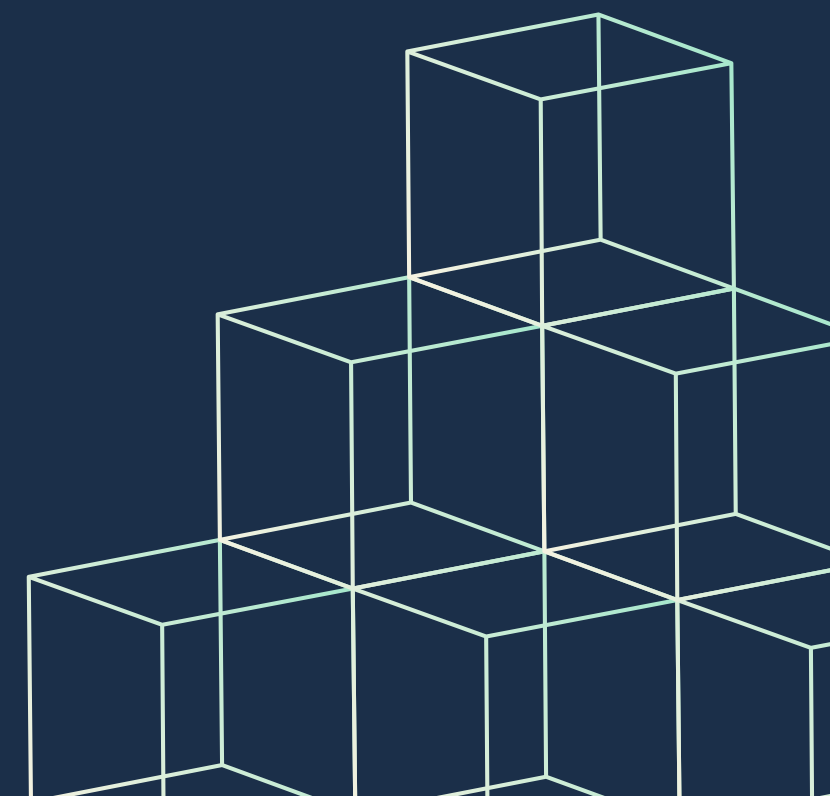
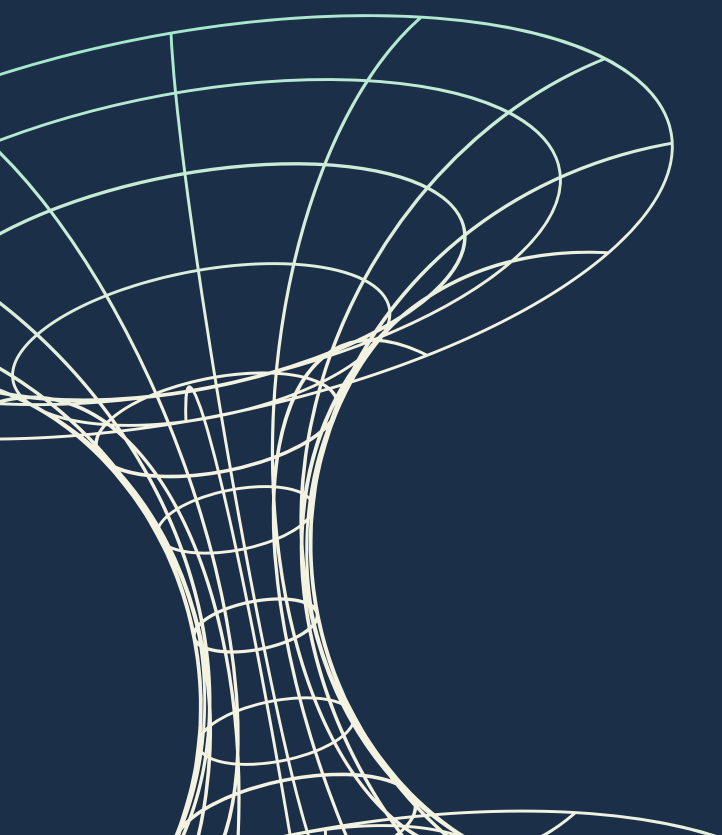


