

Relações entre preço de rendas e densidade populacional

João Oliveira - 56908
Faculdade de Ciências
fc56908@alunos.fc.ul.pt

Tiago Garcia - 54938
Faculdade de Ciências
fc54938@alunos.fc.ul.pt

ABSTRACT

Este relatório detalha o projeto desenvolvido na área de Integração e Processamento Analítico de Informação, cujo objetivo central é a exploração de correlações entre indicadores demográficos e do mercado imobiliário. Para tal, procedeu-se à recolha, integração e análise de múltiplos datasets provenientes de fontes diversas. O estudo foca-se especificamente na investigação das relações entre: o preço médio das rendas e o preço médio de compra de habitação; a densidade populacional e a densidade de alojamentos numa determinada área; o preço médio das rendas e a densidade de alojamentos; a densidade populacional e o preço médio das rendas; e a densidade populacional e o preço médio de compra de habitação. Através do processamento de dados, procuramos obter insights sobre as dinâmicas que interligam a demografia e as condições do mercado habitacional.

ACM Reference Format:

João Oliveira - 56908 and Tiago Garcia - 54938. 2025. Relações entre preço de rendas e densidade populacional. In . ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 QUESTÕES A RESPONDER

Com base nos conjuntos de dados selecionados e nos objetivos delineados para o projeto de Integração e Processamento Analítico de Informação, definiram-se as seguintes questões de investigação centrais que irão basear a nossa análise:

(1) Relação Preço Venda vs. Preço Arrendamento:

- Qual é a correlação quantitativa entre o preço médio de venda de habitação (€) e o preço médio de arrendamento por metro quadrado (€/m²) nos diferentes municípios de Portugal?
- Municípios com preços de venda mais elevados tendem sistematicamente a ter preços de arrendamento por m² mais elevados?

(2) Relação Densidade Populacional vs. Densidade de Alojamentos:

- Existe uma correlação estatisticamente significativa entre a densidade populacional (habitantes por unidade de área, e.g., km²) e a densidade de alojamentos (alojamentos/m²) nos municípios portugueses?

- Esta relação é linear? Um aumento na densidade populacional corresponde a um aumento proporcional na densidade de alojamentos?
- Como se comparam os municípios das áreas metropolitanas (Lisboa, Porto) com os restantes em termos da relação entre estas duas densidades?

(3) Impacto da Densidade de Alojamentos no Arrendamento:

- Qual é o impacto da densidade de alojamentos (alojamentos/m²) no preço médio de arrendamento (€/m²) municipal?
- Uma maior densidade de oferta (alojamentos) está associada a preços de arrendamento mais baixos ou mais altos nesse município?

(4) Impacto da Densidade Populacional no Arrendamento:

- De que forma a densidade populacional influencia o preço médio de arrendamento por metro quadrado (€/m²) nos municípios?
- A densidade populacional é um preditor mais forte do preço de arrendamento do que a densidade de alojamentos?

(5) Impacto da Densidade Populacional no Preço de Venda:

- Qual é a relação entre a densidade populacional de um município e o preço médio de venda de habitação (€)?
- É possível modelar o preço médio de venda de habitação utilizando a densidade populacional como uma das variáveis explicativas?

2 FONTES DE DADOS

A seleção dos conjuntos de dados (*datasets*) para este estudo recorreu a diversas fontes e plataformas, incluindo uma pesquisa de repositórios como pordata.pt, ine.pt, o portal de dados abertos dados.gov.pt e o motor de busca datasetsearch.research.google.com. Os conjuntos de dados utilizados foram:

- portugal_listings.csv**: Contém informação sobre os preços de venda de habitação em Portugal, agregada por município.
- rendasm2.json**: Apresenta os valores de arrendamento por metro quadrado (€/m²) em Portugal, por município.
- densidadePopulacional.json**: Disponibiliza dados relativos à densidade populacional (habitantes por unidade de área) em Portugal, por freguesia.
- densidadealojamentosm2.json**: Contém informação sobre a densidade de alojamentos por metro quadrado, por município.
- Areas_Freg_Conc_Dist_Pais_CAOP2019.xls**: Este dataset tens os codigos geográficos de NUTSIII para podermos fazer as relações entre freguesiaa, municípios e concelhos dos outros datasets.

