

# 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 é a exploração de correlações entre indicadores demográficos e do mercado imobiliário. Para tal, procedeu-se à recolha, integração e análise de múltiplos datasets. O estudo foca-se 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. Com isto, procuramos obter insights sobre as dinâmicas que interligam a demografia e o do mercado habitacional.

## 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:

- (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 (€/m<sup>2</sup>) nos diferentes municípios de Portugal?
  - Municípios com preços de venda mais elevados tendem sistematicamente a ter preços de arrendamento por m<sup>2</sup> mais elevados?
- (2) **Relação Densidade Populacional vs. Densidade de Alojamentos:**
  - Existe uma correlação estatisticamente significativa entre a densidade populacional (Nº/km<sup>2</sup>) e a densidade de alojamentos (Nº/km<sup>2</sup>)?
  - Esta relação é linear? Maior densidade populacional corresponde a maior 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 (Nº/km<sup>2</sup>) no preço médio de arrendamento (€/m<sup>2</sup>) 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 (€/m<sup>2</sup>) nos municípios?
  - A densidade populacional é um indicador 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.
- **rendasm2.json**: Apresenta os valores de arrendamento (€/m<sup>2</sup>), por município.
- **densidadePopulacional.json**: Dados relativos à densidade populacional (Nº/km<sup>2</sup>), por município.
- **densidadealojamentosm2.json**: Contém informação sobre a densidade de alojamentos (Nº/km<sup>2</sup>), por freguesia.
- **Areas\_Freg\_Conc\_Dist\_Pais\_CAOP2019.xls**: Contém os códigos geográficos, para podermos fazer as relações entre freguesia, municípios e concelhos dos outros datasets.

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

## 3 CARACTERIZAÇÃO DOS DADOS E DATA CLEANING


### 3.1 Areas\_Freg\_Conc\_Dist\_Pais\_CAOP2019.xls → concelhosTratado.csv

O dataset original continha 14 colunas com informações diversas sobre concelhos portugueses. Foram aplicadas as seguintes transformações:

#### Transformações:

- Selecionadas apenas 3 colunas relevantes:
  - **DICO** → **MunicipalityCode** (código do concelho)
  - **CONCELHO\_DSG** → **Municipality** (nome do concelho)
  - **NUTSIII\_DSG** → **Region** (região NUTS III)
- Colunas normalizadas criadas:
  - **Municipality\_norm** (nome do concelho normalizado)
  - **Region\_norm** (nome da região normalizado)
- **MunicipalityCode** formatado com 4 dígitos (preenchido com zeros à esquerda)

| Municipality      |        |
|-------------------|--------|
| MunicipalityCode  | int    |
| Municipality      | String |
| Municipality_norm | String |
| Region            | String |
| Region_norm       | String |




### 3.2 Portugal\_listings.csv → listingsTratado.csv

O dataset original continha várias colunas sobre listagens de propriedades. Foram aplicadas as seguintes transformações:

**Transformações:**

- Removidas colunas irrelevantes (**Unnamed**, **Town**)
- **City** renomeado para **Municipality**
- Normalização de texto:
  - **District** → **District\_norm**
  - **Municipality** → **Municipality\_norm**
- **Price** convertido para float
- Removidos registos com **Price** ≤ 0 ou nulos

| Listings          |        |
|-------------------|--------|
| District          | String |
| Price             | float  |
| Municipality_Norm | String |
| Municipality      | String |
| District_Norm     | String |



### 3.3 Rendasm2.json → rendasTratado.csv

Dataset original continha valores de rendas a nível de freguesia.


**Transformações:**

- Extraído **MunicipalityCode** dos últimos 4 dígitos de **geocod**
- Normalizado **geodsg** → **FreguesiaNorm**
- **valor** convertido para float → **RentValue**
- Junção com dados de concelhos para adicionar:
  - **Municipality**

- **Municipality**
- **Municipality\_norm**

- Valores nulos substituídos pela média por concelho
- Agregado por concelho (média) → **RentValueMean**

| Rendas            |        |
|-------------------|--------|
| MunicipalityCode  | String |
| Municipality      | String |
| Municipality_norm | String |
| RentValueMean     | Float  |




### 3.4 DensidadePopulacional.json → densidadePopTratado.csv

Dataset original já estava a nível de concelho.

**Transformações:**

- Extraído **MunicipalityCode** dos últimos 4 dígitos de **geocod**
- Normalizado **geodsg** → **Municipality\_norm**
- **valor** convertido para float → **Density**
- **geodsg** renomeado para **Municipality**

| PopulationDensity |        |
|-------------------|--------|
| geocod            | String |
| Municipality      | String |
| Municipality_Norm | String |
| Density           | Float  |



### 3.5 Densidadealojamentosm2.json → densidadeAlojTratado.csv

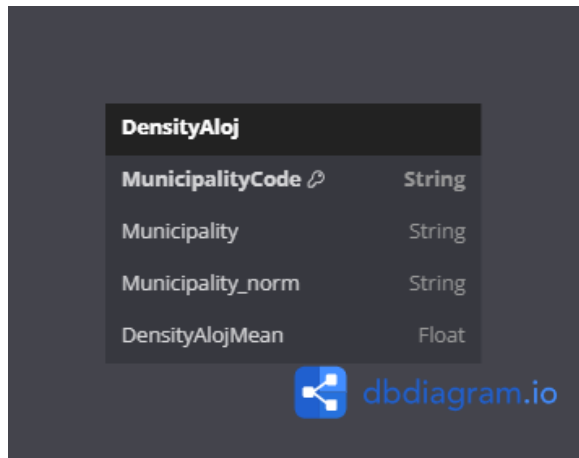
Dataset original continha densidade de alojamentos a nível de freguesia.

