EMC Education Services
Data Science & Big Data Analytics

Final Presentation for Analysts

Have you been wondering what the outcome of the Game of Thrones?

Fedorov Evgeniy
Osipov Samir
Panteleeva Natalia
Bachurin Maksim

19.01.2017

Situation & Project Goals

Situation

1. Game of thrones is one of the most popular tv show based on the book series «A Song of Fire and Ice» with large number of characters and battles between them.

2. The need of the audience (there may be different reasons) to get some special tool that provides them to be able to predict something they wants.

3. Possibility to predict outcomes of the next series could be usable for some persons or organizations (for example: betting shop).

Goals of "The game of thrones"

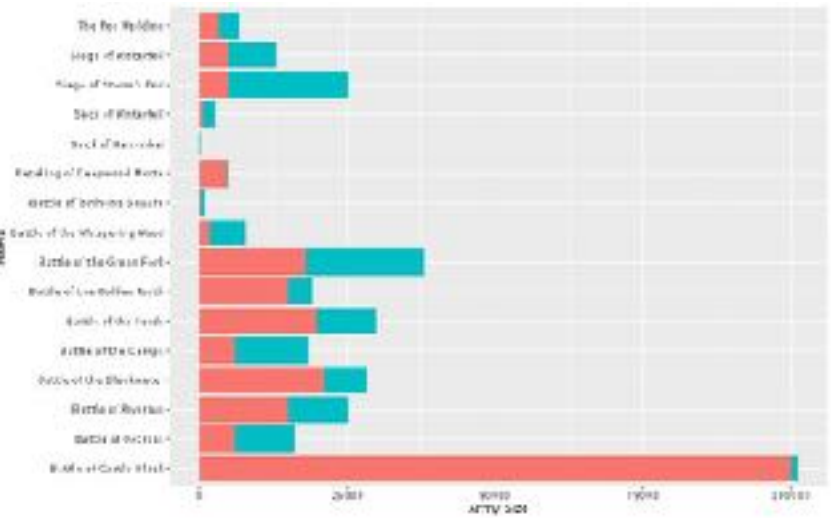
1. Find, prepare and visualize data.

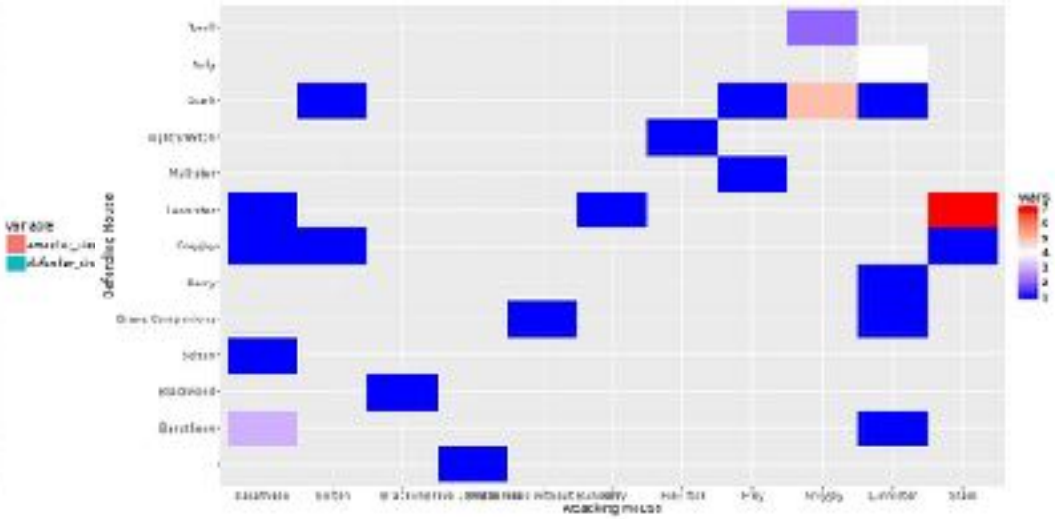
2. Define «the best» house, king's with largest attacker sizes etc.