MERCADO IMOBILIÁRIO

Arthur D. Ogg

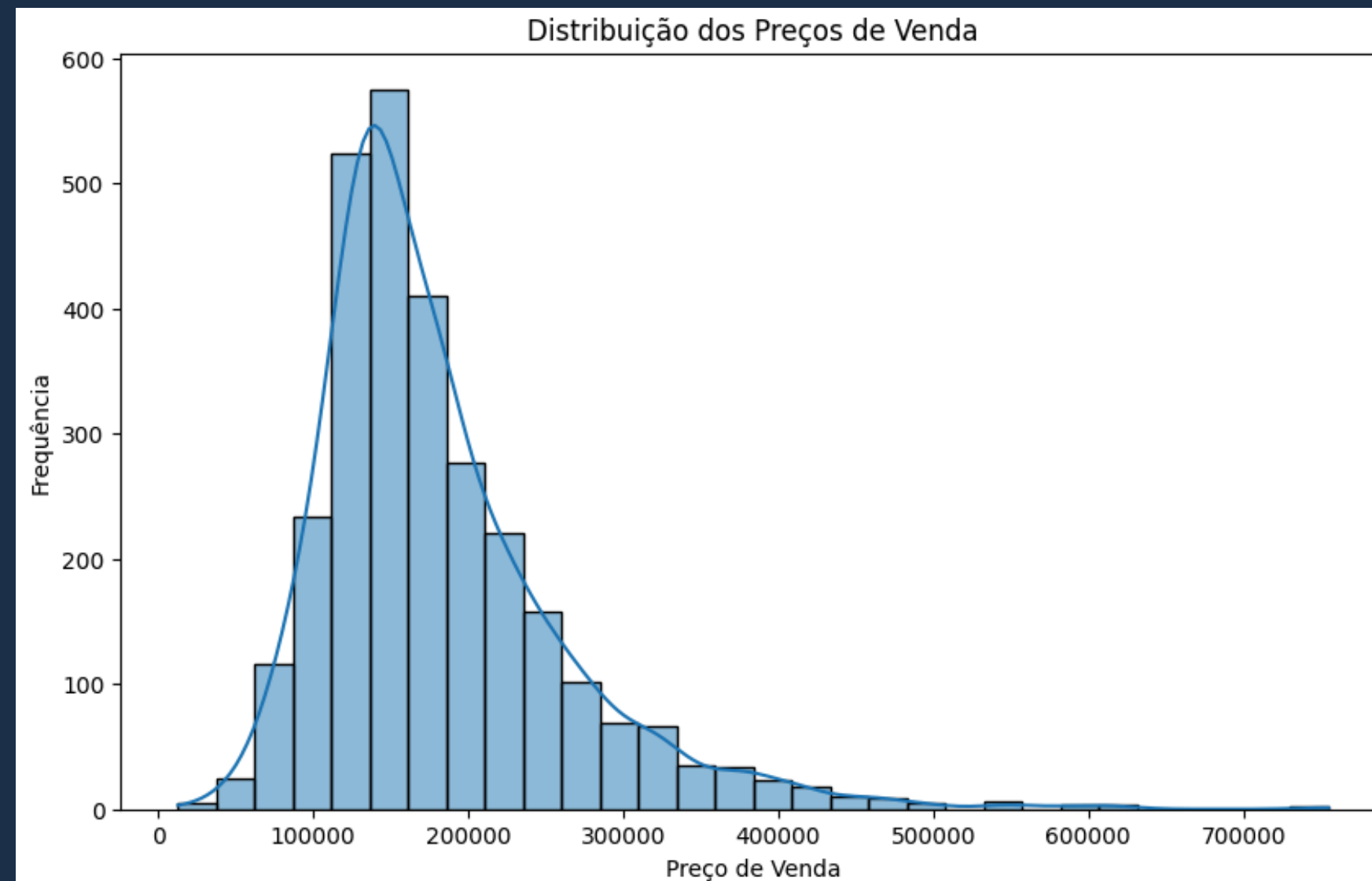
Gabriela Schmitt

Vinicius Dionizio



Análise Exploratória

Histograma da variável SalePrice

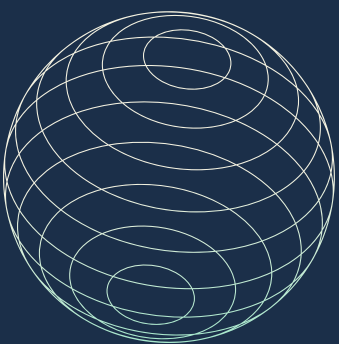


A distribuição dos preços de venda é assimétrica à direita (right-skewed), o que indica que há mais imóveis com preços mais baixos e um número crescente de imóveis com preços mais altos que estão se distanciando dessa média.

