



Análise de Fatores de Valorização Imobiliária em Boston

Uma Análise Exploratória do Dataset "Boston
Housing"

Data Science Aplicada ao Mercado
Imobiliário

Introdução e Estrutura da Análise

Esta apresentação detalha a análise exploratória de dados (EDA) para identificar e quantificar os fatores determinantes do valor dos imóveis residenciais na área de Boston, utilizando uma metodologia rigorosa de Data Science.



Objetivo Central

Compreender os principais fatores que influenciam o Valor Mediano dos Imóveis (**MEDV**), servindo como base para modelos preditivos.



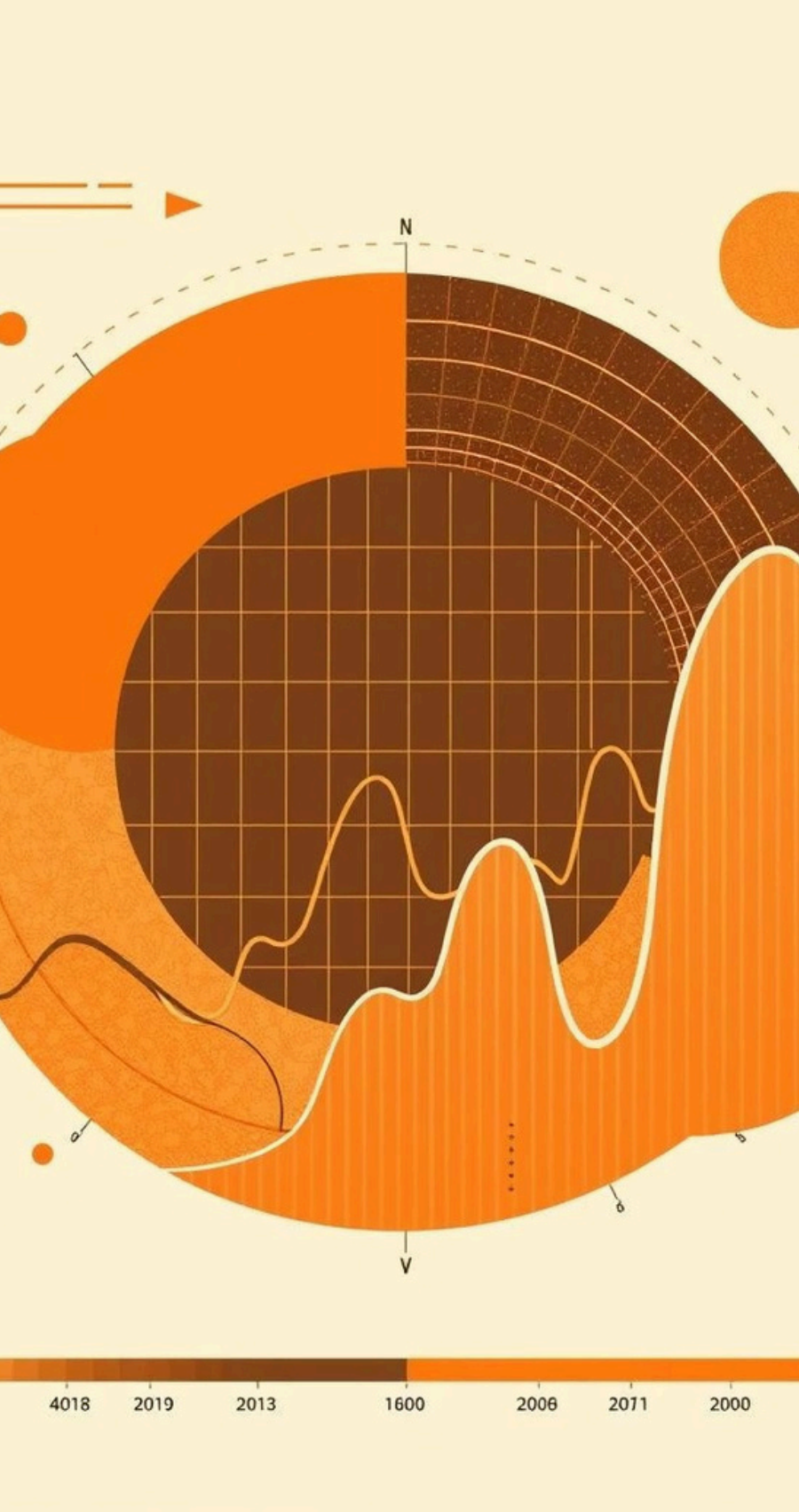
O Dataset

Análise baseada no renomado "Boston Housing Dataset", composto por 394 registros e 14 variáveis socioeconômicas e físicas.



