

Prémio Arquivo.pt

Descrição Sumária do Trabalho

Identificação

- Título: Habitação.NET: Uma visão do Mercado de Habitação em Portugal
- Área temática: Habitação, Estatística, Informática, Economia Imobiliária, Ciências Sociais
- Candidato: Diogo Gonçalves
- Email: contacto@habitacao.net

Descrição do Trabalho

A habitação, enquanto direito consagrado na Constituição da República Portuguesa¹, é uma das principais preocupações dos cidadãos portugueses sendo um tema amplamente debatido pela sociedade e comunicação social, reflexo do crescimento da pressão imobiliária. Como tal, torna-se importante disponibilizar recursos e ferramentas públicas que permitam ao cidadão compreender e contextualizar a evolução dos preços de compra e arrendamento da habitação ao longo do tempo.

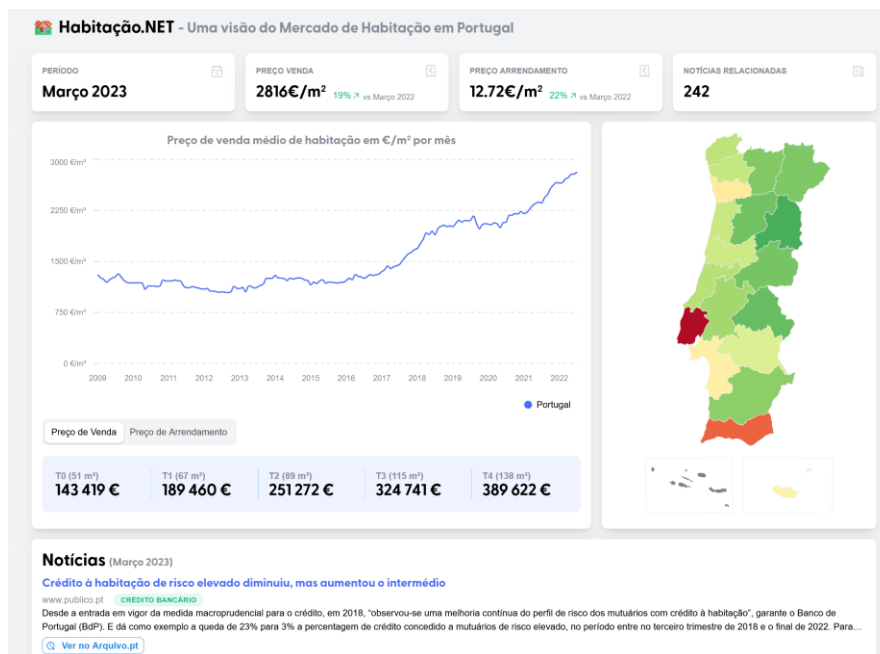
O projeto Habitação.NET é uma ferramenta que permite ao utilizador explorar, de forma interativa, a evolução do valor médio do mercado habitacional e arrendatário português, contextualizado com notícias publicadas sobre o tema e as políticas de habitação. Através da coleção preservada pelo Arquivo.pt foi possível agregar mais de 1.5 milhões de anúncios únicos de diversos portais imobiliários e 15 mil notícias, arquivados desde 2001. Com a publicação deste *website*, é possível não só visualizar a tendência dos preços da habitação, como também comparar valores entre distritos, municípios ou freguesias, permitindo aos utilizadores fazer uma análise própria e descentralizada, dando visibilidade a localidades menos debatidas em comparação com os grandes centros urbanos.

O projeto desenvolvido é composto por duas componentes principais:

- (1) Um conjunto de dados curado, livre e acessível a toda a comunidade social e científica, contendo o histórico de preços da habitação em Portugal durante os últimos 24 anos, permitindo uma análise detalhada sobre o tema.
- (2) Ferramenta Habitação.NET, acessível em <https://www.habitacao.net/>, permitindo navegar, comparar e analisar uma versão agregada dos dados recolhidos num website interativo.

