



UNIVERSIDADE PRESBITERIANA MACKENZIE  
FACULDADE DE COMPUTAÇÃO E INFORMÁTICA  
TECNOLOGIA EM CIÊNCIAS DE DADOS

## **PROJETO APLICADO I**

**PROFESSORES:** THIAGO GRAZIANI TRAUE, VINICIUS PIRO BARRAGAM

**GRUPO:**

EULLER NOGUEIRA – 10416092 – 10416092@mackenzista.com.br

São Paulo

2024

# SUMÁRIO

1. GLOSSÁRIO
2. APRESENTAÇÃO DOS DADOS
  - a. INFORMAÇÕES DO DATASET
  - b. METADADO
3. OBJETIVOS E METAS
4. GITHUB
5. CRONOGRAMA DE ATIVIDADES
6. ELABORAÇÃO DA PROPOSTA ANÁLITICA
  - a. APRESENTAÇÃO DA EMPRESA
  - b. ANÁLISE EXPLORATÓRIA DOS DADOS
7. ELABORAÇÃO DO ESBOÇO DO STORYTELLING
  - a. INTRODUÇÃO
  - b. ENTENDENDO OS DADOS
  - c. CÁLCULO DA VOLATILIDADE
  - d. MODELOS UTILIZADOS PARA PREVISÃO
  - e. CONCLUSÃO

# 1. GLOSSÁRIO

**Dataset** - Conjunto estruturado de dados, frequentemente organizado em tabelas, que pode ser analisado para extrair informações ou insights. Utilizado em estatísticas, ciência de dados e aprendizado de máquina.

**Metadados** - Dados que descrevem e fornecem informações sobre outros dados, incluindo detalhes como origem, formato, data de criação e estrutura. Facilitam a organização, busca e entendimento dos dados.

**Github** - Plataforma de hospedagem de código-fonte e controle de versão baseada no sistema Git. Permite colaboração entre desenvolvedores, gerenciamento de projetos e compartilhamento de software, oferecendo recursos como repositórios, pull requests e issue tracking.

**Análise Exploratória de Dados (AED)** - Processo inicial de análise de dados que visa resumir suas principais características, frequentemente utilizando visualizações gráficas. O objetivo é identificar padrões, tendências, anomalias e hipóteses para orientar investigações mais profundas.

**Séries Temporais** - Conjunto de dados coletados ou observados em intervalos regulares de tempo. Utilizadas para analisar tendências, ciclos e padrões ao longo do tempo, as séries temporais são comuns em economia, finanças, meteorologia e outras áreas que envolvem monitoramento de variáveis ao longo do tempo.

**Média Móvel** - Técnica estatística que suaviza flutuações em séries temporais ao calcular a média de um conjunto de valores ao longo de um intervalo de tempo deslizante.

# 1. APRESENTAÇÃO DOS DADOS

Os dados que selecionamos oferecem uma visão detalhada do desempenho diário das ações da empresa Microsoft entre o período de 04/01/2015 até 04/01/2021 (Precificadas em dólares americanos). O preço de abertura (Open) indica o valor da ação no início do dia de negociação, fornecendo uma base para avaliar o comportamento inicial. O preço máximo (High) representa o valor mais alto alcançado pela ação durante o dia, enquanto o preço mínimo (Low) mostra o valor mais baixo registrado e, finalmente, o preço de fechamento (Close) é o valor da ação ao final do dia, refletindo a última negociação e frequentemente utilizado para análises e previsões. Esses dados são cruciais para entender a volatilidade e a tendência das ações, ajudando investidores e analistas a tomar decisões informadas e a identificar padrões de mercado.