A escolha destes *datasets* específicos foi motivada pela necessidade de responder às questões centrais do projeto.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

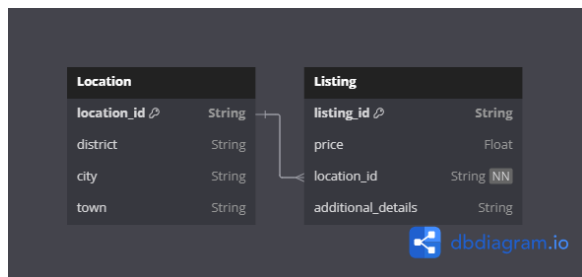
3 CARACTERIZAÇÃO DOS DADOS

3.1 Portugal_listings.csv

Tivemos de eliminar algumas colunas do dataset, pois não nos interessavam. Ficando só com as colunas **Price**, **city**, **District** e **Town**.

Descrição dos Atributos:

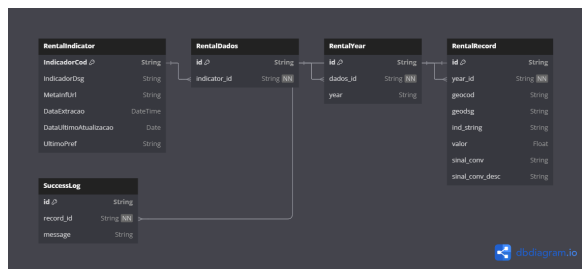
- **Preço: Numérico (Decimal)** - Deve ser um valor positivo (ex: > 0). Representa o preço do imóvel.
- **Distrito: Texto/Categórico** - Tem de ser um distrito válido em Portugal. Não são permitidos valores nulos.
- **Concelho: Texto/Categórico** - Tem de ser um concelho válido dentro do distrito especificado. Não são permitidos valores nulos.
- **Freguesia: Texto/Categórico** - Tem de ser uma freguesia válida dentro do concelho especificado. Não são permitidos valores nulos.



3.2 Rendasm2.json

Descrição dos Atributos:

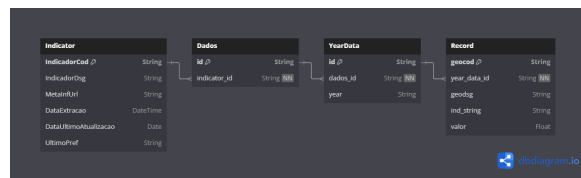
- **IndicadorCod: Texto** - Identificador único para o indicador.
- **IndicadorDsg: Texto** - Descrição do indicador (ex: "Rendas médias mensais (€/m²)").
- **DataExtracao: Texto** - Timestamp da extração de dados (ex: "2025-03-18T16:02:16.973Z").
- **DataUltimoAtualizacao: Texto** - Data da última atualização (ex: "2024-03-19").
- **UltimoPref: Texto** - Período de referência dos dados (ex: "2023").
- **geocod: Texto** - Código geográfico da região (ex: "070513").
- **geodsg: Texto** - Descrição geográfica ou nome da região (ex: "Torre de Coelheiros").
- **ind_string: Texto** - Valor da renda como texto (ex: "10,5").
- **valor: Decimal** - Valor da renda como valor numérico (ex: 10.5). Deve ser não-negativo.



3.3 DensidadePopulacional.json

Descrição dos Atributos:

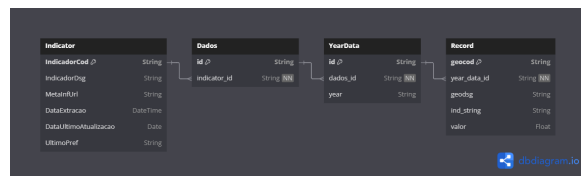
- **IndicadorCod: Texto** - Identificador único para o indicador (ex: "0008337"). Não Nulo.
- **IndicadorDsg: Texto** - Descrição do indicador (ex: "Densidade populacional (N.º/ km²)"). Não Nulo.
- **DataExtracao: DataHora** - Data e hora em que os dados foram extraídos. Não Nulo.
- **DataUltimoAtualizacao: Data** - Data em que os dados foram atualizados pela última vez. Não Nulo.
- **UltimoPref: Inteiro** - Ano dos dados (ex: 2022). Não Nulo, tem de ser um ano válido.
- **geocod: Texto** - Identificador único para a região (ex: "1500802"). Não Nulo.
- **geodsg: Texto** - Nome da região (ex: "Alcoutim"). Não Nulo.
- **ind_string: Texto** - Densidade populacional como texto (ex: "4,3"). Opcional, pode ser derivado de valor.
- **valor: Decimal** - Densidade populacional como valor numérico (ex: 4.3). Não Nulo, tem de ser maior do que 0.



