

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Thalisson Flores Forte

**ANÁLISE DE ASSOCIAÇÃO DE SÉRIES TEMPORAIS EM DADOS DE
CRIPTOMOEDAS, REDES SOCIAIS E GOOGLE TRENDS**

Santa Maria, RS
2022

Thalisson Flores Forte

**ANÁLISE DE ASSOCIAÇÃO DE SÉRIES TEMPORAIS EM DADOS DE
CRIPTOMOEDAS, REDES SOCIAIS E GOOGLE TRENDS**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Ciência da Computação**.

ORIENTADOR: Prof. Joaquim Vinicius Carvalho Assunção

503
Santa Maria, RS
2022

Thalisson Flores Forte

**ANÁLISE DE ASSOCIAÇÃO DE SÉRIES TEMPORAIS EM DADOS DE
CRIPTOMOEDAS, REDES SOCIAIS E GOOGLE TRENDS**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Ciência da Computação**.

DocuSigned by:

Joaquim Vinicius Carvalho Assunção

A4EAD45E45C041F...

Joaquim Vinicius Carvalho Assunção, Dr. (UFSM)
(Presidente/Orientador)

DocuSigned by:

Giovanni Rubert Librelotto

6D09C08B859B499...

Giovani Rubert Librelotto, Dr. (UFSM)

DocuSigned by:

Renata Rojas Guerra

4D85EBBE48BA454...

Renata Rojas Guerra, Dr. (UFSM)

Santa Maria, RS
2022

RESUMO

ANÁLISE DE ASSOCIAÇÃO DE SÉRIES TEMPORAIS EM DADOS DE CRIPTOMOEDAS, REDES SOCIAIS E GOOGLE TRENDS

AUTOR: Thalisson Flores Forte

ORIENTADOR: Joaquim Vinicius Carvalho Assunção

Através da introdução de novas soluções tecnológicas, permitiu-se a evolução e a democratização dos mecanismos financeiros nas últimas décadas, possibilitando também, a existência de novos tipos de ativos digitais. Paralelamente, percebeu-se uma enorme disseminação das moedas digitais, tendo sua visibilidade potencializada nas redes sociais. Assim, tendo percebido a crescente quantidade de buscas e postagens sobre tais moedas, este trabalho propõe a realização de um estudo do mercado de criptomoedas através de associações de séries temporais e comparação de técnicas, utilizando variáveis como o valor de fechamento das criptomoedas, volume de tweets e volume de pesquisas. A partir da extração e manipulação desses dados, espera-se observar possíveis associações e tendências do mercado.

Palavras-chave: Análise de séries temporais. Criptomoedas. Google Trends. Tendências de mercado.

ABSTRACT

TIME SERIES ASSOCIATION ANALYSIS IN CRYPTOCURRENCY DATA, SOCIAL NETWORKS AND GOOGLE TRENDS

AUTHOR: Thalisson Flores Forte

ADVISOR: Joaquim Vinicius Carvalho Assunção

Through the introduction of new technological solutions, the evolution and democratization of financial mechanisms in recent decades were allowed, also enabling the existence of new types of digital assets. At the same time, there was a huge spread of digital currencies, with their visibility enhanced on social networks. Thus, having noticed the increasing amount of searches and posts about such coins, this work proposes to carry out a study of the cryptocurrency market through associations of time series and comparison of techniques, using variables such as the closing value of cryptocurrencies, tweet volume, and search volume. From the extraction and manipulation of these data, it is expected to observe possible associations and market trends.

Keywords: Time series analysis. Cryptocurrencies. Google Trends. Market Trends.

LISTA DE FIGURAS

Figura 2.1 – Funcionamento da <i>blockchain</i> para transações com criptomoedas	14
Figura 2.2 – Diferença entre Distância Euclidiana e <i>Dynamic Time Warping</i> (DTW). ..	16
Figura 2.3 – Exemplo da matriz de distância que calcula o <i>Dynamic Time Warping</i> ...	17
Figura 2.4 – Exemplo do algoritmo de <i>Dynamic Time Warping</i>	18
Figura 2.5 – Exemplo da aplicação do algoritmo de DTW sem e com a presença do <i>Symbolic Aggregate Approximation</i> , respectivamente.	19
Figura 2.6 – Exemplo do resultado da <i>Symbolic Aggregate Approximation</i> (SAX).	20
Figura 4.1 – Fluxo de manipulação dos dados	27
Figura 4.2 – Testes de diferentes configurações de SAX + DTW	28
Figura 4.3 – Aplicação das técnicas	29
Figura 5.1 – Volume de Tweets: Quantidade de Tweets por dia durante todo o intervalo	30
Figura 5.2 – Volume de Pesquisa: <i>Search Volume Index</i> por dia durante todo o intervalo	31
Figura 5.3 – Bitcoin no evento B: visualização geral dos dados	54

LISTA DE ABREVIATURAS E SIGLAS

<i>ADA</i>	Cardano
<i>BTC</i>	Bitcoin
<i>ETH</i>	Ethereum
<i>LTC</i>	Litecoin
<i>XRP</i>	Ripple
<i>ETF</i>	Exchange Traded Fund
<i>FI</i>	Fundo de investimento
<i>SVI</i>	Search Volume Index
<i>SAX</i>	Symbolic Aggregate approximation
<i>DTW</i>	Dynamic Time Warping
<i>P2P</i>	Peer-to-peer

SUMÁRIO

1	INTRODUÇÃO	9
1.1	JUSTIFICATIVA	10
1.2	OBJETIVOS	10
1.2.1	Objetivo Geral	10
1.2.2	Objetivos Específicos	11
1.3	ORGANIZAÇÃO DO TEXTO	11
2	REFERENCIAL TEÓRICO	12
2.1	INVESTIMENTOS DIGITAIS	12
2.1.1	Criptomoeda	12
2.1.2	Exchanges	13
2.2	SOLUÇÕES COMPUTACIONAIS	13
2.2.1	Blockchain	13
2.3	REPRESENTAÇÃO, DISTÂNCIA E CORRELAÇÃO EM SÉRIES TEMPORAIS	14
2.3.1	Correlação de Pearson	14
2.3.2	Distância Euclidiana	15
2.3.3	Dynamic Time Warping	16
2.3.4	Symbolic Aggregate Approximation	19
3	TRABALHOS RELACIONADOS	21
4	METODOLOGIA E DESENVOLVIMENTO DA PROPOSTA	23
4.1	DEFINIÇÃO DO INTERVALO DE TEMPO E EVENTOS RELEVANTES	24
4.2	SELEÇÃO, EXTRAÇÃO E MANIPULAÇÃO DE DADOS	25
4.2.1	Valores de fechamento das moedas	25
4.2.2	Volume de pesquisas	26
4.2.3	Volume de Tweets	26
4.2.4	Manipulação dos dados	27
4.3	CONFIGURAÇÕES DO SYMBOLIC AGREGGATE APPROXIMATION	28
4.4	APLICAÇÃO DAS TÉCNICAS	28
5	RESULTADOS E DISCUSSÃO	30
5.1	ANÁLISE DO VOLUME DE DADOS	30
5.2	ANÁLISE DESCRITIVA DOS DADOS	32
5.2.1	Evento A: Proibição de criptomoedas na China	32
5.2.2	Evento B: Bitcoin como moeda oficial em El Salvador	35
5.2.3	Evento C: Máxima histórica do Bitcoin	38
5.2.4	Evento D: Intervalo aleatório	40
5.2.5	Evento E: Intervalo Aleatório	43
5.2.6	Evento F: Atualização da rede Cardano	45
5.3	INTERPRETAÇÃO DOS RESULTADOS	47
5.3.1	Interpretações da análise entre valor da moeda e <i>Search Volume Index</i>	47
5.3.1.1	<i>Medida de correlação</i>	47
5.3.1.2	<i>Medida de distância</i>	48
5.3.2	Interpretações da análise entre valor da moeda e volume de tweets	49
5.3.2.1	<i>Medida de correlação</i>	49
5.3.2.2	<i>Medida de distância</i>	50
5.3.3	Interpretações da análise entre <i>Search Volume Index</i> e volume de tweets	51
5.3.3.1	<i>Medida de correlação</i>	51

5.3.3.2	<i>Medida de distância</i>	52
5.3.4	Análise da melhor técnica	53
5.3.5	Análise da influência dos eventos relevantes nos resultados	55
6	CONCLUSÃO	56
	REFERÊNCIAS BIBLIOGRÁFICAS	57

1 INTRODUÇÃO

Juntamente com o avanço da tecnologia nas últimas décadas surgiram novas percepções do mercado financeiro. A partir da inclusão dos conhecimentos de *blockchain*, foi possível a criação de um sistema de dinheiro eletrônico ponto-a-ponto, o Bitcoin (NAKAMOTO, 2008), que de forma descentralizada movimenta atualmente bilhões de dólares por dia. De acordo com o site *CoinMarketCap*, existem mais de 13.000 outras criptomoedas gerando, ao todo, uma capitalização de mercado de mais de 16 trilhões de reais, sendo que o Bitcoin é responsável por cerca de 7 trilhões desse total. (CoinMarketCap, 2021)

Paralelamente, ao longo das últimas décadas, a internet se fez cada vez mais presente em todas as parcelas da população mundial. Atualmente, as redes sociais permitem que pessoas compartilhem seus desejos e sentimentos de forma fácil e prática, mantendo o usuário por longos períodos de tempo através da utilização de mecanismos de recompensa para as interações realizadas na rede. Dessa forma, gera-se uma massiva quantidade de dados diariamente, tornando possível a análise e descoberta de tendências e padrões, utilizados posteriormente, para auxiliar nos processos de tomada de decisão. Outro fator importante é o vínculo dos conhecimentos computacionais e estatísticos. Atualmente são diversas as aplicações que unem ambas as áreas, principalmente ao tratar sobre ciência de dados (HAND, 2015), uma área ampla que permite a extração de conhecimento a partir de dados.

Este trabalho propõe, então, o estudo do mercado de criptomoedas a partir de medidas de distância e correlações entre dados, além de comparações entre diferentes abordagens de análise de séries temporais. A análise de medidas de distância baseia-se na hipótese de que essa pode ser uma métrica melhor que a correlação na detecção de eventos relevantes. Partindo do fato que as séries possuem níveis diferentes, um evento pode engatilhar uma mudança positiva (quando o crescimento da série com maior volume for maior) ou negativa (quando o aumento da série com menor volume for maior) da distância, onde, em ambos os casos, a correlação pode se manter similar e a distância demonstrar diferenças significativas. Além disso, tais mudanças podem ser melhor obtidas utilizando o alinhamento temporal, dado que os eventos podem gerar movimentos com diferentes momentos de início para cada série.

Serão utilizadas como variáveis o valor de fechamento das moedas escolhidas, o volume de tweets extraídos do Twitter e o volume de pesquisa extraído do Google Trends, para que seja feito o estudo de tendências e associações desse mercado.

1.1 JUSTIFICATIVA

A introdução da tecnologia em diversos contextos sociais tem ganhado força na última década (MATA et al., 2018). No cenário econômico, por exemplo, vem possibilitando transações sem a necessidade de uma terceira parte confiável, descentralizando o sistema financeiro através das *criptomoedas e blockchains* (NAKAMOTO, 2008).

Atualmente, com a disponibilidade de carteiras criptografadas e com a implementação de Exchange Traded Funds (ETFs) e Fundos de investimento (FIs) baseados em criptomoedas, este mercado vem tornando-se cada vez mais seguro e acessível aos investidores comuns, permitindo ao mercado um maior potencial de volume de negociação.

De forma concomitante, nos últimos anos, percebeu-se um aumento significativo quanto à utilização da internet e das redes sociais (KEMP, 2021). A rede social Twitter, sendo uma das principais nos dias atuais, possibilita uma rápida interação permitindo que com poucas palavras seus usuários possam expressar suas opiniões, indignações e desejos. Com isso, percebe-se uma massiva quantidade de informações sendo publicadas e pesquisadas diariamente pelos usuários, de modo a, por muitas vezes, expressar novas tendências e sentimentos relativos aos aspectos do mundo real. Além disso, atualmente percebe-se o grande impacto dos chamados "influencers" do mercado financeiro no geral (SIERRA, 2021), o que justifica ainda mais a necessidade de se observar os dados gerados nas redes sociais.

Ademais, o Google, maior site de busca do mundo, oferece ferramentas para a análise das pesquisas realizadas por determinada palavra, possibilitando analisar a popularidade de certos assuntos em determinados momentos. Assim, permite uma extração das tendências como forma de suporte às análises do mercado de criptomoedas.

Dessa forma, torna-se viável o estudo e a aplicação de análises para verificar a associação entre dados relativos ao volume de publicações de usuários do Twitter, popularidade de uma criptomoeda através da quantidade de buscas realizadas e as possíveis tendências do valor da criptomoeda, permitindo a extração de associações existentes entre os diferentes dados.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo deste trabalho é o desenvolvimento de um estudo do mercado de criptomoedas através da análise de associação, correlação e comparação de técnicas, buscando encontrar padrões e tendências através de dados a respeito do volume de tweets e do Se-

arch Volume Index (SVI), além de verificar a influência de eventos externos como gatilhos para movimentos nas séries de dados.

1.2.2 Objetivos Específicos

De forma mais específica, este projeto busca pelo(a):

- Análise de associação e correlação entre os dados;
- Análise da influência de eventos externos nos dados;
- Comparação de diferentes técnicas.