## a. INFORMAÇÕES DO DATASET

Colunas: 6 colunas

Linhas: 1511 linhas

## b. METADADOS

	ATRIBUTO	DESCRIÇÃO	TIPO	TAM
1	DATE	Data e hora do registro	DATETIME	-
2	OPEN	Preço de abertura do	FLOAT	16

		ativo		
<b>3</b>	CLOSE	Preço de fechamento do ativo	FLOAT	16
<b>4</b>	LOW	Preço mínimo alcançado	FLOAT	16
<b>5</b>	VOLUME	Quantidade Total	INT	19

## 2. OBJETIVO E METAS

O objetivo principal do projeto é compreender a evolução dos preços, avaliando tendências e padrões de comportamento ao longo do tempo. Buscamos identificar como os preços de abertura, fechamento, máximo e mínimo se comportam em relação às flutuações diárias, calculando a volatilidade para entender o risco associado. Além disso, pretendemos analisar o desempenho das ações comparando o retorno diário e observar como o volume de negociações pode influenciar os preços. A identificação de padrões técnicos e a comparação entre diferentes ações visam revelar oportunidades de investimento e otimizar estratégias. Finalmente, o uso de modelos de previsão ajudará a antecipar futuros movimentos de mercado, permitindo ajustes proativos nas decisões de investimento.

## 3. GITHUB

Link para o repositório no Github:

<https://github.com/nog7/PROJETO-APLICADO-1>

## 5. CRONOGRAMA DE ATIVIDADES

CRONOGRAMA PROJETO APLICADO I							
TAREFA	PRIORIDADE	STATUS	DATA DE INÍCIO	DATA DE CONCLUSÃO	% CONCLUÍDO	CONCLUÍDO?	RESPONSÁVEL
ETAPA 1							
FORMAÇÃO DE GRUPOS	Normal	Concluído	8/8/2024	8/8/2024	100%	SIM	EULLER
DEFINIÇÃO DAS PREMISSAS DO PROJETO	Normal	Concluído	25/8/2024	25/8/2024	100%	SIM	EULLER
OBJETIVOS E METAS	Normal	Concluído	26/8/2024	26/8/2024	100%	SIM	EULLER
criação do cronograma de atividades	Normal	Concluído	26/8/2024	26/8/2024	100%	SIM	EULLER
criação do repositório	Normal	Concluído	2/9/2024	2/9/2024	100%	SIM	EULLER
ETAPA 2							
ELABORAÇÃO DA PROPOSTA ANALÍTICA	Normal	Concluído	14/09/2024	30/09/2024	100%	SIM	EULLER
AJUSTES NOS OBJETIVOS E CRONOGRAMA	Normal	Concluído	10/09/2024	10/09/2024	100%	SIM	EULLER
PREPARAÇÃO, TRATAMENTO E TRANSFORMAÇÕES DA BASE DE DADOS	Normal	Concluído	17/09/2024	22/09/2024	100%	SIM	EULLER
ELABORAÇÃO DA ANÁLISE EXPLORATÓRIA EM PYTHON	Normal	Concluído	17/09/2024	22/09/2024	100%	SIM	EULLER
ESTRUTURAÇÃO DO REPOSITÓRIO	Normal	Concluído	30/09/2024	30/09/2024	100%	SIM	EULLER
DESCRIÇÃO DA BASE TEÓRICA DO MÉTODO ANALÍTICO	Normal	Concluído	25/09/2024	28/09/2024	100%	SIM	EULLER
ETAPA 3							
AJUSTE NO CÓDIGO DE ANÁLISE EXPLORATÓRIA	Normal	Concluído	12/10/2024	25/10/2024	100%	SIM	EULLER
CONSOLIDAÇÃO DOS RESULTADOS ANALÍTICOS	Normal	Concluído	25/10/2024	29/10/2024	100%	SIM	EULLER
ELABORAÇÃO DO ESBOÇO DA APRESENTAÇÃO	Normal	Concluído	28/10/2024	30/10/2024	100%	SIM	EULLER
ETAPA 4							
FINALIZAÇÃO DA ANÁLISE EXPLORATÓRIA	Normal	A Realizar	-	-	0%	NÃO	EULLER
CONCLUSÃO DE APRESENTAÇÃO	Normal	A Realizar	-	-	0%	NÃO	EULLER
ELABORAÇÃO DO DOCUMENTO FINAL	Normal	A Realizar	-	-	0%	NÃO	EULLER

Figura 1 - Cronograma de atividades.

