
Análise de indicadores municipais portugueses

ALUNOS: JOÃO SILVA (A51557) E
SOFIA MACHADO (A52152)

UNIDADE CURRICULAR:

ELEMENTOS DE INTELIGÊNCIA
ARTIFICIAL E CIÊNCIA DE DADOS

DOCENTE: JOÃO C. NEVES

ANO LETIVO: 2024/2025

DATA: 31 DE MAIO DE 2025



Departamento de
Informática

Objetivo do projeto



Analisar indicadores socioeconómicos municipais de Portugal, com foco em três variáveis:

Taxa de Desemprego

Ganho Médio Mensal por Trabalhador

População Residente



Identificar padrões e relações entre os indicadores, recorrendo a técnicas de ciência de dados.



Aplicar métodos de aprendizagem não supervisionada para caracterizar e agrupar os municípios de acordo com as suas características socioeconómicas.



Dados utilizados

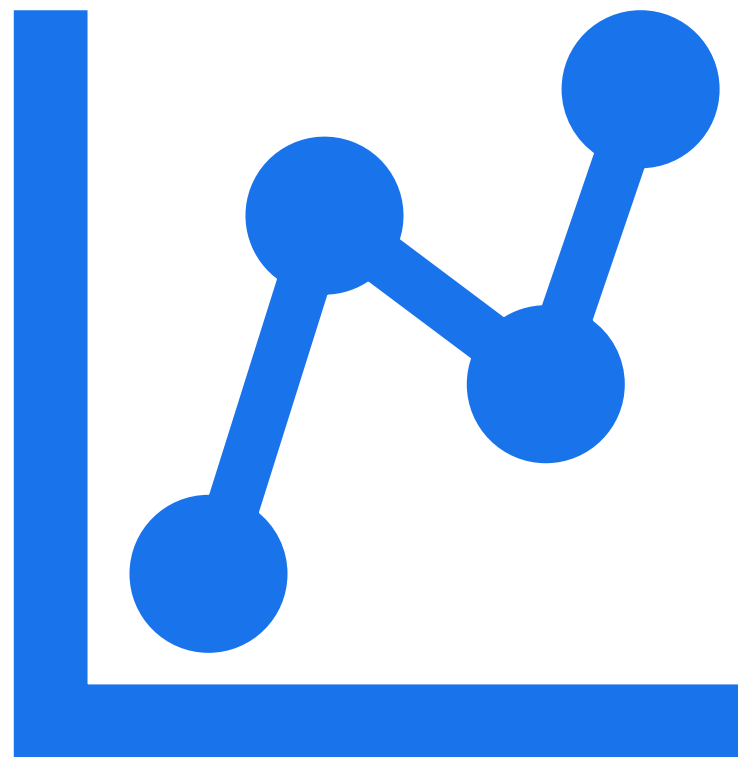
- **Fonte dos Dados:**
 - **PORDATA** – Base de dados estatísticos da Fundação Francisco Manuel dos Santos.
- **Indicadores Recolhidos:**
 - **Taxa de Desemprego (%)**
 - **Ganho Médio Mensal por Trabalhador (€)**
 - **População Residente (n.º de habitantes)**
- **Ficheiros CSV:**
 - `Desemprego.csv`
 - `Ganho_medio_mensal.csv`
 - `populacao_residente.csv`

