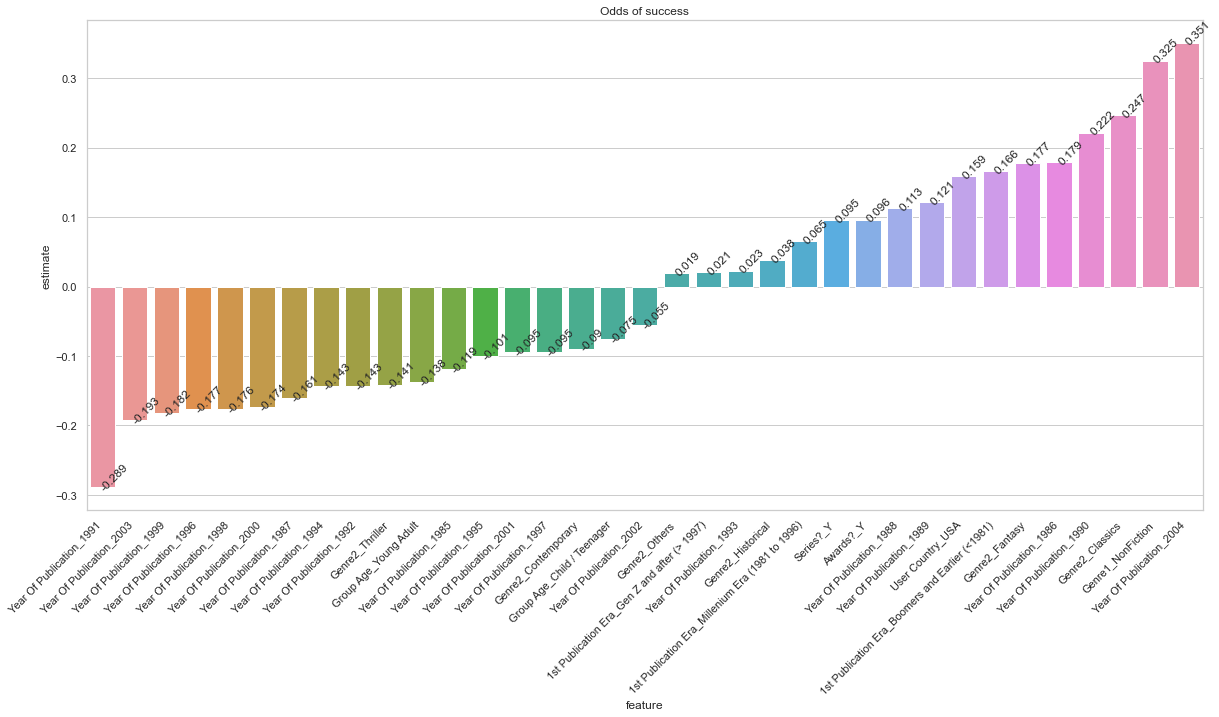
* Rating, Location, Age and other attributes: <http://www2.informatik.uni-freiburg.de/~cziegler/BX/>
* Genre, Setting, and other attributes: <https://zenodo.org/record/4265096#.YXlibBwpCHs>

## Model

The following models were evaluated:

* Logistic Regression
* KNN
* SVM
* AdaBoost
* Random Forest
* MLP

After having evaluate ROC AUC curve, classification report and confusion matrix of the models built, Logistic Regression is the chosen one with 56% accuracy and 70% precision to correctly classify Successful books and for its performance and simplicity for application and explanation for the following explanation:



# Conclusions

* Logistic Regression is the model of choice for its simplicity and good performance in correctly classify successful books
* After modeling, we can see, among other characteristics:
  + Books published in 2004 have the most odds of success with more than 35% compared to Before 1985
  + Non-Fiction have the best odds of being evaluated with almost 33% more odds than Fiction
  + Young Adults are the most difficult group to please with almost 3% less odds of success than an Adult/Senior
  + Classics still makes the best reading having almost 25% more odds of success than a Biography
  + Thriller has almost 18% odds of being successful than Biography
  + Recently published books have not performed well

# Next steps

Increase accuracy:

* Consolidate publisher to add to model
* Gather more data from other countries to expand User Country feature