



Escrever título (escolher no final)

Universidade Federal da Paraíba - CCEN

Gabriel de Jesus Pereira

11 de maio de 2024

Índice

Resumo	3
Capítulo 1	4
Introdução	4
Objetivos	4
Objetivo Geral	4
Objetivos Específicos	4
Organização do Trabalho	4
Capítulo 2	5
Recursos Computacionais	5
Linguagem de Programação R	5
Quarto	5
Linguagem de Programação Python	5
Web Scraping	5
Metodologia	6
Os dados e o procedimento adotado para sua obtenção	6
Capítulo 4	7
Resultados	7
Capítulo 5	8
Conclusão e Discussão	8

Resumo

Capítulo 1

– GABRIEL, NÃO FAÇA ESSA PARTE AGORA! –

Introdução

– Fazer antes da conclusão –

Objetivos

Objetivo Geral

Objetivos Específicos

Organização do Trabalho

Capítulo 2

– Fazer depois da metodologia –

Recursos Computacionais

Linguagem de Programação R

Quarto

Linguagem de Programação Python

Web Scraping

Metodologia

Os dados e o procedimento adotado para sua obtenção

O Web scraping, também conhecido como extração de dados da web, é uma técnica utilizada para o processo de coleta de dados estruturados da web de maneira automatizada. É um processo que vem sendo constantemente utilizado por instituições públicas e privadas que passam cada vez mais a construir produtos que utilizam algoritmos de aprendizagem de máquinas, observa ofertas e descontos, faz análise de mercado ou monitoração de marcas.

Neste projeto, para fins de estudo e análise do mercado imobiliário, os dados foram coletados por meio de extração de dados do site do Zap Imóveis. O Zap Imóveis é um site do Grupo OLX que reúne ofertas do mercado imobiliário e que funciona como uma plataforma dinâmica para facilitar a conexão entre quem deseja alugar, comprar ou vender um imóvel; podendo servir também para corretores ou outros profissionais do setor de imóveis. Este projeto foi possível graças as informações que foram coletadas do site do Zap Imóveis em dois períodos do ano de 2023. O primeiro deles, essas informações foram coletadas utilizando variados pacotes para raspagem de dados e proxies rotativas da linguagem de programação R a fim de evitar ser bloqueado pelos mecanismos de segurança do site. Na segunda etapa, os dados foram coletados empregando a linguagem de programação Python com as bibliotecas Scrapy e Playwright, que serve para web crawling e web scraping, e o Playwright que serve testes na em aplicativos da web, mas que neste caso foi utilizado para manejar páginas dinâmicas.

Capítulo 4

Resultados

Capítulo 5

Conclusão e Discussão