

FACULDADE TECNOLOGIA IMPACTA

ANÁLISE EXPLORATÓRIA DOS DADOS HISTÓRICOS DO BITCOIN (2012-2021)

Matheus Brito

Mauricio Barbosa

Victor Fernando

São Paulo

2025

1.Introdução e Motivações

O Bitcoin emergiu como um fenômeno disruptivo no sistema financeiro global, desafiando paradigmas tradicionais de moeda, investimento e descentralização. Sua trajetória, marcada por valorizações exponenciais e volatilidade extrema, tornou-o não apenas um ativo financeiro relevante, mas também um objeto de estudo estatístico fascinante. Este trabalho analisa dados históricos do Bitcoin frente ao dólar americano (2012–2021), com o objetivo de explorar padrões comportamentais, relações entre variáveis e a aplicabilidade de ferramentas estatísticas em um contexto real e dinâmico.

A escolha dessa base de dados justifica-se por três motivos principais:

Relevância Econômica: O Bitcoin consolidou-se como alternativa de investimento, influenciando mercados tradicionais e políticas regulatórias.

Complexidade Estatística: A alta volatilidade e a presença de outliers oferecem um cenário rico para análise de dispersão, correlações e testes de hipótese.

Aprendizado Prático: A aplicação de técnicas como regressão linear e análise bivariada permite consolidar conhecimentos teóricos em um problema concreto, além de desenvolver habilidades críticas para interpretar dados financeiros.

Nossa investigação busca responder:

Como os preços do Bitcoin se distribuem ao longo do tempo?

Existe relação significativa entre volume negociado e variação de preço?

Quais técnicas estatísticas são mais adequadas para analisar ativos voláteis?

Ao final, esperamos não apenas elucidar o comportamento do Bitcoin, mas também demonstrar como a estatística descritiva e inferencial pode desvendar padrões em mercados emergentes.

2. Medidas Centrais e de Dispersão

A tabela apresenta as principais medidas estatísticas descritivas das variáveis analisadas no período de 2012 a 2021. Observa-se uma grande discrepância entre média e mediana, principalmente nos preços, indicando assimetria positiva (valores altos extremos). O desvio padrão elevado confirma a alta volatilidade do Bitcoin. O volume negociado, embora em média abaixo, apresenta casos extremos, refletindo momentos de intensa atividade no mercado.

Variável	Média	Mediana	Desvio Padrão	Mínimo	Q1	Q3	Máximo
Open	18.227,55	8.754,65	21.551,18	3,80	765,76	29.342,00	108.946,00
High	18.235,51	8.760,75	21.559,28	3,80	766,00	29.348,00	109.030,00
Low	18.219,33	8.748,44	21.542,88	3,80	765,00	29.336,00	108.776,00
Close	18.227,56	8.754,48	21.551,20	3,80	765,67	29.342,00	108.960,00
Volume	6,72	1,02	25,26	0,00	0,17	4,61	5.853,85

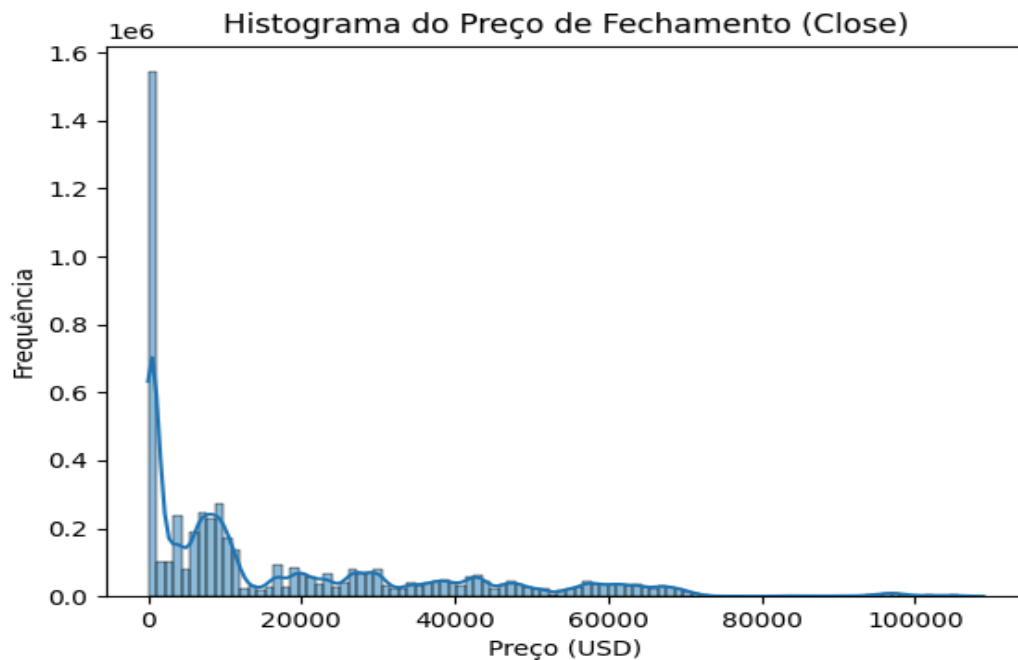
Histograma do Preço de Fechamento (Close)