Executive Summary

Purposes of our project have been reached.

- **The way to compare different houses according to its parameters was chosen**





Approach

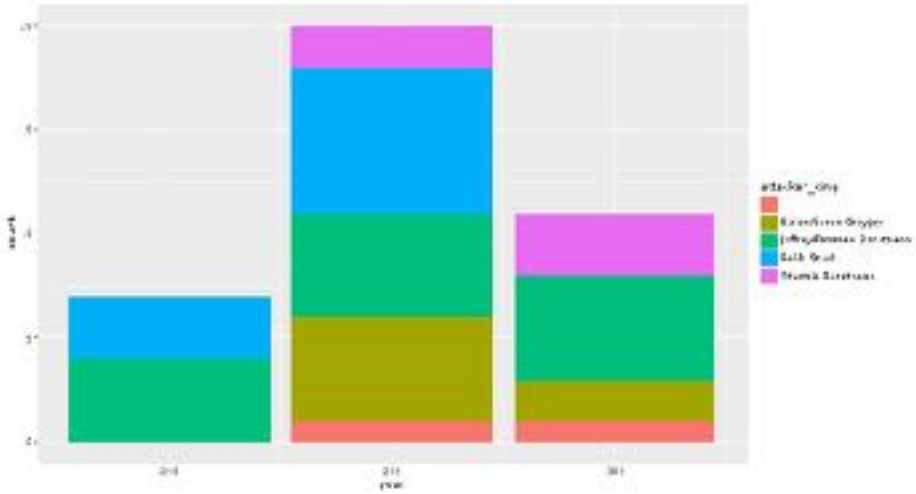
- Considered 12 houses of the world of «Game of thrones» using dataset from the kaggle.com
- Model was realized in RStudio
  - Minimizes variable transformations and binning
  - Good for continuous predictions
- Worked with dataset from kaggle.com to simulate model performance
- Dataset comparison method was used
- The model can be rapidly scored in the database over large datasets

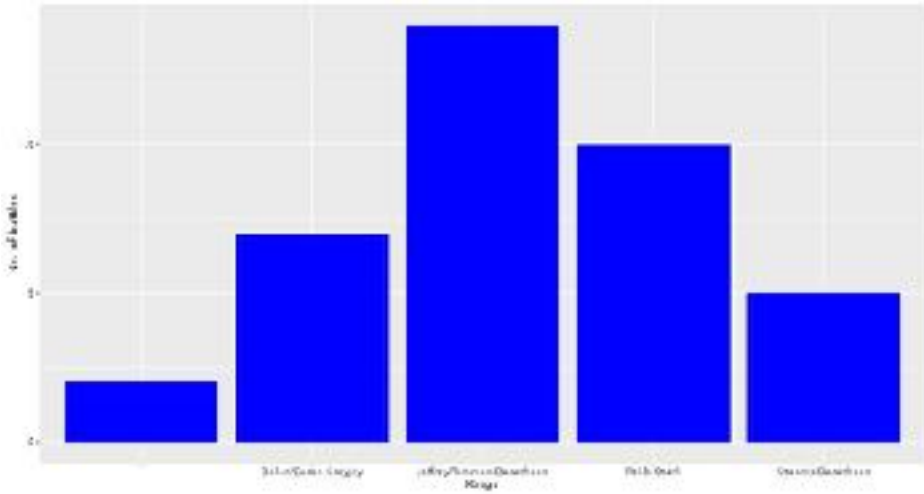
Model Description

- **Overview of Basic Methodology:** predict the possible future king of «Västerås» based on Data
- **Model:** comparison model
- **Scope**:
  - 12 houses including its army size (amount of warriors, battles, battle results, years)
  - 39 data entires
- **Sampling**
  - Training sample: 39
  - Testing sample: 39
- **The model developed has low errors**
  - We created a model with all variables
  - Some variables can be added to the model

