

Análise Inferencial de Dados – Entrega 1

Descrição do projeto:

Este projeto de análise de dados tem como objetivo investigar o desempenho de diversas empresas, índices e contratos futuros ao longo dos últimos 20 anos, com o intuito de compreender como os comportamentos dos mercados financeiros se relacionam e funcionam. Além disso, buscamos explorar a possibilidade de realizar previsões baseadas em eventos históricos. Os dados utilizados estão armazenados em um banco de dados MySQL e incluem informações detalhadas sobre o fechamento diário de várias empresas, organizadas por ramo de atuação. Utilizando uma API desenvolvida em Node.js, extraímos dados específicos — como a data e o valor de fechamento de um determinado símbolo dentro de um período definido — para serem analisados em Python através de um Jupyter Notebook. As análises realizadas englobam desde cálculos estatísticos básicos, como média, quartis e desvio padrão, até técnicas mais avançadas, como regressão linear, intervalos de confiança e testes de hipóteses. O objetivo final é identificar padrões, tendências e correlações significativas que possam fornecer insights valiosos sobre o comportamento do mercado, auxiliando na tomada de decisões estratégicas e financeiras.

Captura dos dados:

Para capturar os dados, usamos a biblioteca requests para chamar a nossa API:

```
requests.get(url, params=params)
```

E guardar eles em um dataframe da biblioteca Pandas:

```
data = response.json()
df = pd.DataFrame(data)
```

Para a demonstração das entregas, selecionamos o valor do fechamento das ações da Apple entre Jan/2022 e Dez/2022, e obtemos os seguintes resultados.

Média: 152.94

A média dos preços de fechamento da Apple Inc. durante o período analisado é de \$152,94. Este valor representa o preço médio pelo qual as ações fecharam no mercado, servindo como uma medida de tendência central. Isso indica que, em média, os investidores estavam dispostos a pagar cerca de \$152,94 por ação nesse período, refletindo a percepção geral do valor da empresa no mercado.

O valor foi obtido a partir do seguinte método:

```
media = df['Fechamento'].mean()
```

Moda: [134.19, 139.91, 145.26, 146.88, 147.4, 148.78, 152.6, 163.73, 166.55, 169.6]

A moda é o valor ou valores que ocorrem com maior frequência no conjunto de dados. Neste caso, temos múltiplas modas, indicando que os preços \$134,19, \$139,91, \$145,26, \$146,88, \$147,40, \$148,78, \$152,60, \$163,73, \$166,55 e \$169,60 ocorreram com a mesma frequência máxima. A presença de várias modas sugere que não houve um único preço dominante, mas sim vários preços de fechamento que se repetiram ao longo do período, refletindo a volatilidade natural do mercado de ações.

Os valores foram obtidos a partir do seguinte método:

```
moda = df['Fechamento'].mode()
```

Quartis:

1º Quartil (25%): 143.11

2º Quartil (Mediana, 50%): 152.19

3º Quartil (75%): 163.73

Os quartis dividem o conjunto de dados em quatro partes iguais, proporcionando insights sobre a distribuição dos preços de fechamento. O primeiro quartil (\$143,11) indica que 25% dos preços foram menores ou iguais a esse valor. A mediana (\$152,19) mostra que metade dos preços estava abaixo desse valor e a outra metade acima. O terceiro quartil (\$163,73) revela que 75% dos preços foram menores ou iguais a esse valor. Essa distribuição sugere que a maioria dos preços de fechamento está concentrada entre \$143,11 e \$163,73.

Os valores foram obtidos a partir do seguinte método:

```
quartis = df['Fechamento'].quantile([0.25, 0.5, 0.75])
```

Distância Interquartil (IQR): 20.62

A distância interquartil (IQR) é a diferença entre o terceiro e o primeiro quartil: \$163,73 - \$143,11 = \$20,62. O IQR mede a dispersão dos 50% centrais dos dados, mostrando que há uma variação de \$20,62 nos preços de fechamento médios. Um IQR maior indica maior dispersão entre os preços médios, enquanto um IQR menor sugere que os preços estão mais concentrados.

O valor foi obtido a partir do seguinte método:

```
iqr = quartis[0.75] - quartis[0.25]
```

Variância: 161.94

A variância é uma medida da dispersão dos dados em relação à média. Com uma variância de 161,94, isso indica que os preços de fechamento têm um grau moderado de dispersão ao redor da média. Uma variância mais alta significaria que os preços estão mais espalhados, enquanto uma variância mais baixa indicaria que os preços estão mais próximos da média.