## 6. ELABORAÇÃO DA PROPOSTA ANALÍTICA

### a. APRESENTAÇÃO DA EMPRESA

**Nome da Empresa:** New York Stock Exchange (NYSE).

**Missão/Visão/Valores:** A missão da NYSE é utilizar tecnologia e informações para solucionar os problemas dos seus clientes, com foco em eficiência e inovação

**Segmento de atuação:** A NYSE é uma das maiores bolsas de valores do mundo, atuando no segmento de mercados financeiros e de capitais.

**Número de colaboradores:** A NYSE faz parte do Intercontinental Exchange (ICE), que tem aproximadamente 9.000 colaboradores ao redor do mundo.

**Iniciativas na área de Data Science:** A NYSE investe fortemente em iniciativas digitais, incluindo a criação de plataformas de liderança em tecnologia e capital, promovendo debates sobre economia digital, fintechs e análise de dados.

**Trabalhos em destaque:** A NYSE destaca-se por seus eventos globais, como o lançamento do plano "Abenomics" do Japão e suas colaborações com bolsas internacionais, como a de Singapura, além de suas contribuições para o desenvolvimento de mercados financeiros globais.

## **b. ANÁLISE EXPLORATÓRIA DE DADOS**

Iniciamos a elaboração de nossa proposta analítica, realizando a análise exploratória dos dados. Esse processo visa realizar as verificações iniciais e o tratamento necessário na base, a fim de torná-la apta para as análises e insights. Utilizamos a linguagem Python, mais especificamente as bibliotecas Pandas para manipulação do dataset e análise exploratória inicial, Matplotlib para plotar os gráficos baseados nas análises e manipulações com Pandas, e Streamlit para fornecer uma interface gráfica mais robusta ao código.

**HEAD** – O head (Figura 2) representa as 5 primeiras linhas do nosso dataset e nos dá uma noção inicial sobre a estrutura e o conteúdo dos dados. Através dessa visualização já podemos inferir algumas coisas, como, tipos de dados, estrutura das colunas, contexto do dataset, etc.

## Head:

Date	Open	High	Low	Close	Volume
2015-04-01 16:00:00	40.6	40.76	40.31	40.72	36,865,322
2015-04-02 16:00:00	40.66	40.74	40.12	40.29	37,487,476
2015-04-06 16:00:00	40.34	41.78	40.18	41.55	39,223,692
2015-04-07 16:00:00	41.61	41.91	41.31	41.53	28,809,375
2015-04-08 16:00:00	41.48	41.69	41.04	41.42	24,753,438

Figura 2 - Head

**TAIL** – O tail (Figura 3) exibe as últimas 5 linhas do nosso dataset e é uma ferramenta igualmente útil para entender a parte final dos dados. Ele também nos permite inferir algumas coisas, como os dados recentes, por exemplo.

## Tail:

Date	Open	High	Low	Close	Volume
2021-03-25 16:00:00	235.3	236.94	231.57	232.34	34,061,853
2021-03-26 16:00:00	231.55	236.71	231.55	236.48	25,479,853
2021-03-29 16:00:00	236.59	236.8	231.88	235.24	25,227,455
2021-03-30 16:00:00	233.53	233.85	231.1	231.85	24,792,012
2021-03-31 16:00:00	232.91	239.1	232.39	235.77	43,623,471

Figura 3 - Tail

**Estatísticas Descritivas** – As estatísticas descritivas (Figura 4) são ferramentas essenciais na análise de dados, pois fornecem um resumo numérico das características principais de um dataset. Elas ajudam a entender a distribuição,



centralidade e dispersão dos dados. Aqui estão os principais componentes das estatísticas descritivas:

