**Project: “Magic: The Gathering”**

**Picchia – Petroni – Redaelli**

**Task 1: Data Collection**

The Magic: The Gathering (MTG) API can be used to access the API endpoints, in order to get information on MTG cards and sets. All API access is performed over HTTPS.

**API endpoints format:** *https://api.magicthegathering.io/<version>/<resource>*

**Rate limits:** Third-party applications are officially throttled to 5000 requests per hour.

We initially create a dataset containing all the cards in the database. Each record represents an individual card; attributes are defined below.

**API mandatory parameters selection:**

*version*: **v1** (the only version available to date)

*resource*: **cards**

*https://api.magicthegathering.io/v1/cards*

By analyzing the “*link”* response header – after making a sample GET request to the API above – it is possible to identify the total number of API pages available: 623 (to date). The target page to be retrieved can be specified with an optional parameter *“page”* in the URL.

*https://api.magicthegathering.io/v1/cards?page=1*

We make a GET request to all the database pages available and store the retrieved data.

The rate limits allow us to build the whole dataset in one single code execution.

The dataset retrieved is made up of 62261 observations; parameters are as follows.

We build our final dataset by sampling 10K observations at random, without replacement.

A preliminary data cleaning is also performed: observations for which the *“multiverseid”* parameter is NULL (special editions of cards already included in the database) are excluded, preventing the inclusion of duplicated or extremely similar rows.

|  |  |
| --- | --- |
| **Parameter name** | **Description** |
| **artist** | The artist of the card. This may not match what is on the card as MTGJSON corrects many card misprints |
| **cmc** | Converted mana cost. Always a number |
| **colorIdentity** | The card’s color identity, by color code. [“Red”, “Blue”] becomes [“R”, “U”]. A card’s color identity includes colors from the card’s rules text |
| **colors** | The card colors. Usually this is derived from the casting cost, but some cards are special (like the back of dual sided cards and Ghostfire) |
| **flavor** | The flavor text of the card |
| **foreignNames** | Foreign language names for the card, if this card in this set was printed in another language. An array of objects, each object having ‘language’, ‘name’ and ‘multiverseid’ keys. Not available for all sets |
| **hand** | Maximum hand size modifier. Only exists for Vanguard cards |
| **id** | A unique id for this card. It is made up by doing an SHA1 hash of setCode + cardName + cardImageName |
| **imageUrl** | The image url for a card. Only exists if the card has a multiverse id |
| **layout** | The card layout. Possible values: normal, split, flip, double-faced, token, plane, scheme, phenomenon, leveler, vanguard, aftermath |
| **legalities** | Which formats this card is legal, restricted or banned in. An array of objects, each object having ‘format’ and ‘legality’ |
| **legality** | The legality of the card for a given format, such as Legal, Banned or Restricted |
| **life** | Starting life total modifier. Only exists for Vanguard cards |
| **loyalty** | The loyalty of the card. This is only present for planeswalkers |
| **manaCost** | The mana cost of this card. Consists of one or more mana symbols. (use cmc and colors to query) |
| **multiverseid** | The multiverseid of the card on Wizard’s Gatherer web page. Cards from sets that do not exist on Gatherer will NOT have a multiverseid. Sets not on Gatherer are: ATH, ITP, DKM, RQS, DPA and all sets with a 4 letter code that starts with a lowercase ‘p’ |
| **name** | The card name. For split, double-faced and flip cards, just the name of one side of the card. Basically each ‘sub-card’ has its own record |
| **number** | The card number. This is printed at the bottom-center of the card in small text. This is a string, not an integer, because some cards have letters in their numbers |
| **originalText** | The original text on the card at the time it was printed. This field is not available for promo cards |
| **originalType** | The original type on the card at the time it was printed. This field is not available for promo cards |
| **power** | The power of the card. This is only present for creatures. This is a string, not an integer, because some cards have powers like: “1+\*” |
| **printings** | The sets that this card was printed in, expressed as an array of set codes |
| **rarity** | The rarity of the card. Examples: Common, Uncommon, Rare, Mythic Rare, Special, Basic Land |
| **rulings** | The rulings for the card. An array of objects, each object having ‘date’ and ‘text’ keys |
| **set** | The set the card belongs to (set code) |
| **setName** | The set the card belongs to |
| **subtypes** | The subtypes of the card. These appear to the right of the dash in a card type. Usually each word is its own subtype. Example values: Trap, Arcane, Equipment, Aura, Human, Rat, Squirrel, etc. |
| **supertypes** | The supertypes of the card. These appear to the far left of the card type. Example values: Basic, Legendary, Snow, World, Ongoing |
| **text** | The oracle text of the card. May contain mana symbols and other symbols |
| **toughness** | The toughness of the card. This is only present for creatures. This is a string, not an integer, because some cards have toughness like: “1+\*” |
| **type** | The card type. This is the type you would see on the card if printed today. Note: The dash is a UTF8 ‘long dash’ as per the MTG rules |
| **types** | The types of the card. These appear to the left of the dash in a card type. Example values: Instant, Sorcery, Artifact, Creature, Enchantment, Land, Planeswalker |
| **variations** | If a card has alternate art (for example, 4 different Forests, or the 2 Brothers Yamazaki) then each other variation’s multiverseid will be listed here, NOT including the current card’s multiverseid |
| **watermark** | The watermark on the card. Note: Split cards don’t currently have this field set, despite having a watermark on each side of the split card |