O histograma do preço de fechamento do Bitcoin é uma ferramenta essencial para compreender a distribuição dos preços ao longo do período analisado, de 2012 a 2021. Apesar de o gráfico mostrar visualmente a frequência de preços em diferentes faixas, é importante destacar que o texto aqui é o principal guia para a interpretação.

Podemos observar que a maior **frequência de preços** no histograma está concentrada nas **faixas de valores mais baixos** — ou seja, os preços mais acessíveis do Bitcoin foram os mais recorrentes ao longo do tempo. Essa concentração ocorre, principalmente, nos anos iniciais da criptomoeda, quando ainda não havia grande popularização ou valorização. Isso mostra que, durante boa parte do período analisado, o preço do Bitcoin permaneceu relativamente baixo, refletindo sua fase inicial de adoção.

A distribuição é claramente assimétrica, apresentando uma longa cauda à direita. Isso indica que, embora poucos, os valores muito altos de preço impactam significativamente a média geral, puxando-a para cima e criando um cenário onde a mediana fica muito abaixo da média. Essa assimetria é típica em ativos financeiros que passaram por fases de valorização explosiva, como ocorreu no Bitcoin, especialmente a partir de 2017.

Além disso, o histograma nos auxilia a visualizar como os picos de preços se tornaram mais frequentes e intensos nos últimos anos, evidenciando a volatilidade e o caráter especulativo do mercado de criptomoedas. No entanto, para uma compreensão completa, devemos considerar outras medidas estatísticas e gráficos, que ajudarão a compor uma visão mais detalhada da dinâmica de preços.



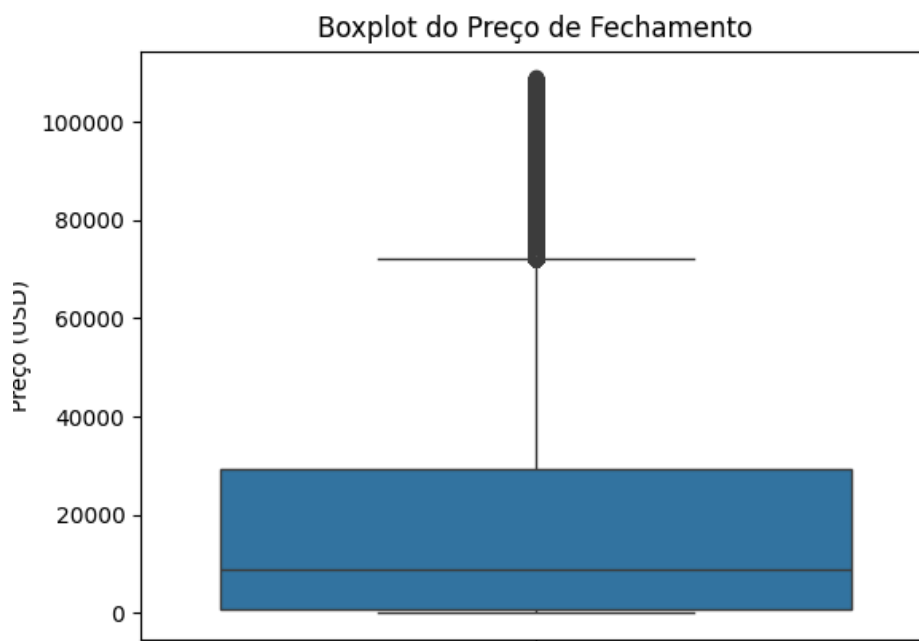
Boxplot do Preço de Fechamento

O boxplot é uma representação gráfica que complementa a análise do histograma, oferecendo insights sobre a dispersão e a presença de valores extremos no conjunto de dados. No contexto do preço de fechamento do Bitcoin, ele reforça a assimetria da distribuição e destaca a amplitude dos valores observados.