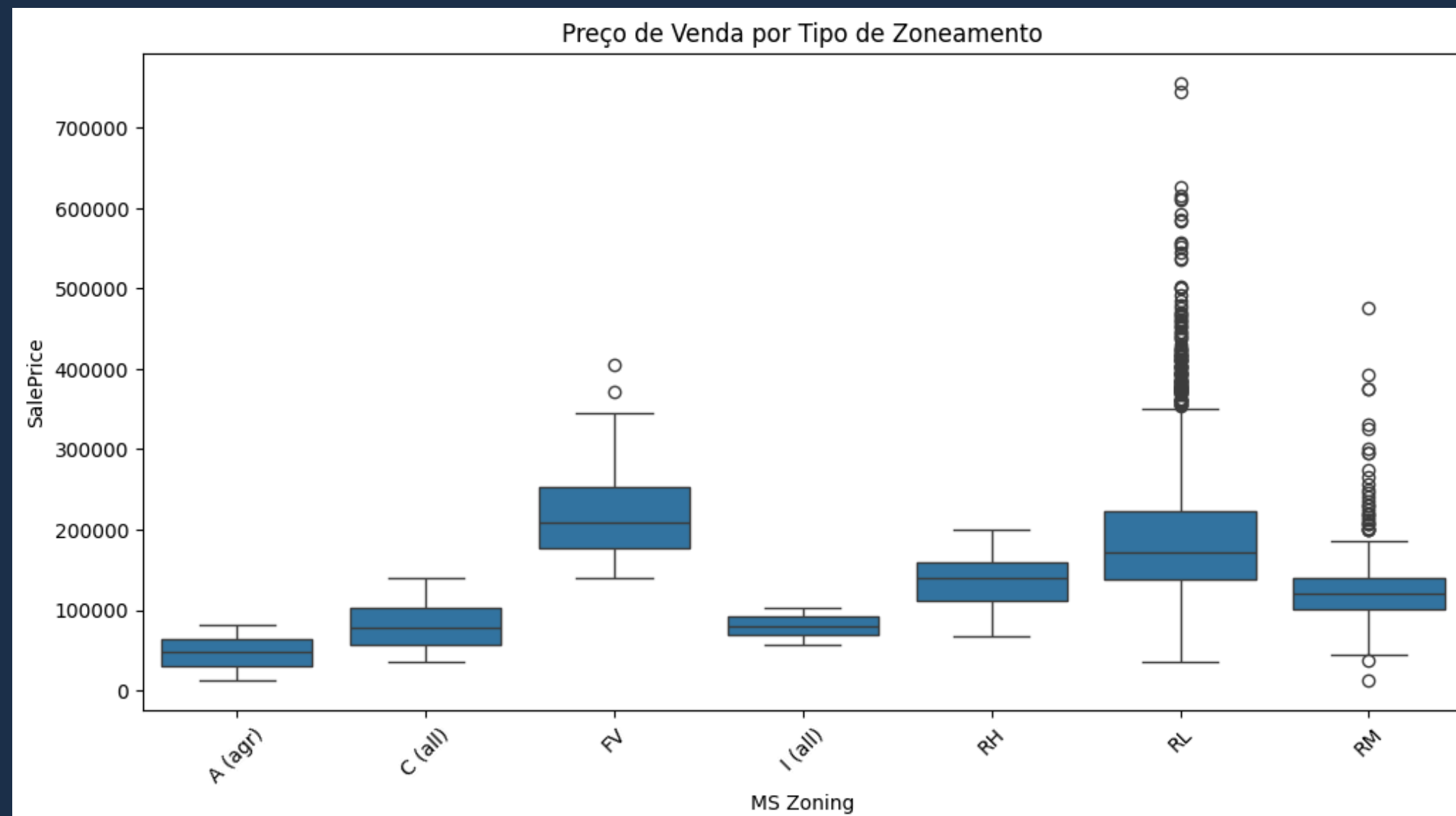
Há um pico claro em torno de \$100,000 a \$200,000, que é onde a maior parte dos imóveis se concentra. Isso sugere que essa faixa de preço é a mais comum para as propriedades neste conjunto de dados.

