

# Exploring the relationship between rental/sale prices and population density in Portugal

Integração e Processamento Analítico de Informação  
Group 2  
João Oliveira - 56908  
Tiago Garcia - 54938

# Questions we want to answer

- Sale Price vs. Rental Price Relationship
- Population Density vs. Housing Density Relationship
- Impact of Housing Density on Rental and Sale Prices
- Impact of Population Density on Rental Prices
- Impact of Population Density on Sale Prices

# Data Sources

kaggle™



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



# Data characterization

- **portugal\_listings.csv**: Contains information on housing sale prices by municipality.
- **rendasm2.json**: Displays rental values (€/m<sup>2</sup>) by parish.
- **densidadePopulacional.json**: Data on population density (inhabitants/km<sup>2</sup>) by municipality.
- **densidadealojamentosm2.json**: Contains information on housing unit density (units/km<sup>2</sup>) by parish.
- **Areas\_Freg\_Conc\_Dist\_Pais\_CAOP2019.xls**: Includes geographic codes to establish relationships between parishes, municipalities, and districts.

# Data preparation

Cleaning, transforming and aligning datasets for integration

## Tools used

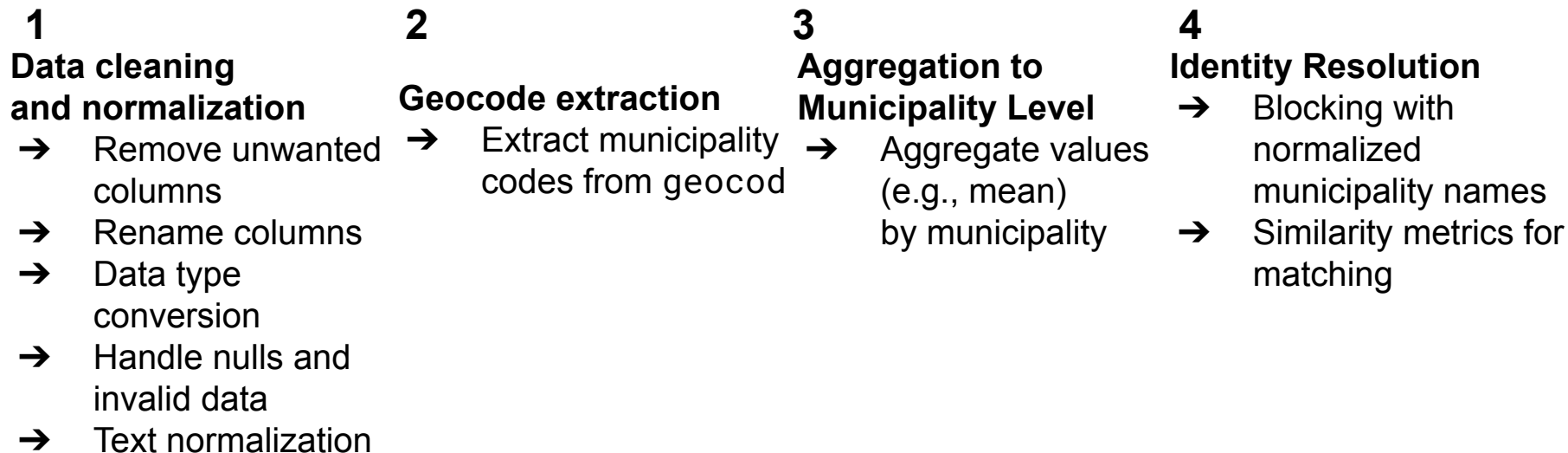


Rapid *Fuzz*



Unicode 1.3.8

# Workflow



# Final Structure

PopulationDensity	
geocod	String
Municipality	String
Municipality_Norm	String
Density	Float

DensityAloj	
MunicipalityCode	String
Municipality	String
Municipality_norm	String
DensityAlojMean	Float