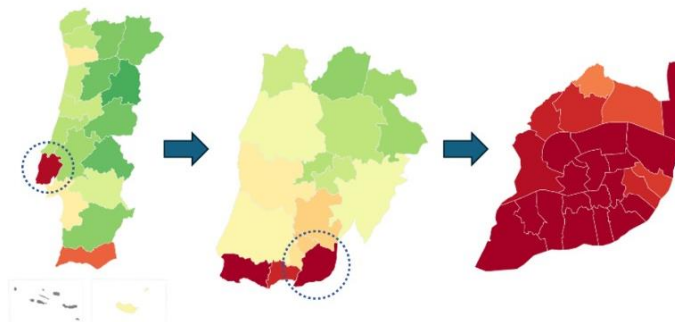
¹ Constituição da República Portuguesa, *Artigo 65.º*

Ferramenta Habitação.NET



A ferramenta está dividida em quatro áreas principais, interligadas entre si, que permitem interagir com os preços da habitação em Portugal entre 2009 e 2023².

- Um campo de informações gerais, que informa o utilizador, relativamente ao período selecionado (mês/ano), sobre o preço de venda e de arrendamento em €/m² com uma comparação em percentagem face ao ano anterior e sobre a quantidade de notícias relacionadas com o tema da habitação redigidas nesse mês;
- Um gráfico de linhas que permite consultar o preço médio da habitação em €/m² por mês no panorama nacional. Ao navegar por esta linha, o utilizador consegue selecionar o período usado para apresentar as restantes estatísticas, alterando os valores apresentados no *heat map* de Portugal e, na área de notícias.



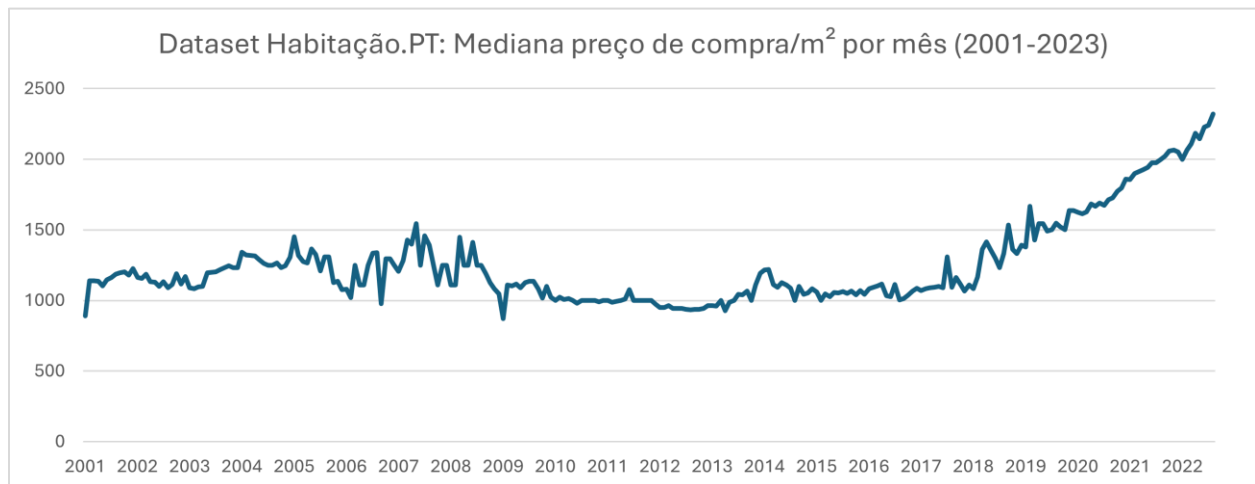
² O conjunto de dados produzido está compreendido entre 2001 e 2023. No entanto, os dados entre 2001 e 2008 foram omitidos da plataforma por serem insuficientes para se produzir uma análise detalhada ao nível local.