1.3 ORGANIZAÇÃO DO TEXTO

O trabalho está estruturado da seguinte forma: no Capítulo 2 estão descritos os conceitos e as tecnologias empregadas no desenvolvimento do trabalho; no Capítulo 3 são analisados outros trabalhos relacionados a este; no Capítulo 4 é apresentada a forma com que o trabalho foi desenvolvido, esclarecendo todos os passos para a obtenção dos objetivos propostos; no Capítulo 5 são apresentados os experimentos e resultados, além das interpretações dos resultados, e por fim, no Capítulo 6, as conclusões do trabalho.

2 REFERENCIAL TEÓRICO

Neste capítulo são vistos conceitos pertinentes, como investimentos digitais, soluções computacionais e medidas em séries temporais, ao entendimento do desenvolvimento do trabalho.

2.1 INVESTIMENTOS DIGITAIS

Além do progresso no mercado convencional de ações, o avanço dos meios tecnológicos possibilitaram a criação de ativos digitais. Tais ativos, como, por exemplo, as criptomoedas, contam com uma ampla diversidade de utilização, possibilitando que futuramente integrem um sistema complexo de transações baseado apenas nos mecanismos digitais. Além disso, as *exchanges* são as responsáveis por permitir que pessoas comuns possam investir nesses ativos (comprar e vender).

2.1.1 Criptomoeda

Criptomoeda é o principal exemplo de ativo digital. Pode ser vista como um mecanismo de pagamento e troca de recursos online. A primeira criptomoeda criada foi o *Bitcoin*, em 2008 (NAKAMOTO, 2008). Sua principal característica é a descentralização, ou seja, diferentemente das moedas convencionais, não há instituições ou governos exercendo controle direto sobre a moeda, tendo sua segurança garantida por mecanismos de criptografia e pela tecnologia de *blockchain*.

Assim, criptomoedas permitem transações diretamente entre duas partes, sem que seja necessária a presença de intermediários ou terceiros de confiança (como é o caso dos bancos tradicionais). Entretanto, a confirmação de uma transação com a inclusão do registro na *blockchain* pode ser feita de formas diferentes, dependendo da tecnologia utilizada pela criptomoeda.

A forma de validação de transações mais conhecida é a *Proof Of Work* (FINNEY, 2004), utilizado pelo *Bitcoin*, onde é necessário um grande poder computacional para a resolução e descoberta de um *hash* válido para a inclusão do novo bloco representativo da transação e posterior consenso da rede (tal abordagem de inclusão de um novo bloco é melhor descrita na seção Blockchain). Essa metodologia acaba trazendo alguns problemas a respeito da demora de validação de uma transação, além do grande gasto de energia elétrica para a manutenção dos equipamentos computacionais necessários na realização dos cálculos e descoberta dos blocos (também chamado de processo de mineração).

Outro mecanismo de validação de transações reconhecidamente importante é o *Proof Of Stake*, que utiliza a ideia de participação (*stake*) para eleger de forma aleatória quem será o responsável pela validação de determinado bloco, oferecendo, assim como o *Proof Of Work*, uma recompensa para que essa ação seja feita. A eleição do nó responsável pela validação de um bloco pode levar em consideração diversos fatores, para que tal escolha seja feita de forma democrática entre todos os nós da rede.

2.1.2 Exchanges

As *exchanges* são corretoras responsáveis por intermediar as transações de criptoativos. Elas permitem com que os usuários possam operar através de uma plataforma digital, comprando e vendendo determinadas criptomoedas. Assim, intermediando as negociações entre investidores, resolve o problema de confiança das negociações *Peer-to-peer* (P2P), e cobra pequenas taxas para a realização desses serviços (REIS, 2021). As *exchanges*, além de oferecerem uma plataforma de trocas entre pessoas, podem também oferecer um sistema de armazenamento de criptomoedas, funcionando como uma carteira digital.

Além disso, cabe mencionar que as *exchanges* têm forte influência no processo de popularização dos investimentos em criptoativos. Muitas possuem comunidades e blogs para trocas de conhecimento entre os usuários, além de estarem em boa parte das redes sociais, local em que conseguem atingir um público ainda maior de potencial investidores. Dessa forma, trazendo mais investidores novos ao mercado, aumenta-se ainda mais o volume de negociação das criptomoedas, ocasionando uma oscilação positiva.

2.2 SOLUÇÕES COMPUTACIONAIS

2.2.1 Blockchain

A tecnologia de *blockchain* é a responsável por permitir o funcionamento do *Bitcoin* e de todas as demais criptomoedas. Seu funcionamento baseia-se na resolução de um dos problemas clássicos da computação, o Problema dos Generais Bizantinos (LAMPORT; SHOSTAK; PEASE, 1982), através da validação das transações e consenso da rede. Cada transação é assinada em um bloco que se vincula aos demais blocos existentes na rede através de um número *hash* específico (NAKAMOTO, 2008). Sendo assim, a transação só será efetuada após a descoberta do valor que torna determinado bloco válido, e qualquer alteração precisa passar por todo esse processo novamente. Dessa forma,

torna-se difícil (teoricamente impossível) fraudar uma transação, uma vez que por ser uma rede distribuída, a alteração do *hash* pode ser facilmente identificada por todos os demais usuários dessa rede, inutilizando o bloco fraudado e não realizando a operação.

Figura 2.1 – Funcionamento da *blockchain* para transações com criptomoedas



Fonte: Adaptado e traduzido de (BOTT et al., 2015)

2.3 REPRESENTAÇÃO, DISTÂNCIA E CORRELAÇÃO EM SÉRIES TEMPORAIS

2.3.1 Correlação de Pearson

Correlação de Pearson (também chamado de coeficiente de correlação produto-momento) é um recurso utilizado para medir a associação linear entre 2 variáveis diferentes (PEARSON, 1895, p. 240-242). O método traça uma linha com o melhor ajuste possível aos conjuntos de dados, e o resultado indica se existe algum grau de correlação entre os dados baseado nas distâncias dos pontos até essa linha. O coeficiente resultante é sempre um valor entre -1 e 1, onde valores abaixo de 0 representam uma correlação negativa, acima de 0 representam uma correlação positiva e 0 indica a não existência de correlação

entre os dados. Para que a fórmula seja corretamente utilizada, alguns cuidados com a linearidade dos valores e a presença de outliers¹ devem ser tomados, uma vez que podem afetar os dados e corromper os resultados.

Supondo 2 conjuntos de valores, x e y , com n elementos cada, o Coeficiente de Correlação de Pearson (ρ) é calculado pela fórmula demonstrada abaixo.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (2.1)$$

Diferentes autores, como (SILVA, 2018), atribuem formas diferentes de interpretação aos resultados obtidos pela fórmula. Na tabela abaixo, algumas interpretações para os resultados do coeficiente de correlação de Pearson sugeridos pelo autor.

Tabela 2.1 – Interpretação dos Coeficientes de Pearson

Força de Associação	Coeficiente de Pearson (ρ)	
	Positiva	Negativa
Nula	0	0
Quase nula	0,10 a 0,25	-0,10 a -0,25
Fraca	0,25 a 0,50	-0,25 a -0,5
Moderada	0,50 a 0,75	-0,50 a -0,75
Forte	0,75 a 1	-0,75 a -1

Fonte: Próprio autor

2.3.2 Distância Euclidiana

Distância Euclidiana é, de forma simples, a distância entre 2 pontos. Seu cálculo é baseado no Teorema de Pitágoras e tal distância pode ser calculada em espaços com muitas dimensões. Entretanto, para este trabalho, utilizaremos a Distância Euclidiana considerando um espaço unidimensional, uma vez que estaremos analisando 2 pontos relativos ao mesmo momento no eixo de tempo (BATISTA; WANG; KEOGH, 2011).

Dados os pontos $P = p_1 \dots p_n$ e $Q = q_1 \dots q_n$ de 2 séries quaisquer, a fórmula que define a Distância Euclidiana é:

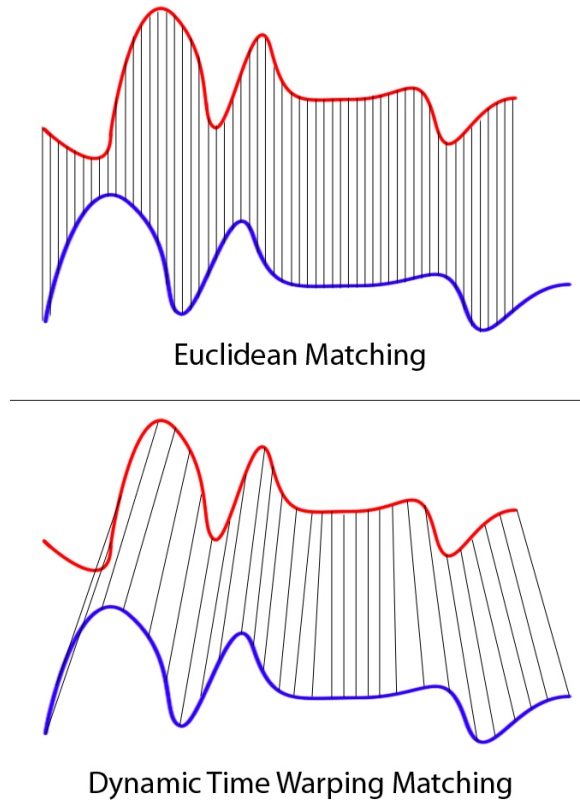
$$\sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.2)$$

Entretanto, a Distância Euclidiana acaba manifestando o problema de alinhamento temporal, como representado na Figura 2.2, não sendo a melhor abordagem para ser

¹Valores que destoam da sequência esperada dos dados, possuindo um afastamento discrepante, muitas vezes causados por movimentos atípicos ou por erros nos dados.

usada em séries temporais. Dessa forma, surge a técnica de *Dynamic Time Warping*.

Figura 2.2 – Diferença entre Distância Euclidiana e *Dynamic Time Warping* (DTW).

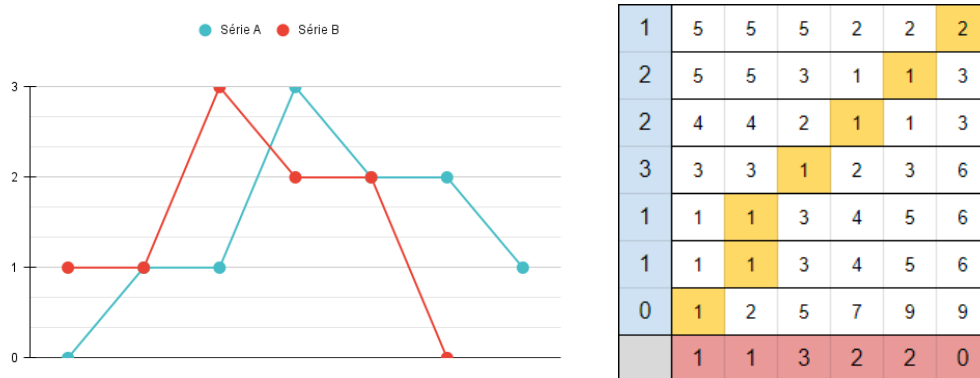


Fonte: (XANTACROSS, 2011)

2.3.3 Dynamic Time Warping

Como visto, a Distância Euclidiana calcula a diferença dos pontos das 2 séries em linha reta, ou seja, não se adapta aos possíveis deslocamentos temporais de valores, afetando os resultados de correlações baseadas nessa abordagem. Assim, o DTW (BERNDT; CLIFFORD, 1994) busca trazer uma abordagem que resolva tal problema, adaptando as variações das séries para um reflexo mais fidedigno dos movimentos correlatos.

Então, para que o alinhamento entre as séries seja feito, é necessário a criação de uma matriz de distâncias entre os pontos dessa série. Logo, supondo uma série A com m valores e uma série B com n valores, uma matriz de dimensão $[m, n]$ será criada, onde cada elemento representa a distância dos pontos levando em consideração a distância anterior já calculada, como mostra a Figura 2.3.

Figura 2.3 – Exemplo da matriz de distância que calcula o *Dynamic Time Warping*

Fonte: Próprio autor

Obtendo a distância entre as séries de dados é importante que o resultado seja normalizado para que se consiga garantir uma métrica comparável independente da proporção de distância entre as séries. Para isso, obtém-se a distância máxima possível para aquele intervalo de dados e divide-se a distância encontrada, resultando em um valor entre 0 e 1 que expressa a associação entre os dados, onde 0 seria a maior associação possível (nenhuma distância entre as séries) e 1 a menor associação possível (distância máxima entre as séries). De forma similar a correlação, podemos assumir uma interpretação aos resultados obtidos pelo *Dynamic Time Warping* e que pode também ser utilizada para as demais medidas de distância.

Na tabela abaixo, algumas interpretações para os resultados das distâncias, sugeridos pelo autor.