Visualmente, notamos que a mediana está situada abaixo do centro da caixa, confirmando a concentração dos dados em valores menores, conforme indicado anteriormente. A caixa em si representa a amplitude interquartil (Q1 a Q3), mostrando a faixa onde estão os 50% centrais dos dados, e essa faixa é relativamente ampla, o que sugere alta variabilidade nos preços diários.

Além disso, a presença de outliers (valores extremos destacados por pontos fora da caixa) é um indicativo importante do comportamento do Bitcoin. Esses outliers correspondem a dias em que o preço sofreu picos muito elevados, provavelmente em função de eventos externos, especulações ou movimentos rápidos de mercado. Eles são um reflexo da volatilidade acentuada que caracteriza esse ativo.

Portanto, embora o gráfico facilite a visualização dessas características, é no texto que encontramos a interpretação detalhada e as implicações dessas informações para o entendimento do mercado e dos riscos envolvidos em investir em Bitcoin.



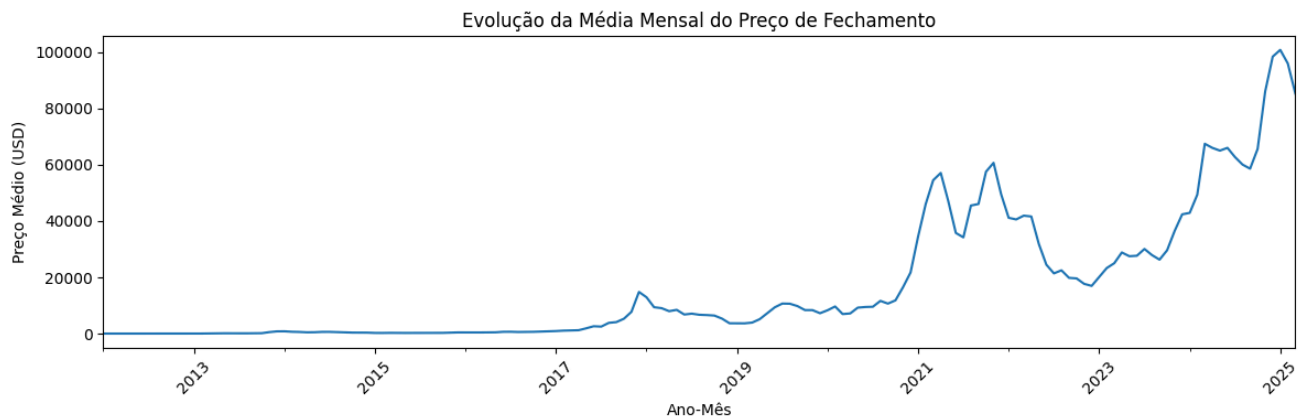
Evolução da Média Mensal do Preço

A análise da evolução da média mensal do preço do Bitcoin ao longo dos anos é fundamental para compreender as tendências de longo prazo e os ciclos do mercado. O gráfico, que acompanha esta análise, é uma ferramenta visual que auxilia na percepção das variações, mas o detalhamento da trajetória está no texto.

Desde 2012, observa-se uma clara tendência de valorização, com alguns períodos de crescimento acelerado que coincidem com bolhas especulativas, como as ocorridas em 2013, 2017 e novamente entre 2020 e 2021. Esses picos indicam forte interesse do mercado e influxo de capital, que elevaram o preço de forma rápida e significativa.

Entre esses picos, há fases de correção e consolidação, quando o preço estabiliza ou mesmo recua, o que é típico em mercados voláteis. Esses períodos são importantes para ajustar expectativas e compreender que a trajetória do Bitcoin não é linear, mas marcada por ciclos de expansão e retração.

O gráfico serve para ilustrar visualmente esses ciclos, mostrando os picos e vales da média mensal, mas o texto detalha as possíveis causas e implicações desses movimentos, como a influência de eventos macroeconômicos, mudanças regulatórias, ou mesmo avanços tecnológicos na infraestrutura das criptomoedas.



