**TRANSFORMAÇÃO E ANÁLISE DE DADOS IMOBILIÁRIOS UTILIZANDO O POWER BI E O PROCESSO CRISP-DM**

André Carvalho

Pedro Henrique Antunes Mello

Letícia Araújo Costa

**Resumo**

Este artigo apresenta o processo de preparação, transformação e análise de dados imobiliários aplicando o modelo CRISP-DM. Utilizando dados históricos de vendas de imóveis, o trabalho foca na criação de novas variáveis e categorização de preços para facilitar a análise visual e a interpretação dos dados no Power BI. Por meio da construção de um dashboard, buscou-se permitir uma análise detalhada sobre tendências de preço e vendas, aplicando ferramentas interativas para tomada de decisão. O processo de transformação dos dados incluiu a criação de colunas calculadas e ajustes nos intervalos de preço para categorizar imóveis, além de uma segmentação fictícia por cidade para análise geográfica.

**Palavras-chave:** CRISP-DM, Power BI, Dados Imobiliários, Dashboard, Transformação de Dados.

### **Introdução**

No contexto do mercado imobiliário, o uso de dados históricos para análise e previsão de tendências de vendas e preços é uma prática essencial para a tomada de decisões estratégicas. Com o aumento da quantidade de dados disponíveis, cresce também a necessidade de ferramentas e métodos eficazes para preparar, transformar e visualizar essas informações de maneira acessível e informativa. Este trabalho aplica o modelo CRISP-DM (Cross Industry Standard Process for Data Mining) na preparação de dados de vendas de imóveis, utilizando a plataforma Power BI para visualização interativa.

O objetivo do estudo é apresentar um processo estruturado para a criação de um painel de controle no Power BI, que permite uma análise abrangente sobre as tendências de vendas de imóveis ao longo do tempo, categorizando-os por faixa de preço, tipo e localização. A metodologia CRISP-DM foi escolhida para orientar cada etapa da análise, garantindo uma abordagem sistemática e reprodutível.

### **1. Entendimento do Negócio**

O setor imobiliário é influenciado por variáveis como localização, tipo de imóvel, número de quartos e condições econômicas. Este estudo busca explorar dados históricos de vendas de imóveis, categorizando-os com o objetivo de entender tendências de preço e frequência de vendas em diferentes localidades.

**Objetivo Geral:** Preparar um conjunto de dados de vendas de imóveis para visualização no Power BI, facilitando a compreensão e análise de padrões de preços, tipos de imóveis, sazonalidades e distribuições geográficas.

**Objetivos Específicos:**

* Classificar imóveis em diferentes categorias de preço.
* Criar variáveis temporais e geográficas para análise.
* Desenvolver um dashboard para tomada de decisão com gráficos interativos que permitam uma análise detalhada e personalizada.

### **2. Entendimento dos Dados**

Foram utilizados dois conjuntos de dados principais sobre vendas de imóveis, com registros de 2007 a 2019. Estes dados incluem variáveis como preço de venda, tipo de imóvel, número de quartos, código postal e data de venda. Abaixo está uma descrição detalhada de cada conjunto de dados:

**Conjunto de Dados 1 (“ma\_lga”)**

* **DataVenda:** Data de venda do imóvel.
* **MovelPrecoMediano:** Preço médio do imóvel.
* **TipoImovel:** Tipo de imóvel (casa ou apartamento).
* **Quartos:** Número de quartos do imóvel.

**Conjunto de Dados 2**

* **DataVenda:** Data de venda do imóvel.
* **CEP:** Código postal.
* **Preço:** Preço de venda do imóvel.
* **TipoImovel:** Tipo de imóvel (casa, apartamento, etc.).
* **Quartos:** Número de quartos do imóvel.

Os dados foram traduzidos para português, permitindo uma compreensão mais clara. Foram criadas novas colunas com informações temporais e de categorização de preços, necessárias para uma análise mais detalhada.