- Um heat map de Portugal representado por distritos que, conjugado com o gráfico, reflete numa gradação de cor o preço médio da habitação no período selecionado. À medida que o utilizador navega pelo mapa, uma linha extra surge no gráfico, permitindo comparar o panorama nacional com o panorama distrital, filtrável até ao nível do município/freguesia;
- Uma área de notícias que, em conjunto com o gráfico, contextualiza o histórico de preços da habitação/arrendamento com as fontes de informação facultadas pelos vários órgãos de comunicação social integradas na linha temporal disponibilizada. As notícias estão categorizadas de forma a facilitar a identificação do seu tema;



Conjunto de Dados sobre Habitação

Foi produzido um conjunto de dados com os preços da habitação em Portugal, compreendidos entre o período de 2001 a 2023, através da recolha de conteúdos arquivados pelo Arquivo.PT. Esses dados foram processados e disponibilizados nas seguintes variantes:



- Lista de 1.6 milhões de páginas (URLs) que contêm anúncios imobiliários válidos arquivadas em Arquivo.pt.
- Conjunto de dados estruturado com 30 milhões de anúncios (1.5 milhões únicos), recolhidos nas páginas mencionadas no ponto anterior. Contém informação base dos imóveis como tipologias, preço e localização.

| Campo | Exemplo | Descrição |
|----------------|--------------|--|
| id_anuncio | 1156 | Identificador único do anúncio |
| data | 20030325 | Data em que foi encontrado em Arquivo.pt |
| tipo_venda | venda | venda ou arrendamento |
| tipo_habitacao | moradia | Tipo de imóvel (apartamento, moradia, etc) |
| area | 170 | Área do imóvel em m2 |
| preco | 93275 | Preço de venda ou arrendamento do imóvel |
| preco_m2 | 549 | Preço de venda/Área do Imóvel |
| quartos | 3 | Tipologia (T0, T1, T2, ...) |
| distrito | Portalegre | Localização: Distrito |
| municipio | Monforte | Localização: Município |
| freguesia | Santo Aleixo | Localização: Freguesia |

- 15 mil notícias arquivadas em Arquivo.pt, relacionadas com o tema da habitação.

Este conjunto de dados pode ser consultado diretamente através dos ficheiros disponibilizados, programaticamente através de uma API com os dados processados e resumidos, ou diretamente na página Habitação.NET. O processo para produção deste conjunto de dados é descrito em maior detalhe na área de comentários adicionais.

Objetivos

Os principais objetivos deste trabalho são:

- Desenvolver uma ferramenta intuitiva e dinâmica que permita um cidadão comum, jornalistas, decisores políticos ou profissionais da área explorar e analisar a evolução do estado do mercado habitacional e comparar a tendência dos preços da habitação e arrendamento entre distritos, municípios ou freguesias;
- No contexto da habitação, oferecer um ponto de vista mais próximo da perceção do utilizador, utilizando como referência os valores publicados em plataformas imobiliárias³.
- Disponibilizar acesso rápido a um arquivo de notícias sobre habitação, permitindo analisar os diversos fatores que influenciam o mercado e aceder a informação sobre políticas de habitação.
- Facultar um conjunto de dados aberto e detalhado, permitindo à comunidade científica realizar estudos que complementem estes dados com outras fontes de informação e observar tendências.

Resultados Atingidos

Como já referido, foi criado e disponibilizado um *dataset* que centraliza 1.5 milhões de anúncios de compra e arrendamento recolhidos e processados de diversas páginas web imobiliárias, condensados num website único.

³ Em contraste com outras métricas como o Índice de Preços da Habitação (IPHab), calculada com base nas transações de imóveis.

Estes dados confirmam um aumento do preço médio da habitação, com um crescimento mais acentuado a partir de 2016, ficando claro que os distritos com maior pressão imobiliária são Lisboa, Faro, Porto e Setúbal. Comparando com outros distritos, é possível observar que os demais se mantêm relativamente estáveis, como Guarda ou Portalegre. Com a granularidade oferecida pela ferramenta é possível também observar de que forma cada município e freguesia influenciam o preço médio do distrito. A título de exemplo, municípios do mesmo distrito podem apresentar valores muito dispares, como Espinho (Aveiro) que comparativamente aos municípios vizinhos, apresenta um preço médio com uma diferença maior que 2000€/m².

Adicionalmente, é possível comparar os valores médios de compra e arrendamento de habitação, e observar qual a opção financeiramente mais favorável em cada período.

Originalidade e caráter inovador

Pelo que foi possível apurar pelo autor, é a primeira vez que se utiliza o Arquivo.pt para recolher o histórico dos portais imobiliários, juntando mais de duas décadas de anúncios que quando agregados e relacionados permitem dar uma visão global sobre a evolução do mercado imobiliário português.

