

Relatório de Análise de Estatísticas Municipais: Caixas Multibanco e Depósitos Bancários

Licenciatura em Inteligência Artificial e Ciências de Dados

27 de maio de 2025

1 Recolha de Dados

Os dados utilizados neste projeto foram extraídos da plataforma **PORDATA** (<https://www.pordata.pt/>), que disponibiliza estatísticas a nível municipal. Os temas selecionados para análise foram:

- Caixas Multibanco
- Depósitos Bancários

Estes dados foram reunidos em arquivos CSV e contêm informações económicas por município e ano.

2 Integração de Dados

Os dados recolhidos foram integrados num único conjunto estruturado, no formato tabular, com as seguintes colunas principais:

- **Ano**
- **Município**
- **CaixasMultibanco**
- **DepositosBancarios**

O ficheiro final foi guardado como `dados_municipais_integrados_limpos.csv` - já está incluído dentro do pacote o ficheiro limpo

3 Análise Exploratória de Dados

Foram utilizadas as bibliotecas `pandas`, `matplotlib` e `seaborn` para a análise exploratória dos dados.

Verificação de valores nulos

Foi utilizada a função `df.isnull().sum()` para identificar valores em falta, que foram posteriormente tratados.

Estatísticas descritivas

A função `df.describe()` permitiu obter medidas como média, desvio padrão, quartis, entre outras.

Correlação entre variáveis

Um mapa de calor com `seaborn.heatmap` revelou correlação positiva entre as variáveis `CaixasMultibanco` e `DepositosBancarios`.

Identificação de outliers

Foram gerados `boxplots` para ambas as variáveis, auxiliando na detecção visual de outliers.

4 Limpeza e Pré-processamento de Dados

A limpeza dos dados incluiu:

- Remoção de linhas com valores nulos usando `df.dropna()`.
- Remoção de outliers com base na técnica do IQR (intervalo interquartil), através da função `remover_outliers(col)`.

O conjunto limpo foi então utilizado nas fases seguintes.

5 Análise Descritiva

Foi empregue o algoritmo **K-Means** para realizar a análise não supervisionada. O procedimento incluiu:

- Agrupamento dos dados por município com `groupby` e cálculo das médias.
- Normalização dos dados com `StandardScaler`.
- Aplicação do algoritmo `KMeans(n_clusters=3)`.
- Visualização dos clusters com `seaborn.scatterplot`.

Foram identificados três grupos de municípios com características semelhantes quanto à presença de caixas multibanco e ao volume de depósitos bancários.

6 Funções e Scripts Utilizados

As principais funções e scripts utilizados no projeto incluem:

- Leitura e análise de dados: `pandas.read_csv`, `df.describe()`, `df.isnull()`
- Visualizações: `matplotlib.pyplot`, `seaborn.heatmap`, `seaborn.boxplot`, `seaborn.scatterplot`
- Limpeza: `df.dropna()`, função personalizada `remover_outliers`
- Clusterização: `StandardScaler`, `KMeans`
- Interface de utilizador para análise municipal: função `analisar_municipio(nome)`

7 Conclusão

O projeto permitiu identificar padrões e agrupar municípios com base em duas estatísticas bancárias. A utilização de métodos de análise de dados e aprendizagem não supervisionada foi eficaz para gerar conhecimento relevante a partir dos dados municipais.

Fonte dos Dados

- Plataforma PORDATA: <https://www.pordata.pt/DB/Municipios/Ambiente-de-Consulta>