

# Desafio Quantitas

Eduardo Renani Ribeiro

<https://github.com/EduardoRenani/Desafio-Quantitas>

## 1. Programa:

### 1.1 Extração de dados:

Foram criados quatro arquivos em Python (com o nome das quatro revendedoras) que utilizam a biblioteca Selenium para entrar nos sites e extrair todo o conteúdo HTML dos mesmos. Como alguns sites não tinham explicitamente o número total de páginas a serem extraídas, o programa procura pelo botão de próxima página e o clica, utilizando um “*try-catch*” para quando chegar na última página e não achar o botão de próximo. Para contornar problemas de conexão, foi implementado um “*retry-catch*” que procura os elementos que contém as informações dos carros antes de lê-los e no caso de não achar, atualiza a página e procura de novo até achá-los. Ainda nesses arquivos, acontece a parte de limpeza e padronização dos dados. Como alguns sites tem os nomes dos veículos fora do padrão, há alguns “*if-else*’s” que visam reescrever o nome de alguns modelos conforme o padrão da tabela fiipe, visto que mais a frente será necessário o “*match*” entre os dados extraídos e os dados fiipe.

### 1.2 Dados Fiipe:

Foi criado um arquivo para se comunicar com uma API (<http://fiipeapi.appspot.com/>) que disponibiliza os dados fiipe de todos os modelos de veículo a partir de uma comunicação simples e direta. Porém, esta API é desenhada para funcionar com um número baixo de requisições, o que obrigou o programa a trabalhar com delay explícito nas iterações a fim de não ter a conexão interrompida com o servidor da API.

Uma vez tendo todos os dados de carros de pequeno porte em mãos, o programa modifica alguns nomes de fabricantes e carros para que fiquem em conformidade com o arquivo de dados gerado na etapa 1.1 e os salva em formato csv. Após isso, realiza-se casamento entre o par (FABRICANTE+MODELO+ANO, preço fiipe) e a tupla (FABRICANTE+MODELO+ANO, preço, quilometragem) que constam nos arquivos *fiipe.csv* e *carros.csv* respectivamente. Com isso, os dados são salvos num terceiro arquivo (*carrosEFiipe.csv*).

### 1.3 Geração do Database:

Por uma questão de organização e possibilidade de rodar apenas a geração do database sem precisar passar pelos passos 1.1 e 1.2, foi criado o terceiro arquivo *geraDatabase.py*, que apenas abre o arquivo *carrosEFiipe.csv* e realiza o cálculo de desconto sob fiipe e desconto por quilometragem. Estes cálculos têm como objetivo avaliar a porcentagem de desconto da revendedora em relação ao preço fiipe e relativizar isso à

rodagem do veículo, visto que o uso gera depreciação e isto é embutido no preço de maneira diferente por cada empresa.

## 2. Resultados:

Os resultados gerados possuem algumas discrepâncias notadas em rápida avaliação do banco de dados. Isso se dá pela falta de informação de alguns sites em relação aos seus veículos. Um exemplo rápido é comparar o preço de dois carros do mesmo modelo, ano e quilometragem, mas câmbios automático e manual, motores 1.0 e 1.4, além de especificações de série como ar condicionado, vidro elétrico e outras características diferentes. Os dois veículos terão valor fiipe igual segundo o programa, pois os dados diferentes não foram especificados e os dois veículos foram processados como idênticos.

Ainda pelo banco de dados é possível notar carros com valores muito acima ou muito abaixo do praticado pela média das empresas, e, buscando informações destes veículos em específico nos sites, pode-se confirmar que se trata de uma oportunidade real de compra de um ativo abaixo do preço de mercado ou de um ativo supervalorizado.

