

API - Web scraping - Project Ironhack Data Bootcamp

Miguel Ángel Ávalos

Data Part Time Barcelona Dic 2019

Project Description

El objetivo de este proyecto consiste en obtener información a través de API y páginas web y obtener tablas para su posterior análisis. La temática escogida ha sido el tratar de obtener información sobre las cartas presentes en un nuevo formato del juego de cartas coleccionables Magic the Gathering, llamado *Pioneer*.

Por ello, en primer lugar y mediante **APIs** se ha obtenido el listado de las cartas posibles válidas en el formato así como los diferentes sets y expansiones que lo componen.

Complementariamente, mediante **web scrapping** se han obtenido diferentes datos sobre los tres grandes eventos mundiales de este formato: los torneos de Bruselas, Nagoya y Phoenix. Los datos obtenidos hacen referencia a los mazos (es decir el conjunto de cartas que cada jugador ha utilizado en el torneo), así como los resultados de los emparejamientos de cada ronda del torneo.

APIs

Las APIs han sido las siguientes:

- <https://scryfall.com/docs/api> : Obtención del listado de sets o colecciones que componen el formato *Pioneer*
- <https://docs.magicthegathering.io/> : Obtención del nombre de las cartas de cada uno de los sets.

WEB

La web consultada para obtener los resultados de los diferentes torneos y de los mazos utilizados por cada jugador son las siguientes:

Por ejemplo en el caso de Bruselas:

- Listado de los resultados de cada jugador en cada ronda del torneo (de 1 a 16 en caso de Bruselas)
<https://magic.gg/news/players-tour-brussels-2020-round-1-results>
- Listado de las listas de los mazos utilizados por cada jugador en el torneo:
<https://magic.wizards.com/en/events/coverage/players-tour-brussels-2020-decklists>

Trabajo realizado

En una primera fase, se ha obtenido mediante llamadas a las APIs los sets o colecciones los cuales son válidos para jugar en el formato Pioneer mediante el uso de la librería *json* y su posterior conversión a *dataframe* mediante *pandas*.

Posteriormente, mediante el uso de las librerías *requests* y *Beautiful Soup* se han parseado los datos para posteriormente realizar una serie de transformaciones y uniones de diferentes datos. Se ha obtenido el listado de copias de las cartas jugadas y se ha añadido a la información previa sobre las cartas de las APIs.

Además, con los resultados de los enfrentamientos entre jugadores y tras varias transformaciones se ha obtenido el % de juegos ganados (win rate) de cada uno de los mazos entre sí.

Resultados

Se adjuntan las siguientes tablas:

- *pioneer_sets.csv*: datos del listado de sets legales en el formato *Pioneer* junto con el número
- *pioneer_cards.csv*: datos con varias estadísticas sobre las diferentes cartas, junto con el número total de copias jugadas en cada torneo más el total acumulado.
- *deck_results.csv*: datos con los resultados de cada enfrentamiento de cada tipo de mazo obtenido a través de las decklist de cada jugador de cada torneo.