

Unlocking the Forge: Using Machine Learning to understand the game Keyforge.

Higher Diploma in Science in Data Analytics

**Proposal**

Word Count: 994

Barry Sheppard

Barry.Sheppard@gmail.com

07/09/2019

Contents

[Contents 2](#_Toc515325985)

[1. Introduction 3](#_Toc515325986)

[2. Project Scope and Objectives 3](#_Toc515325987)

[3. Student’s Learning Objectives 4](#_Toc515325988)

[4. Technical Specification of the Project 4](#_Toc515325989)

[5. Project Plan 4](#_Toc515325990)

[6. Conclusion 5](#_Toc515325991)

[7. References / Bibliography 5](#_Toc515325992)

# Introduction

KeyForge (“KeyForge,” 2019) is a collectable card game developed by Fantasy Flight Games (“Fantasy Flight Games,” 2019), similar to Pokemon (“Pokémon Trading Card Game,” 2019), Yu-Gi-Oh (“Yu-Gi-Oh!,” 2019), and Magic the Gathering (“Magic: The Gathering,” 2019). Where it sets itself apart is in the sales model. Rather than selling players of the game random cards which are then combined to create decks, the company sells pre-generated random decks which cannot be changed. These decks are all unique meaning no two decks contain the same combination of cards. The model means a player can buy a single deck and be ready to play. It also means the more dedicated players will buy multiple decks to get the best deck they can (D’Anastasio, 2018).

From a data analysis perspective, this provides a discrete problem to solve. If we were to buy two decks, could we determine which was the better deck? First, we would need some data on which to base our decisions. Luckily, after a player buys a deck, they register it on the companies site (“KeyForge Master Vault,” 2019) which shows a full list of all the cards in that deck. This listing of cards will be our predictor variables. Good cards or card combinations in the deck will improve the chance of the deck winning. Secondly, when players enter into tournaments, the wins and losses of each deck are recorded. So we have an outcome variable, decks with higher win rates have demonstrated better performance than those with low ones.

If we can extract this data from the company’s website, clean that data for use, determine the best machine learning model, we can then create a predictor to select the best deck from those presented. A system like this would allow players of the game select the most competitive deck they have and could also be used by the developer of the game to determine any balance issues.

# Project Scope and Objectives

The scope of this project is narrow due to the available data. While there are over 1.25 million registered decks, the info available on those decks is minimal. We merely know the 36 cards in each deck and the number of wins and losses during tournaments. We do not know the skill of the players involved, the power of the deck's played against, or how lucky the player was in their draw of the random cards. Based on that, the project scope and objectives are similar to most data analysis projects.

1. Extract the data from the website.
2. Clean and structure the data for analysis.
3. Create descriptive plots and detail any insights based on the data.
4. Generate machine learning models based on the data.
5. Test and improve the models.
6. Using the final model, create a program that will select the best deck from a range.
7. Detail any insights developed from the model.

# Student’s Learning Objectives

During the course, we have completed extensive work in machine modelling methods, and we have completed models in Python programming. Our experience putting those machine learning methods into practice using Python is however minimal as the majority of work was completed using R.

The primary learning objective during this project is to become familiar with machine learning methods within Python. Before this, data will need to be extracted and cleaned. Rather than looking for an existing dataset, a component of this project will focus on using Python to complete that element.

# Technical Specification of the Project

Based on the current scope, the expected technology is minimal with the initial extract completed with a Python script and the data analysis element completed in Jupyter Notebook (“Project Jupyter,” 2019) or Google Colab (“Google Colab,” 2019).

# Project Plan

This project proposal is due for submission on the 10th of September with supervisor assigned by the end of that week. The initial meeting with supervisor is due the week after which sets the start point for the project as the 23rd of September. The interim report submission is due on the 25th of October leaving roughly 4 weeks of work followed by a week of preparation of the report. The interim presentation is due to occur the week of the 28th of October with the next major milestone the final report submission just over 5 weeks later on the 13th of December. The below table details a proposed project plan:

|  |  |
| --- | --- |
| **Week Starting** | **Activity** |
| 16/09/2019 | Meeting with Supervisor |
| 23/09/2019 | - Initial project setup and planning |
| 30/09/2019 | - Extracting the data |
| 07/10/2019 | - Data preparation and cleaning |
| 14/10/2019 | - Initial insights based on the existing data |
| 21/10/2019 | Interim Report Submission |
| 28/10/2019 | Interim Report Presentation |
| 04/11/2019 | - Basic machine learning models |
| 11/11/2019 | - More advanced machine learning |
| 18/11/2019 | - Testing and tuning the models |
| 25/11/2019 | - Use the model to create a predictor |
| 02/12/2019 | - Extra week for inevitable delays |
| 09/12/2019 | Final Report Submission |
| 16/12/2019 | Final Presentation |

# Conclusion

While in structure, this project seems straightforward, it includes many challenges.

Extraction of the data will require web scraping as the data is not readily available. Next, the data must be structured in a format suitable for analysis. The anticipated extract format will have around 1.25M rows, one for each deck registered with a list of ids for the cards. This data will need to be restructured to have a column for every potential card (~500 cards in total) along with deck metadata, including the win rate.

The large number of features also poses a challenge. Issues like this have theoretical solutions but must also be considered within the context of the subject. For example, as decks are limited to 60 cards in total, the majority of features will be 0, so feature selection may not be appropriate. Resolving this may require some more advanced data machine techniques closer to deep learning.

With this considered, the proposal should provide a challenging but achievable piece of work that will provide excellent development opportunities.

# References / Bibliography

D’Anastasio, C., 2018. KeyForge Is A Whole New Kind Of Card Game [WWW Document]. URL <https://kotaku.com/keyforge-is-breaking-the-rules-of-collectible-card-game-1830544174> (accessed 9.10.19).

Fantasy Flight Games [WWW Document], 2019. URL <https://www.fantasyflightgames.com/en/index/> (accessed 9.10.19).

Google Colab [WWW Document], 2019. URL <https://colab.research.google.com> (accessed 9.10.19).

KeyForge Master Vault [WWW Document], 2019. URL <https://www.keyforgegame.com/> (accessed 9.10.19).

KeyForge [WWW Document], 2019. URL <https://www.fantasyflightgames.com/en/products/keyforge/> (accessed 9.10.19).

Magic: The Gathering [WWW Document], 2019. . MAGIC: THE GATHERING. URL <https://magic.wizards.com/en> (accessed 9.10.19).

Pokémon Trading Card Game [WWW Document], 2019. URL <https://www.pokemon.com/us/pokemon-tcg/> (accessed 9.10.19).

Project Jupyter [WWW Document], 2019. URL <https://www.jupyter.org> (accessed 9.10.19).

Yu-Gi-Oh! [WWW Document], 2019. URL <https://www.yugioh-card.com/en/> (accessed 9.10.19).