

Documento Detalhado: Como Fazer a Manutenção do Projeto

A manutenção eficaz deste projeto é fundamental para garantir sua longevidade, confiabilidade e a precisão contínua dos dados coletados.

1. Atualização e Ajuste dos Scrapers

É crucial testar os scripts de scraping periodicamente. Sites web frequentemente mudam seus layouts, a estrutura HTML e os seletores CSS utilizados para identificar e extrair dados. Ao detectar que um scraper parou de funcionar ou está extraíndo informações incorretas, é necessário ajustar os seletores nas funções de extração. Também é importante validar se os mecanismos de paginação (como botões "próxima página" ou "ver mais") ainda funcionam conforme esperado. Manter uma lista atualizada de User-Agents ajuda a simular a aparência de navegadores reais e reduzir a probabilidade de bloqueios. Ajustar os delays entre as requisições é outra medida importante para evitar que o comportamento do scraper seja detectado como automatizado, prevenindo bloqueios de IP.

2. Manutenção do Ambiente de Execução

O ambiente onde os scripts são executados também requer atenção. É necessário manter o navegador Chrome instalado localmente atualizado para garantir a compatibilidade com a versão do `undetected_chromedriver` utilizada. A geração de logs detalhados durante a execução dos scripts pode ajudar a identificar erros frequentes, especialmente aqueles relacionados à inicialização do driver. Em casos de travamentos, pode ser necessário eliminar processos CHROME e do driver que ficaram ativos, utilizando comandos como `taskkill` (em Windows) ou comandos similares em outros sistemas operacionais. Considerar o uso de ambientes virtuais Python (`venv` ou `conda`) é uma boa prática para isolar as dependências do projeto e evitar conflitos com outras instalações Python no sistema.

3. Gerenciamento do Cache de Geocodificação

O arquivo `geocode_cache.json`, que armazena os resultados das geocodificações, deve ser revisado regularmente. É aconselhável fazer backups deste arquivo antes de realizar qualquer limpeza ou edição manual. Se houver uma mudança significativa na forma como a geocodificação é realizada ou no serviço utilizado, pode ser necessário resetar o cache para garantir a consistência dos novos dados. É vital controlar e

monitorar o número de requisições enviadas ao serviço de geocodificação para não violar seus termos de uso e limites.

4. Atualização do Script de Processamento

O script `Processamento.py` pode precisar de ajustes à medida que a base de dados cresce ou se a estrutura dos dados brutos coletados pelos scrapers mudar. As heurísticas de detecção de duplicidade podem precisar ser refinadas para manter a precisão. É importante monitorar erros de parsing JSON e melhorar o tratamento de exceções para tornar o script mais robusto a dados inconsistentes. Se novos campos de informação forem capturados pelos scrapers, o script de processamento precisará ser expandido para incluir a normalização desses novos campos. O ajuste do nível de paralelismo (multiprocessing) pode ser necessário dependendo dos recursos de CPU e memória disponíveis no ambiente de execução.

5. Manutenção dos Mapas

No script `Mapa.py`, as faixas de preço e as cores da legenda, definidas no dicionário `FAIXAS_PRECO_GLOBAL_CONFIG`, devem ser ajustadas periodicamente para refletir as mudanças no mercado imobiliário. É importante verificar se os mapas gerados apresentam os clusters de imóveis corretamente e se as informações nos popups dos marcadores estão sendo exibidas sem erros. Manter a dependência da biblioteca Folium atualizada ajuda a evitar incompatibilidades e garante o acesso a novas funcionalidades ou correções.

6. Backup e Versionamento

Manter backups regulares dos arquivos JSON gerados em cada etapa do processo é uma prática de segurança essencial. O uso de um sistema de controle de versão como o Git é altamente recomendado para rastrear todas as mudanças feitas no código-fonte do projeto. Documentar as modificações realizadas nos scrapers, especialmente quando seletores são ajustados, facilita a auditoria e a compreensão do histórico de alterações.

7. Execução Recomendada para Atualização dos Dados

Para realizar uma atualização completa dos dados, a sequência de execução recomendada é a seguinte: Primeiro, remova ou archive os arquivos JSON antigos dos diretórios de dados brutos e do diretório de resultados para evitar a mistura de dados de diferentes execuções. Em seguida, execute os scripts de scraping desejados para

coletar os novos dados brutos. Após a conclusão dos scrapers, execute o script `Processamento.py` para harmonizar, normalizar, geocodificar e remover duplicados dos novos dados. Finalmente, execute o script `Mapa.py` para gerar as visualizações de mapa com as informações atualizadas. Durante e após a execução, inspecione os logs gerados, verifique se houve erros e valide a qualidade e a validade dos arquivos de saída antes de considerar a atualização concluída.

Conclusão

Este projeto, estruturado em módulos distintos para scraping, processamento e mapeamento, constitui um ecossistema completo para a coleta sistemática e análise de dados do mercado imobiliário. Aderir a práticas rigorosas de manutenção, conforme detalhado, é indispensável para assegurar a durabilidade do projeto, a confiabilidade dos dados coletados e a precisão das análises derivadas ao longo do tempo.