A linha azul suave sobreposta (KDE - Kernel Density Estimate) ajuda a visualizar a densidade dos dados, mostrando que a frequência de preços diminui gradualmente à medida que você se afasta do pico. Isso confirma a assimetria e indica que, embora haja imóveis mais caros, eles são muito menos frequentes.

À medida que os preços aumentam além de \$200,000, a frequência de imóveis vendidos diminui rapidamente. Isso pode indicar que há menos compradores dispostos a pagar preços mais altos ou que esses imóveis são menos comuns.



Boxplot de SalePrice por MS Zoning



O boxplot permite observar a distribuição dos preços de venda dos imóveis para diferentes tipos de zoneamento.

Cada caixa representa os quartis, com a linha no meio representando a mediana.

****RL**** (Residential Low Density): Apresenta a maior mediana e uma faixa interquartil (IQR) relativamente ampla, indicando uma alta variação de preços e a presença de imóveis de maior valor.

****RM**** (Residential Medium Density): Tem uma mediana mais alta em comparação com outros tipos de zoneamento, o que sugere que as propriedades nessa categoria também têm preços de venda consideráveis.

****RH**** (Residential High Density) e **FV** (Floating Village): Mostram preços de venda mais baixos em comparação com RL e RM, mas ainda assim têm uma faixa de preços variados.

****C**** (Commercial) e **I** (Industrial): Mostram uma faixa de preços que é geralmente mais baixa, e os boxplots têm menos variabilidade, indicando preços mais consistentes.

****A**** (Agricultural): Tem a mediana mais baixa entre todas as categorias, sugerindo que os preços dos imóveis nesta categoria são significativamente mais baixos.

Portanto a variação nos preços de venda de acordo com o zoneamento sugere que o tipo de zoneamento tem um impacto significativo sobre o preço dos imóveis.

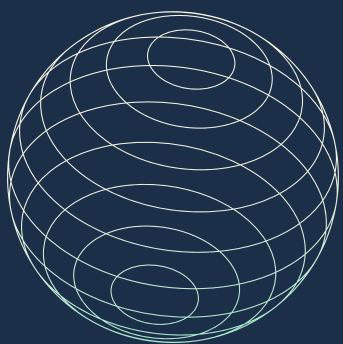
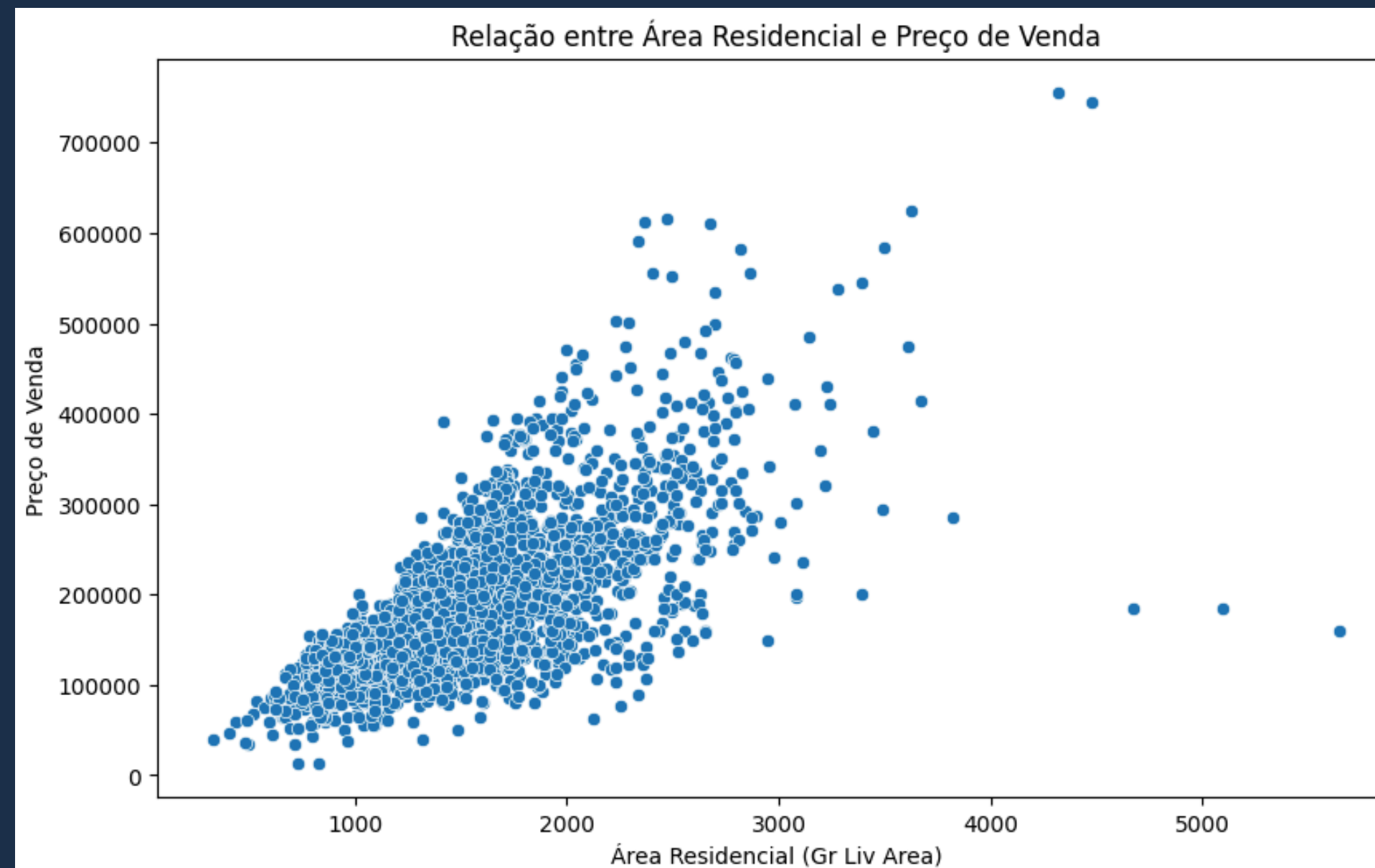
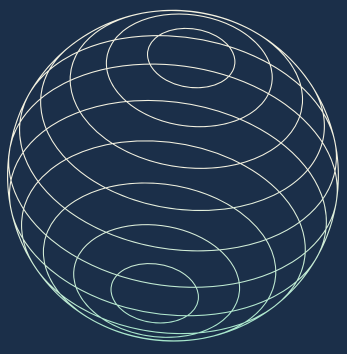