Metodologia Adotada

Análise Descritiva, Análise de Correlação (Heatmap) para mapear relações, e Testes de Hipóteses para validar inferências estatísticas.



Análise Exploratória dos Dados (EDA)

Avaliando a Distribuição do Valor dos Imóveis (MEDV)

Estatísticas Chave do Imóvel

Aanálise descritiva inicial revela informações importantes sobre a amostra:

Métrica	Valor Médio
Valor Mediano (MEDV)	\$21.050
Média de Quartos (RM)	6.28
Taxa de Criminalidade (CRIM)	Alta Variação

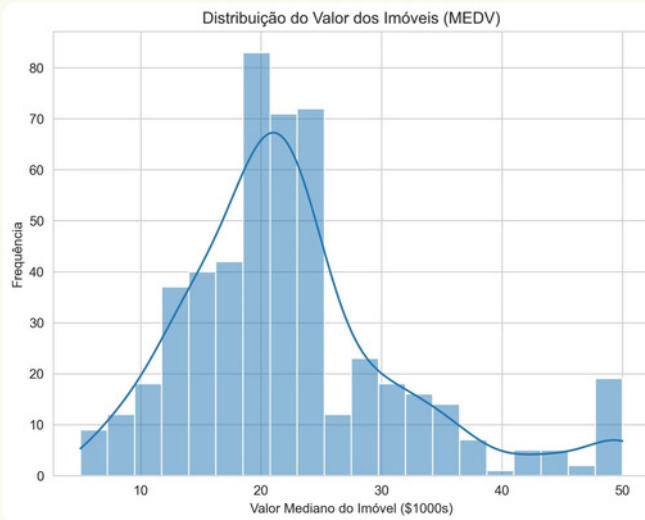
A alta variação na taxa de criminalidade por bairro (**CRIM**) sugere uma heterogeneidade significativa na qualidade das vizinhanças de Boston.

Distribuição do Preço (MEDV)

O valor dos imóveis apresenta uma distribuição predominantemente concentrada, mas com forte assimetria à direita (cauda longa).

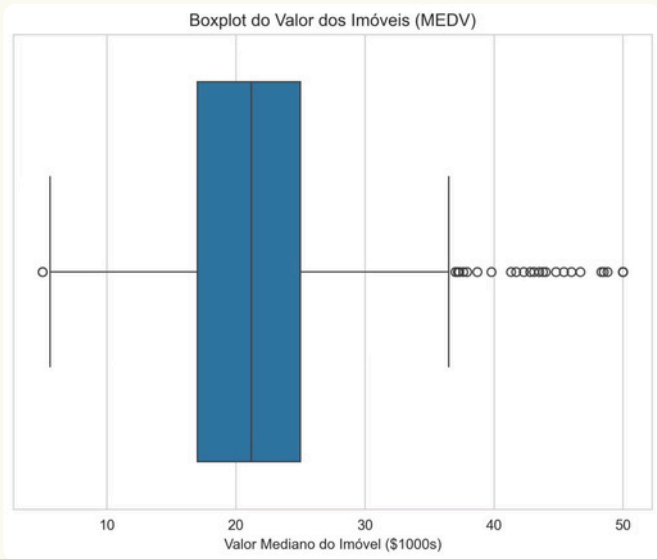
Histograma

Distribuição assimétrica à direita de MEDV.