Tabela 2.2 – Interpretação das distâncias entre séries

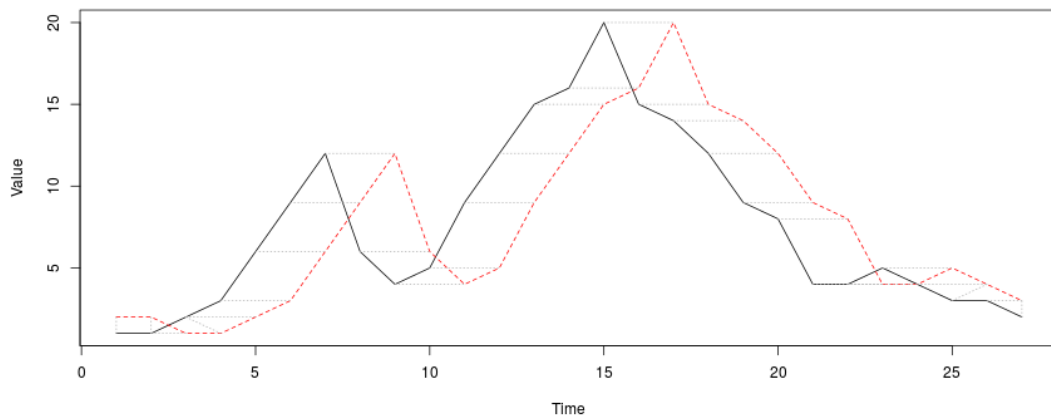
Força de associação	Distância entre as séries
Nula	1
Quase nula	0,875 a 1
Muito fraca	0,75 a 0,875
Fraca	0,5 a 0,625
Moderada fraca	0,375 a 0,5
Moderada	0,25 a 0,375
Moderada forte	0,125 a 0,25
Forte	0 a 0,125

Fonte: Próprio autor

Em resumo, o DTW pode ser compreendido como um algoritmo de minimização de custo, onde o menor custo representará uma correlação ajustada aos deslocamentos temporais de 2 séries. Para exemplificar tais ajustes, na Figura 2.4 foram geradas séries

semelhantes com um deslocamento de 2 unidades de tempo. A linha cinza tracejada representa a melhor combinação entre os pontos das séries.

Figura 2.4 – Exemplo do algoritmo de *Dynamic Time Warping*



Fonte: Próprio autor. Gerado via screenshot da ferramenta *TS Analysis Tool* (ASSUNÇÃO, 2021)

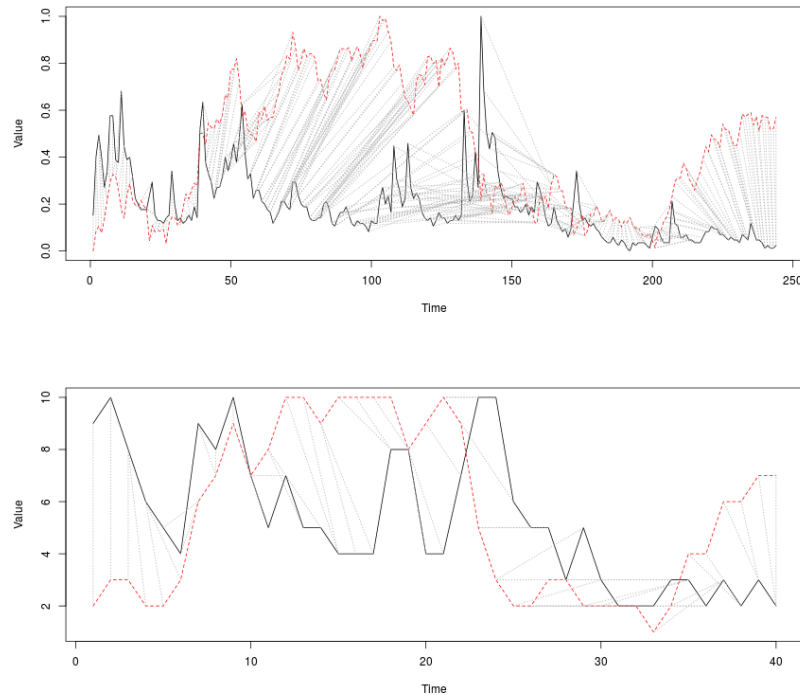
Assim, percebido tal problema de deslocamento temporal, torna-se claro que uma melhor combinação, diferentemente daquela calculada pela Distância Euclidiana, pode ser obtida através do DTW, baseando-se nas diferenças absolutas para realizar a minimização do custo. Entretanto, no prosseguimento do processo de análise, surge outro impasse relativo à grande quantidade de dados.

Quantidades excessivas de dados, como, por exemplo, os valores de uma criptomoeda durante anos, podem ocasionar dificuldades no processo de análise e descoberta de conhecimento quando busca-se uma visão mais ampla dos fenômenos que ocorreram nesse intervalo de tempo.

Jessica Lin e outros autores trazem uma abordagem de transformação e simplificação de conjuntos de dados numéricos para valores discretos, com a criação do *Symbolic Aggregate Approximation* (LIN et al., 2003). Entretanto, simplificando os dados a partir dessa transformação, suprime-se a exatidão das variações em troca do benefício de se ter um panorama mais amplo e mais simplificado dos movimentos.

Na Figura 2.5, o algoritmo de *Dynamic Time Warping* foi aplicado nos conjuntos contendo os valores da criptomoeda *Bitcoin* (tracejado vermelho) e o *Search Volume Index* dessa mesma moeda (linha preta). O gráfico superior mostra o DTW aplicado sobre os valores desses indicadores apenas normalizados. Já o gráfico inferior mostra a mesma abordagem, porém aplicada nos valores após o processo de *Symbolic Aggregate Approximation*.

Figura 2.5 – Exemplo da aplicação do algoritmo de DTW sem e com a presença do *Symbolic Aggregate Approximation*, respectivamente.



Fonte: Próprio autor. Gerado via screenshot da ferramenta *TS Analysis Tool* (ASSUNÇÃO, 2021)

2.3.4 Symbolic Aggregate Approximation

Symbolic Aggregate Approximation, ou SAX, é um método que permite a transformação dos valores de uma série temporal em um conjunto discreto de símbolos, formando uma *string*². É um algoritmo de baixa complexidade computacional que se baseia, e utiliza, o *Piecewise Aggregate Approximation* (PAA) (KEOGH et al., 2001). Antes de ser discretizada, a série temporal passa por um processo de normalização. Esse processo é realizado a partir da normalização *Z-Score*:

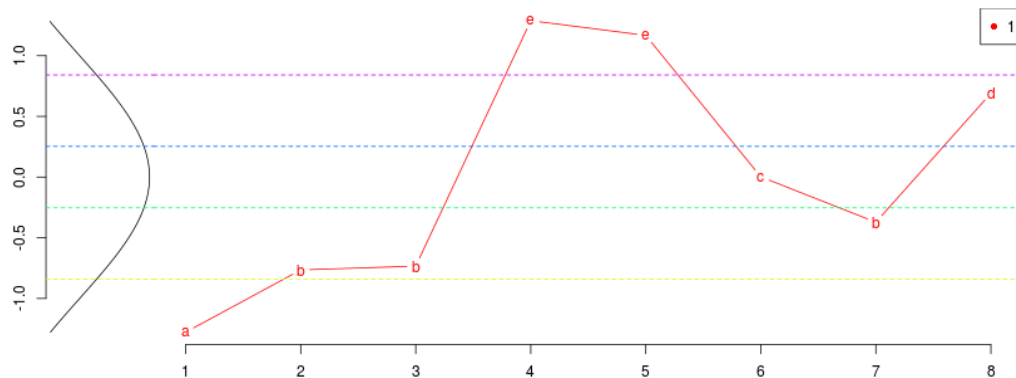
$$z = \frac{\text{valor} - \mu}{\sigma} \quad (2.3)$$

Estando o conjunto de valores normalizados, a série terá uma distribuição gaussiana (ou também chamada de distribuição normal), onde a média é 0 e o desvio padrão é 1. Assim, o método SAX realiza a divisão em N segmentos, também chamado de *tamanho da janela*, calcula o valor médio para cada intervalo e ajusta o resultado de acordo com o

²Sequência de caracteres

símbolo que melhor representa a posição do valor calculado. A partir da definição de áreas de tamanhos iguais sobre a curva da distribuição gaussiana, o método permite que todos os símbolos (representados por letras do alfabeto) sejam equiprováveis.

Figura 2.6 – Exemplo do resultado da *Symbolic Aggregate Approximation (SAX)*.



Fonte: Próprio autor. Gerado via screenshot da ferramenta *TS Analysis Tool* (ASSUNÇÃO, 2021)

No exemplo da Figura 2.6, os valores foram divididos em 8 segmentos (eixo horizontal) e foram definidos 5 símbolos/letras (eixo vertical) para representarem esses dados. A representação simbólica para a série utilizada é: abbeecbd.

A abordagem de *Symbolic Aggregate Approximation* é utilizada em diversas aplicações uma vez que permite a definição do número de segmentos (tamanho da janela) e número de símbolos (tamanho do alfabeto), adaptando sua utilização e permitindo um maior ou menor grau de detalhamento. Pode ser utilizada para associar séries temporais, que em grande maioria são contínuas, a outros formalismos que adotam dados discretos, e além disso, as sequências geradas permitem com que estados sejam mapeados diretamente de acordo com probabilidades e distribuições (ASSUNÇÃO; FERNANDES; VINCENT, 2019).

3 TRABALHOS RELACIONADOS

Diante da notável disseminação dos conceitos e estudos sobre análise de dados, extração de padrões e mercado financeiro nos últimos anos, percebe-se uma vasta quantidade de trabalhos relacionados a essas áreas.

Mesmo com a evolução dos algoritmos de análise de sentimento, o trabalho de Jethin Abraham e outros autores (ABRAHAM et al., 2018), utilizando *Valence Aware Dictionary for sEntiment Reasoning* (VADER), demonstrou em seu resultado uma fraca correlação entre sentimentos das publicações no *Twitter* e a queda no valor do *Bitcoin*, contrastando com o fato de que, tanto o volume de *Tweets*, quanto o *Search Volume Index* (SVI), demonstraram fortes correlações com os mesmos movimentos. Além disso, é mostrado que os sentimento das publicações no *Twitter* relativas às criptomoedas tendem a ser positivas independentemente das flutuações no preço, demonstrando interesses além da oportunidade momentânea de investimento.

No trabalho de Foteini Kollintza e outros autores (KOLLINTZA-KYRIAKOULIA et al., 2018), foi estudado a existência de correlação entre os preços de empresas americanas, notícias e opinião pública, buscando ao fim, o aumento da acurácia dos modelos de predição a partir das correlações encontradas. Para tal realização, os autores utilizaram técnicas de análise de séries temporais como o *Symbolic Aggregate Approximation* e *Dynamic Time Warping*. Testes de comparação foram realizados entre padrões encontrados repetidamente nas séries e intervalos aleatórios escolhidos, a fim de validar tais padrões demonstrando seu maior grau de correlação. Além disso, diferentes métodos de regressão foram analisados, onde os resultados obtidos com a utilização de *Support Vector Machines* foram melhores. Por fim, após descobertos e validados os padrões, esses foram incluídos nos modelos de regressão e demonstraram melhores resultados que os testes realizados com os intervalos aleatórios, demonstrando a influência de certos padrões de movimento que acontecem repetidamente nas séries.

(BARBER; ODEAN, 2008) citam em seu trabalho a importância do mecanismo de atenção para a tomada de decisão dos investidores. Diante da vasta quantidade de oportunidades, investidores tendem a priorizar determinados conjuntos de investimentos, uma vez que a análise de todas as opções é muito custosa (e em teoria inviável, dado que existem milhares de oportunidades, e durante o decorrer do tempo de análise manual mais opções surgem). Como citado nesse trabalho, "quando as alternativas são muitas e os custos de busca altos, a atenção pode afetar a escolha mais profundamente do que as preferências". Dessa forma, um conjunto mais restrito de ações, que chamam uma maior atenção e que possuem um grau de atratividade maior (seja por grandes movimentos no valor, maior repercussão nas redes sociais, etc.) tornam-se prioridades e facilitam o processo de escolha, uma vez que compõem um leque menor e limitado de opções.

Tabela 3.1 – Síntese dos resultados de trabalhos relacionados

	Trabalho relacionado	Contribuição para este trabalho
1	(ABRAHAM et al., 2018) Cryptocurrency price prediction using tweet volumes and sentiment analysis	Sentimentos de publicações no Twitter relativas às criptomoedas tendem a serem positivas, não demonstrando boas correlações; melhores correlações foram encontradas usando o volume de tweets
2	(KOLLINTZA-KYRIAKOULIA et al., 2018) Measuring the impact of financial news and social media on stock market modeling using time series mining techniques	Padrões de movimento no valor das ações americanas que acontecem repetidamente expressam melhores correlações que intervalos aleatórios de tempo
3	(BARBER; ODEAN, 2008) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors	Investidores priorizam um conjunto restrito de opções baseado nos mecanismos de atenção, seja por grandes movimentos no valor, grande repercussão, etc

Fonte: Próprio autor

Dessa forma, com as contribuições citadas, torna-se atrativa a investigação a respeito das possíveis associações de dados do mercado de criptomoedas, do volume de tweets e, também, do volume de pesquisas. Assim, baseando-se em um conjunto mais restrito de moedas escolhidas de acordo com sua capitalização e volume de transações, e definindo eventos externos que podem expressar melhores intervalos de análise, diferentemente de escolhas aleatórias de tempo, para realizar a busca por tais associações.

4 METODOLOGIA E DESENVOLVIMENTO DA PROPOSTA

Para este projeto serão analisadas 5 criptomoedas diferentes: Cardano (ADA), Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) e Ripple (XRP). A escolha dessas criptomoedas baseia-se no alto volume de negociação diário, o que demonstra uma constante movimentação de valores e um certo grau de interesse por parte dos investidores.

Para verificar a existência de padrões correlatos aos movimentos no valor, além dos valores de fechamento das moedas, são utilizadas 2 fontes diferentes de dados. A primeira é o volume de pesquisa (SVI), fornecido pelo site de busca *Google*, e a segunda é o volume de *Tweets* a respeito de cada criptomoeda.

Além disso, para realizar o processo de manipulação dos dados, será utilizado a linguagem de programação R (R Core Team, 2021).