Gráfico de dispersão entre Gr Liv Area e SalePrice

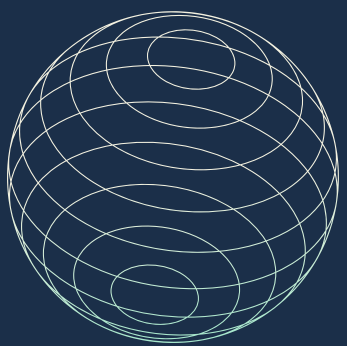




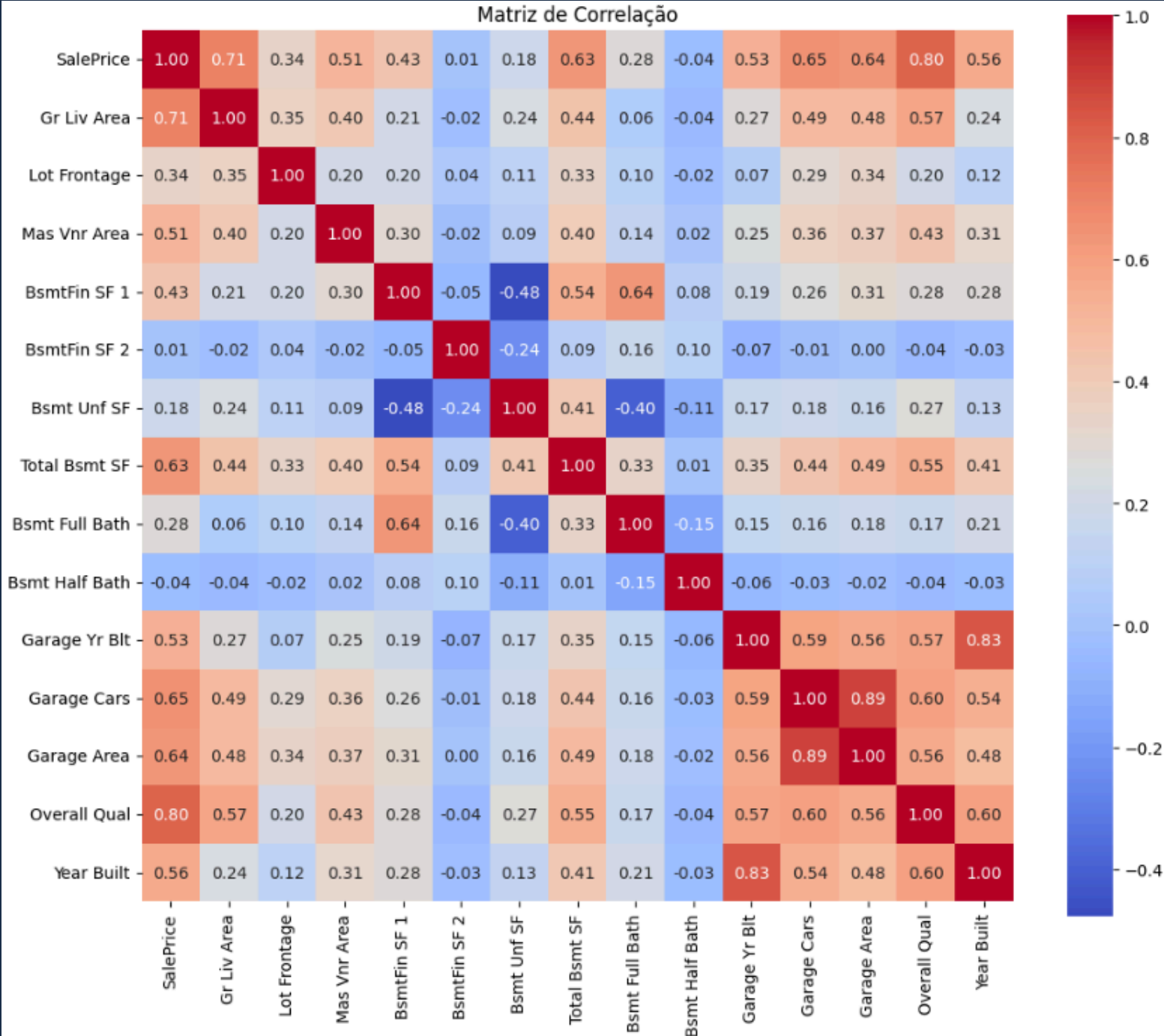
Existe uma clara correlação positiva entre a área residencial e o preço de venda. À medida que a área aumenta, o preço de venda tende a aumentar também. Isso sugere que, em geral, casas maiores são vendidas por preços mais elevados.

Apesar da dispersão dos pontos, há uma tendência linear razoavelmente forte no aumento do preço com o aumento da área. Isso indica que o Gr Liv Area é um bom preditor para o preço de venda.

A maior concentração de pontos está entre 1000 e 2500 de Gr Liv Area e de 100.000 a 300.000 de SalePrice. Isso mostra que a maioria das casas analisadas tem essas características.



Mapa de calor da matriz de correlação



****Garage Area e Garage Cars: 0,89.**** -

Isso indica que a área total da garagem (Garage Area) e o número de carros que a garagem comporta (Garage Cars) têm uma correlação positiva muito forte.

****Garage Yr Blt e Garage Cars / Garage Area: 0,59 e 0,56, respectivamente.**** -

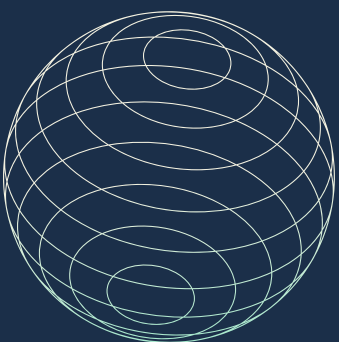
Isso sugere que a idade da garagem tem uma relação moderada com o tamanho e a capacidade da garagem.

****Total Bsmt SF e BsmtFin SF 1: 0,54.**** -

A área total do porão e a área finalizada do porão (tipo 1) também apresentam uma forte correlação positiva.