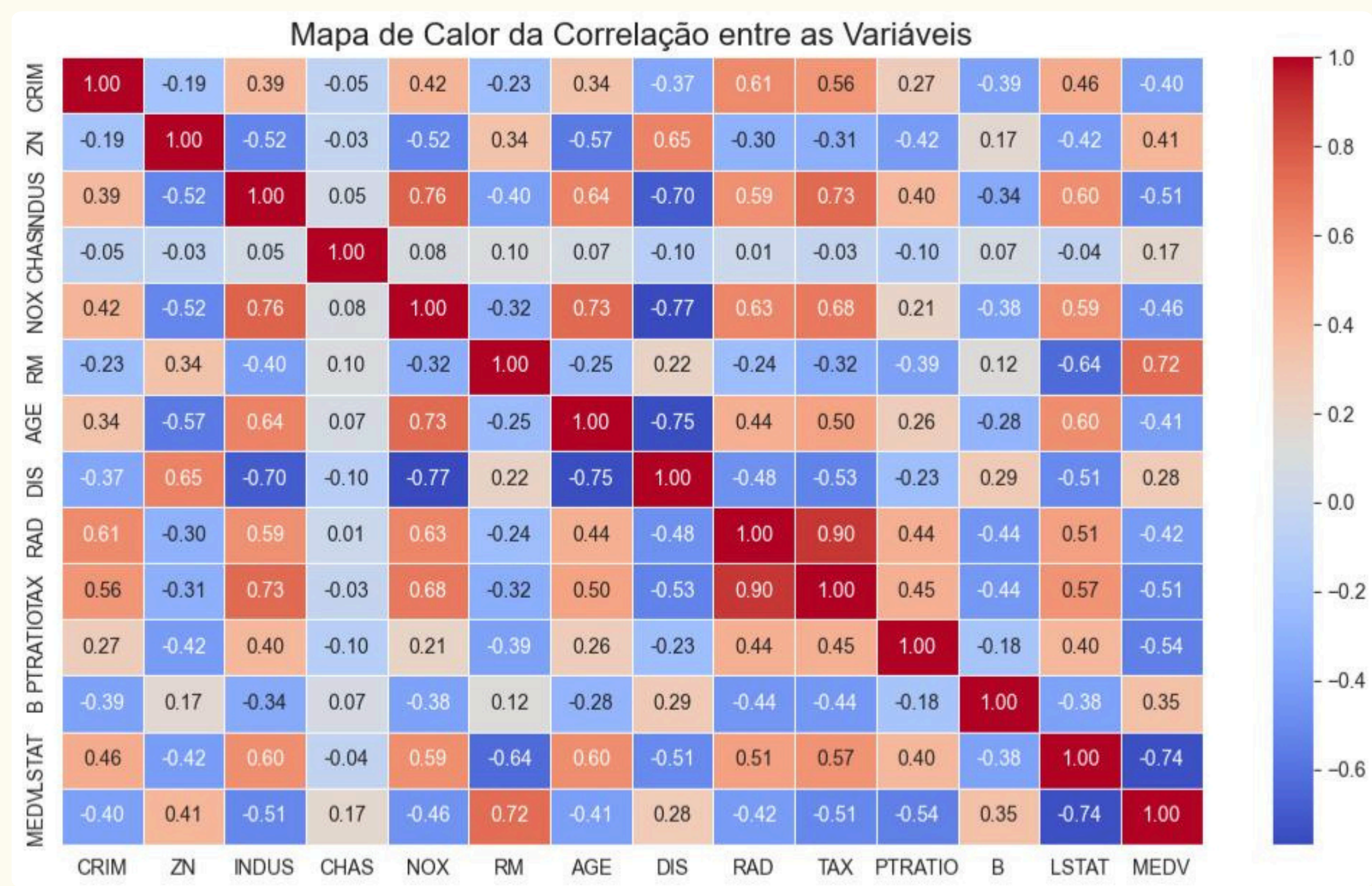
Boxplot

Mostra outliers altos claramente visíveis.



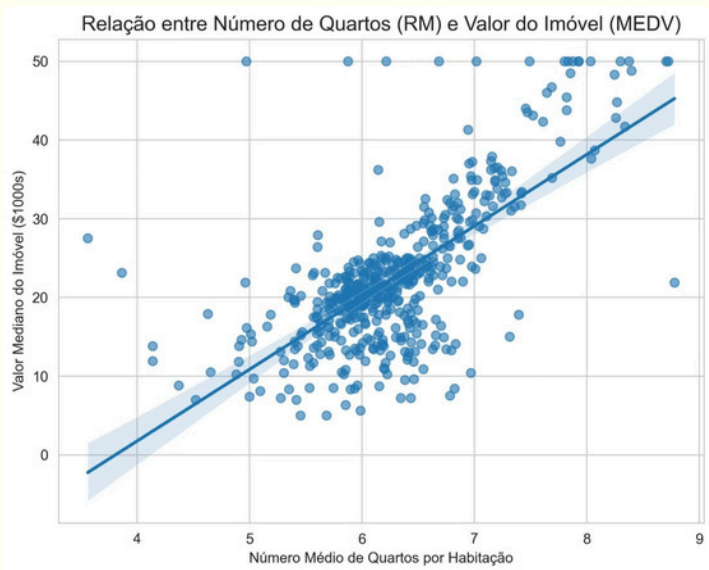
A maioria dos preços se concentra entre **\$16.800 e \$25.000**, mas a presença de uma "cauda" de imóveis de altíssimo valor (outliers) indica um mercado de luxo distinto.