3.4 Densidadealojamentosm2.json

Descrição dos Atributos:

- **IndicadorCod: Texto** - Identificador único para o indicador.
- **IndicadorDsg: Texto** - Descrição do indicador (ex: "Densidade de alojamentos (N.º/ km²)").
- **DataExtracao: Texto** - Timestamp da extração de dados (ex: "2025-03-18T16:02:16.973Z").
- **DataUltimoAtualizacao: Texto** - Data da última atualização (ex: "2024-03-19").
- **UltimoPref: Texto** - Período de referência dos dados (ex: "2021").
- **geocod: Texto** - Código geográfico da região (ex: "070513").
- **geodsg: Texto** - Descrição geográfica ou nome da região (ex: "Torre de Coelheiros").
- **ind_string: Texto** - Densidade de alojamentos como texto (ex: "2,1").
- **valor: Decimal** - Densidade de alojamentos como valor numérico (ex: 2.1). Deve ser não-negativo.

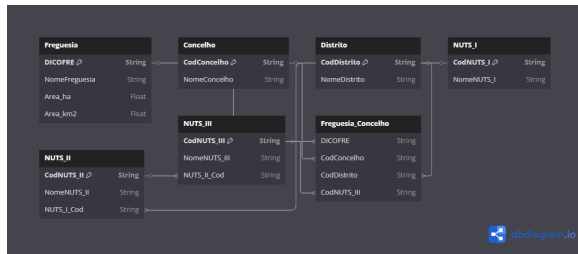


3.5 Areas_Freg_Conc_Dist_Pais_CAOP2019.xls

Descrição das Tabelas e Atributos:

- **Indicator:**

- **IndicadorCod: Texto** - Código único do indicador (chave primária).
- **IndicatorDsg: Texto** - Descrição do indicador (ex: "Densidade Populacional").
- **MetaInfUrl: Texto** - URL com metadados sobre o indicador.
- **DataExtracao: DateTime** - Timestamp da extração dos dados.
- **DataUltimoAtualizacao: Date** - Data da última atualização.
- **UltimoPref: Texto** - Período de referência (ex: "2023").
- **Region:**
 - **geocod: Texto** - Código geográfico único (chave primária).
 - **geodsg: Texto** - Nome da região.
- **HousingDensity:**
 - **IndicadorCod: Texto** - Código do indicador (referência a Indicator).
 - **geocod: Texto** - Código da região (referência a Region).
 - **valor: Decimal** - Valor da densidade habitacional.
 - **ind_string: Texto** - Valor em formato textual.
- **RentalPrices:**
 - **IndicadorCod: Texto** - Código do indicador (referência a Indicator).
 - **geocod: Texto** - Código da região (referência a Region).
 - **valor: Decimal** - Valor médio de renda (€/m²).
 - **ind_string: Texto** - Valor em formato textual.
- **PropertyListings:**
 - **ListingID: Texto** - Identificador único do imóvel (chave primária).
 - **Price: Decimal** - Preço do imóvel.
 - **geocod: Texto** - Código da região (referência a Region).
- **ParishAreas:**
 - **DICOFRE: Texto** - Código da freguesia (chave primária).
 - **NomeFreguesia: Texto** - Nome da freguesia.
 - **Area_ha: Decimal** - Área em hectares.
 - **Area_km2: Decimal** - Área em km².
 - **CodNUTS_III: Texto** - Código NUTS III (referência a Region.geocod).
- **NUTS_III:**
 - **CodNUTS_III: Texto** - Código NUTS III (chave primária).
 - **NomeNUTS_III: Texto** - Nome da região NUTS III.



4 MODELO INTEGRADO

4.1 Plano de Integração e Justificação de Escolhas

O modelo integrado foi concebido para agregar dados de diversas fontes (*Portugal_listings.csv*, *Rendasm2.json*,