Data comparison

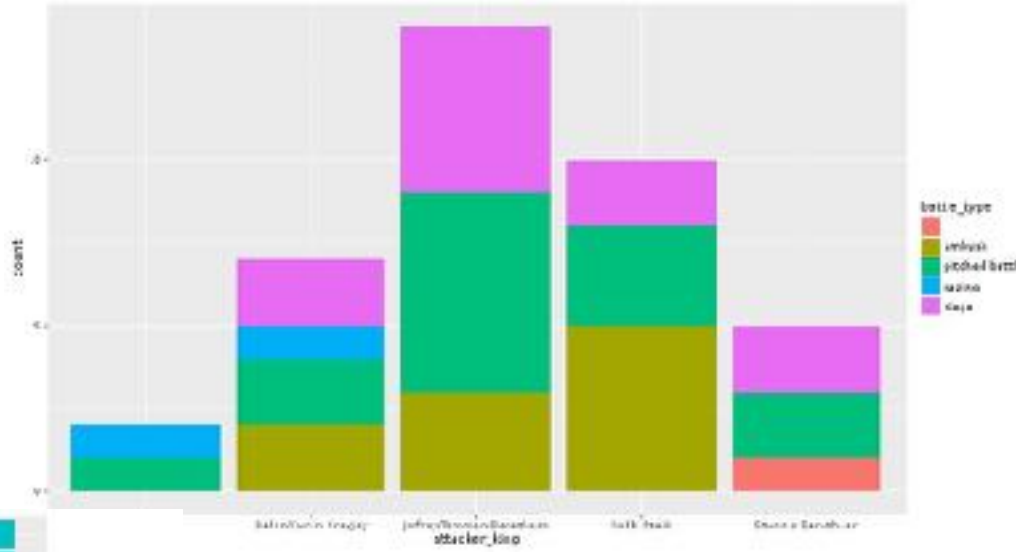
Plots show the dependance between kings and number of battles, year and count

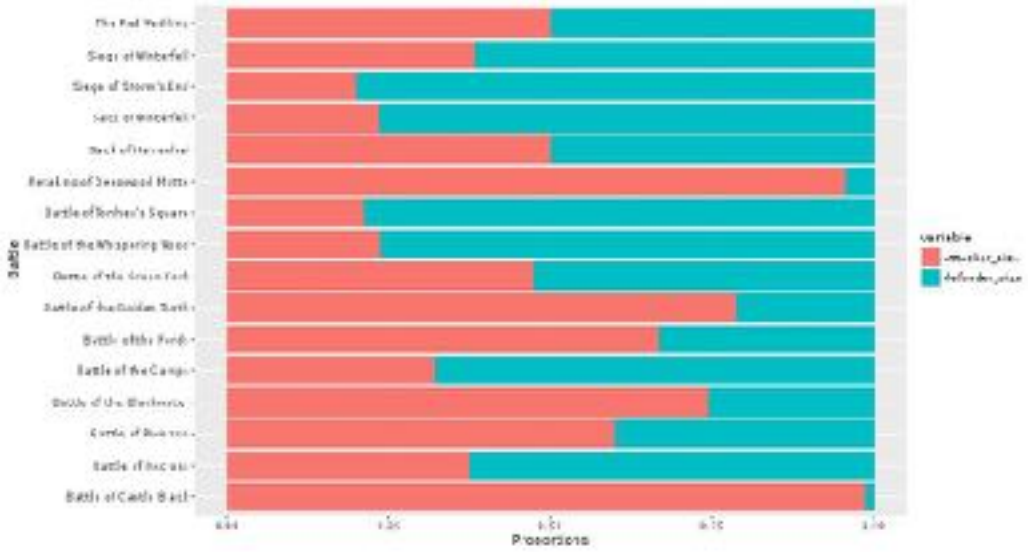




Data comparison

- attacker king/count
- battle/proportions





Model Details

- *Our model threw out 7 variables based in training*
- *Variables with highest impact:*
- *1) kings with highest number of battles;*
- *2) year with highest count of battles ;*
- *3) kings with highest count of attacks;*
- *Some points were loosed in the model:*
- *1) were expected more complicated model, but for «Game of thrones» it was enough.*