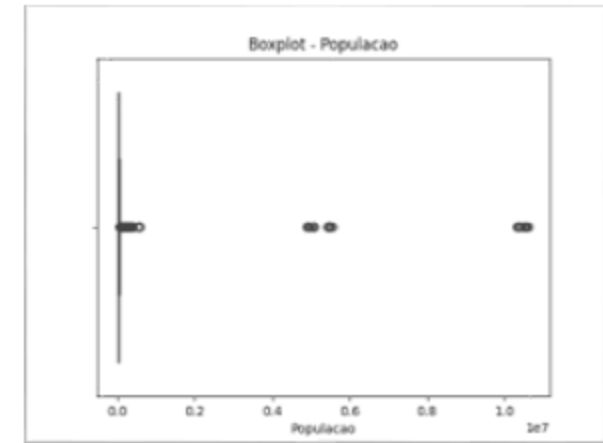
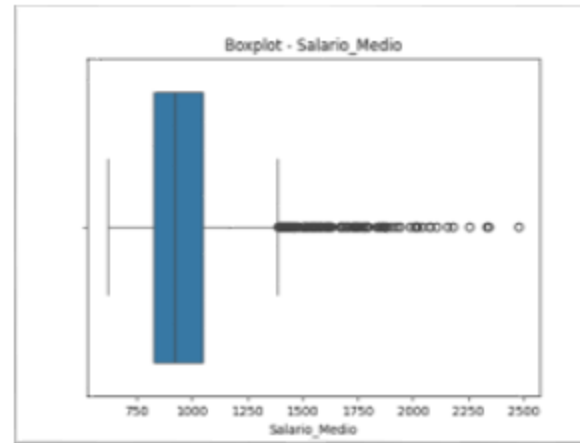
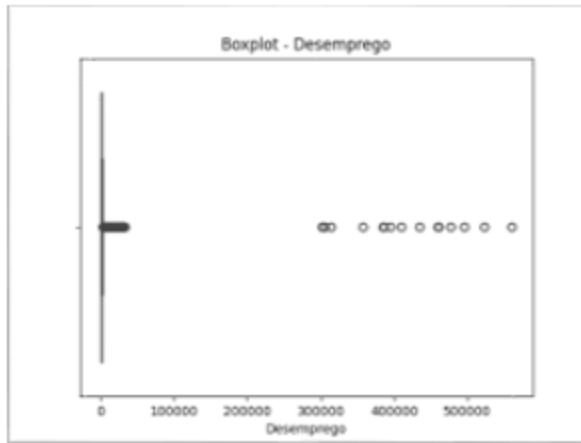


Introdução

- Antes de aplicarmos qualquer modelo, começámos por explorar e compreender o comportamento dos dados brutos. A análise exploratória permitiu identificar problemas que poderiam distorcer os resultados das técnicas de clustering, como outliers e diferentes escalas entre variáveis.