Análise de Correlação: Encontrando Conexões Chave



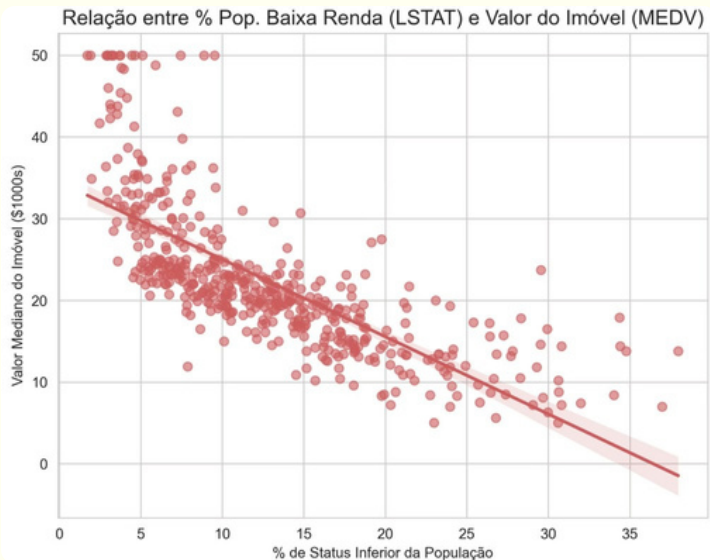
O mapa de calor de correlação revela a força e a direção do relacionamento linear entre o valor do imóvel (MEDV) e as demais variáveis. Destacamos as duas variáveis mais críticas.

Relação Positiva Forte (RM vs. Preço)



O número médio de quartos por residência (**RM**) é o fator mais correlacionado positivamente com o preço (**r = +0.72**). Mais quartos, maior o valor. Sendo a correlação estatisticamente significativa ao nível de 5%.

Relação Negativa Forte (LSTAT vs. Preço)



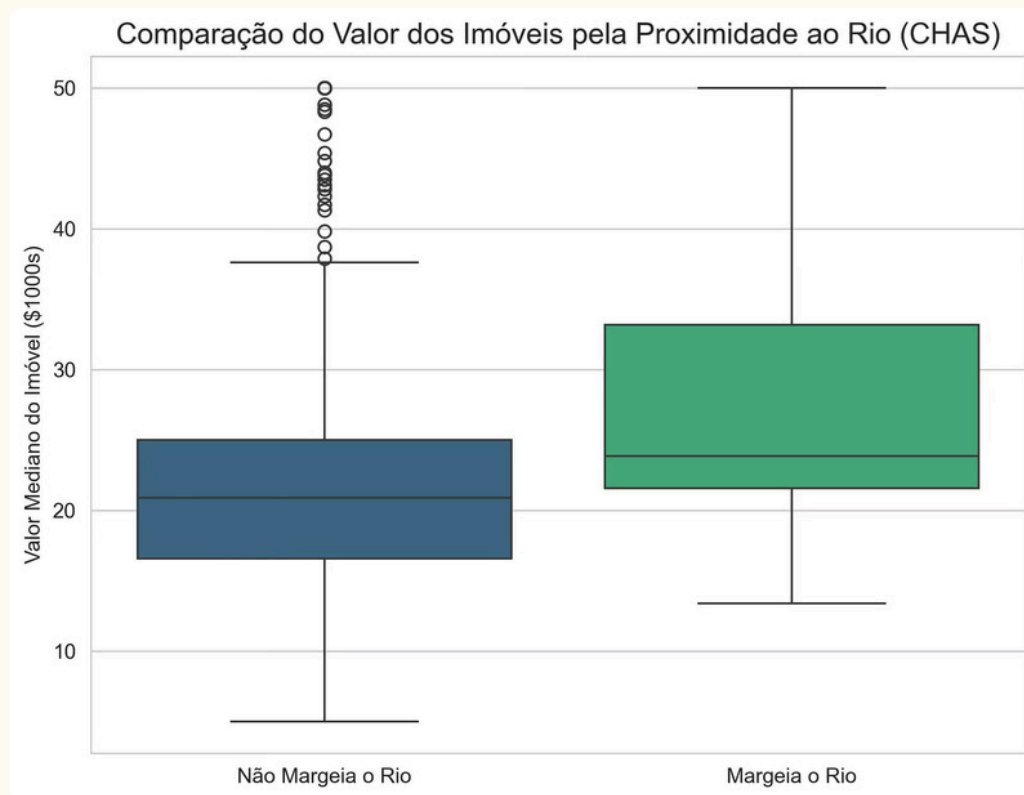
O percentual de população de baixo status socioeconômico (**LSTAT**) tem a correlação negativa mais forte (**r = -0.74**). Pior o status, menor o preço. Sendo a correlação estatisticamente significativa ao nível de 5%.

Validando Hipóteses: Localização vs. Tamanho

Para validar as influências observadas na EDA, realizamos testes de hipóteses comparando subgrupos do dataset.