database.csv

File Edit View Insert Format Data Tools Add-ons Help

All changes saved in Drive

100% - \$ % .0 .00 123 - Arial - 10 - B I S A

fx 0.525133691373085

	A	B	C	D	E	F	G	H	I	J	K
1	Revendedor	Fabricante	Modelo	Ano	Kilometragem	Preço	Preço Fiipe	%Desconto-I	%Desconto/1000Kilometros		
2247	Movida	RENAULT	DUSTER	2018	26747	58360	67897	14.05%	0.5251336914		
2416	Movida	RENAULT	DUSTER	2018	28922	58360	67897	14.05%	0.4856438121		
2512	Movida	RENAULT	DUSTER	2018	29827	58360	67897	14.05%	0.4709090779		
2555	Unidas	RENAULT	DUSTER	2018	26752	59490	67897	12.38%	0.4628262388		
2627	Movida	RENAULT	DUSTER	2018	35006	57160	67897	15.81%	0.4517284807		
2709	Movida	RENAULT	DUSTER	2018	35947	57160	67897	15.81%	0.4399037199		
2770	Movida	RENAULT	DUSTER	2018	32529	58360	67897	14.05%	0.4317945274		
2777	Movida	RENAULT	DUSTER	2018	36718	57160	67897	15.81%	0.4306669279		
2796	Movida	RENAULT	DUSTER	2018	36985	57160	67897	15.81%	0.4275579658		
2816	Movida	RENAULT	DUSTER	2018	37309	57160	67897	15.81%	0.4238450529		
2827	Movida	RENAULT	DUSTER	2018	33325	58360	67897	14.05%	0.4214810051		
2858	Movida	RENAULT	DUSTER	2018	38058	57160	67897	15.81%	0.4155037948		
2865	Movida	RENAULT	DUSTER	2018	38209	57160	67897	15.81%	0.4138617881		
2867	Movida	RENAULT	DUSTER	2018	38250	57160	67897	15.81%	0.4134181832		
2873	Unidas	RENAULT	DUSTER	2018	30029	59490	67897	12.38%	0.4123206916		
2876	Movida	RENAULT	DUSTER	2018	38424	57160	67897	15.81%	0.4115461008		
2878	Movida	RENAULT	DUSTER	2018	34168	58360	67897	14.05%	0.4110824425		
2890	Movida	RENAULT	DUSTER	2018	34289	58360	67897	14.05%	0.4096318453		
2919	Movida	RENAULT	DUSTER	2018	39263	57160	67897	15.81%	0.402752112		
2921	Movida	RENAULT	DUSTER	2018	39281	57160	67897	15.81%	0.4025675608		
2948	Movida	RENAULT	DUSTER	2018	39353	57160	67897	15.81%	0.4018310445		
2960	Movida	RENAULT	DUSTER	2018	39477	57160	67897	15.81%	0.4005688972		
3003	Movida	RENAULT	DUSTER	2018	40260	57160	67897	15.81%	0.3927785928		
3008	Unidas	RENAULT	DUSTER	2018	31670	59490	67897	12.38%	0.3909567228		
3009	Unidas	RENAULT	DUSTER	2018	31672	59490	67897	12.38%	0.3909320357		
3117	Movida	RENAULT	DUSTER	2018	41538	57160	67897	15.81%	0.3806942614		
3155	Unidas	RENAULT	DUSTER	2018	33135	59490	67897	12.38%	0.3736718484		
3186	Movida	RENAULT	DUSTER	2018	43034	57160	67897	15.81%	0.3674604142		
3199	Movida	RENAULT	DUSTER	2018	43403	57160	67897	15.81%	0.3643364419		
3210	Unidas	RENAULT	DUSTER	2018	34333	59490	67897	12.38%	0.3606334936		
3231	Movida	RENAULT	DUSTER	2018	44300	57160	67897	15.81%	0.3569594123		
3236	Movida	RENAULT	DUSTER	2018	44382	57160	67897	15.81%	0.3562999104		
3288	Movida	RENAULT	DUSTER	2018	48105	56560	67897	16.70%	0.3470949653		
3289	Unidas	RENAULT	DUSTER	2018	35909	59490	67897	12.38%	0.3448061924		
3311	Unidas	RENAULT	DUSTER	2018	36090	59490	67897	12.38%	0.3430769546		
3319	Movida	RENAULT	DUSTER	2018	48810	56560	67897	16.70%	0.3420817111		
3351	Unidas	RENAULT	DUSTER	2018	36809	59490	67897	12.38%	0.3363757231		
3451	Unidas	RENAULT	DUSTER	2018	39171	59490	67897	12.38%	0.3160928818		
3480	Unidas	RENAULT	DUSTER	2018	40040	59490	67897	12.38%	0.3092327956		
3682	Movida	RENAULT	DUSTER	2018	59057	56560	67897	16.70%	0.282728003		

### 3. Arquivos de dados:

- *carrosLocaliza.csv*, *carrosMovida.csv*, *carrosLocamerica.csv* e *carrosUnidas.csv*: Possuem todos os veículos das revendedoras em questão.
- *carros.csv*: arquivo com todos os dados dos quatro arquivos acima compilados em apenas um.
- *Fipe.csv*: possui informações apenas de fabricante e modelo e suas respectivas codificações fiipe.
- *carrosEFipe.csv*: possui os pares (FABRICANTE+MODELO+ANO, preço fiipe) para cada modelo de veículo de determinado ano.
- *Database.csv*: arquivo final com os dados requisitados e dados fiipe

### 4. Utilização:

O programa conta com as opções de atualização apenas do banco de um dos sites, de todos os sites, ou usar os bancos que já estão nos arquivos para gerar o database. A atualização dos quatro bancos pode demorar um tempo considerável, visto que em um dos sites (Localiza) há quase 7 mil resultados e apenas 10 carros por página, o que obriga o programa a no mínimo 700 iterações (clicando na página, esperando carregar e lendo o conteúdo). A opção de gerar database irá apenas atualizar os dados fiipe se comunicando com a API e comparar com os dados das revendedoras já salvos em arquivo csv correspondente.

```
C:\Users\eduar\Anaconda3\python.exe C:/Users/eduar/PycharmProjects/webScraping/main.py
1 - Atualizar banco Unidas
2 - Atualizar banco Locamerica
3 - Atualizar banco Movida
4 - Atualizar banco Localiza
5 - Atualizar todos os bancos de revendedoras
6 - Apenas Extrair os dados FIPE e gerar Database a partir de bancos atuais das revendedoras
7 - Atualizar todos os bancos, extrair dados FIPE e gerar database
8 - Sair
```