Durante o desenvolvimento, algumas questões serão de extrema importância para definir os rumos e possíveis resultados do trabalho:

1. Existe volume de *Tweets* e de pesquisa suficiente para que a análise seja realizada?
2. Existem associações entre os dados analisados?
 - (a) Existe associação entre os movimentos no valor e o *Search Volume Index* de determinada criptomoeda?
 - (b) Existe associação entre os movimentos no valor e o volume de *Tweets* a respeito de determinada criptomoeda?
 - (c) Existe associação entre os movimentos do volume de *Tweets* e do *Search Volume Index* de determinada criptomoeda?
3. Qual técnica proporciona um melhor resultado para a associação entre valor, *Search Volume Index* e volume de *Tweets*?
4. Existe influência dos eventos externos para o aumento ou diminuição das associações?

Todas as questões levantadas serão respondidas no desenvolver do trabalho, e podem ser encontradas no capítulo 5 de experimentos e resultados. Para a Pergunta 1, podem ser encontrados na seção 5.1 de análise do volume de dados. Para a Pergunta 2 (itens a, b e c) podem ser encontradas na subseção 5.3 da parte de análise dos dados. Para a Pergunta 3, podem ser encontrados na seção 5.3.4 de análise das técnicas. Para a Pergunta 4 podem ser encontradas na seção 5.3.5 de análise de influência dos eventos.

4.1 DEFINIÇÃO DO INTERVALO DE TEMPO E EVENTOS RELEVANTES

O intervalo de tempo escolhido para a análise foi do dia 01/01/2021 até o dia 15/09/2021, abrangendo todo o ano de 2021 até o início deste trabalho. Para que a definição e divisão dos eventos potencialmente relevantes seja feita corretamente, é necessário entender a existência de eventos externos que ocorreram dentro do intervalo de tempo escolhido.

Foi entre essas datas que o *Bitcoin* atingiu sua máxima histórica anterior, chegando a alcançar o valor de R\$ 367.279,33, no dia 14 de abril, demonstrando o grande entusiasmo do mercado de criptomoedas.

Além disso, durante esse intervalo de tempo, ocorreram diversas tentativas de intervenção por parte do governo Chinês sobre o *Bitcoin*, o que causou muita discussão nas redes sociais, além de impulsionar a quantidade de pesquisas a respeito do tema. No dia 21 de junho, o Banco Popular da China (PBOC) emitiu um comunicado alertando e proibindo os bancos de transacionar criptomoedas. Após tal aviso, diversos bancos como o Banco Agrícola da China (AGRÍCOLA, 2021), o Banco Industrial e Comercial da China (COMERCIAL, 2021) e o Banco de Construção da China (CONSTRUÇÃO, 2021), divulgaram seus comunicados informando a proibição e o interrompimento das transações envolvendo criptomoedas. Com receio das próximas ações do governo Chinês, o mercado reagiu de forma pessimista derrubando o preço do *Bitcoin* e do *Ethereum* em cerca de 10% e 15%, respectivamente.

Outro evento importante para a movimentação do mercado de criptomoedas, foi a adoção do *Bitcoin* como moeda legal no país de *El Salvador* (SALVADOR, 2021). O acontecimento se deu no dia 07 de setembro, fazendo com que o mercado reagisse sutilmente de forma positiva. Entretanto, 1 dia após a adoção dessa medida, a moeda acumulava uma queda de 11%, retornando dos \$46.000. Tal queda pode ter sido impulsionada por problemas ocasionados pela alta demanda da população, uma vez que servidores precisaram ser desligados por conta do alto número de pedidos de registro de usuários no país. Além disso, a carteira digital criada pelo Governo salvadorenho, a Chivo, não foi ofertada por todas as plataformas disponíveis, ocasionando certa descrença nos investidores. Muitos protestos também ocorreram nesse dia, dado que uma parcela da população temia a instabilidade financeira que poderia ser gerada por essa medida. El Salvador, na busca pela adoção da moeda, distribuiu cerca de 30 dólares em *bitcoins* para cada cidadão.

Tais eventos citados afetam principalmente o Bitcoin, entretanto, no mesmo intervalo de tempo, também encontram-se grandes repercussões a respeito de outras moedas. Durante o mês de agosto, uma das principais e mais esperadas atualizações da rede Cardano deu início em sua fase de testes final, a atualização Alonzo Purple, a qual disponibiliza contratos inteligentes na blockchain Cardano (ALKUDMANI, 2021).

Diversos eventos externos podem ser mencionados, entretanto, para este trabalho,

os eventos acima citados basearam as escolhas de divisão dos intervalos de tempo. Além disso, intervalos aleatórios de tempo foram escolhidos para possibilitar a comparação dos resultados obtidos a fim de validar ou não algumas das hipóteses levantadas.

Tendo como referência o dia em que cada evento ocorreu, os intervalos de tempo iniciam 1 semana antes da data do evento e finalizam 1 semana após a mesma data, gerando diversas divisões do conjunto de dados. A tabela abaixo identifica cada um dos eventos com um nome e com o intervalo de tempo relativo.

Tabela 4.1 – Eventos escolhidos para análise

	Acontecimento	Nome do evento	Intervalo de tempo (2021)
1	Proibição do Banco Popular da China	Evento A	17/06 - 01/07
2	Adoção do Bitcoin como moeda oficial de El Salvador	Evento B	31/08 - 14/09
3	Máxima histórica do valor do Bitcoin	Evento C	07/04 - 21/04
4	Evento aleatório	Evento D	08/02 - 22/02
5	Evento aleatório	Evento E	13/07 - 27/07
6	Atualização Alonzo Purple da Cardano	Evento F	01/08 - 15/08

Fonte: Próprio autor

4.2 SELEÇÃO, EXTRAÇÃO E MANIPULAÇÃO DE DADOS

A extração dos dados ocorreram de forma separada, uma vez que cada fonte exigiu uma abordagem diferente para a obtenção dos *datasets*.

4.2.1 Valores de fechamento das moedas

Para a extração dos valores de fechamento, foi utilizado o site (Crypto Data Download, 2021). Todos os *datasets* disponibilizados fornecem diversas informações a respeito

das moedas, como valor de abertura, valor mais alto atingido no dia, valor mais baixo atingido no dia, valor de fechamento, volume de transações, etc.

Para extrair apenas a data e o valor de fechamento, foi desenvolvido um código em *R* que possibilitou gerar um arquivo *CSV* a partir daquele obtido através do site.

Ao final, na geração do arquivo de dados relativos aos valores das criptomoedas, foram obtidos 258 registros (linhas) contendo data e o respectivo valor de fechamento.

4.2.2 Volume de pesquisas

Para a extração do *Search Volume Index*, fora utilizada a ferramenta Google Trends (Google, 2021). Especificando a palavra e o intervalo de tempo desejado, são exibidos valores entre 0 e 100 para cada dia. Esses valores definem o volume de pesquisas e ficam limitados ao número 100, onde 100 é o dia em que mais ocorreram pesquisas.

A ferramenta adapta a exibição dos dados de acordo com o intervalo de tempo selecionado. Quando o intervalo é maior que 270 dias, a exibição do *SVI* é agrupada e reflete um índice semanal, e não mais diário. Entretanto, não houveram problemas com os *datasets* escolhidos, uma vez que o intervalo escolhido não ultrapassa os limites do formato de exibição por dia.

Dado que a exibição dos dados depende do intervalo escolhido, foi necessário, para cada um dos eventos, gerar uma nova consulta aos volumes de pesquisa, gerando também, arquivos específicos para cada evento com seu respectivo volume de pesquisas dentro do intervalo selecionado.

Dessa forma, para cada uma das moedas, foram obtidos 6 arquivos com 15 registros (linhas) contendo data e o respectivo *Search Volume Index*.

4.2.3 Volume de *Tweets*

Para a coleta do volume de *Tweets*, fora utilizado o site (BitInfoCharts, 2021). Os dados relativos ao volume de *Tweets* são exibidos em um gráfico. Para conseguir coletar esses dados, foi necessária a execução de um comando no console do navegador, o que gerou um arquivo com os dados exibidos.

O arquivo gerado possui informações do número de *Tweets* diários sobre determinada criptomoeda num intervalo maior que o escolhido para este trabalho, portanto, foi necessário o ajuste do intervalo de tempo. Além disso, foram necessárias algumas correções relativas aos meses do ano, uma vez que o *dataset* iniciava a contabilização dos meses a partir do número zero (representando o mês de janeiro) e finalizava no número onze (representando dezembro).

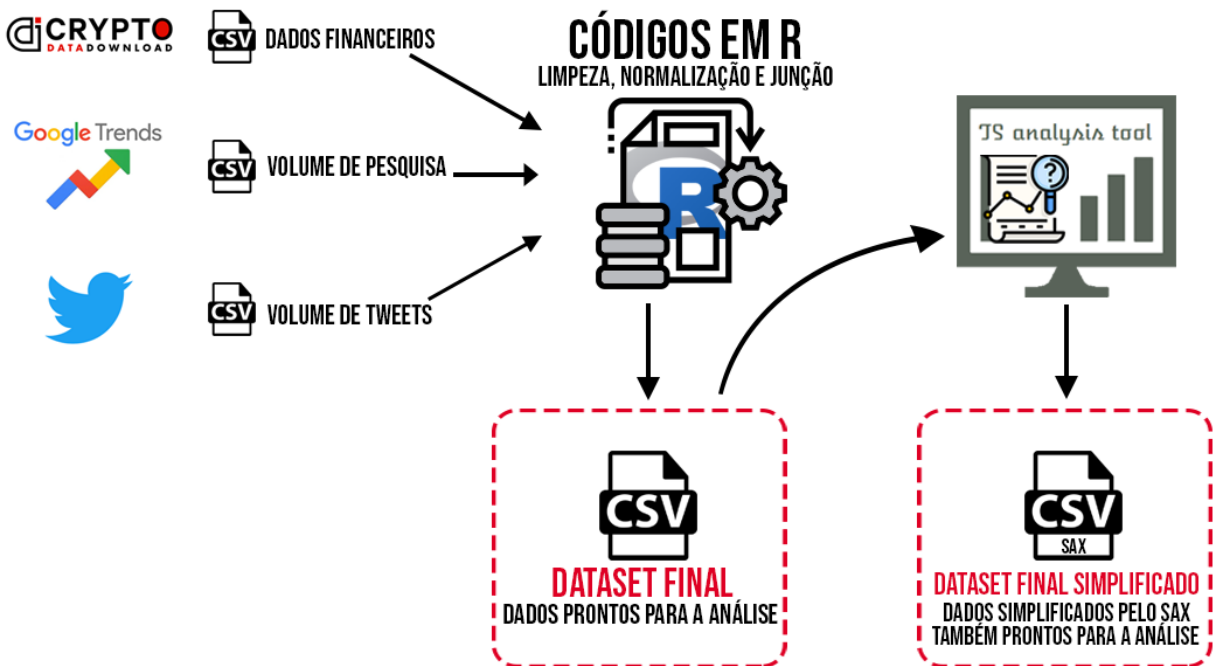
Ao final, na geração do arquivo de dados relativos ao volume de Tweets, foram obtidos 258 registros (linhas) contendo data e a respectiva quantidade de Tweets.

4.2.4 Manipulação dos dados

Tendo obtido os 3 arquivos (de dados financeiros, de volume de pesquisa e de volume de Tweets), foram necessárias algumas manipulações para capturar o intervalo de cada um dos eventos mencionados, normalizar os valores e unir os dados em um arquivo final.

O arquivo final é uma junção dos dados brutos normalizados de um determinado intervalo de tempo. Após gerado, foi feita a aplicação do SAX utilizando a ferramenta *TS Analysis Tool* (ASSUNÇÃO, 2021), e dando origem a um novo arquivo contendo os dados simplificados de acordo com a técnica de *Symbolic Aggregate Approximation*. Abaixo, um fluxograma simplificado do processo de manipulação dos dados até a criação dos *datasets* usados nas análises.

Figura 4.1 – Fluxo de manipulação dos dados



Fonte: Próprio autor.

4.3 CONFIGURAÇÕES DO SYMBOLIC AGREGGATE APPROXIMATION

Como mencionado no capítulo de Trabalhos Relacionados, o SAX admite diferentes configurações para o tamanho da janela e tamanho do alfabeto, resultando em maiores ou menores graus de simplificação dos dados. Para definir se determinada configuração era interessante para o aumento da associação dos dados, foram realizados pequenos testes separadamente e alguns resultados foram avaliados.

A figura abaixo demonstra os resultados obtidos no dataset da criptomoeda Bitcoin. Para cada configuração foram testadas as relações: moeda x volume de pesquisa, moeda x volume de Tweets, volume de pesquisa x volume de Tweets, respectivamente dentro de cada um dos eventos. As menores distâncias indicam as melhores associações.