Hipótese 1: Localização Privilegiada Gera Mais Valor?

Comparamos imóveis próximos ao Rio Charles (**CHAS=1**) com os demais (**CHAS=0**).

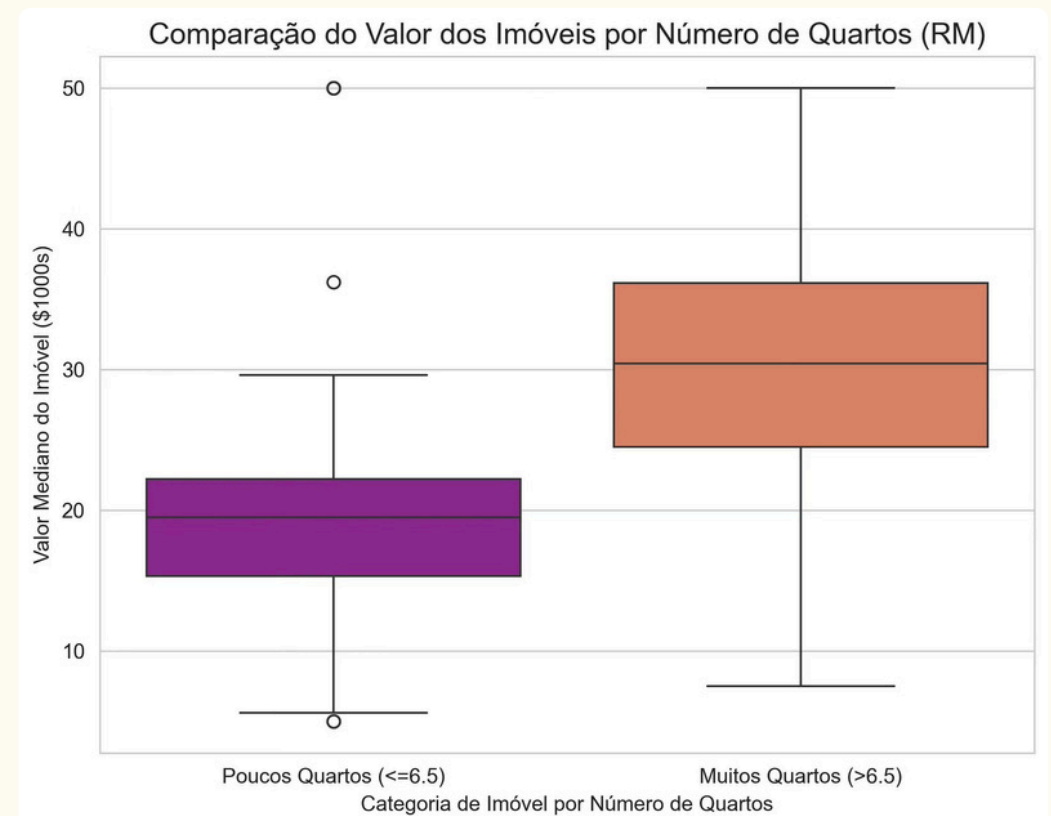


O valor mediano à beira do rio é de **\$23.300**, significativamente maior do que os **\$20.800** dos demais. A localização premium é um diferencial de valor.

📌 Conclusão: Hipótese Confirmada.

Hipótese 2: Imóveis Maiores São Consistentemente Mais Caros?

Dividimos os imóveis em dois grupos com base no número de quartos (RM > 6.5 vs. RM < 6.5).



O valor mediano de imóveis maiores que 6.5 quartos é de **\$30.100**, em contraste com os **\$19.450** dos imóveis menores. O tamanho é um fator primário de precificação.

📌 Conclusão: Hipótese Confirmada.

Conclusões Finais e Próximos Passos

Os resultados da análise exploratória fornecem uma base sólida para entender a dinâmica de preços no mercado imobiliário de Boston.



Tamanho é Crucial

O número de quartos (**RM**) é o principal fator positivo. Esta é a característica física mais importante que impulsiona a valorização.



A Vizinhança Importa

O perfil socioeconômico (**LSTAT**) do bairro é o fator negativo mais forte. A qualidade da vizinhança penaliza ou valoriza o preço final.



Localização Premium

Características únicas, como a proximidade ao Rio Charles, geram um maior valor agregado, confirmando o peso de uma localização específica.

Próximos Passos: Construção de Modelos Preditivos

O próximo estágio é utilizar estes **insights** para desenvolver um modelo de **Machine Learning** (Regressão) capaz de prever o valor dos imóveis com alta precisão com base nas características identificadas.