**Readme:**

This folder contains all the explanations for the Jupyter notebooks, excel spreadsheets and the interim csv files made during my dissertation about RPGLite Log analysis.

The Original JSON file that was downloaded from http://researchdata.gla.ac.uk/1070/ is not included.

Screendata is omitted as well, due to its large size over 100MB and github policies.

**Jupyter notebook files**

**Players.ipynb** : The Jupyter notebook that contains EDA such as health points, win vs skill points, player demographics and the preliminary data wrangling processes. Used for data preprocessing and the whole research overall, especially Chapter 2 background section, Chapter 3, and Chapter 4 Sections 2 and 3.

**ML\_new.ipynb** and **ML\_old.ipynb** : Used for machine learning methodology in Chapter 4 Section 4. ML\_new is the file that contains machine learning model with new skill points, and ML\_old is the file that contains machine learning with old skill point metrics.

**MLModel.ipynb** : Used in Chapter 4 Sections 4 and 5. Previous attempts at modeling and encoding (one-hot, popularity, win rate). Later half of the file contains attempts at linear regression and multivariate linear regression.

**ML\_newtarget.ipynb** : Used in Chapter 4 Section 4. First attempt to use machine learning model with player 1 / player 2 as the target variable

**Linreg\_winpopturn.ipynb** : Multivariate linear regression for finding popularity with win rate and turn length as the variables. (Used for Chapter 4 Section 6)

**CSV files**

The files noted here are directly or indirectly affiliated with the Jupyter notebooks mentioned above.

**Games.csv** : CSV file containing the games data from JSON. Used in **Players.ipynb.** Added into SQLite DB for further usage.

**Players.csv** : CSV file containing the players data from JSON. Used in **Players.ipynb.** Added into SQLite DB for further usage.

**Merged.csv** : CSV file containing modified data from Players.ipynb such as win rates and character select / wins. Used in **Players.ipynb** and **MLModel.ipynb.** Added into SQLite DB for further usage.

**Rpglite\_wrangled.csv** : contains the data of match results with character pairs represented in alphabet initials and the winner replaced as ID. Later deprecated as the ML model started to use ‘player 1/player2’ as the target variable(check games\_all.csv). Used in **MLModel.ipynb** as a part of interim research

**Games\_all.csv** : used as a one-fits-all data csv for ML modeling in ML\_new.ipynb and ML\_old.ipynb. Games\_old, prediction\_popularity, prediction\_onehot are all deprecated versions of games\_all.csv.

**Prediction\_popularity.csv,prediction\_onehot.csv :** interim products that led to games\_all.csv

**Winpopturn.csv**: csv with the popularity, winrate and average turns of the character pairs. Used for regression in **Linreg\_winpopturn.ipynb**

**Afterpatch.csv** : csv that contains all the character combinations after patch 1.2. Used for correlational analysis and regression in **MLModel.ipynb**

**Winrates.csv**: csv that contains all the users’ win rates of character pairs.

**Games\_newmmr.csv :** interim product that contains game data for calculation with the character pair win rates from winrates.csv used to encode character pair parameters.

**Players2.csv :** players with skill points encoded according to the equation in Figure 4.8

**Rpglite\_postpatch.csv :** csv containing activities of players after the patch 1.2