Estatísticas descritivas: ↔					
	Open	High	Low	Close	Volume
count	1,511	1,511	1,511	1,511	1,511
mean	107.386	108.4375	106.2945	107.4221	30,198,625.0814
std	56.6913	57.3823	55.9772	56.7023	14,252,659.0873
min	40.34	40.74	39.72	40.29	101,612
25%	57.86	58.06	57.42	57.855	21,362,129
50%	93.99	95.1	92.92	93.86	26,629,615
75%	139.44	140.325	137.825	138.965	34,319,615.5
max	245.03	246.13	242.92	244.99	135,227,059

Figura 4 - Estatísticas Descritivas

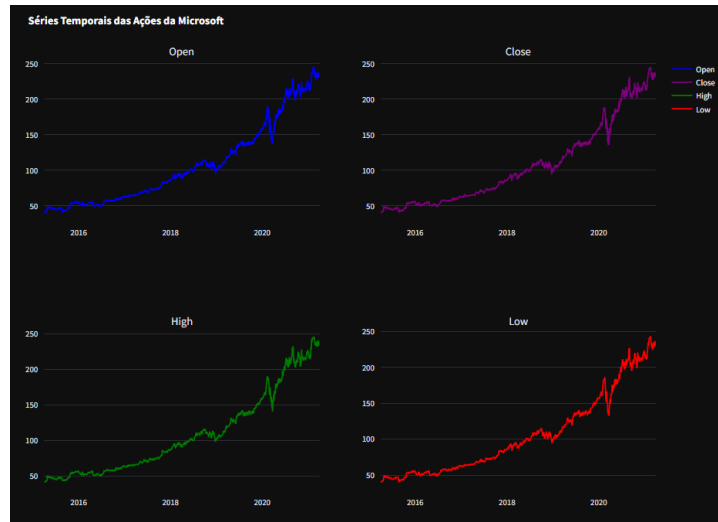
**Verificação de Valores Ausentes** – A verificação de valores ausentes é uma etapa crucial no processo de limpeza e preparação de dados. Dados faltantes podem impactar significativamente as análises e modelagens, levando a resultados imprecisos ou enviesados.

Verificação de valores ausentes:	
	0
Open	0
High	0
Low	0
Close	0
Volume	0

Figura 5 - Verificação de Valores Ausentes

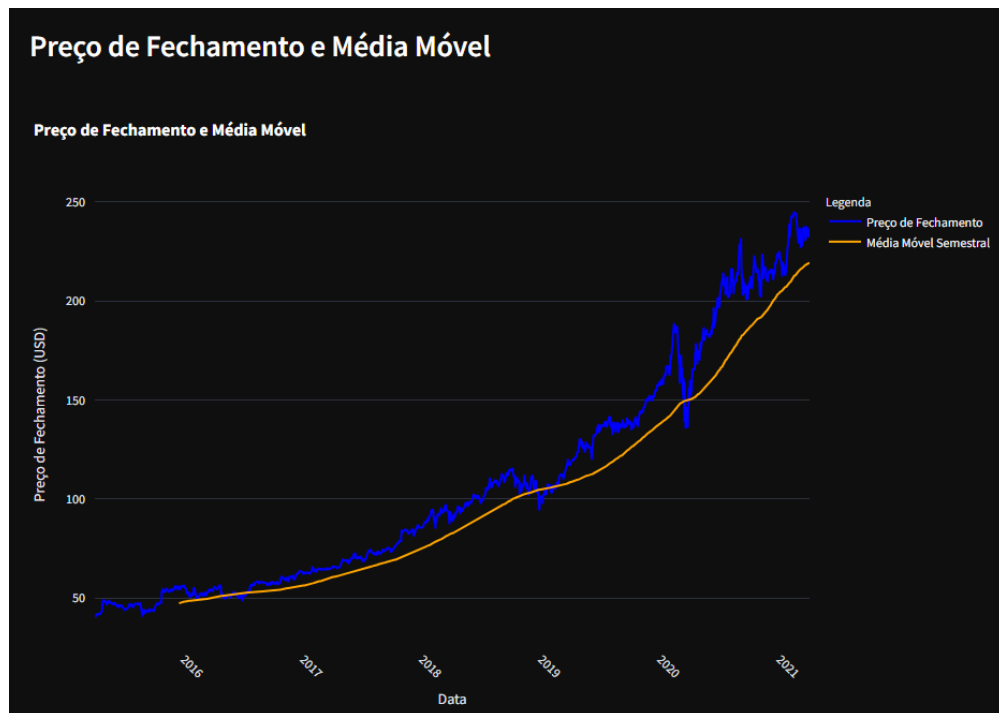
**Gráficos de séries temporais** - Os gráficos de séries temporais (Figura 6) servem para ilustrar as variáveis open, close, high e low dos preços das ações da empresa, permitindo visualizar claramente as tendências de longo prazo no comportamento do mercado. Esses gráficos são essenciais para