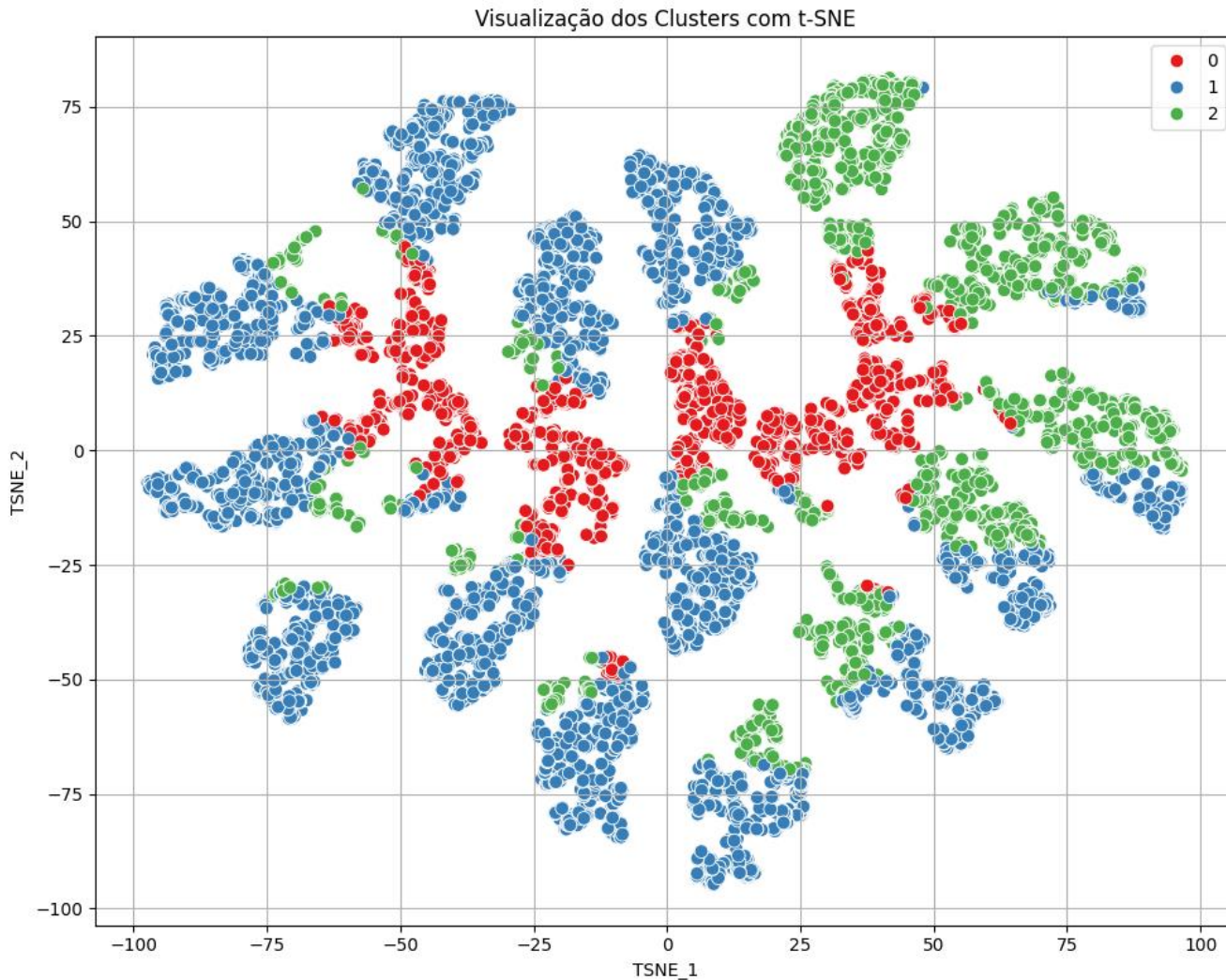
Interpretação de dados





Boxplots

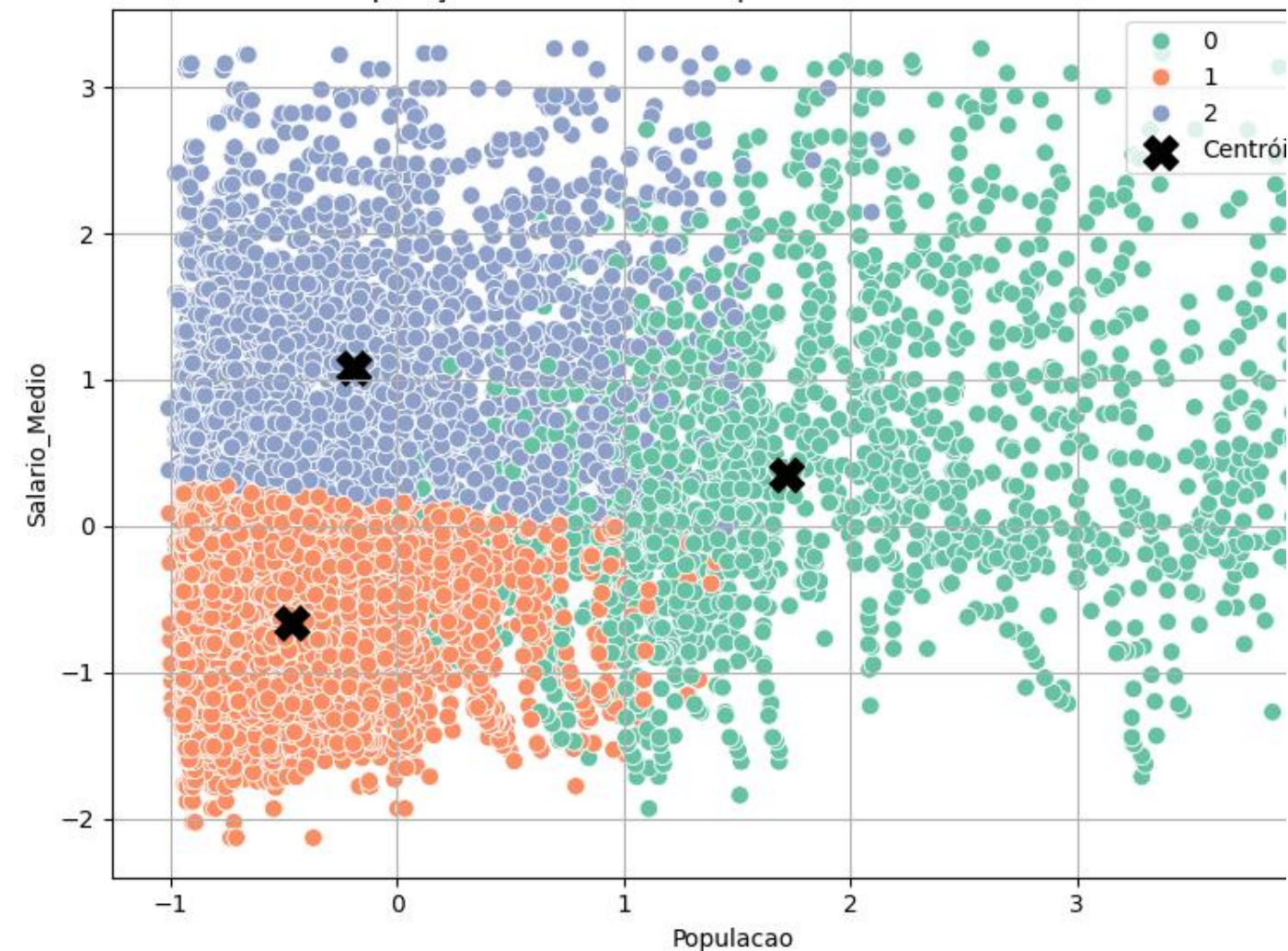
• Os gráficos apresentados correspondem aos dados antes da normalização. Observa-se a presença de diversos outliers, os quais podem comprometer a integridade das análises subsequentes. Esta visualização evidencia a necessidade de um pré-processamento adequado, incluindo a normalização e o eventual tratamento de valores discrepantes.