Figura 4.2 – Testes de diferentes configurações de SAX + DTW

SAX	Janela	Alfabeto	Evento A			Evento B			Evento C			Evento D			Evento E		
Configuração 1	10	4	0,47	0,37	0,13	0,20	0,28	0,20	0,28	0,15	0,15	0,27	0,37	0,10	0,12	0,13	0,10
Configuração 2	15	4	0,41	0,24	0,17	0,18	0,11	0,10	0,31	0,12	0,13	0,17	0,30	0,11	0,14	0,13	0,19
Configuração 3	10	10	0,23	0,17	0,11	0,09	0,08	0,05	0,14	0,11	0,07	0,13	0,16	0,08	0,09	0,08	0,09
Configuração 4	15	10	0,25	0,19	0,11	0,11	0,08	0,12	0,19	0,12	0,10	0,20	0,23	0,21	0,19	0,14	0,10

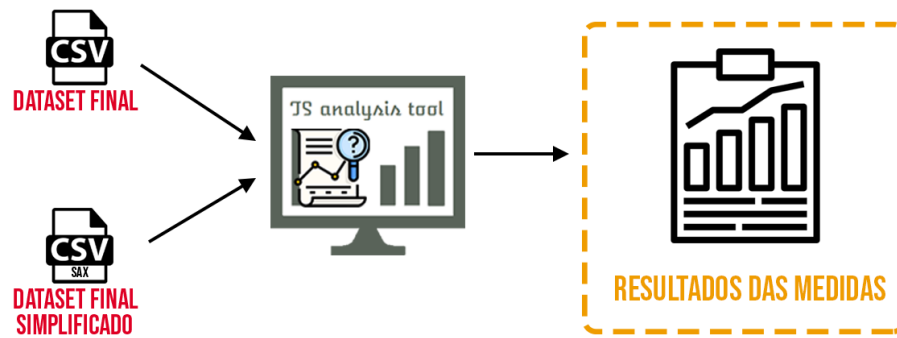
Fonte: Próprio autor.

Os resultados, em sua maioria, demonstraram uma melhora na associação com as configurações 3 e 4, sendo estas as escolhidas para o restante da análise, onde a configuração 3 possui janela 10 e alfabeto 10 e a configuração 4 possui janela 15 e alfabeto 10.

4.4 APLICAÇÃO DAS TÉCNICAS

Tendo gerado os datasets com a organização dos dados para a análise, os arquivos foram individualmente carregados dentro da ferramenta *TS Analysis Tool* (ASSUNÇÃO, 2021) onde as técnicas foram aplicadas, retornando o resultado de cada uma das medidas. Os resultados foram extraídos manualmente para uma tabela, facilitando a visualização e a interpretação dos dados.

Figura 4.3 – Aplicação das técnicas



Fonte: Próprio autor.

5 RESULTADOS E DISCUSSÃO

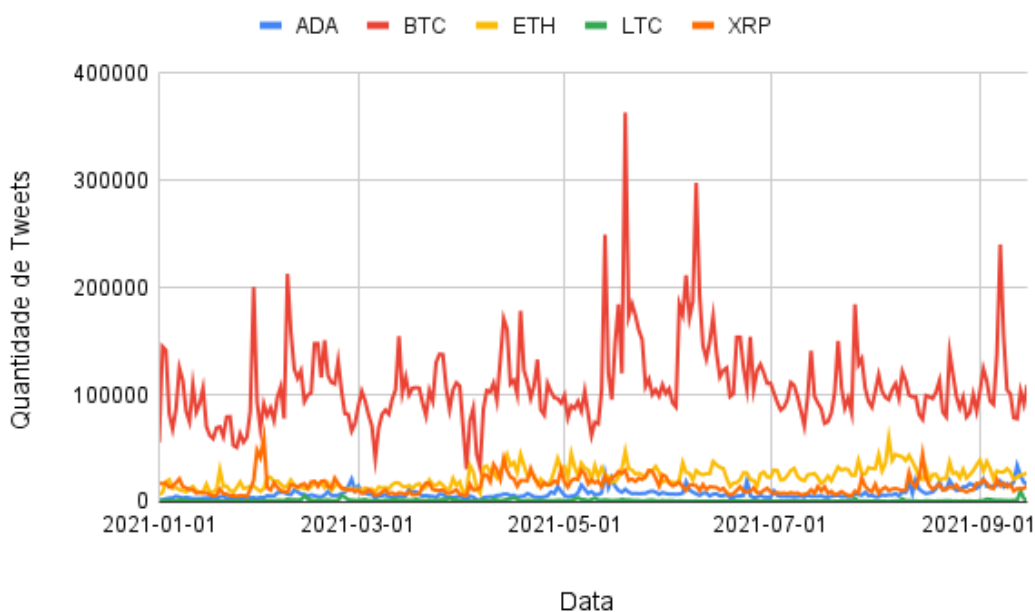
As análises realizadas visam responder as perguntas elencadas na Seção de Metodologia e foram definidas a partir dos eventos descritos e selecionados anteriormente. Cada um dos eventos foi analisado separadamente, tendo seus resultados divididos pelas moedas, ou seja, para cada evento capturou-se os resultados de correlação e distância entre os dados de cada uma das moedas.

5.1 ANÁLISE DO VOLUME DE DADOS

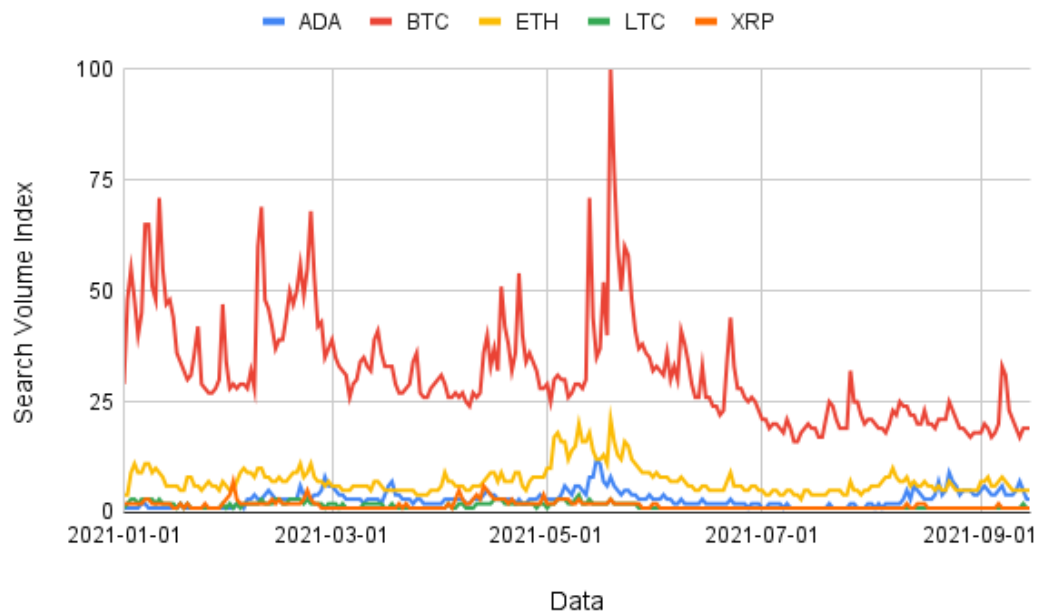
Para que uma análise seja realizada corretamente é necessário que o volume de dados seja suficiente. Sendo assim, um possível impeditivo para o desenvolvimento do trabalho era o volume de dados, tanto de pesquisa, quanto de tweets, uma vez que os mesmos poderiam não fornecer informações suficientes para a análise.

Assim, para responder a pergunta 1 foram gerados os gráficos abaixo contendo o volume de tweets, e também, o *Search Volume Index* (índice do volume de pesquisa) a respeito de cada uma das moedas, demonstrando os resultados obtidos.

Figura 5.1 – Volume de Tweets: Quantidade de Tweets por dia durante todo o intervalo



Fonte: Próprio autor.

Figura 5.2 – Volume de Pesquisa: *Search Volume Index* por dia durante todo o intervalo

Fonte: Próprio autor.

De acordo com os dados levantados, todas as criptomoedas demonstraram a existência suficiente de informações. Para o volume de tweets, todas demonstraram boa média de publicações relacionadas a elas diariamente, não existindo dias sem tweets relacionados. Para o *Search Volume Index*, de maneira análoga, o índice demonstrou bom volume de pesquisa a respeito de cada criptomoeda diariamente. Para os valores de fechamento, todas as moedas possuíam informações completas para o intervalo de tempo escolhido, e dessa forma, confirmando que os dados disponíveis são suficientes para a análise. Além disso, e por ser a principal moeda atualmente, o Bitcoin demonstrou prevalência em relação às demais moedas, tanto em volume de tweets, quanto no índice de volume de pesquisa. Abaixo, a tabela da média dos valores levantados dentro do intervalo de tempo de análise.

Tabela 5.1 – Tabela da média diária de volume de Tweets e *Search Volume Index*

	ADA	BTC	ETH	LTC	XRP
Volume de Tweets	7893,02	108676,45	22973,92	1437,12	14801,33
Search Volume Index	3,04	31,88	7,22	1,49	1,61

Fonte: Próprio autor.

5.2 ANÁLISE DESCRITIVA DOS DADOS

Para verificar a associação dos dados, os datasets gerados foram carregados na ferramenta *TS Analysis Tool* e os resultados obtidos pelas medidas foram separados pelos eventos em tabelas externas.

Para que a leitura dos dados seja feita corretamente, é importante lembrar que a Correlação de Pearson mede a força de correlação dos dados, ou seja, quanto mais próximos de -1 ou 1, mais correlacionados estarão os dados (mais associados). Entretanto, as demais medidas utilizadas são medidas de distância, que ficam limitadas de 0 a 1, onde quanto mais próximo de 1, mais distante estão os dados (menos associados).

Os resultados devem ser interpretados levando em consideração a tabela sugerida de interpretações da correlação de Pearson (Tabela 2.1) e a tabela sugerida de interpretações das distâncias entre as séries (Tabela 2.2).

Tabela 5.2 – Tabela de auxílio à leitura dos resultados

		Intervalo dos resultados	Menor associação	Maior associação
Medida de Correlação	Correlação de Pearson	[-1, 1]	Valores próximos de 0	Valores próximos de -1 ou 1
Medidas de Distância	Distância Euclidiana DTW SAX + DTW	[0, 1]	Valores próximos de 1	Valores próximos de 0

Fonte: Próprio autor.

5.2.1 Evento A: Proibição de criptomoedas na China

O evento A refere-se à proibição das transações de criptomoedas na China. Por ser um evento que movimentou negativamente as criptomoedas, espera-se uma correlação negativa entre o valor das moedas e os volumes. Espera-se, também, que haja correlação moderada ou forte entre o *Search Volume Index* e o volume de tweets, uma vez que são métricas que devem aumentar quanto maior for a repercussão de um evento. Para as medidas de distância, espera-se valores próximos a 0, representando uma boa associação entre os dados.

Abaixo as tabelas com os resultados obtidos a partir das análises dos dados e algumas informações relevantes extraídas dos dados.

Tabela 5.3 – Resultados entre valor das moedas e *Search Volume Index*

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	-0,672	-0,558	-0,513	-0,347	-0,295
Distância Euclidiana	0,537	0,499	0,498	0,460	0,474
DTW	0,346	0,306	0,244	0,181	0,254
SAX (Configuração 3) + DTW	0,156	0,233	0,239	0,144	0,206
SAX (Configuração 4) + DTW	0,219	0,252	0,185	0,156	0,222

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e *Search Volume Index* que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação negativa em todas as moedas, e em sua maioria, correlação de força moderada.
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. A média das distâncias calculadas pelo DTW foi de 0,266, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada.

Tabela 5.4 – Resultados entre valor das moedas e volume de Tweets

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,067	-0,529	0,305	-0,227	-0,324
Distância Euclidiana	0,526	0,521	0,345	0,480	0,441
DTW	0,431	0,213	0,186	0,410	0,195
SAX (Configuração 3) + DTW	0,150	0,167	0,156	0,217	0,139
SAX (Configuração 4) + DTW	0,219	0,185	0,144	0,222	0,181

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e volume de tweets que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação negativa em maior parte das moedas, entretanto a força de correlação foi fraca, exceto Bitcoin (BTC) com correlação moderada.
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,287, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada.

Tabela 5.5 – Resultados entre *Search Volume Index* e volume de Tweets

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,259	0,690	-0,069	0,173	0,247
Distância Euclidiana	0,350	0,300	0,416	0,389	0,366
DTW	0,164	0,176	0,220	0,211	0,188
SAX (Configuração 3) + DTW	0,156	0,106	0,156	0,206	0,122
SAX (Configuração 4) + DTW	0,159	0,115	0,156	0,163	0,152

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre *Search Volume Index* e volume de tweets que podem ser citadas:

1. Diferentemente do resultado esperado, a maior parte das moedas não demonstrou correlação de força significativa, exceto o Bitcoin (BTC) que demonstrou correlação positiva moderada.
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,192, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação moderada forte.

Os testes demonstraram a correlação negativa esperada para as moedas no caso da análise entre valor e *Search Volume Index*. Entretanto, a mesma correlação negativa não foi evidenciada em todas as moedas no caso da análise entre valor e volume de tweets, tendo apenas o Bitcoin demonstrado correlação negativa para esse caso. A análise entre *Search Volume Index* e volume de tweets indicou uma correlação de força moderada apenas para o Bitcoin.

Em todos os testes verificou-se a redução da distância utilizando o DTW em relação à distância euclidiana. Além disso, a simplificação usando o SAX (Configuração 3) garantiu uma distância ainda menor entre as séries de dados para todos os casos.

5.2.2 Evento B: Bitcoin como moeda oficial em El Salvador

O evento B refere-se à adoção do Bitcoin como moeda oficial de El Salvador. Mesmo que tal evento demonstre uma grande euforia para o mercado de criptomoedas, a adoção foi marcada por diversos problemas, tanto pela alta demanda gerada, como também por protestos e descrença da população, principalmente por conta do medo da instabilidade financeira. Dessa forma, espera-se encontrar uma correlação fraca entre o valor e os volumes, tanto de pesquisa, quanto de tweets, uma vez que os problemas podem acabar limitando o indicativo de euforia do mercado.