O valor foi obtido a partir do seguinte método:

```
variancia = df['Fechamento'].var()
```

Desvio Padrão: 12.73

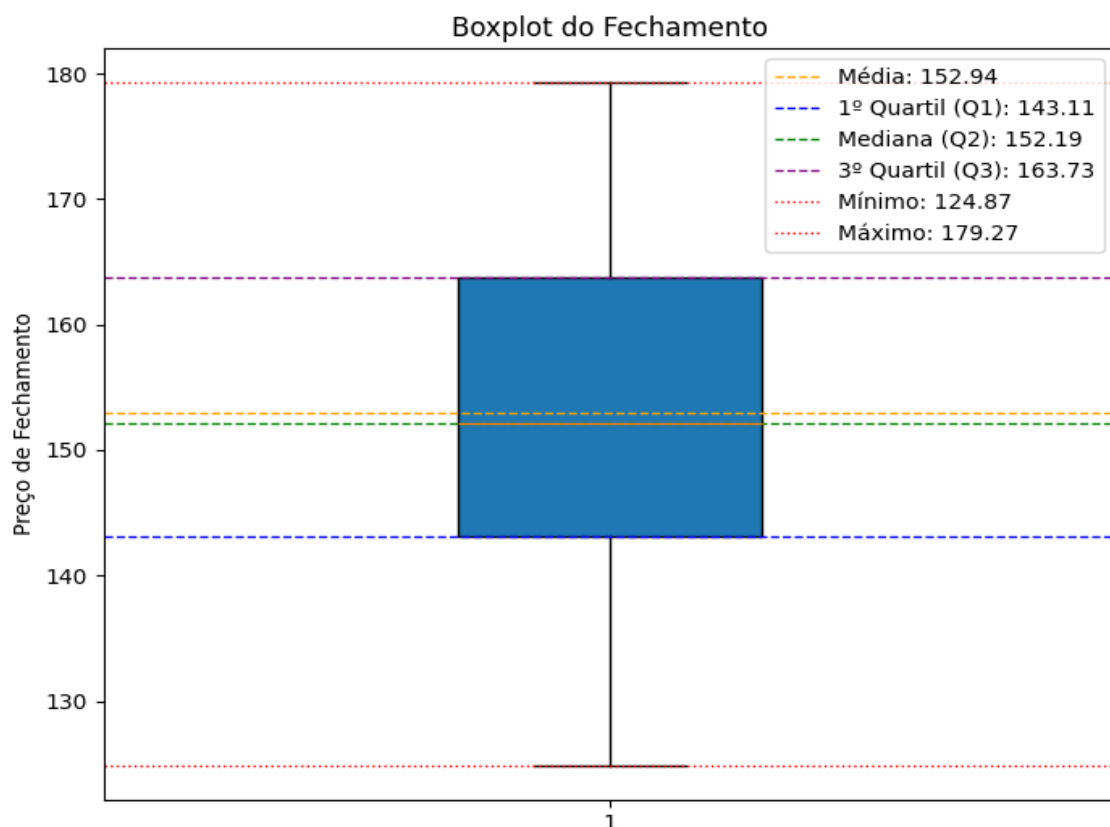
O desvio padrão, que é a raiz quadrada da variância, é de \$12,73. Isso significa que, em média, os preços de fechamento desviam-se da média em \$12,73. O desvio padrão fornece uma compreensão mais intuitiva da variabilidade dos preços, indicando que os preços flutuaram aproximadamente \$12,73 acima ou abaixo da média durante o período analisado.

O valor foi obtido a partir do seguinte método:

```
desvio_padrao = df['Fechamento'].std()
```

Boxplot

A distribuição parece ser ligeiramente simétrica, com poucos valores extremos. A presença de valores nos extremos sugere uma dispersão moderada dos preços, com uma tendência central bem definida ao redor da mediana e da média.



Histograma

A concentração de frequências em torno da média sugere estabilidade no preço das ações durante o período analisado. Essa estabilidade pode indicar uma percepção consistente de valor por parte do mercado. A forma simétrica, sem caudas longas, reforça a ideia de que os preços não oscilaram amplamente em relação à média.

