

AES Crawler

Asset Estate Scraper

Pedro Henrique Dias Nobrega

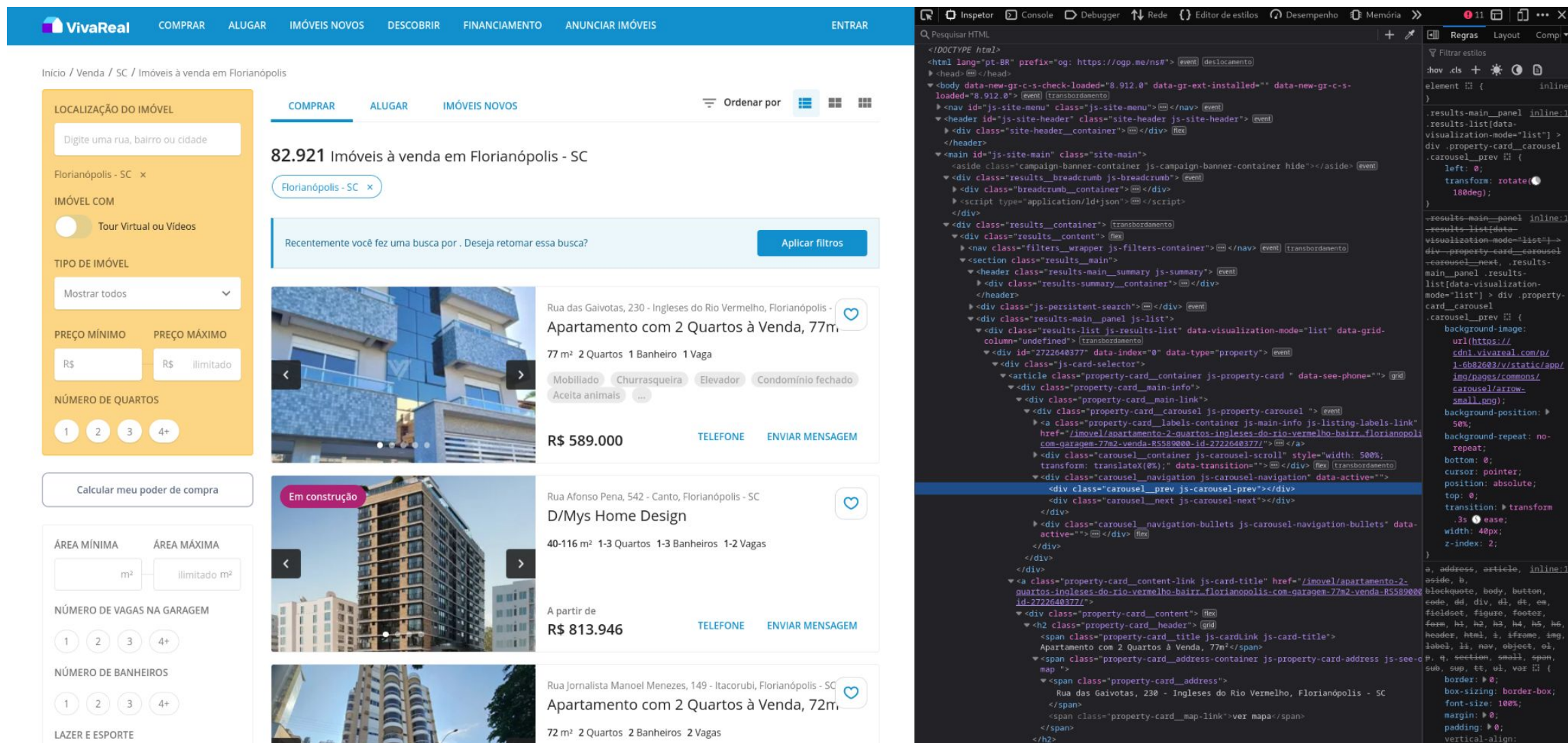
Introdução

- Apresentar :
 - Crawler e extractor;
 - Anúncio de Imóveis:
 - Vários imóveis são anunciados em diversos portais por diversas imobiliárias, este projeto busca unir todos esses dados em uma única fonte de dados;
 - Se espera a obtenção de dados imobiliários relevantes para tomada de decisão em relação à investimentos.
 - Esse domínio foi escolhido pois é uma área que afeta a todos na ilha e ter uma fonte de informação centralizada pode auxiliar a tomada de decisão.

Objetivo

- O objetivo do trabalho é coletar dados imobiliários;
- Receber como entrada lista de anúncios de imóveis à venda;
- Gerar como saída uma lista de imóveis já processada.