Espera-se, também, que haja correlação de força significativa entre o *Search Volume Index* e o volume de tweets, uma vez que são métricas que devem aumentar quanto maior for a repercussão de um evento. Para as medidas de distância, espera-se valores próximos a 0, representando uma boa associação entre os dados.

Abaixo as tabelas com os resultados obtidos a partir das análises dos dados e algumas informações relevantes extraídas dos dados.

Tabela 5.6 – Resultados entre valor das moedas e *Search Volume Index* para o Evento B

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,050	-0,215	0,671	0,382	0,338
Distância Euclidiana	0,451	0,492	0,292	0,312	0,365
DTW	0,236	0,176	0,121	0,130	0,166
SAX (Configuração 3) + DTW	0,122	0,094	0,078	0,089	0,111
SAX (Configuração 4) + DTW	0,204	0,115	0,115	0,085	0,159

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e *Search Volume Index* que podem ser citadas:

1. Assim como o resultado que se esperava encontrar, a maior parte das moedas evidenciou correlação fraca. Foi evidenciada correlação negativa para o Bitcoin, entretanto, tal correlação não demonstra uma força significativa de associação.
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,166, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada forte.
3. Ambas as configurações de SAX demonstraram boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW.

Tabela 5.7 – Resultados entre valor das moedas e volume de Tweets para o Evento B

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	-0,277	0,050	0,409	-0,037	0,529
Distância Euclidiana	0,548	0,412	0,360	0,415	0,310
DTW	0,390	0,159	0,168	0,222	0,157
SAX (Configuração 3) + DTW	0,183	0,078	0,089	0,083	0,106
SAX (Configuração 4) + DTW	0,296	0,081	0,111	0,063	0,137

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e volume de Tweets que podem ser citadas:

1. Assim como o resultado que se esperava encontrar, foi evidenciada correlação fraca ou quase nula com a maior parte das moedas, exceto Ripple (XRP).
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,219, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada forte.
3. Ambas as configurações de SAX demonstraram boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW.

Tabela 5.8 – Resultados entre *Search Volume Index* e volume de Tweets para o Evento B

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,757	0,888	0,595	0,864	0,486
Distância Euclidiana	0,227	0,141	0,286	0,180	0,305
DTW	0,147	0,054	0,167	0,108	0,148
SAX (Configuração 3) + DTW	0,094	0,050	0,128	0,056	0,122
SAX (Configuração 4) + DTW	0,152	0,122	0,152	0,070	0,141

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre *Search Volume Index* e volume de Tweets que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação positiva (moderada e forte) em todas as moedas.
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de

0,125, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação forte.

3. SAX (Configuração 3) demonstrou boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW.

Os testes demonstraram que a correlação entre os valores das moedas e os volumes foram fracos, assim como o esperado. Entretanto, métricas de distância demonstraram força significativa de associação.

Além disso, os testes entre volume de pesquisa e volume de tweets demonstraram bons resultados (força significativa de associação), tanto de correlação, quanto de distância, assim como o esperado e demonstrando uma possível associação desses fatores ao evento.

5.2.3 Evento C: Máxima histórica do Bitcoin

O evento C refere-se à máxima histórica do valor do Bitcoin. Por ser um evento que movimentou positivamente as criptomoedas, em específico o Bitcoin, espera-se uma correlação positiva entre o valor e os volumes, tanto de pesquisa, quanto de tweets. Espera-se, também, que haja correlação de força significativa entre o *Search Volume Index* e o volume de tweets, uma vez que são métricas que devem aumentar quanto maior for a repercursão de um evento. Para as medidas de distância, espera-se valores próximos a 0, representando uma boa associação entre os dados.

Abaixo as tabelas com os resultados obtidos a partir das análises dos dados e algumas informações relevantes extraídas dos dados.

Tabela 5.9 – Resultados entre valor das moedas e *Search Volume Index* para o Evento C

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,800	-0,087	0,799	0,894	0,740
Distância Euclidiana	0,222	0,464	0,203	0,145	0,275
DTW	0,102	0,232	0,138	0,062	0,126
SAX (Configuração 3) + DTW	0,022	0,139	0,117	0,078	0,050
SAX (Configuração 4) + DTW	0,104	0,185	0,081	0,044	0,104

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e *Search Volume Index* que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação positiva (forte) na maior parte das moedas, exceto Bitcoin que obteve uma correlação praticamente nula;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,132, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação moderada forte.

Tabela 5.10 – Resultados entre valor das moedas e volume de Tweets para o Evento C

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,484	0,328	0,312	0,689	0,251
Distância Euclidiana	0,319	0,387	0,335	0,248	0,357
DTW	0,117	0,184	0,180	0,113	0,211
SAX (Configuração 3) + DTW	0,122	0,106	0,167	0,111	0,139
SAX (Configuração 4) + DTW	0,100	0,122	0,159	0,141	0,178

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e volume de tweets que podem ser citadas:

1. Diferentemente do resultado esperado, foi evidenciada correlação fraca na maior parte das moedas, exceto na moeda Litecoin (LTC);
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,161, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação moderada forte.

Tabela 5.11 – Resultados entre *Search Volume Index* e volume de Tweets para o Evento C

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,544	0,698	0,662	0,528	0,642
Distância Euclidiana	0,308	0,227	0,240	0,263	0,230
DTW	0,160	0,124	0,159	0,104	0,128
SAX (Configuração 3) + DTW	0,156	0,072	0,128	0,117	0,139
SAX (Configuração 4) + DTW	0,156	0,096	0,137	0,130	0,163

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre *Search Volume Index* e volume de tweets que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação positiva (moderada) em todas as moedas;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,135, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação moderada forte.

Os testes demonstraram força de correlação significativa entre os valores das moedas e o volume de pesquisas. Entretanto, os mesmos resultados não foram encontrados nos testes entre valor e volume de tweets.

Os testes entre *Search Volume Index* e volume de tweets demonstraram bons resultados (força significativa de associação) em correlação e também em medida de distância.

Além disso, forças significativas de associação utilizando o DTW foram encontradas, com uma média geral de 0,143 unidades de medida, correspondendo a um nível de associação significativa dos dados.

5.2.4 Evento D: Intervalo aleatório

O evento D inicialmente foi escolhido como um intervalo aleatório de 08/02 a 22/02. Entretanto, foi identificado um evento potencialmente relevante no primeiro dia do intervalo de análise (dia 08/02). Mesmo assim, como tal evento está deslocado do ponto médio do intervalo (dia 15/02), é importante que a análise seja normalmente realizada para que se possa verificar se um evento mais afastado, ainda assim, pode causar boas associações entre os dados.

Espera-se que não sejam encontradas boas correlações e medidas de distância, uma vez que trata-se de um intervalo, em boa parte, sem nenhum evento relevante.

Abaixo as tabelas com os resultados obtidos a partir das análises dos dados e algumas informações relevantes extraídas dos dados.

Tabela 5.12 – Resultados entre valor das moedas e *Search Volume Index* para o Evento D

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,645	0,172	0,162	0,664	0,228
Distância Euclidiana	0,338	0,370	0,389	0,293	0,477
DTW	0,213	0,217	0,191	0,200	0,391
SAX (Configuração 3) + DTW	0,106	0,128	0,167	0,083	0,244
SAX (Configuração 4) + DTW	0,111	0,200	0,167	0,100	0,215

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e *Search Volume Index* que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação fraca na maior parte das moedas, exceto Cardano (ADA) e Litecoin (XRP);
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,242, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada forte.
3. Ambas as configurações de SAX demonstraram boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW.

Tabela 5.13 – Resultados entre valor das moedas e volume de Tweets para o Evento D

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,187	-0,161	-0,152	0,037	0,360
Distância Euclidiana	0,399	0,446	0,462	0,540	0,286
DTW	0,259	0,305	0,181	0,400	0,163
SAX (Configuração 3) + DTW	0,122	0,156	0,272	0,189	0,072
SAX (Configuração 4) + DTW	0,215	0,226	0,163	0,230	0,178

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e volume de tweets que podem ser citadas:

1. Assim como o resultado esperado, nenhuma correlação de força significativa foi encontrada;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,262, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada.

Tabela 5.14 – Resultados entre *Search Volume Index* e volume de Tweets para o Evento D

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,732	0,632	0,194	0,471	0,540
Distância Euclidiana	0,220	0,235	0,347	0,338	0,461
DTW	0,127	0,142	0,155	0,213	0,315
SAX (Configuração 3) + DTW	0,078	0,078	0,189	0,172	0,206
SAX (Configuração 4) + DTW	0,119	0,215	0,137	0,259	0,226

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre *Search Volume Index* e volume de tweets que podem ser citadas:

1. Diferentemente do resultado que se esperava encontrar, foi evidenciada correlação positiva (moderada e forte) na maior parte das moedas;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,190, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada forte.

Os testes demonstraram a não-correlação do valor e dos volumes, tanto de pesquisa, quanto de tweets, em sua maioria. Testes envolvendo o *Search Volume Index* e volume de tweets demonstraram boas correlações e uma medida menor de distância utilizando DTW (0,190 unidades de medida). A média da distância utilizando DTW resultou em 0,232 unidades de medida.

5.2.5 Evento E: Intervalo Aleatório

O evento E foi escolhido como um intervalo aleatório de 13/07 a 27/07. Espera-se que não sejam encontradas boas correlações e medidas de distância, uma vez que trata-se de um intervalo sem nenhum grande evento relevante encontrado.

Abaixo as tabelas com os resultados obtidos a partir das análises dos dados e algumas informações relevantes extraídas dos dados.

Tabela 5.15 – Resultados entre valor das moedas e *Search Volume Index* para o Evento E

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	-0,116	0,437	0,442	0,424	0,032
Distância Euclidiana	0,559	0,293	0,340	0,281	0,358
DTW	0,387	0,163	0,207	0,166	0,203
SAX (Configuração 3) + DTW	0,156	0,094	0,072	0,172	0,117
SAX (Configuração 4) + DTW	0,167	0,185	0,189	0,152	0,185

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e *Search Volume Index* que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação fraca nas moedas, além de Ripple (XRP) e Cardano (ADA) com correlação praticamente nula;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,225, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada forte.

Tabela 5.16 – Resultados entre valor das moedas e volume de Tweets para o Evento E

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,218	0,485	0,644	0,264	0,038
Distância Euclidiana	0,446	0,274	0,237	0,527	0,408
DTW	0,287	0,152	0,100	0,405	0,248
SAX (Configuração 3) + DTW	0,161	0,078	0,050	0,194	0,133
SAX (Configuração 4) + DTW	0,144	0,141	0,085	0,170	0,189

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e volume de tweets que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação fraca na maior parte das moedas, exceto Ethereum (ETH) que obteve correlação positiva moderada;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,239, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada forte.
3. Ambas as configurações de SAX demonstraram boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW.

Tabela 5.17 – Resultados entre *Search Volume Index* e volume de Tweets para o Evento E

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,715	0,759	0,275	0,574	0,148
Distância Euclidiana	0,219	0,193	0,339	0,387	0,459
DTW	0,116	0,095	0,161	0,307	0,296
SAX (Configuração 3) + DTW	0,122	0,094	0,078	0,133	0,128
SAX (Configuração 4) + DTW	0,141	0,096	0,152	0,274	0,144

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre *Search Volume Index* e volume de tweets que podem ser citadas:

1. Diferentemente do resultado que se esperava encontrar, foi evidenciada correlação positiva moderada na maior parte das moedas, exceto Ethereum (ETH) e Ripple (XRP).
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,195, onde, de acordo com a tabela de interpretação e diferentemente do esperado, demonstra uma associação moderada forte.

Os testes envolvendo valor e volumes, tanto de pesquisa, quanto de tweets, resultaram em correlações fracas, em sua maioria, como era esperado. Entretanto, testes envolvendo *Search Volume Index* e volume de tweets demonstraram, em sua maioria, boas correlações e distâncias menores entre as séries.

A média das distâncias calculadas com o DTW, de todos os testes realizados, foi de 0,22 unidades de medida. Entretanto, pode-se observar que as médias encontradas nas análises envolvendo o valor ficaram ambas acima de 0,22, enquanto a análise entre *Search Volume Index* e volume de tweets demonstrou uma média abaixo de 0,195. Assim, é possível inferir que exista associação maior desses fatores (*Search Volume Index* e volume de tweets) mesmo em intervalos aleatórios de tempo.

5.2.6 Evento F: Atualização da rede Cardano

O evento F refere-se à atualização *Alonzo Purple* da rede Cardano. Por ser uma atualização muito aguardada por conta da implementação dos contratos inteligentes na rede Cardano, espera-se encontrar uma correlação positiva entre o valor e os volumes, tanto de pesquisa, quanto de tweets. Para as medidas de distância, espera-se valores próximos a 0, representando uma boa associação entre os dados.

Abaixo as tabelas com os resultados obtidos a partir das análises dos dados e algumas informações relevantes extraídas dos dados.

Tabela 5.18 – Resultados entre valor das moedas e *Search Volume Index* para o Evento F

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,955	0,644	0,348	0,800	0,962
Distância Euclidiana	0,114	0,276	0,368	0,215	0,098
DTW	0,061	0,163	0,185	0,111	0,053
SAX (Configuração 3) + DTW	0,039	0,117	0,178	0,044	0,033
SAX (Configuração 4) + DTW	0,030	0,137	0,178	0,081	0,037

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e *Search Volume Index* que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação positiva (forte) na maior parte das moedas;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,115, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação forte.
3. Ambas as configurações de SAX demonstraram boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW.