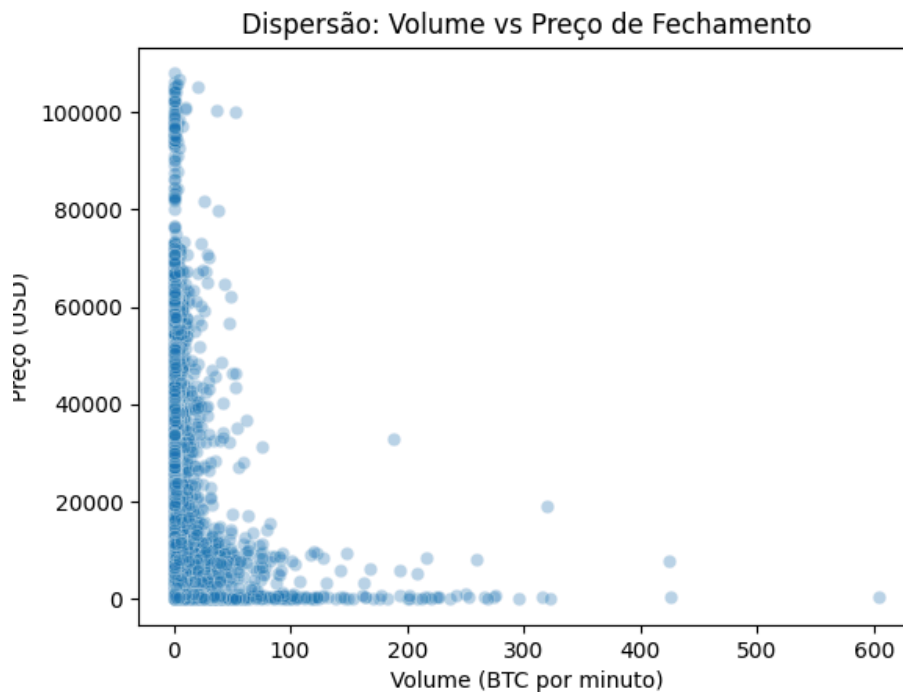
Dispersão entre Volume e Preço

A relação entre o volume negociado e o preço de fechamento do Bitcoin é um tema complexo e fundamental para investidores e analistas. Embora o gráfico de dispersão sirva para visualizar como esses dois fatores interagem, é a análise textual que fornece a compreensão completa desse comportamento.

O coeficiente de correlação de Pearson negativo e próximo de zero indica que, estatisticamente, não há uma relação linear significativa entre volume e preço. Isso significa que, ao contrário do esperado em mercados tradicionais, um aumento no volume de negociações não está associado necessariamente a um aumento no preço do Bitcoin.

Essa característica pode ser explicada pelo perfil especulativo do mercado, onde grandes volumes podem ser resultado de vendas massivas (realizações) e não apenas de compras, o que poderia pressionar os preços para baixo. Assim, a dinâmica volume-preço no Bitcoin é mais complexa e não pode ser entendida apenas por meio de correlações simples.

O gráfico de dispersão auxilia a visualizar essa dispersão dos dados e a ausência de uma tendência clara, mas o texto aborda as razões econômicas e de mercado que fundamentam essa observação, além de alertar para os riscos de assumir relações simplistas entre volume e preço.