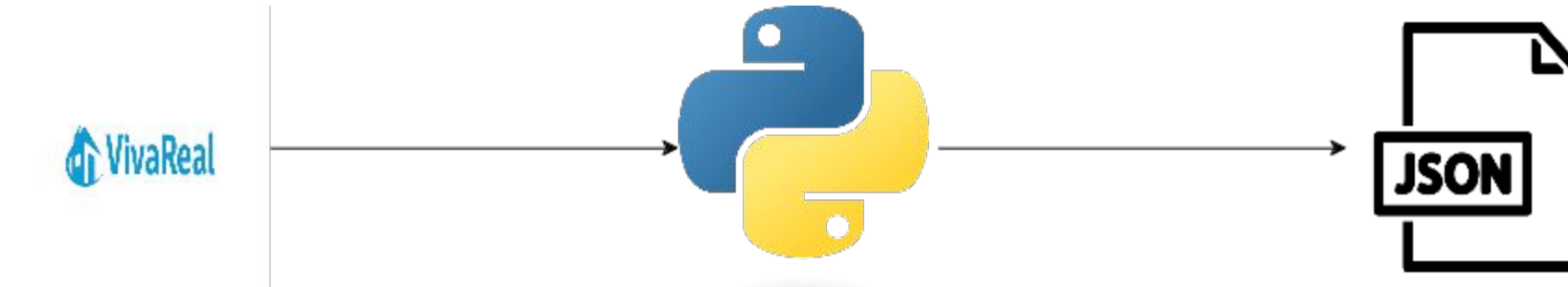
Exemplo dos dados de entrada



Exemplo dos dados de saída

```
{
  "title": "Apartamento com 2 Quartos \u00e0 Venda, 77m\u00b2",
  "url": "https://www.vivareal.com.br/imovel/apartamento-2-quartos-ingleses-do-rio-vermelho-bairros-florianopolis-com-garagem-77m2-venda-RS589000-id-2722640377/",
  "endereco": "Rua das Gaivotas, 230 - Ingleses do Rio Vermelho, Florian\u00f3polis - SC",
  "area": "77",
  "area_label": "m\u00b2",
  "quartos": "2",
  "quartos_label": "Quartos",
  "banheiros": "1",
  "banheiros_label": "Banheiro",
  "vagas": "1",
  "vagas_label": "Vaga",
  "preco": "R$ 589.000",
  "fonte": "https://www.vivareal.com.br/venda/santa-catarina/florianopolis/?pagina=1#onde=Santa%20Catarina,Florian%C3%B3polis,,,,city,BR%3ESanta%20Catarina%3ENULL%3EFlorianopolis,-27.594808,-48.556915,&itl_id=1000183&itl_name=vivareal_-_botao-cta_buscar_to_vivareal_resultado-pesquisa"
},
{
  "title": "RESIDENCIAL BACK GREEN TOWERS - 92m\u00b2",
  "url": "https://www.vivareal.com.br/moveis-lancamento/residencial-back-green-towers-92m-2569358347/",
  "endereco": "Rua Trajano Margarida, 180 - Trindade, Florian\u00f3polis - SC",
  "area": "92",
  "area_label": "m\u00b2",
  "quartos": "3",
  "quartos_label": "Quartos",
  "banheiros": "1",
  "banheiros_label": "Banheiro",
  "vagas": "2",
  "vagas_label": "Vagas",
  "preco": "R$ 1.319.000",
  "fonte": "https://www.vivareal.com.br/venda/santa-catarina/florianopolis/?pagina=1#onde=Santa%20Catarina,Florian%C3%B3polis,,,,city,BR%3ESanta%20Catarina%3ENULL%3EFlorianopolis,-27.594808,-48.556915,&itl_id=1000183&itl_name=vivareal_-_botao-cta_buscar_to_vivareal_resultado-pesquisa"
},
}
```

Descrição sobre o trabalho



Descrição sobre o trabalho

- Informações:
 - AES - Asset Estate Scraper
 - Infraestrutura tecnológica usada:
 - Python;
 - BeautifulSoup;
 - requests;
 - pandas.
- Abordagem e métodos usados
 - Extraídos do website Vivareal e Armazenados localmente
 - Investidores de imóveis ou construtoras, para avaliar a viabilidade de um investimento;
 - Dados coletados de Florianópolis em português;
 - Os dados têm origem em HTML e são armazenados em JSON;

Descrição sobre o trabalho

- Informações:
 - Abordagem e métodos usados
 - Foram coletados os seguintes atributos:

```
{
  "Qualitativo": [
    "title",
    "url",
    "endereco",
    "area_label",
    "quartos_label",
    "banheiros_label",
    "vagas_label",
    "fonte"
  ],
  "Quantitativo": [
    "area",
    "quartos",
    "banheiros",
    "vagas",
    "preco"
  ]
}
```

- A Seed utilizada foi

"https://www.vivareal.com.br/venda/santa-catarina/florianopolis/#onde=,Santa%20Catarina,Florian%C3%B3polis,,,,city,BR%3ESanta%20Catarina%3ENULL%3EFlorianopolis,-27.594808,-48.556915,&itl_id=1000183&itl_name=vivareal_-_botao-cta_buscar_to_vivareal_resultado-pesquisa"

e as URL alcançadas foram as URL dos anúncios, como:

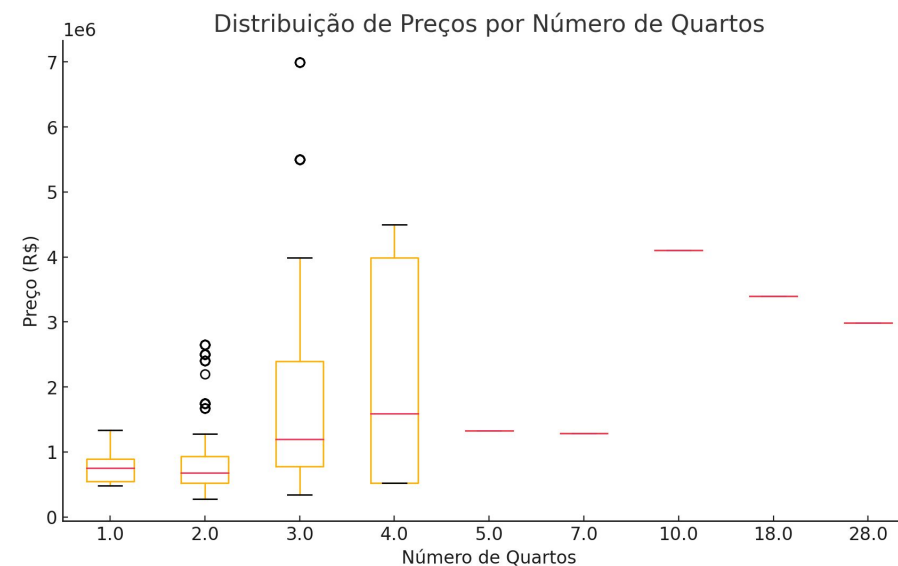
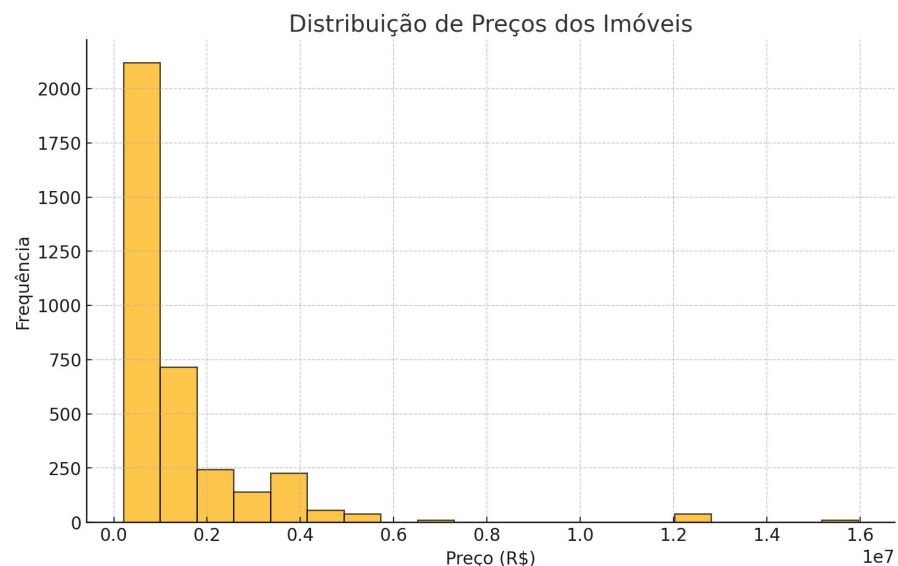
"https://www.vivareal.com.br//imovel//apartamento-2-quartos-ingleses-do-rio-vermelho-bairros-florianopolis-com-garagem-77m2-venda-RS-589000-id-2722640377//"

- Foram coletados de forma pontual;
- Foram feitas duas grandes coletas, dia 09/06 e dia 20/06;

Análise dos resultados finais

Análises Estatísticas:

- Distribuição de Preços:**
 - O preço dos imóveis varia consideravelmente, indo de R\$ 320.000 a R\$ 2.798.027.
- Distribuição de Área:**
 - As áreas dos imóveis variam de 10m² até 900m².
- Correlação entre Preço e Área:**
 - Uma correlação positiva é esperada entre o preço do imóvel e sua área total, embora imóveis em áreas mais valorizadas ou com mais comodidades possam quebrar essa correlação.
- Número de Quartos e Preços:**
 - Imóveis com mais quartos tendem a ter um preço mais alto, mas a localização e o padrão de construção também influenciam significativamente o preço.



Casos de falha

- Onde o trabalho falhou?

- O crawler falhou em não coletar imagens dos anúncios;
- Não foi possível avaliar cada anúncio de forma individual, pois o website limitava quantas requisições poderiam ser feitas;
- Foram minerados apenas os dados necessários, porém os anúncios poderiam apresentar mais informações ;

Conclusão

- Lidar com o rate limit do website;
- Teria tentado utilizar uma ferramenta de crawler mais poderosa para extrair dados de páginas individuais para cada anúncio;
- Além disso, gostaria de experimentar utilizar um LLM, como o ChatGPT, como crawler, ao invés da utilização do BeautifulSoup;
- É possível obter uma lista dos imóveis de forma mais direta e fazer análises mais rápidas sobre viabilidade de investimento;
- Falta integrar com mais website de vendas de imóveis e cruzar os dados de anúncios repetidos.