

Análise da API da Marvel

A análise foi feita em uma parte dos dados pois a API restringe o limite de 100 linhas por requisição, então por fins de praticidade foram limitados a 10 requisições com limite de 100.

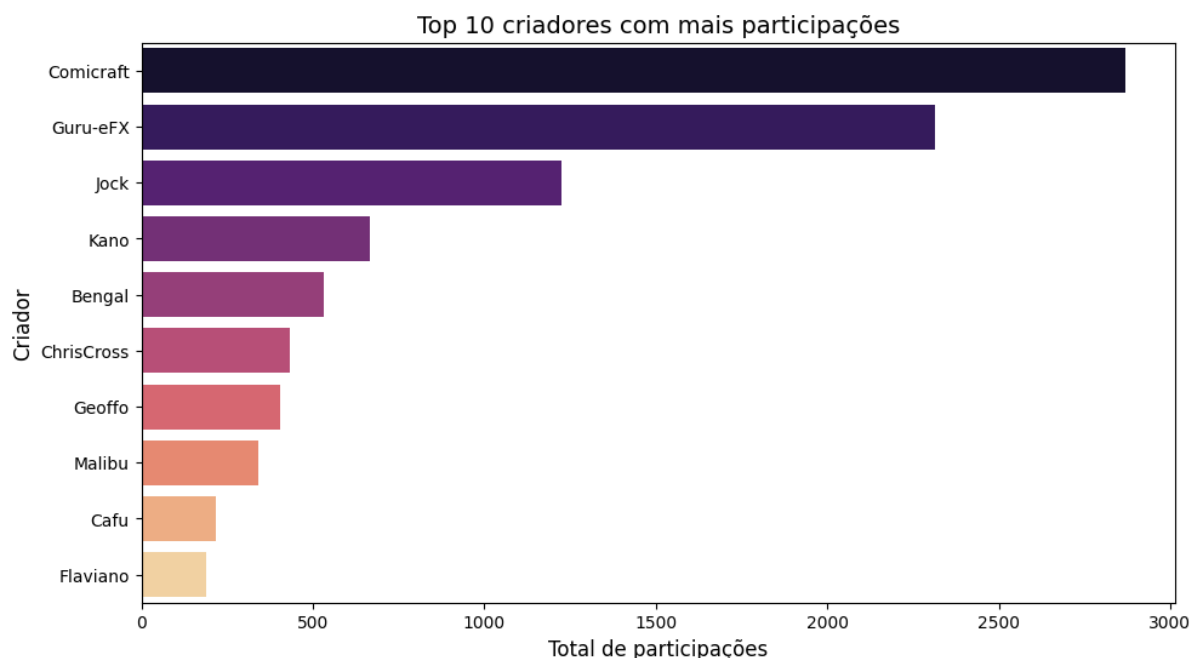
Para o consumo dessa API necessita de uma 'apikey' definida por timestamp + Private Key + Public Key em hash MD5

(<https://developer.marvel.com/documentation/authorization>)

Devido à limitação das requests foi implementado 'offset' que coleta os dados da API por turnos e os agrupa em um único Data Frame.

O primeiro endpoint a ser consumido foi o de Creators, então foi criado uma nova feature no Data Frame "participations" que é a soma de comics, series, stories, events.

O gráfico a seguir mostra os criadores com mais participações na amostra de dados coletados.



Endpoint COMICS

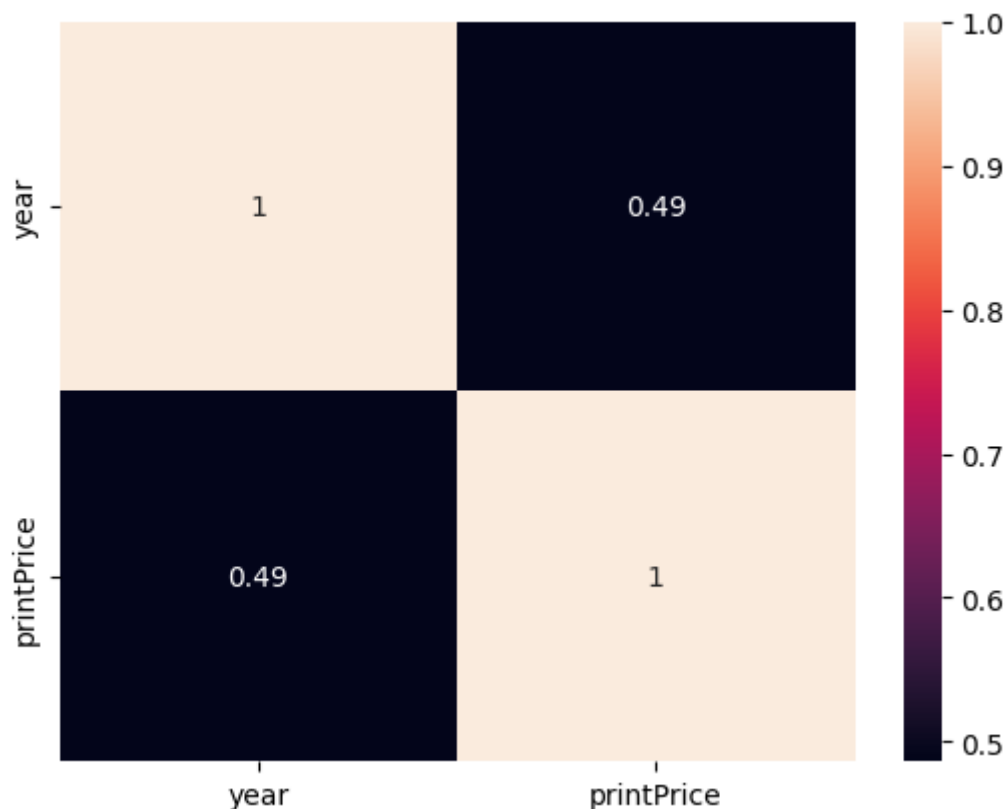
Usando as mesmas técnicas já citadas anteriormente a respeito do consumo da API, agora foi coletado o endpoint comics.