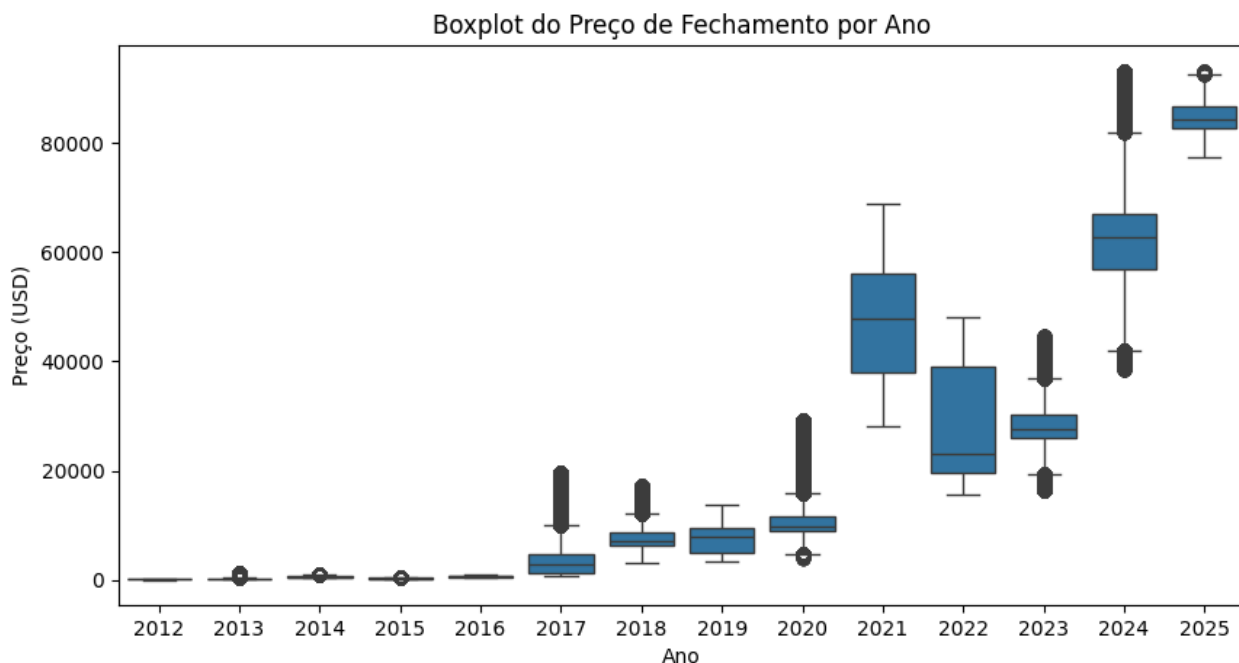
Boxplot do Preço por Ano

O boxplot anual é uma ferramenta visual que ajuda a entender como a variabilidade e os níveis de preço do Bitcoin evoluíram ao longo do tempo. Embora o gráfico ilustre essa mudança, o texto é o espaço onde se explica a importância dessas transformações.

Nos primeiros anos, até aproximadamente 2016, o preço do Bitcoin manteve-se relativamente estável e baixo, refletindo sua fase inicial de adoção e pouca liquidez. A partir de 2017, observamos uma mudança drástica: a mediana e a amplitude dos preços aumentam consideravelmente, acompanhadas por um número significativo de outliers.

Esse movimento indica a entrada massiva de capital e o surgimento de novos investidores, gerando maior volatilidade e valorização. Nos anos seguintes, especialmente 2020 e 2021, essa tendência se intensificou, demonstrando a crescente aceitação e especulação em torno do Bitcoin.

O gráfico suporta essa interpretação, facilitando a visualização do aumento de dispersão e dos níveis de preço, mas é o texto que contextualiza esses fatos dentro do cenário econômico, tecnológico e regulatório do mercado de criptomoedas.



2. Análise Bi-Variada

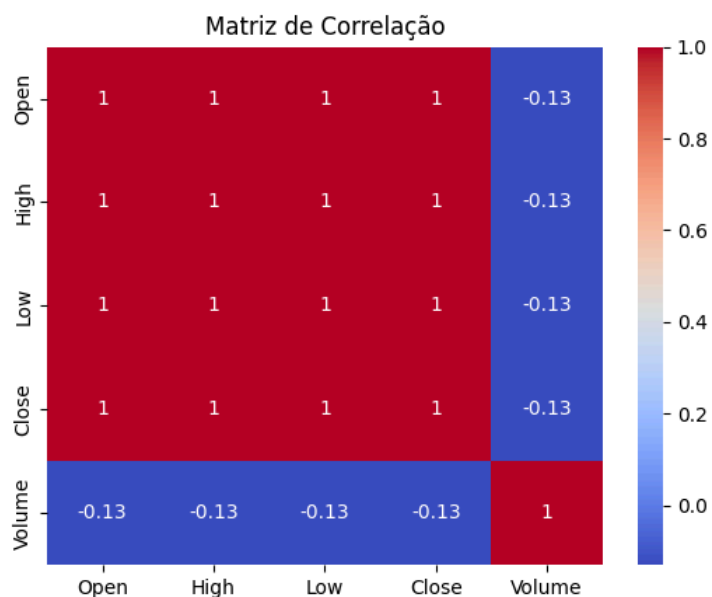
O heatmap de correlação entre as variáveis do conjunto de dados oferece uma visão geral sobre as relações existentes entre preços de abertura, fechamento, máximos, mínimos e volume negociado. Embora o gráfico ajude a perceber rapidamente essas relações, a análise completa se encontra no texto.