Distribuição de Municípios no Espaço t-SNE

- Identificação de clusters bem definidos.
- Cada ponto: um município num determinado ano.
- Clusters indicam grupos de municípios com características socioeconômicas semelhantes:
 - Altas taxas de desemprego e baixos salários.
 - Zonas urbanas com maior rendimento e menor desemprego.

População vs Salário Médio por Cluster com Centróides

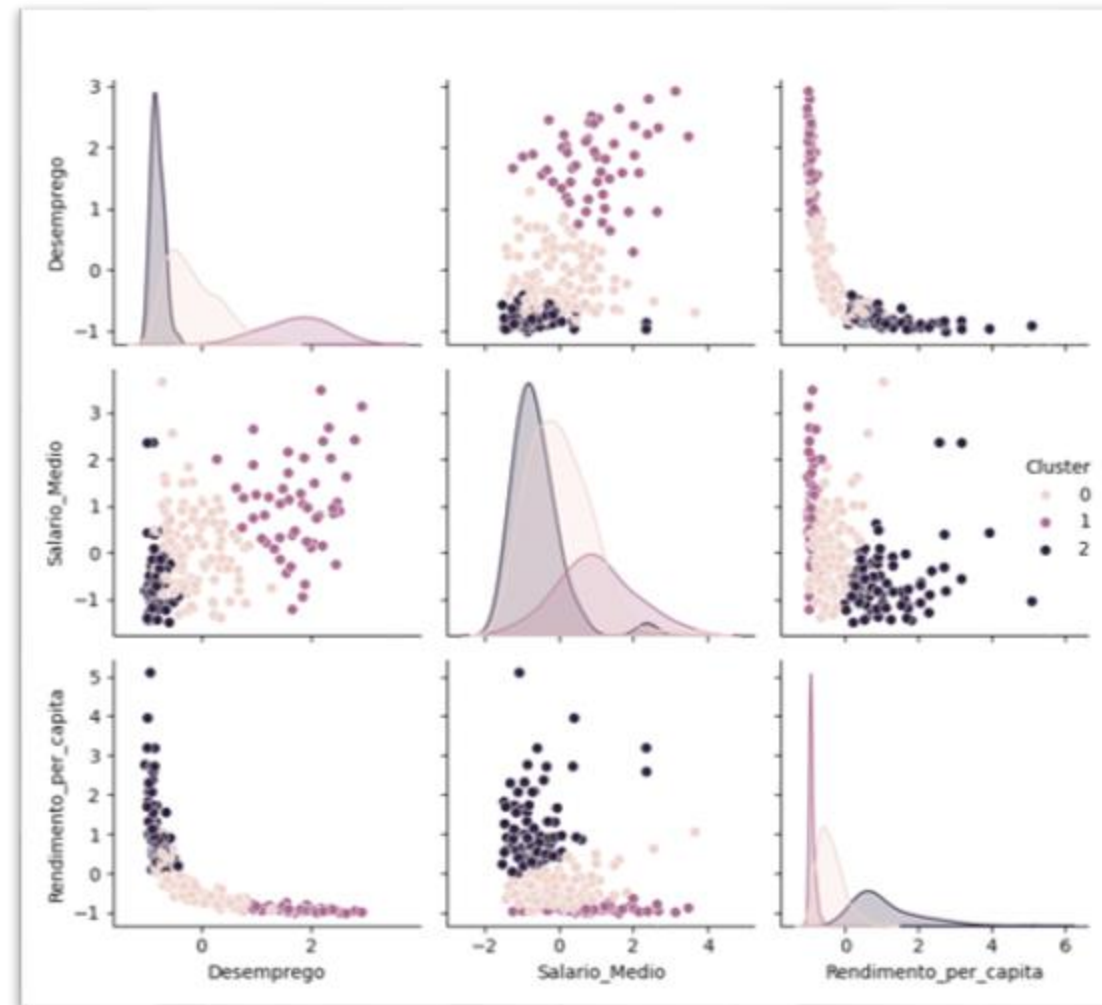


Relação entre População e Salário Médio

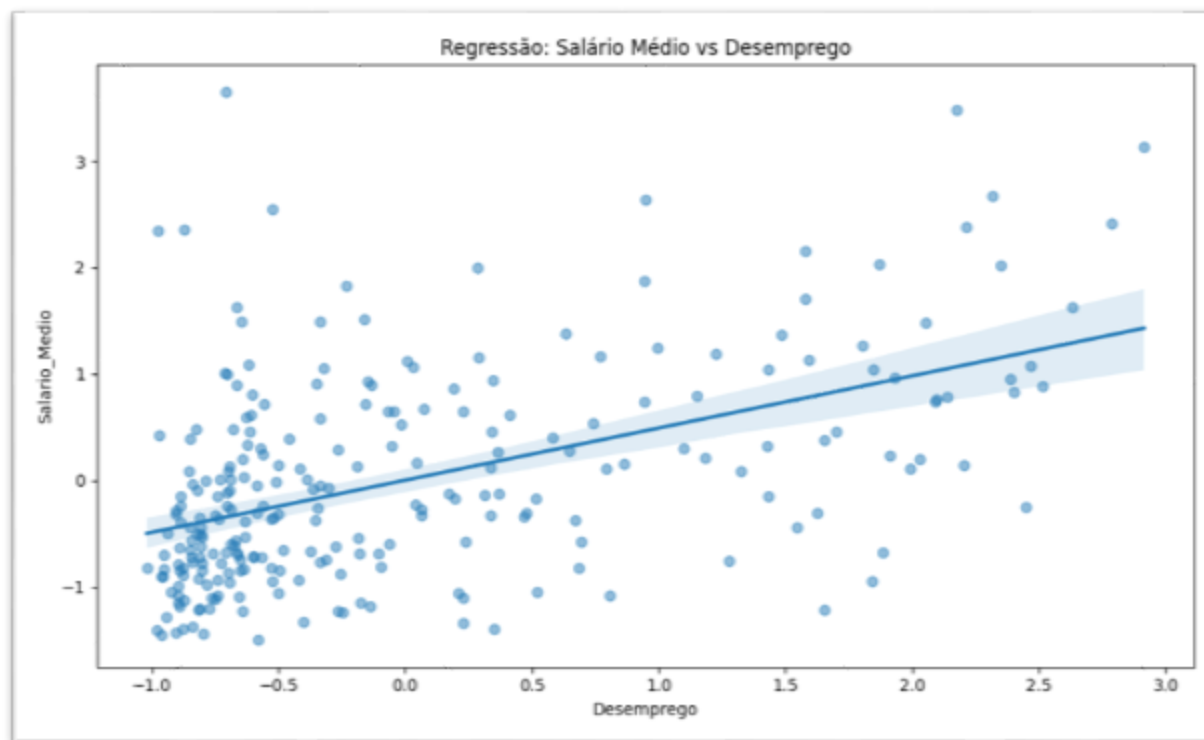
- Dispersão dos municípios com centróides destacados.
- Confirma a distinção entre os clusters.
- Municípios mais populosos tendem a apresentar salários médios mais elevados

Caracterização dos Clusters

- **Cluster 0:** Salários médios altos, baixo desemprego e maior rendimento per capita (regiões urbanas/desenvolvidas).
- **Cluster 1:** Valores intermédios nos três indicadores (equilíbrio relativo).
- **Cluster 2:** Desemprego elevado, salários baixos e menor rendimento per capita (zonas mais vulneráveis). As distribuições KDE nas diagonais mostram as densidades dentro de cada cluster, e os gráficos de dispersão evidenciam correlações:
 - **Salario_Medio** e **Rendimento_per_capita** mostram relação positiva clara;
 - Desemprego parece inversamente relacionado com os outros dois indicadores nos clusters mais extremos.

