

Ciência de Dados e Analytics

Sprint: Engenharia de Dados

Aluno: Roberto Ritschel da Silva

1. Objetivo

Obter e analisar os dados de negociação do mercado futuro de Bitcoin através da API da Binance, afim de identificar padrões e tendências. Abaixo, segue as perguntas das quais esse estudo deve responder:

- Qual o topo do preço no período estudado, ou seja, qual o seu valor máximo?
- Qual o fundo do preço no período estudado, ou seja, qual o seu valor mínimo?
- Há algum tipo de relação entre o indicador do long/short ratio e o preço de fechamento do ativo?
- Há algum tipo de relação entre o indicador do open interest e o preço de fechamento do ativo?

2. Coleta de Dados

Os dados foram obtidos utilizando API Request em Python, com chave de autenticação própria e de uso pessoal. Para obter o token de autorização, basta criar uma conta na Binance e seguir o passo a passo disponível em:

<https://www.binance.com/en/support/faq/how-to-create-api-keys-on-binance-360002502072>

Já para utilizar a API, a documentação da mesma encontra-se disponível em:

<https://binance-docs.github.io/apidocs/spot/en/#general-info>

O período escolhido para estudos foi de 2 semanas de dados, entre o período de 11/06/2024 a 26/06/2024, na qual obtive-se os seguintes dados:

- Preço de abertura, fechamento, máximo e mínimo para o intervalo gráfico de 5 minutos.
- Indicador de Long/Short Ratio, ou seja, qual a proporção de posições em long (compradas) no mercado em relação as posições de short (vendidas).
- Indicador de Open Interest, ou seja, qual a quantidade total de contratos em aberto no mercado e seu respectivo volume financeiro.

Os dados foram obtidos e então exportados para arquivos xlsx, na qual foram transformados em csv e então submetidos à plataforma Databricks utilizando código no jupyter notebook, disponível no arquivo MVP Final.ipynb em anexo à este documento.