Tabela 5.19 – Resultados entre valor das moedas e volume de Tweets para o Evento F

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,716	-0,278	-0,187	0,263	0,780
Distância Euclidiana	0,267	0,467	0,470	0,441	0,221
DTW	0,113	0,296	0,273	0,275	0,082
SAX (Configuração 3) + DTW	0,039	0,256	0,250	0,106	0,033
SAX (Configuração 4) + DTW	0,059	0,252	0,163	0,107	0,093

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre valor das moedas e volume de tweets que podem ser citadas:

1. Diferentemente do resultado que se esperava encontrar, foi evidenciada correlação fraca com a maior parte das moedas, sendo algumas delas correlação negativa. Entretanto, Cardano (ADA) e Ripple (XRP) demonstraram correlação positiva moderada e forte, respectivamente.
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,208, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação moderada forte.
3. SAX (Configuração 3) demonstrou boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW

Tabela 5.20 – Resultados entre *Search Volume Index* e volume de Tweets para o Evento F

Medidas	ADA	BTC	ETH	LTC	XRP
Correlação de Pearson	0,837	0,435	0,675	0,643	0,832
Distância Euclidiana	0,202	0,313	0,227	0,360	0,169
DTW	0,071	0,196	0,164	0,211	0,083
SAX (Configuração 3) + DTW	0,072	0,167	0,133	0,072	0,067
SAX (Configuração 4) + DTW	0,052	0,133	0,130	0,163	0,104

Fonte: Próprio autor.

A partir dos resultados, pode-se identificar informações importantes entre *Search Volume Index* e volume de tweets que podem ser citadas:

1. Assim como o resultado esperado, foi evidenciada correlação positiva, em sua maioria associações fortes e moderadas, na maior parte das moedas;
2. Distâncias menores ao usar a técnica DTW, demonstrando a existência de deslocamentos temporais entre as séries. Média das distâncias calculadas pelo DTW foi de 0,145, onde, de acordo com a tabela de interpretação e assim como o esperado, demonstra uma associação moderada forte;
3. SAX (Configuração 3) demonstrou boa simplificação dos dados, reduzindo ainda mais a distância encontrada pelo DTW

Os testes envolvendo valor e *Search Volume Index* indicaram boas correlações na maioria das moedas, além de menores distâncias usando o DTW. Nos testes de valor e

volume de tweets, as correlações, em sua maioria, foram fracas e até mesmo negativas. Entretanto, Cardano e Ripple tiveram boas correlações, além de menores distâncias.

Os testes entre *Search Volume Index* e volume de tweets também indicaram boas forças de associações, tanto na correlação, quanto nas medidas de distância, reforçando a ideia de que haja uma associação entre esses fatores. A média das distâncias calculadas com o DTW, de todos os testes realizados, foi de 0,156 unidades de medida.

5.3 INTERPRETAÇÃO DOS RESULTADOS

Nesta seção serão apresentadas interpretações dos dados baseadas nas tabelas de interpretação de correlação e de medida de distância, buscando compreender e comparar os resultados obtidos.

5.3.1 Interpretações da análise entre valor da moeda e *Search Volume Index*

Nesta subseção serão analisados e interpretados os resultados obtidos através das análises entre valor da moeda e *Search Volume Index*.

5.3.1.1 Medida de correlação

Em todos os eventos a medida de correlação utilizada (correlação de Pearson) foi de acordo com o resultado esperado, entretanto, as correlações nem sempre demonstraram um forte grau de associação, apenas tendo a mesma tendência (positiva ou negativa).

Para o evento B (adoção do Bitcoin como moeda oficial de El Salvador), a correlação do Bitcoin, mesmo que negativa, não demonstrou força de associação significativa. A partir disso, podemos inferir que existiu um balanço entre a euforia positiva da adoção e a decepção da população já no primeiro dia, ocasionando o resultado fraco para a força de associação.

Abaixo, a tabela comparativa com as médias das correlações de Pearson para cada evento.

Tabela 5.21 – Média das correlações de Pearson

	Média da correlação de Pearson
Evento A	-0,477
Evento B	0,245
Evento C	0,629
Evento D	0,374
Evento E	0,244
Evento F	0,742

Fonte: Próprio autor.

Dessa forma, analisando a média das correlações para cada evento, alguns eventos pré-definidos, como o evento A, evento C e o evento F, demonstraram um grau um pouco maior de correlação. Porém, não se percebe uma regra forte que diferencie todos os eventos pré-definidos dos intervalos aleatórios. Logo, entende-se que nem todos os eventos geram uma força significativa de correlação entre os dados.

5.3.1.2 Medida de distância

Todos os eventos demonstraram um aumento da associação dos dados ao comparar distância euclidiana ao DTW, demonstrando que os dados não devem ser medidos de forma linear e direta, uma vez que os eventos comumente são associados com dias de distância, exigindo alinhamento temporal das séries de dados.

Os eventos B, C, D e E, com a utilização do DTW, demonstraram associação moderada forte. O evento A demonstrou associação moderada, e o evento F, associação forte entre os dados. De forma geral, todas as análises demonstraram um certo grau de associação entre os dados, mesmo com intervalos aleatórios, diferentemente do esperado.

Entretanto, de acordo com a tabela abaixo, percebe-se que os eventos aleatórios demonstraram uma maior média analisando os resultados da técnica de DTW. O evento A também resultou em uma média maior devido a inversão entre a tendência de queda do valor e a tendência de alta do *Search Volume Index*.

Além disso, foi necessário a análise das médias obtidas com o SAX (configuração 3, que obteve as menores medidas) para a verificação da porcentagem de redução da distância para cada um dos eventos.

Tabela 5.22 – Média das distâncias obtidas com DTW e SAX

	Média do DTW	Média do SAX (Configuração 3)	Porcentagem de redução da distância usando SAX
Evento A	0,266	0,196	26,32%
Evento B	0,166	0,099	40,36%
Evento C	0,132	0,081	38,64%
Evento D	0,242	0,146	39,67%
Evento E	0,225	0,122	45,78%
Evento F	0,115	0,082	28,70%

Fonte: Próprio autor.

Dessa forma, percebe-se que para os eventos aleatórios e para o evento A (que possui tendência inversa entre valor e *Search Volume Index*), a média obtida com a técnica de DTW manteve-se acima de 0,2, enquanto os demais eventos demonstraram uma média menor de distância entre os dados, representando melhores associações para aqueles eventos conhecidos e pré-definidos.

Além disso, a porcentagem de redução expressa pela utilização do SAX demonstrou que tal diminuição ocorreu para todos os eventos, sem que existisse uma regra clara que diferenciasse os eventos aleatórios dos eventos pré-definidos. Dessa forma, pode-se inferir que o SAX apenas realizou uma simplificação geral dos dados e não teve influência no processo de melhora das associações.

5.3.2 Interpretações da análise entre valor da moeda e volume de tweets

Nesta subseção serão analisados e interpretados os resultados obtidos através das análises entre valor da moeda e volume de Tweets.

5.3.2.1 Medida de correlação

O evento A, mesmo expressando correlações negativas em todas as moedas, não demonstrou força significativa de correlação.

Os eventos B, C, D e E também não demonstraram força significativa de correlação para a maioria das moedas, exceto Litecoin que obteve correlação de 0,689 no evento C, e Bitcoin que obteve correlação de 0,644 no evento E.

O evento F demonstrou força significativa de correlação para as moedas Cardano e Ripple, obtendo de correlação 0,716 e 0,780 respectivamente. As demais moedas não

apresentaram forças significativas de correlação.

Abaixo, uma tabela comparativa com as médias das correlações de Pearson para cada evento.

Tabela 5.23 – Média das correlações de Pearson

	Média da correlação de Pearson
Evento A	-0,142
Evento B	0,135
Evento C	0,413
Evento D	0,054
Evento E	0,33
Evento F	0,259

Fonte: Próprio autor.

Assim, percebe-se que os eventos não demonstraram força significativa de correlação, não existindo uma regra forte que possa diferenciar os intervalos aleatórios dos eventos pré-definidos. Além disso, os resultados permitem inferir que os movimentos no volume de tweets não são correlacionados aos movimentos nos valores das moedas, mesmo com a existência de eventos relevantes.

5.3.2.2 Medida de distância

Como visto anteriormente, todos os eventos demonstraram um aumento da associação dos dados ao comparar a distância euclidiana ao DTW, demonstrando, mais uma vez, que os dados não devem ser medidos de forma linear e direta, exigindo alinhamento temporal das séries de dados.

Entretanto, em todos os eventos, ao analisar os resultados do DTW, poucas associações fortes foram encontradas, demonstrando um fraco grau de associação entre as séries analisadas. Apenas para o evento F foi encontrada uma distância abaixo de 0,1 unidades de medida na moeda Ripple (XRP).

Abaixo, a média das distâncias obtidas com DTW e SAX.

Tabela 5.24 – Média das distâncias obtidas com DTW e SAX

	Média do DTW	Média do SAX (Configuração 3)	Porcentagem de redução da distância usando SAX
Evento A	0,287	0,166	42,16%
Evento B	0,219	0,108	50,68%
Evento C	0,161	0,129	19,88%
Evento D	0,262	0,162	38,17%
Evento E	0,239	0,123	48,54%
Evento F	0,208	0,137	34,13%

Fonte: Próprio autor.

De acordo com os resultados obtidos pelo DTW, os eventos aleatórios, bem como o evento A (que possui tendência inversa entre valor e volume de tweets) obtiveram médias moderadamente maiores do que os demais eventos. As médias dos intervalos com eventos relevantes e de mesma tendência demonstraram médias de DTW abaixo dos 0,22, enquanto os eventos aleatórios juntamente ao evento A demonstraram médias maiores que 0,238.

Como já visto, mais uma vez, a porcentagem de redução expressa pela utilização do SAX demonstrou que tal diminuição ocorreu para todos os eventos, sem que existisse uma regra clara que diferenciasse os eventos aleatórios dos eventos pré-definidos. Dessa forma, pode-se inferir que o SAX apenas realizou uma simplificação geral dos dados e não teve influência no processo de melhora das associações.

5.3.3 Interpretações da análise entre *Search Volume Index* e volume de tweets

Nesta subseção serão analisados e interpretados os resultados obtidos através das análises entre *Search Volume Index* e volume de tweets.

5.3.3.1 Medida de correlação

Todos os eventos demonstraram boas correlações, exceto o evento A.

O evento A demonstrou força significativa de correlação apenas para Bitcoin, as demais moedas demonstraram correlações fracas.

Nos eventos B, C, D, E e F foram observados, em sua maioria, correlações moderadas e fortes, demonstrando a associação entre as métricas de *Search Volume Index* e volume de tweets, independente do intervalo de tempo possuir eventos relevantes.

Abaixo, uma tabela comparativa com as médias das correlações de Pearson para cada evento.

Tabela 5.25 – Média das correlações de Pearson

	Média da correlação de Pearson
Evento A	0,26
Evento B	0,718
Evento C	0,615
Evento D	0,514
Evento E	0,494
Evento F	0,684

Fonte: Próprio autor.

Assim, percebe-se que os eventos, em sua maioria, demonstraram força significativa de correlação, permitindo inferir que exista correlação entre as métricas do *Search Volume Index* e volume de tweets em boa parte do intervalo de tempo, mesmo sem a existência de um evento relevante.

5.3.3.2 Medida de distância

Reforçando os encontrados anteriores, todos os eventos demonstraram um aumento da associação dos dados ao comparar a distância euclidiana ao DTW, demonstrando que os dados não devem ser medidos de forma linear e direta, exigindo alinhamento temporal das séries de dados, uma vez que os eventos comumente são associados com dias de distância.

Os eventos, em geral, demonstraram boas medidas de distância. Diversas associações fortes e moderadas fortes foram encontradas nos eventos, o que também causou médias mais baixas das distâncias, como mostra a tabela abaixo.

O evento A demonstrou associações moderadas fortes em todas as moedas. O evento B e C demonstraram, em sua maioria, associações moderadas, além de 2 associações fortes em cada evento. Os eventos D, E e F também demonstraram, principalmente, associações moderadas e fortes.

Tabela 5.26 – Média das distâncias obtidas com DTW e SAX

	Média do DTW	Média do SAX (Configuração 3)	Porcentagem de redução da distância usando SAX
Evento A	0,192	0,149	22,40%
Evento B	0,125	0,09	28,00%
Evento C	0,135	0,122	9,63%
Evento D	0,19	0,144	24,21%
Evento E	0,195	0,111	43,08%
Evento F	0,145	0,102	29,66%

Fonte: Próprio autor.

Entre todas as médias das distâncias obtidas nas análises anteriores, a combinação de *Search Volume Index* e volume de tweets resultou na menor distância média geral, demonstrando a associação maior dessas métricas, independente do intervalo de tempo.

Entretanto, ao se observar apenas os eventos relevantes pré-definidos (eventos A, B, C, F), a média diminuiu mais, indo para 0,149 em comparação a média de 0,193 dos eventos aleatórios (evento D e E).

E mais uma vez, a porcentagem de redução expressa pela utilização do SAX demonstrou que tal diminuição ocorreu para todos os eventos, sem que existisse uma regra clara que diferenciasse os eventos aleatórios dos eventos pré-definidos. Dessa forma, pode-se inferir que o SAX apenas realizou uma simplificação geral dos dados e não teve influência no processo de melhora das associações.