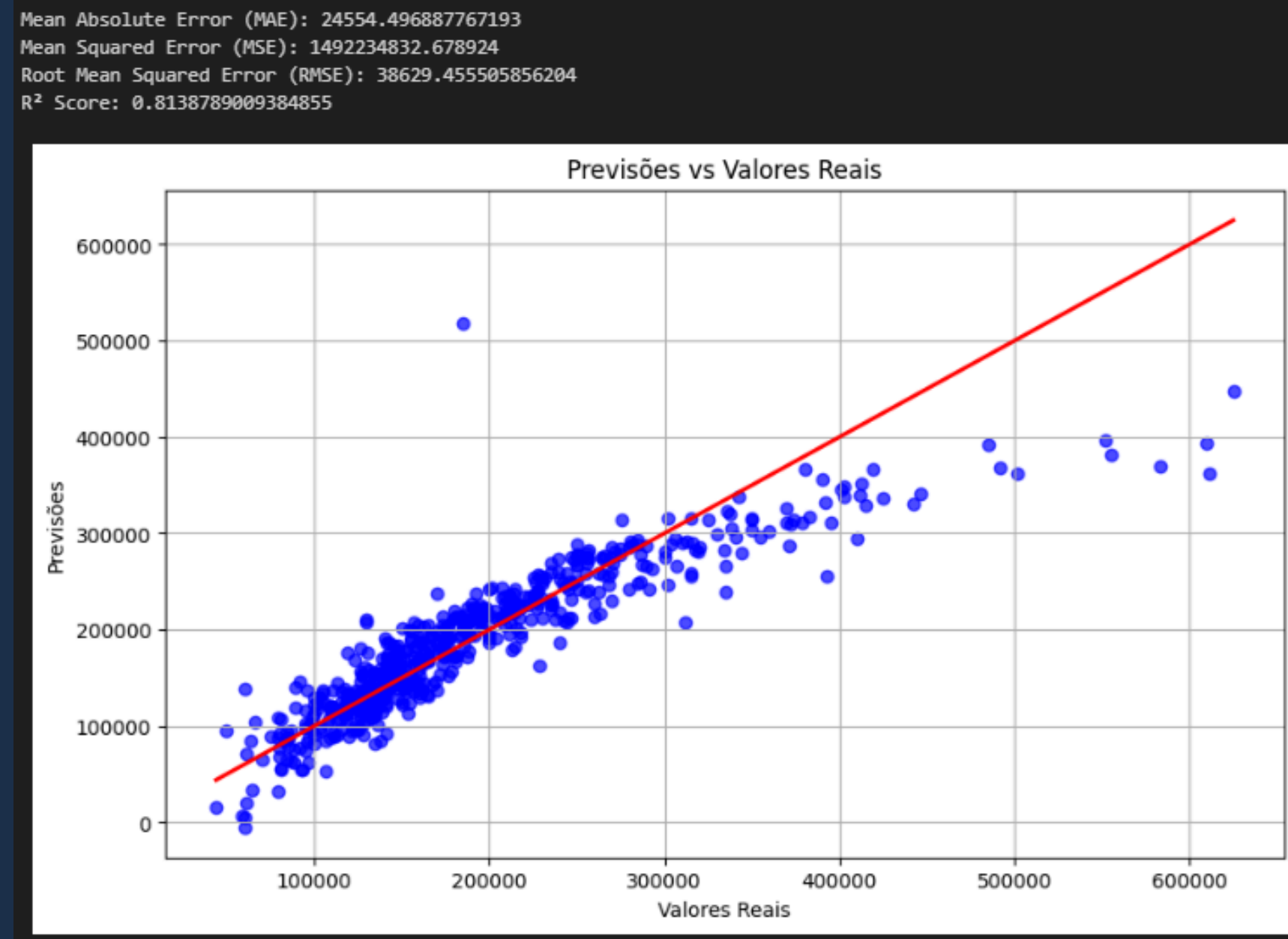
****BsmtFin SF 1 e Bsmt Unf SF: -0,48.**** -

Isso significa que se a área finalizada (tipo 1) aumenta, a área não finalizada tende a diminuir. Faz sentido, pois o aumento na área construída reduz a parte inacabada.