### **3. Preparação dos Dados**

Na fase de preparação, foram realizadas diversas transformações para adaptar os dados ao modelo de análise pretendido.

* **Tradução de Colunas:** A tradução dos nomes das colunas para português visou simplificar a interpretação e o uso no Power BI.
* **Criação de Novas Colunas:**
  + **CategoriaPreco:** Classifica os imóveis em três categorias com base no preço:
    - Baixo: Preço ≤ 667252
    - Médio: 667252 < Preço ≤ 916752
    - Alto: Preço > 916752

**DAX usado para CategoriaPreco:**

let

Preco = [MovelPrecoMediano],

Categoria = if Preco <= 667252 then "Baixo" else if Preco <= 916752 then "Médio" else "Alto"

in

Categoria

* + **AnoVenda e MesVenda:** Extraídas da coluna DataVenda para permitir análises temporais:

Fórmulas DAX:  
DAX  
Copiar código  
AnoVenda = Date.Year([DataVenda])

MesVenda = Date.Month([DataVenda])

* + **Cidade\_Ficticia:** Para facilitar uma análise geográfica, foi criada uma coluna preenchida aleatoriamente com nomes de cidades fictícias do Distrito Federal, como Brasília, Taguatinga, e Ceilândia.

### **4. Modelagem**

Após a preparação dos dados, a modelagem foi orientada para facilitar a visualização e categorização das informações no Power BI. Duas tabelas foram utilizadas e unidas com base na chave "DataVenda", permitindo a visualização de categorias de preços e variáveis temporais.

Os modelos de visualização incluíram:

1. **Classificação de Imóveis por Faixa de Preço:** A categorização Baixo, Médio e Alto auxilia na segmentação de dados para análises comparativas.
2. **Distribuição Geográfica e Temporal:** Foi criada uma análise de variação de preços e volume de vendas por cidade e ano.

### **5. Avaliação do Modelo**

A qualidade das variáveis e das transformações foi validada por meio de testes de coerência, como a distribuição equilibrada dos intervalos de preços. Os gráficos foram revisados para verificar se representavam adequadamente os padrões esperados.

**Principais Observações:**

* A categorização por preço apresentou um bom equilíbrio, facilitando a visualização de padrões por faixa de preço.
* As variáveis de data (AnoVenda e MesVenda) funcionaram bem para identificar sazonalidades e tendências ao longo dos anos.

### **6. Deployment (Implementação)**

O conjunto de dados final foi carregado no Power BI, onde foi criado um dashboard interativo com as seguintes características:

* **Importação de Dados:** Os dados foram importados diretamente de um arquivo .csv.
* **Automação dos Gráficos:** Os gráficos foram configurados para atualização automática, garantindo que novas inserções de dados sejam refletidas automaticamente.
* **Filtros Interativos:** Foram adicionados filtros por ano, categoria de preço e cidade.

### **7. Feedback**

O feedback de usuários ajudará a adaptar o dashboard para diferentes necessidades. A incorporação de sugestões permite ajustes nos gráficos e intervalos de categorização, bem como inclusão de novas variáveis, conforme a evolução das análises.

### **Conclusão**

A aplicação do processo CRISP-DM neste estudo permitiu uma estrutura organizada para o desenvolvimento de um dashboard no Power BI, facilitando a análise e interpretação de dados de vendas de imóveis. As transformações aplicadas aos dados proporcionaram insights valiosos sobre o comportamento do mercado imobiliário. O uso de variáveis como categoria de preço e localização fictícia trouxe uma visão detalhada e fácil de interpretar. Como próximo passo, recomenda-se a implementação de modelos preditivos e a incorporação de variáveis adicionais, como dados econômicos, para aumentar a precisão das análises.

### 

### 

### **Referências**

FONTE DOS DADOS ORIGINAIS: HTAG Holdings. *Property Sales Data*. Kaggle, 2019. Disponível em:<https://www.kaggle.com/datasets/htagholdings/property-sales?select=raw_sales.csv>. Acesso em: 12 nov. 2024.