Demonstra-se assim que é possível utilizar informação arquivada no Arquivo.pt para construir um conjunto de dados de grandes dimensões com utilidade pública e científica, que não tenha a utilização de notícias como fonte principal, mas não esquecendo a sua importância na contextualização da informação.

A publicação destes dados num formato de acesso livre, constitui o maior conjunto de dados público de preços de habitação em Portugal.

Impacto social (aplicação e utilidade social)

A ferramenta Habitação.NET foi desenhada com o objetivo de permitir a qualquer cidadão, independentemente dos seus conhecimentos técnicos, explorar o mercado habitacional português, permitindo-lhe ter uma visão geral ou mais detalhada ao nível da freguesia. Toda a informação é contextualizada com notícias sobre o tema, permitindo avaliar o impacto de decisões políticas ou o sentimento geral sobre a habitação ao longo do tempo.

Para um cidadão comum, esta plataforma pode ser usada com um objetivo mais prático de comparar os custos da sua habitação com o resto do mercado, percebendo por exemplo quais os valores médios em vigor nas várias freguesias do país para comprar uma nova habitação ou avaliar se o custo do seu arrendamento está ajustado, servindo como ferramenta de suporte à tomada de decisão no momento de escolher uma habitação.

Impacto científico (aplicação e utilidade científica)

Com este projeto disponibilizamos informação detalhada de acesso livre (através de uma ferramenta e *dataset*) sobre o estado da habitação em Portugal desde 2001, permitindo a qualquer membro da comunidade científica ou cidadão interessado consultar essa informação e produzir estudos com dados que não são de fácil acesso ou recolha. A disponibilização de dados ao nível da freguesia permite fazer uma análise da habitação mais aprofundada e relacionar com outros conjuntos de dados geográficos.

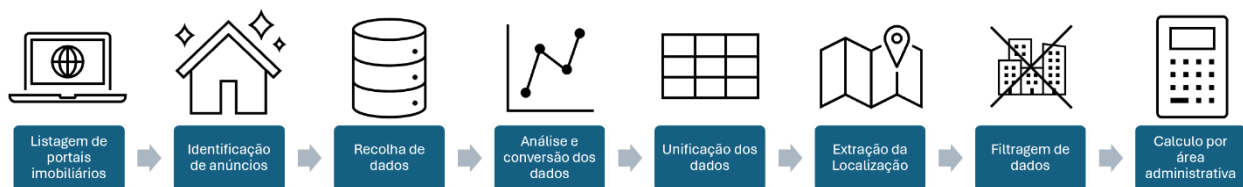
Relevância da utilização do Arquivo.pt

Este trabalho não seria possível concretizar sem o Arquivo.pt. Para garantir uma boa experiência de utilização, os portais imobiliários online apenas mostram aos utilizadores os anúncios ativos naquele momento no mercado. O Arquivo.pt apresenta-se como a única forma de acesso livre que permite recuperar essa informação, possibilitando o acesso a anúncios publicados em plataformas descontinuadas, que desapareceram com o tempo. Além disso, apenas o Arquivo.pt tem disponível publicamente o volume de informação suficiente para produzir uma linha de tempo do estado habitacional ao nível das freguesias, sendo possível recuperar anúncios publicados desde o início do século.

O Arquivo.pt apresenta-se também como uma ferramenta fundamental para recuperar notícias do passado, recorrendo às APIs⁴ de pesquisa de texto disponibilizadas, sem as quais seria impossível recolher notícias sobre temas relacionados com a habitação.