**Transformações:**

- Extraído **MunicipalityCode** dos primeiros 4 dígitos de **geocod**
- Normalizado **geodsg** → **FreguesiaNorm**
- Junção com dados de concelhos para adicionar:
  - **Municipality**

– **Municipality\_norm**

- **valor** convertido para float → **Density**
- Valores nulos substituídos pela média por concelho
- Agregado por concelho (média) → **DensityAlojMean**



## 4 MODELO INTEGRADO

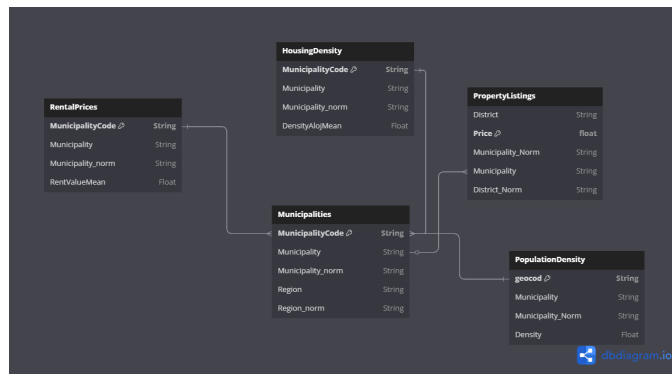


Figure 1: Diagrama do modelo integrado

### 4.1 Plano de Integração e Justificação de Escolhas

O modelo integrado foi concebido para agregar dados de diversas fontes através de chaves geográficas comuns (*geocod/MunicipalityCode*), 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 Municipalities:** Criou-se a entidade **Municipalities** com chave primária *MunicipalityCode*, para

relacionar as tabelas *RentalPrices*, *HousingDensity* e *PopulationDensity*. A tabela *PropertyListings* é relacionada pela entidade *Municipality*.

- **Chaves Primárias Unificadas:** Os campos *geocod/MunicipalityCode* e *Municipality* serve como chave primária de junção entre todas as entidades, garantindo consistência nos relacionamentos. Todas as relações são one-to-many:
  - *RentalPrices* → *Municipalities* (N:1)
  - *PropertyListings* → *Municipalities* (N:1)
  - *PopulationDensity* → *Municipalities* (N:1)
  - *HousingDensity* → *Municipalities* (N:1)

### 4.3 Abordagem de Resolução de Identidades

Para garantir correspondência precisa entre entidades:

- **Chaves Primárias:** Utilização de *geocod* e *MunicipalityCode* como identificador canónico para as regiões de cada dataset.

## 5 FERRAMENTAS UTILIZADAS

- **Pandas** - Biblioteca principal para manipulação de dados tabulares (via *DataFrames*), usada na leitura, limpeza, integração e exportação dos datasets.
- **Unicode** - Utilizada para normalizar nomes de localidades, removendo acentos e padronizando textos para junções consistentes.
- **Rapidfuzz** - Aplicada na resolução de identidades geográficas, utilizando algoritmos de similaridade textual (*token\_sort\_ratio*, *Levenshtein*, *token\_set\_ratio*) para corresponder nomes de municípios.
- **Jupyter Notebook** - Ambientes para desenvolvimento interativo.

O código fonte completo deste projeto, está disponível no repositório público do GitHub:

**Repositório**

## 6 CONCLUSÃO

A primeira fase do projeto permitiu construir um pipeline robusto de integração e preparação de dados sobre o mercado da habitação, com foco nas relações entre preços de habitação, rendas, densidade populacional e densidade de alojamentos.

Foram recolhidos dados de fontes oficiais (INE, PORDATA, DGT) e tratados de forma a uniformizar a granularidade ao nível do concelho. A resolução de identidade foi essencial para cruzar o dataset de listings (sem geocod) com os restantes, recorrendo a nomes normalizados e métricas de similaridade textual (*token\_sort\_ratio*, *Levenshtein*, entre outras).

#### Principais resultados:

- Estrutura relacional unificada com base no *MunicipalityCode*;
- Datasets homogêneos e prontos para análise cruzada;
- Processo de integração replicável e documentado.

#### Próximos passos:

- Consolidar os pares de identidade com limiares de confiança;
- Atribuir códigos aos listings e gerar tabela agregada por concelho;
- Explorar visualizações e correlações entre indicadores demográficos e preços de habitação.