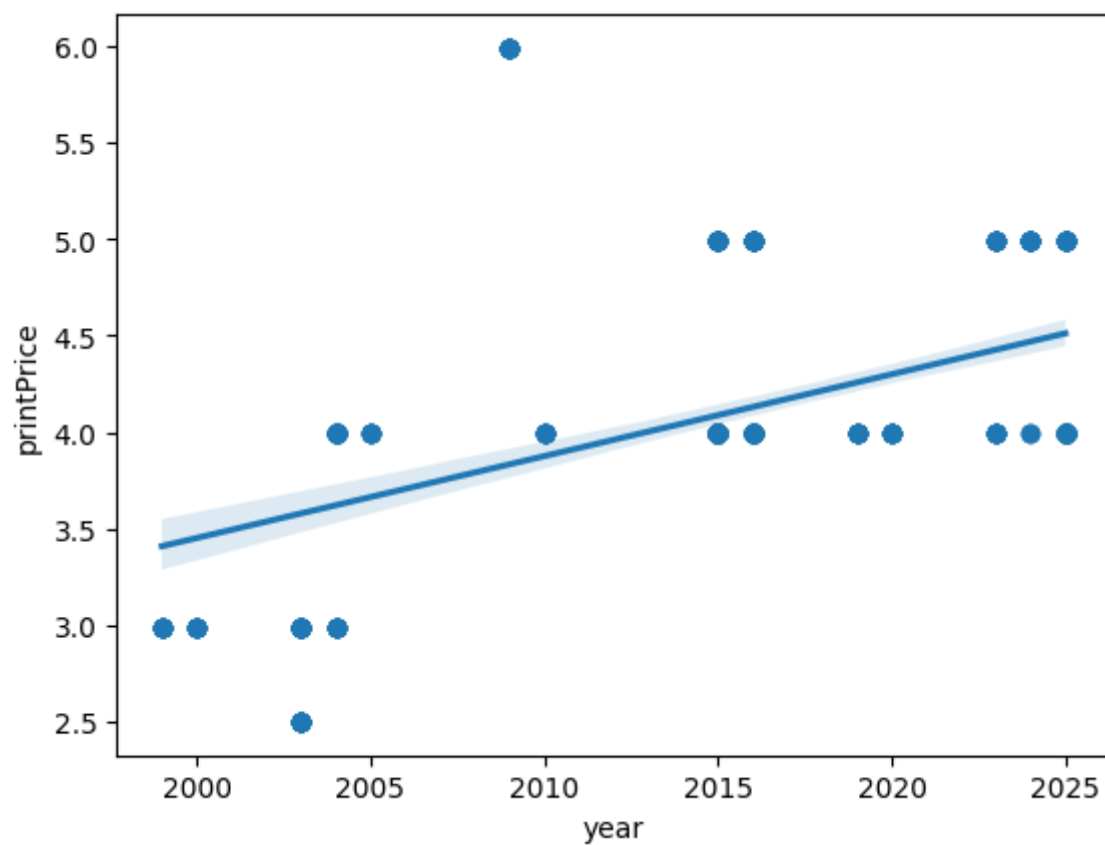
A feature “prices” do Data Frame inicialmente era um dicionário com as chaves “printPrice” e “digitalPurchasePrice”, então foram criadas novas features no Data Frame, sendo essas duas chaves do dicionário. Além de ser removida a coluna ‘prices’, pois era redundante.

Foi percebido que na maioria dos títulos havia o ano de publicação, sendo uma informação valiosa. Para o uso desses dados foi usado regex que coleta a presença de 4 números inteiros em sequência, para então ser criado uma nova feature ‘year’.

Havia títulos que não possuíam o ano de publicação, então essas linhas foram removidas para esta análise.

E com base nessas informações e nos gráficos a seguir, nota-se a correlação positiva moderada entre o ano de publicação e o preço.





Ou seja, a tendência é que com o passar dos anos o preço dos comics aumente, sendo impossível analisar por meio dos dados desta API o motivo aparente.