avaliar a volatilidade das ações, pois mostram as flutuações entre os preços máximos e mínimos ao longo do tempo. Além disso, eles ajudam a identificar pontos de entrada e saída estratégicos para operações de compra e venda



**Figura 6** – Gráficos de séries temporais

**Preço de Fechamento e Média Móvel** - O gráfico de série temporal do preço de fechamento das ações (Figura 7), com a linha de média móvel, é fundamental para analisar o comportamento do ativo ao longo do tempo. Ele mostra as oscilações diárias do preço de fechamento e destaca a tendência geral, com a média móvel suavizando as flutuações. Quando o preço está acima da média, indica tendência de alta; quando está abaixo, sugere tendência de baixa. Os cruzamentos entre o preço e a média móvel servem como sinais importantes para decisões de compra ou venda, auxiliando os investidores em suas escolhas.



**Figura 6** – Preço de Fechamento e Média Móvel

**Mudança Diária no Preço** - O gráfico de mudança diária do preço das ações (Figura 7), calculado pela diferença entre os preços de fechamento e abertura, é crucial para analisar a volatilidade e o comportamento intradiário do mercado. Ele permite visualizar rapidamente a amplitude das oscilações diárias, ajudando investidores a identificar tendências de alta ou baixa, pontos de reversão e padrões de comportamento do preço em um intervalo diário. Esse tipo de análise é fundamental para estratégias de curto prazo, como day trade, além de fornecer insights sobre o sentimento do mercado e possíveis eventos de alta relevância.



Figura 7 – Mudança Diária no Preço

## 7. ELABORAÇÃO DO ESBOÇO DO STORYTELLING

### a. INTRODUÇÃO

**Tema Central:** A importância de compreender a evolução dos preços das ações e como isso impacta decisões de investimento.

**Objetivo:** Apresentar a análise de preços de abertura, fechamento, máximo e mínimo, destacando a volatilidade e o risco associado e mostrar como a análise de dados e modelos de previsão podem oferecer insights valiosos para investidores.

## **b. ENTENDENDO OS DADOS**

### **Apresentação dos Dados:**

1. Descrição das variáveis de interesse: preços de abertura, fechamento, máximo, mínimo e volume de negociações.

### **Análise Exploratória:**

1. Utilização de gráficos de série temporal para visualizar a evolução dos preços ao longo do tempo.
2. Identificação de tendências e flutuações diárias, destacando momentos de alta volatilidade.

## **c. CÁLCULO DA VOLATILIDADE**

### **Cálculo da volatilidade**

1. Explicação do que é volatilidade e como ela é medida (por exemplo, desvio padrão dos retornos).
2. Apresentação dos resultados e o que eles indicam sobre o risco das ações analisadas.

## **d. MODELOS UTILIZADOS PARA PREVISÃO**

### **Importância da Previsão**

1. Discussão acerca de como a previsão pode ajudar a antecipar movimentos de mercado

### **Uso do Modelo SARIMAX**

1. Introdução do modelo SARIMAX e como ele é aplicado para prever futuros movimentos de preço
2. Apresentação dos resultados das previsões

## **e. CONCLUSÃO**

### **Resumo dos Insights Obtidos:**

1. Recapitulação das descobertas sobre a evolução dos preços, volatilidade, e a análise de desempenho das ações.