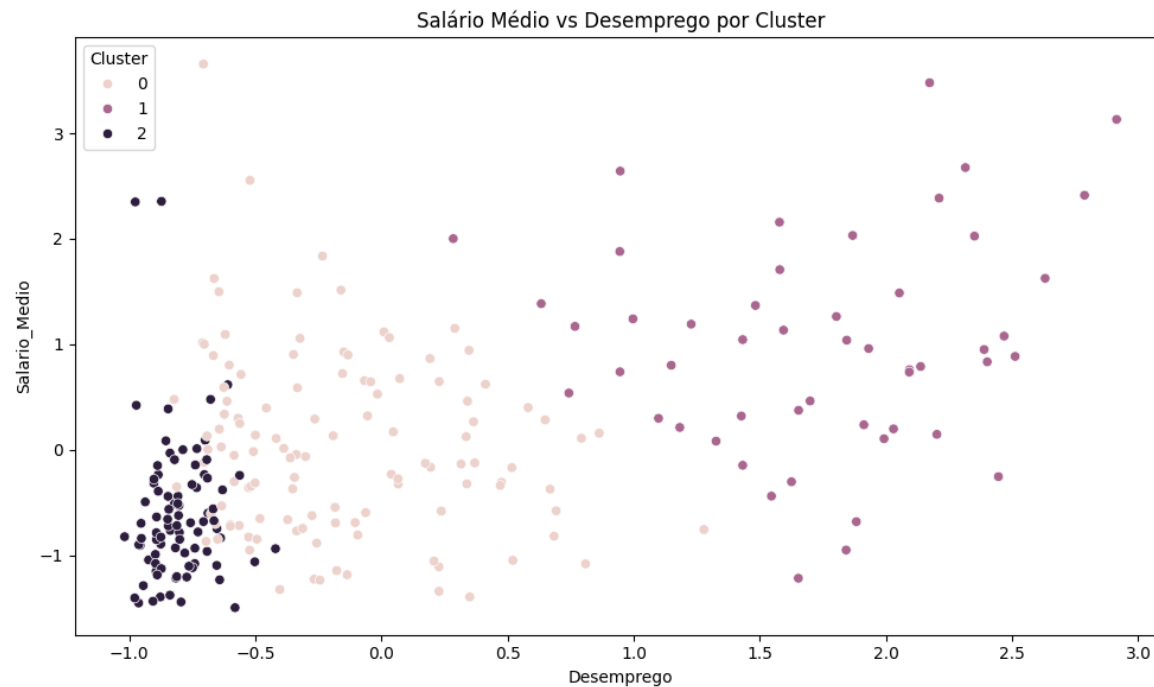


Relação entre Desemprego e Salário Médio



- Tendência negativa entre desemprego e salário médio.
- De acordo com a teoria económica: maior desemprego → menor dinamismo económico e salários mais baixos.
- Modelo linear capta uma direção clara, apesar da dispersão dos pontos

Distribuição dos Clusters no Espaço Socioeconômico



- Clusters ocupam zonas distintas no gráfico.
- Separação clara de regiões com perfis económicos específicos:
 - **Cluster 2:** Baixo desemprego, salários mais altos (desenvolvimento económico).
 - **Cluster 0:** Desemprego elevado, salários baixos (fragilidade económica).
 - **Cluster 1:** Valores intermédios (transição).
- Tendência visível de separação socioeconómica, ainda que não totalmente linear.



Conclusão

- Em suma, as técnicas de aprendizagem não supervisionada permitiram identificar padrões socioeconómicos entre municípios portugueses, destacando desigualdades regionais e relações esperadas entre variáveis económicas. Este tipo de análise pode ser útil para apoiar decisões de políticas públicas e desenvolvimento regional.