Comentários adicionais

A recolha e processamento dos dados para este projeto pode ser descrito pelas seguintes fases:



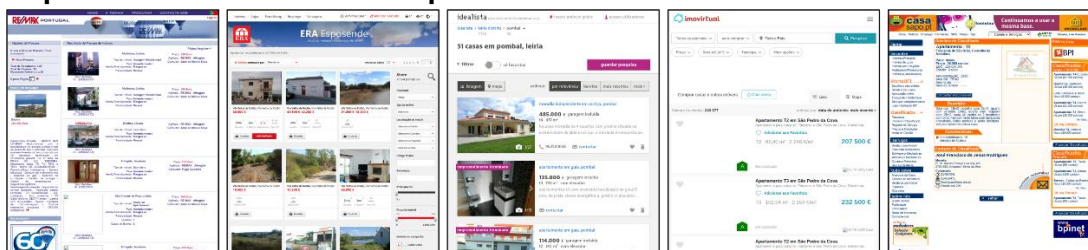
1) Listagem de portais imobiliários

Construção manual de uma lista de portais de classificados relevantes, ativos ou descontinuados, cujos conteúdos se encontrassem preservados. Para isso recorreu-se

⁴ Mais detalhe na área de Comentários Adicionais.

a pesquisas manuais no Arquivo.pt⁵ (usando a API Full-text Search) e aos estudos de mercado periódicos publicados pela Marktest⁶ com referências aos portais imobiliários mais utilizados. Foram selecionadas as plataformas: bpiimobiliario, casasapo, era, idealista, imovirtual, lardocelar, miau, remax, trovit.

2) Mapeamento da estrutura dos portais



Para cada portal imobiliário mapeou-se a estrutura das páginas web, de forma a encontrar as áreas com anúncios relevantes tendo em consideração as suas variações/redesigns ao longo dos anos. Desta estrutura foram consideradas não só as páginas de listagem de anúncio como também as páginas com o detalhe de cada anúncio quando disponível. No total, foram processadas 32 milhões de páginas recorrendo à API CDX do Arquivo.pt.

3) Recolha de dados

Das páginas processadas, foram descarregadas, utilizando a API Wayback do Arquivo.pt, aquelas com anúncios relevantes. No total foram recolhidas 3.1 milhões de páginas (296 GB de dados).

4) Análise e conversão dos dados

Para cada um dos grupos identificados no ponto 2, implementou-se um processo de conversão de dados para extrair a informação relevante para cada anúncio (preço, área, tipologia, localização, descrição do imóvel e outros). Devido às sucessivas reestruturações de alguns portais imobiliários, foi necessário desenvolver mais do que um *parser* por portal. Neste ponto foram rejeitadas as páginas que continham anúncios expirados ou conteúdos inválidos, resultando em 1.6 milhões de páginas.

5) Unificação dos dados

Cada site imobiliário apresenta nomenclaturas diferentes para a mesma informação, por isso foi necessário mapear a informação para um formato único. Em alguns casos foi necessário processar informação não estruturada presente na descrição para extrair elementos relevantes como a tipologia do imóvel ou a sua área.

6) Extração da Geolocalização

A apresentação da geolocalização de um imóvel não é consistente entre as plataformas imobiliárias. Algumas apresentam a morada completa, enquanto outras separam os

⁵ <https://arquivo.pt/page/search?q=apartamento+beja&l=pt&from=20000806&to=20030429>

⁶ <https://www.marktest.com/wap/pesquisa.aspx?what=sites%20imobili%C3%A1rio&s=N>

imóveis por zonas que não têm mapeamento direto para a Carta Administrativa Oficial de Portugal. Considerando ainda as reorganizações administrativas que ocorreram desde o início do século, foi necessário correr um processo para unificar estes dados. Recorreu-se às cartas publicadas pela Direção-Geral do Território e à API OpenStreetMap Nominatim para o mapeamento das localizações.

7) Filtragem de dados

Alguns portais imobiliários contêm anúncios a espaços comerciais e industriais que, estando fora do âmbito do projeto, foram descartados para o cálculo do preço/m² final.

8) Cálculo do preço/m² para cada área administrativa

Para cada mês de dados recolhidos, calculou-se o preço médio por tipologia em cada freguesia. Esses dados foram ponderados com a distribuição do parque habitacional presente no Censos 2021 para cálculo dos valores agregados por municípios, distritos e total para o país.

Recursos complementares

- Portal Habitação.NET - <https://www.habitacao.net/>
Ferramenta interativa com os preços de habitação em Portugal.
- Conjunto de dados Habitação.NET - <https://www.habitacao.net/dados/>
Página com documentação e ligações para os *datasets* produzidos neste projeto.