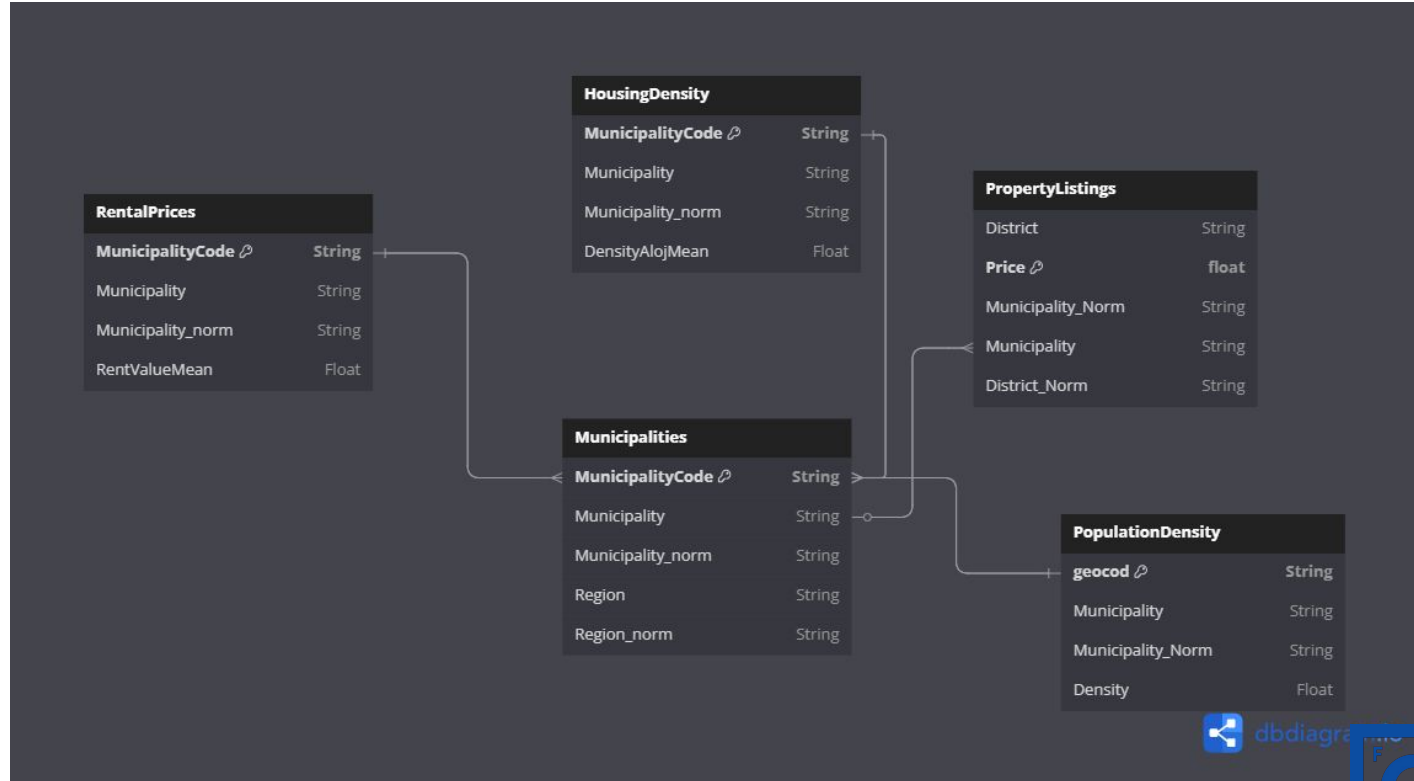
Rendas	
MunicipalityCode	String
Municipality	String
Municipality_norm	String
RentValueMean	Float

Municipalities	
MunicipalityCode	String
Municipality	String
Municipality_norm	String
Region	String
Region_norm	String

Listings	
District	String
Price	float
Municipality_Norm	String
Municipality	String
District_Norm	String



# Integrated Model



# Conclusion and next steps

# Conclusions

- Cleaned and normalized heterogeneous datasets.
- Solved identity resolution issues to allow cross-dataset analysis.
- Built a reusable and documented process.

# Next steps

- Consolidate identity pairs with confidence thresholds.
- We want to improve our similarity metrics, with an algorithm to compare geocodes.
- Explore visualizations and correlations between demographic indicators and housing prices.

Any questions?  
Thank you