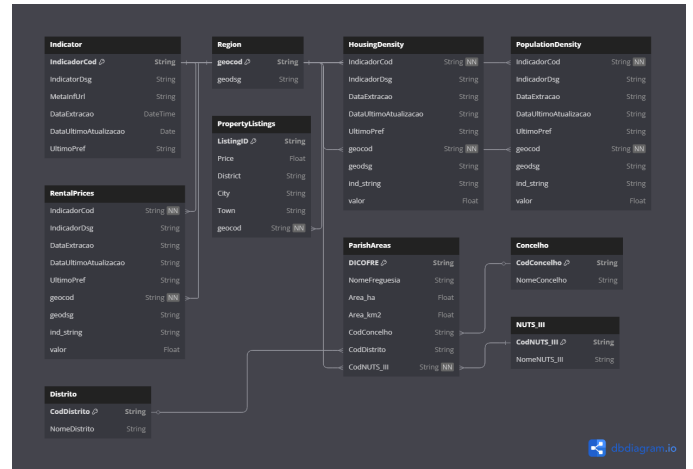


Figure 1: Diagrama do modelo integrado

DensidadePopulacional.json e *Densidadealojamentosm2.json*) através de chaves geográficas comuns (*geocod*), permitindo fazer uma análise entre:

- **Preços de arrendamento** (Rental prices - valor) com **preços de propriedades** (PropertyListings - Price)
- **Densidades populacionais** (PopulationDensity - valor) com **disponibilidade de alojamentos** (HousingDensity)
- Todos os indicadores com a **dimensão geográfica** (Region - geocod)

4.2 Decisões do Esquema Integrado

- **Criação da tabela de Indicadores:** Criou-se a entidade **Indicator** com chave primária *IndicadorCod*. Isto foi feito para armazenar metadados comuns (código, descrição, datas), evitando redundâncias nas tabelas *RentalPrices*, *HousingDensity* e *PopulationDensity*.
- **Criação da tabela de Região:** Criou-se a entidade **Region** para armazenar e relacionar as regiões/municípios de cada dataset utilizando o *geocod* como chave primária. Este *geocod* vai nos permitir comparar regiões/municípios com outras características demográficas.
- **Chaves Primárias Unificadas:** O campo *geocod* serve como chave primária de junção entre todas as entidades, garantindo consistência nos relacionamentos. Todas as relações são one-to-many:
 - *RentalPrices* → *Region* (N:1)
 - *PropertyListings* → *Region* (N:1)
 - *PopulationDensity* → *Region* (N:1)
 - *HousingDensity* → *Region* (N:1)
 - *ParishAreas* → *Region* (N:1)
- **Criação da tabela ParishAreas:** Criou-se a entidade **ParishAreas** para armazenar e relacionar as regiões de cada dataset com as suas freguesias/municípios/concelhos utilizando o *geocod* como chave primária. Fomos criar também as tabelas para *NUTS_III*, *Distrito* e *Concelho* com relações one-to-one, para armazenar os códigos e nomes destes mesmos.

4.3 Abordagem de Resolução de Identidades

Para garantir correspondência precisa entre entidades:

- **Chaves Primárias:** Utilização de *geocod* e *IndicadorCod* como identificador canónico para regiões e metadados respetivamente, proveniente dos códigos oficiais dos datasets.
- **Metadados de Contexto:** Campos como *UltimoPref* (período de referência) e *DataExtracao* permitem validar a temporalidade dos dados.

5 FERRAMENTAS UTILIZADAS

Para o desenvolvimento deste projeto, foram utilizadas diversas ferramentas tecnológicas que permitiram a recolha, processamento e análise dos dados, bem como a criação de visualizações e o desenvolvimento da aplicação final. As principais ferramentas utilizadas incluem:

- **Python/Pandas** - Para realizar o código
- **Excel** - Para visualizar datasets .csv
- **Jupyter Notebook** - Para compilar o projeto
- **Visual Studio Code** - Para visualizar os datasets .json
- **LaTeX** - Para escrever o relatório

- **GitHub** - Plataforma de repositório do trabalho
- **dbdiagram.io** - Para fazer o modelo de integração

O código fonte completo deste projeto, incluindo todos os scripts de processamento de dados, notebooks de análise e documentação adicional, está disponível no repositório público do GitHub:

Repositório

6 CONCLUSÃO

Este trabalho permitiu desenvolver um modelo integrado de dados para análise das relações entre o mercado imobiliário e indicadores demográficos em Portugal. Através da consolidação de múltiplas fontes de dados oficiais, criámos uma estrutura relacional robusta que procura:

- Normalizar a hierarquia geográfica portuguesa (freguesias, concelhos, distritos e NUTS III) numa tabela central **Region**
- Estabelecer relações claras entre indicadores de preços de habitação (venda e arrendamento) e métricas populacionais
- Respeitar as especificidades dos datasets originais enquanto permite cruzamentos de dados