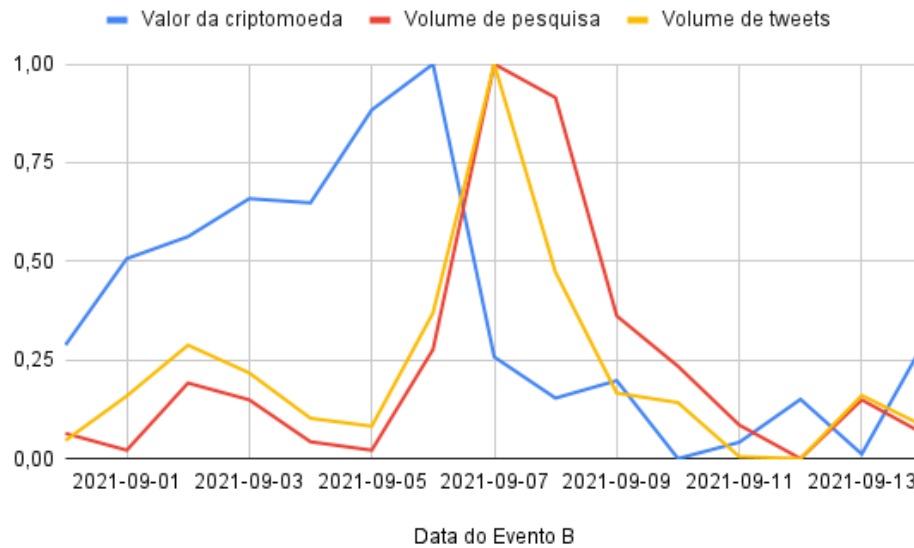
5.3.4 Análise da melhor técnica

Para que a definição da melhor técnica de análise seja feita, é importante perceber e entender o objetivo principal do trabalho.

Tendo por objetivo encontrar a melhor associação entre as séries de dados em busca de possíveis tendências correlatas, o trabalho exige que os dados sejam analisados de forma flexível em relação ao tempo, ou seja, tendências correlatas nem sempre estarão expressas no mesmo alinhamento temporal.

Um dos exemplos para tal caso, é o da moeda Bitcoin nas análises do evento B, como expressa o gráfico abaixo.

Figura 5.3 – Bitcoin no evento B: visualização geral dos dados



Fonte: Próprio autor.

Para tal caso, a correlação de Pearson da medida entre valor e *Search Volume Index* resultou em $-0,215$, ou seja, correlação negativa quase nula. Além disso, a distância euclidiana, que assim como a correlação baseia-se em análises sem o ajuste temporal, resultou em $0,492$, ou seja, uma associação fraca. Entretanto, para a técnica de Dynamic Time Warping, que trabalha com o ajuste temporal das séries, a distância foi de $0,176$, ou seja, uma associação moderada forte.

Além disso, outro fator a ser considerado é a técnica de Symbolic Aggregate Approximation. Como verificado em todos os testes, a porcentagem de diminuição da distância a partir da utilização do SAX ocorreu de modo semelhante em todos os eventos, não tendo uma regra que diferenciasse os eventos relevantes pré-definidos dos intervalos aleatórios. Dessa forma, a técnica mostrou realizar uma simplificação geral dos dados, reduzindo a distância de todos os eventos. Mesmo que alterações nas configurações possam demonstrar melhores resultados, para as configurações analisadas, não foi possível de identificar uma melhora nas associações aos eventos definidos.

Assim, tendo em vista que o trabalho busca pela melhor associação entre os dados, a técnica de Dynamic Time Warping pode ser classificada como a melhor técnica na busca por esse resultado.

5.3.5 Análise da influência dos eventos relevantes nos resultados

Para analisar a influência dos eventos, foi realizada a média dos resultados dos eventos relevantes (eventos A, B, C e F) e, para comparação, a média dos resultados dos eventos aleatórios (eventos D e E).

As tabelas abaixo evidenciam os resultados de acordo com as séries analisadas, buscando demonstrar a influência dos eventos relevantes nos resultados.

Tabela 5.27 – Resultados da análise entre valores das moedas e *Search Volume Index*

	Média da Correlação de Pearson	Média da técnica de DTW
Eventos relevantes	0,523	0,170
Eventos aleatórios	0,309	0,234

Fonte: Próprio autor.

Tabela 5.28 – Resultados da análise entre valores das moedas e volume de tweets

	Média da Correlação de Pearson	Média da técnica de DTW
Eventos relevantes	0,237	0,219
Eventos aleatórios	0,192	0,251

Fonte: Próprio autor.

Tabela 5.29 – Resultados da análise entre *Search Volume Index* e volume de tweets

	Média da Correlação de Pearson	Média da técnica de DTW
Eventos relevantes	0,569	0,149
Eventos aleatórios	0,504	0,193

Fonte: Próprio autor.

Assim, percebe-se uma maior força de correlação de Pearson, e também uma menor distância utilizando a técnica de Dynamic Time Warping, ao utilizar intervalos de tempo que possuam eventos relevantes, evidenciando a influência dos eventos externos nas medidas de associação entre as séries. Mesmo para as combinações que não demonstraram boas associações (como valor e volume de tweets demonstrado anteriormente), ainda assim, as médias demonstraram certa diferenciação dos eventos relevantes para os eventos aleatórios.

6 CONCLUSÃO

Por ser um ambiente descentralizado, o mercado de criptomoedas movimenta-se, em geral, baseado nas expectativas dos investidores. Dessa forma, e assim como apontam os resultados, determinados eventos externos podem gerar impactos, imediatos ou não, ao causar euforia ou descontentamento neste mercado. E assim, demonstram ainda mais o poder das autoridades e dos influenciadores em gerar tais movimentos. Dessa forma, identificar a temperatura do mercado em diferentes abordagens demonstra ser um bom indicador a ser usado na tomada de decisão. Assim como existem índices que buscam analisar os sentimentos do mercado, este trabalho buscou, através de variáveis de volume e valor, identificar as associações implícitas dos dados, buscando também, gerar interesse por novas investigações na busca pela associação dos movimentos do mercado aos eventos relevantes.

As análises de séries temporais permitem, também, diferentes abordagens de acordo com o objetivo de cada trabalho. Neste, por não haver a finalidade de criação de um modelo preditivo, as análises tornam-se mais flexíveis e, de certa forma, resultam em melhores associações, uma vez que fatores como a limitação do deslocamento de tempo podem ser ultrapassados a partir da inclusão de técnicas específicas como o Dynamic Time Warping. Assim, por demonstrarem uma melhor associação em comparação as medidas lineares, os resultados demonstram certa atratividade na utilização de novas técnicas que utilizem ajustes de tempo ao analisar séries temporais em dados sobre o mercado de criptomoedas.

Como trabalhos futuros, na busca pelo aprimoramento das técnicas de associação entre séries, sugere-se estudar as evoluções da técnica de Dynamic Time Warping e de Symbolic Aggregate Approximation, além da realização de testes envolvendo novas configurações e maiores conjuntos de dados. Além disso, a inclusão de mais eventos relevantes e eventos aleatórios podem permitir o aumento da precisão dos resultados, verificando a influência indireta dos fatores externos. Outro ponto a ser analisado é a relação entre a quantidade de eventos relevantes e os movimentos dos dados (valor e volumes). A falta de eventos pode ser um indicativo de diminuição de euforia do mercado no geral, e com isso, pode-se perceber a redução de ambos os volumes (pesquisa e tweets), bem como a queda do valor. Assim, validar a influência dos eventos possibilitará, também, incluir o resultado e calibrar pesos de modelos preditivos com maior assertividade, trazendo avanços para esse tipo de análise.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHAM, J. et al. Cryptocurrency price prediction using tweet volumes and sentiment analysis. **SMU Data Science Review**, v. 1, n. 3, p. 1, 2018.

AGRÍCOLA, B. **Declaração sobre a proibição do uso dos serviços do nosso banco para transações de Bitcoin e outras moedas virtuais**. 2021. Disponível em: <https://www.abchina.com/cn/PersonalServices/SvcBulletin/202106/t20210621_2011168.htm>.

ALKUDMANI, F. **Cardano está prestes a concluir sua atualização mais importante; entenda o que vai mudar**. 2021. Disponível em: <<https://portaldobitcoin.uol.com.br/cardano-esta-prestes-a-concluir-sua-atualizacao-mais-importante-entendo-o-que-vai-mudar/>>.

ASSUNÇÃO, J.; FERNANDES, P.; VINCENT, J. Piecewise aggregation for HMM fitting: A pre-fitting model for seamless integration with time-series data. **Int. J. Softw. Eng. Knowl. Eng.**, v. 29, n. 11&12, p. 1835–1850, 2019. Disponível em: <<https://doi.org/10.1142/S0218194019400242>>.

ASSUNÇÃO, J. V. C. **TS Analysis Tool**. 2021. Disponível em: <<https://joaquim.shinyapps.io/TSmeasTool/>>.

BARBER, B. M.; ODEAN, T. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. **The review of financial studies**, Oxford University Press, v. 21, n. 2, p. 785–818, 2008.

BATISTA, G.; WANG, X.; KEOGH, E. A complexity-invariant distance measure for time series. In: . [S.l.: s.n.], 2011. p. 699–710.

BERNDT, D. J.; CLIFFORD, J. Using dynamic time warping to find patterns in time series. In: SEATTLE, WA, USA: **KDD workshop**. [S.l.], 1994. v. 10, n. 16, p. 359–370.

BitInfoCharts. **Cryptocurrency statistics**. 2021. Disponível em: <<https://bitinfocharts.com/>>.

BOTT, I. et al. **Technology: Banks seek the key to blockchain**. 2015. Disponível em: <<https://www.ft.com/content/eb1f8256-7b4b-11e5-a1fe-567b37f80b64>>.

CoinMarketCap. **CoinMarketCap**. CoinMarketCap, 2021. Disponível em: <<https://coinmarketcap.com/pt-br>>.

COMERCIAL, B. I. e. **A proibição do uso dos serviços do nosso banco para realizar transações em moeda virtual**. 2021. Disponível em: <<http://www.icbc.com.cn/icbc/20210621.htm>>.

CONSTRUÇÃO, B. de. **Anúncio sobre a Proibição do Uso dos Serviços do Nosso Banco para Transações em Moeda Virtual**. 2021. Disponível em: <http://www1.ccb.com.cn/v3/include/notice/20210621_1624266880.html>.

Crypto Data Download. **HISTORICAL DATA**. 2021. Disponível em: <<https://www.cryptodatadownload.com/data/>>.

FINNEY, H. **RPOW - Reusable Proofs of Work**. 2004. Disponível em: <<http://web.archive.org/web/20071222072154/http://rpow.net/>>.

Google. **Google Trends**. 2021. Disponível em: <<https://trends.google.com.br/trends>>.

HAND, D. J. Statistics and computing: the genesis of data science. **Statistics and Computing**, Springer, v. 25, n. 4, p. 705–711, 2015.

KEMP, S. **Digital 2021 April Global Statshot Report**. [S.l.], 2021.

KEOGH, E. et al. Locally adaptive dimensionality reduction for indexing large time series databases. Association for Computing Machinery, New York, NY, USA, v. 30, n. 2, p. 151–162, 2001. ISSN 0163-5808. Disponível em: <<https://doi.org/10.1145/376284.375680>>.

KOLLINTZA-KYRIAKOULIA, F. et al. Measuring the impact of financial news and social media on stock market modeling using time series mining techniques. **Algorithms**, v. 11, p. 181, 11 2018.

LAMPORT, L.; SHOSTAK, R.; PEASE, M. The byzantine generals problem. **ACM Trans. Program. Lang. Syst.**, Association for Computing Machinery, New York, NY, USA, v. 4, n. 3, p. 382–401, jul. 1982. ISSN 0164-0925. Disponível em: <<https://doi.org/10.1145/357172.357176>>.

LIN, J. et al. A symbolic representation of time series, with implications for streaming algorithms. In: **Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery**. New York, NY, USA: Association for Computing Machinery, 2003. (DMKD '03), p. 2–11. ISBN 9781450374224. Disponível em: <<https://doi.org/10.1145/882082.882086>>.

MATA, V. da S. et al. Indústria 4.0: a revolução 4.0 e o impacto na mão de obra. **Revista de Ciências Exatas e Tecnologia**, v. 13, n. 13, p. 17–22, 2018.

NAKAMOTO, S. Bitcoin: A peer-to-peer electronic cash system. **Decentralized Business Review**, p. 21260, 2008.

PEARSON, K. **Proceedings of the Royal Society of London**. Taylor & Francis, 1895. v. 58. 240, 241, 242 p. Disponível em: <<https://books.google.com.br/books?id=60aL0zIT-90C>>.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>.

REIS, T. **Exchange: o que é e como funciona uma corretora de criptoativos?** 2021. Disponível em: <<https://www.suno.com.br/artigos/exchange/>>.

SALVADOR, G. de E. **Los ojos del mundo están sobre El Salvador, que inicia una nueva era en su economía con la circulación del bitcóin**. 2021. Disponível em: <<https://www.presidencia.gob.sv/los-ojos-del-mundo-estan-sobre-el-salvador-que-inicia-una-nueva-era-en-su-economia-con-la-circulacion-del-bitcoin/>>.

SIERRA, D. **O impacto dos influenciadores no mercado financeiro**. 2021. Disponível em: <https://www.jornaldocomercio.com/_conteudo/ge2/artigos/2021/07/802097-o-impacto-dos-influenciadores-no-mercado-financeiro.html>.

SILVA, C. M. L. D. **Estatística Descritiva - Manual de Auto-Aprendizagem**. [S.l.]: Sílabo, 2018. v. 3ª edição.

XANTACROSS. **Difference in matching between Euclidean and Dynamic Time Warping**. 2011. Disponível em: <https://commons.wikimedia.org/wiki/File:Euclidean_vs_DTW.jpg>.