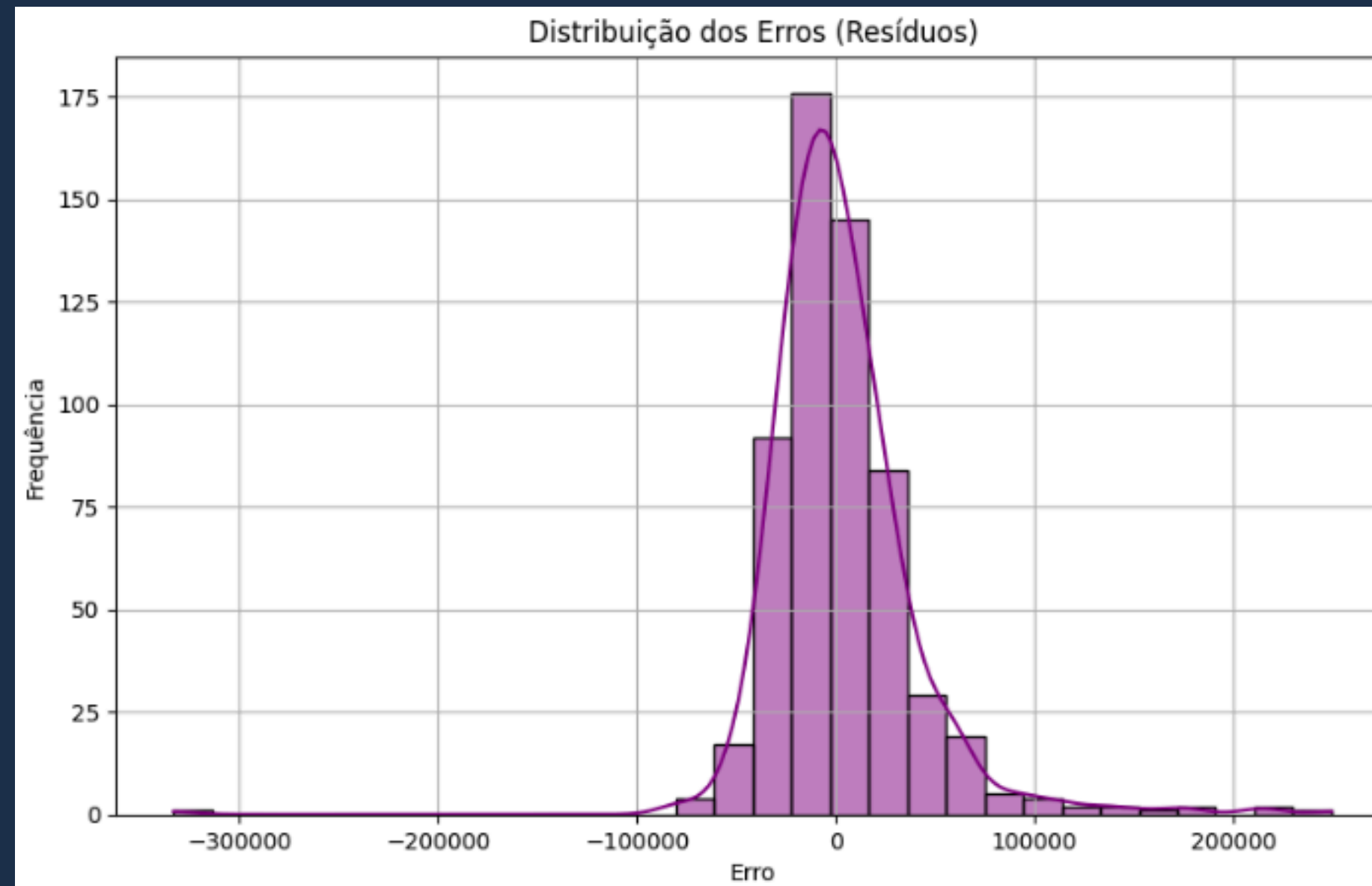


Modelagem Previsional com visualização e métricas

Visualização: Gráfico de dispersão das previsões VS valores reais



Visualização: Distribuição dos erros (resíduos)



Principais Métricas de Avaliação:

- * Mean Absolute Error (MAE): 29.169 — A média dos erros absolutos indica que, em média, as previsões do modelo estão errando por cerca de 29.169 unidades monetárias.
- * Root Mean Squared Error (RMSE): 45.345 — A raiz do erro médio quadrático nos dá uma métrica útil, que representa o erro médio em termos das unidades do preço (ex. dólares), ajudando na interpretação do quão longe estão as previsões.
- * R^2 Score: 0.743 — O modelo conseguiu explicar aproximadamente 74,3% da variabilidade nos preços dos imóveis, o que mostra uma boa capacidade preditiva.

Visualizações:

- * O Gráfico de Previsões vs. Valores Reais mostrou que as previsões do modelo seguem uma boa relação linear com os valores reais, ainda que com alguns outliers que podem ser investigados ou tratados.
- * A Distribuição dos Erros (Resíduos) indicou uma distribuição aproximadamente normal, o que é um bom sinal de que o modelo está ajustado corretamente aos dados.