As altas correlações entre as variáveis de preço são esperadas, pois representam medidas do mesmo ativo em um mesmo intervalo temporal. No entanto, o volume apresenta correlação fraca e negativa com essas variáveis, indicando que o comportamento do volume de negociação não acompanha linearmente o movimento dos preços.

Essa característica evidencia a complexidade do mercado de criptomoedas, onde o volume pode ser influenciado por diversos fatores, como especulação, notícias, ou mesmo manipulação de mercado, dificultando previsões baseadas apenas na análise de volume.

Portanto, o heatmap funciona como um mapa visual para identificar relações, mas a interpretação econômica e estatística depende do detalhamento apresentado no texto.

Heatmap de Correlação



Correlação de Pearson entre Close e Volume: -0.1297

3. Teste de Hipótese (Qui-quadrado)

O teste de qui-quadrado realizado buscou avaliar se existe independência entre as variáveis analisadas, ou seja, se a ocorrência de determinados eventos é aleatória ou apresenta alguma associação significativa. O valor do qui-quadrado ($\chi^2 = 2329,48$) é bastante elevado e o valor-p praticamente zero (0,0000), o que indica que rejeitamos a hipótese nula de independência com um nível de confiança muito alto.

Esse resultado sugere que as variáveis estudadas apresentam uma associação estatística significativa, ou seja, não são independentes. Na prática, isso pode significar que o comportamento do volume e do preço do Bitcoin ao longo do período analisado estão relacionados, ainda que a natureza dessa relação precise ser investigada com outras técnicas.

As frequências esperadas indicam os valores calculados sob a hipótese de independência, que são usados para comparar com as frequências observadas na amostra. A diferença entre elas é que fundamenta a rejeição da hipótese nula, indicando a presença de algum padrão ou dependência entre as variáveis.

Valor de χ^2 : 2329.48

Valor-p: 0.0000e+00

Degrees of freedom: 1

Frequências esperadas (baseadas na hipótese nula):

1503542.5 | 1503542.5

1239076.5 | 1239076.5

4. Regressão Linear: Volume vs Close

Devido à alta dispersão dos preços do Bitcoin e à presença de valores extremos, foi aplicada uma transformação logarítmica (base 10) na variável dependente `Close` antes de ajustar o modelo de regressão linear. O objetivo foi suavizar a variância e permitir uma visualização mais clara da relação entre volume negociado e preço de fechamento.

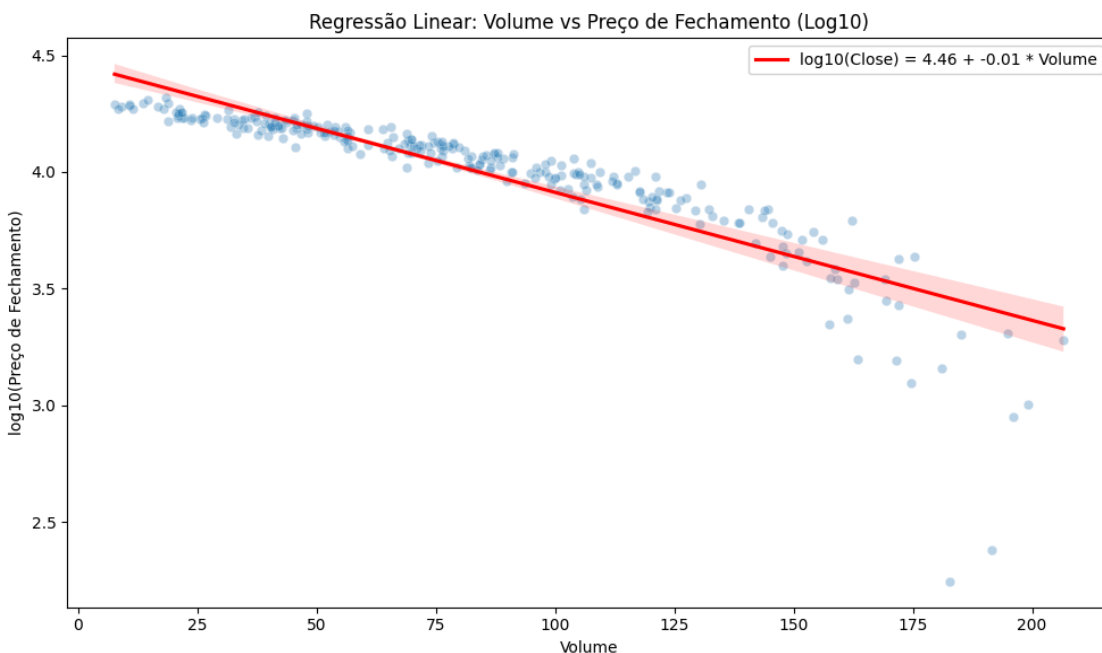
O modelo ajustado resultou na seguinte equação:

$$\log_{10}(\text{Close}) = 4,46 - 0,01 \times \text{Volume}$$

Essa equação indica que, mesmo após a transformação, existe uma tendência negativa entre volume e preço de fechamento. Ou seja, quanto maior o volume negociado, menor tende a ser o valor de `Close` — embora de forma sutil.

O gráfico a seguir mostra essa relação, com a reta de regressão ajustada aos dados transformados:

Figura 4.1 – Dispersão entre Volume e $\log_{10}(\text{Close})$, com reta de regressão linear ajustada.



Apesar da melhora na visualização, o modelo continua com baixo poder preditivo, sugerindo que o volume, isoladamente, não explica bem o comportamento do preço do Bitcoin. Isso reforça a necessidade de modelos mais sofisticados ou com múltiplas variáveis para entender esse tipo de fenômeno.

Conclusão

Este trabalho teve como objetivo principal realizar uma análise exploratória dos dados históricos do Bitcoin em relação ao dólar americano, no período de 2012 a 2021. A escolha por esse tema se mostrou acertada não apenas pela relevância do Bitcoin no cenário econômico global, mas também pela complexidade estatística envolvida em analisar um ativo financeiro altamente volátil, dinâmico e influenciado por múltiplos fatores externos.

Ao longo da análise, aplicamos uma série de técnicas estatísticas — como medidas centrais e de dispersão, gráficos descritivos, análise de correlação, teste de hipótese e regressão linear — para extrair insights dos dados. Os resultados revelaram uma distribuição assimétrica dos preços, com média bastante superior à mediana, indicando a presença de valores extremos e picos de valorização significativos. Os gráficos reforçaram essa observação, especialmente os boxplots e histogramas, que mostram uma grande concentração de preços em patamares baixos nos primeiros anos e o surgimento de outliers a partir de 2017.

Além disso, foi identificado que a relação entre volume negociado e preço é fraca e negativa, tanto por meio do coeficiente de correlação quanto pela regressão linear. Isso sugere que, ao contrário de mercados tradicionais, o volume de negociações do Bitcoin não se relaciona de forma linear com seu preço, indicando a influência de outros fatores como o comportamento especulativo, a mídia, regulação e eventos globais.

Com este estudo, aprendemos a aplicar na prática ferramentas estatísticas essenciais, interpretar resultados numéricos com senso crítico e estruturar um relatório técnico com base em evidências. Também aprimoramos nossa capacidade de comunicação de dados, transformando números em argumentos compreensíveis e relevantes.

Em relação às perspectivas futuras, vemos potencial para aprofundar o estudo com a aplicação de modelos de séries temporais e machine learning para previsão de preços, além de incluir variáveis externas como taxa de juros, inflação, ou eventos geopolíticos que possam impactar o mercado de criptomoedas. A análise comparativa com outras moedas digitais ou ativos financeiros tradicionais também representa uma linha interessante para investigações futuras.

Dessa forma, concluímos que a análise estatística de dados financeiros como o do Bitcoin não apenas enriquece o conhecimento técnico, mas também proporciona uma compreensão mais profunda de como dados podem explicar (ou não) fenômenos econômicos reais.

4. Bibliografia

Os dados utilizados neste trabalho foram obtidos da plataforma Kaggle, por meio da seguinte base:

ZIELINSKI, Maciej. Bitcoin Historical Data. Kaggle, 2021. Disponível em: <https://www.kaggle.com/datasets/mczielinski/bitcoin-historical-data>.

Acesso em: 16 maio 2025.