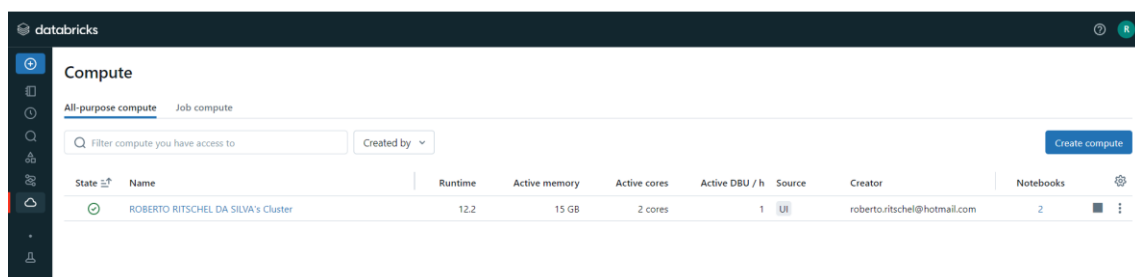


Figura 1: Cluster criado

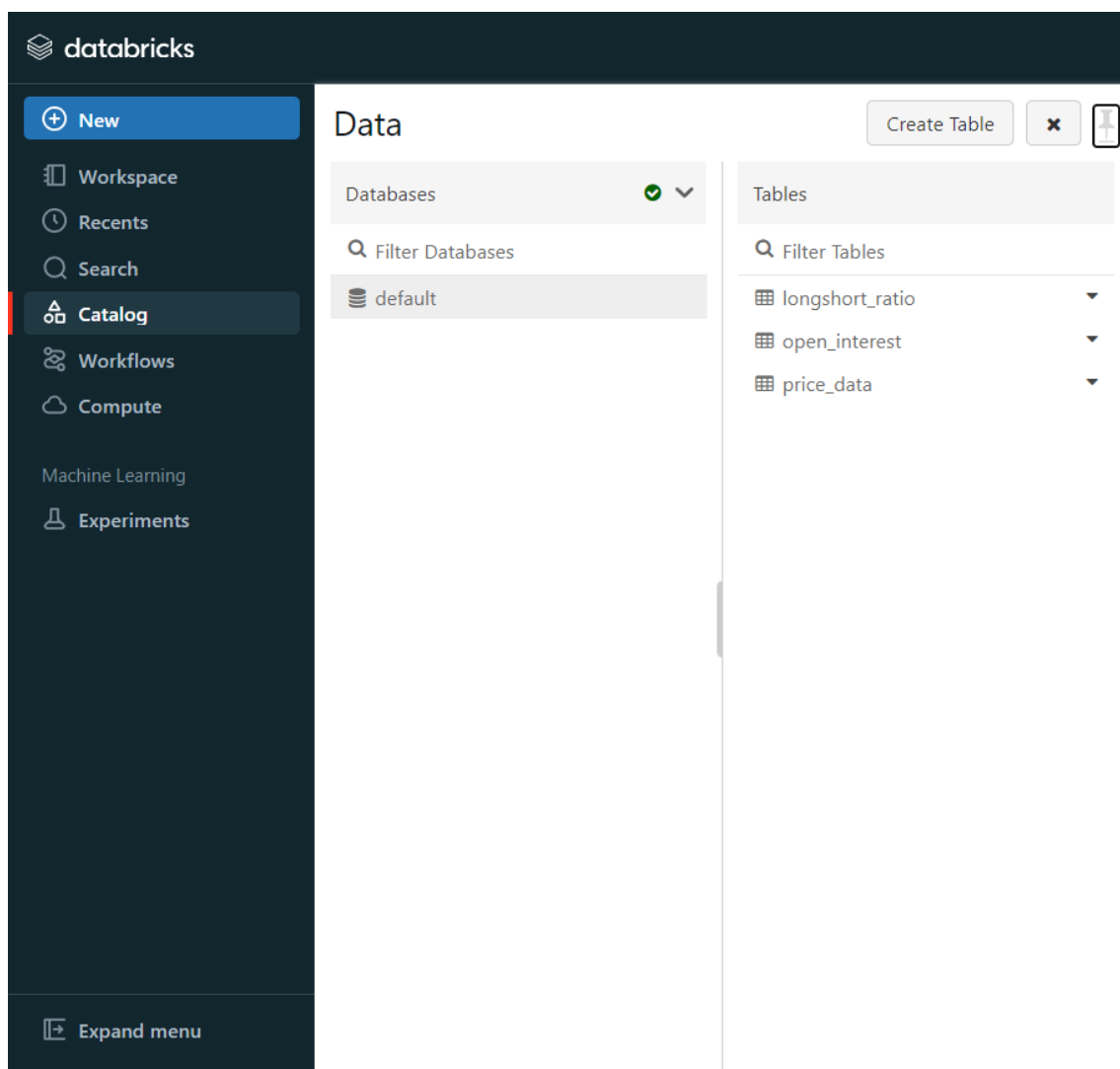


Figura 2: Tabelas disponibilizadas na plataforma “Databricks”

3. Catálogo de Dados

open_interest

Nome da Coluna	Tipo de Dado	Comentário
symbol	texto (str)	Indica qual é o ativo cujo indicador OpenInterest se refere
sumOpenInterest	texto (str)	Indica a quantidade de contratos ativos no momento
sumOpenInterestValue	texto (str)	Indica o valor financeiro desses contratos ativos
Datetime	texto (str)	Indica a data,hora e ms ao qual esse indicador foi extraído

Figura 3: Metadados da Base “open_interest”

longshort_ratio

Nome da Coluna	Tipo de Dado	Comentário
symbol	texto (str)	Indica qual é o ativo cujo indicador OpenInterest se refere
Long Account	texto (str)	Indica o % das posições compradas (em long)
Long/Short Ratio	texto (str)	Indica a relação entre o % das posições compradas e % das posições vendidas
Short Account	texto (str)	Indica o % das posições vendidas (em short)
Datetime	texto (str)	Indica a data,hora e ms ao qual esse indicador foi extraído

Figura 4: Metadados da Base “longshort_ratio”

price_data

*obs: tempo gráfico (candle)= 5min

Nome da Coluna	Tipo de Dado	Comentário
Open Time	texto (str)	Indica qual é o horário de abertura do candle
Open	texto (str)	Indica qual o preço de abertura do candle
High	texto (str)	Indica qual o valor máximo do candle
Low	texto (str)	Indica qual o valor mínimo do candle
Close	texto (str)	Indica qual o valor de fechamento do candle
Volume	texto (str)	Indica o volume financeiro total negociado naquele candle

Figura 5: Metadados da Base “price_data”

4. Modelagem

Foi escolhido o modelo estrela para o relacionamento entre as bases, na qual o diagrama foi realizado através da ferramenta genmymodel, disponível em: <http://www.genmymodel.com/>. Abaixo, segue o resultado do diagrama:

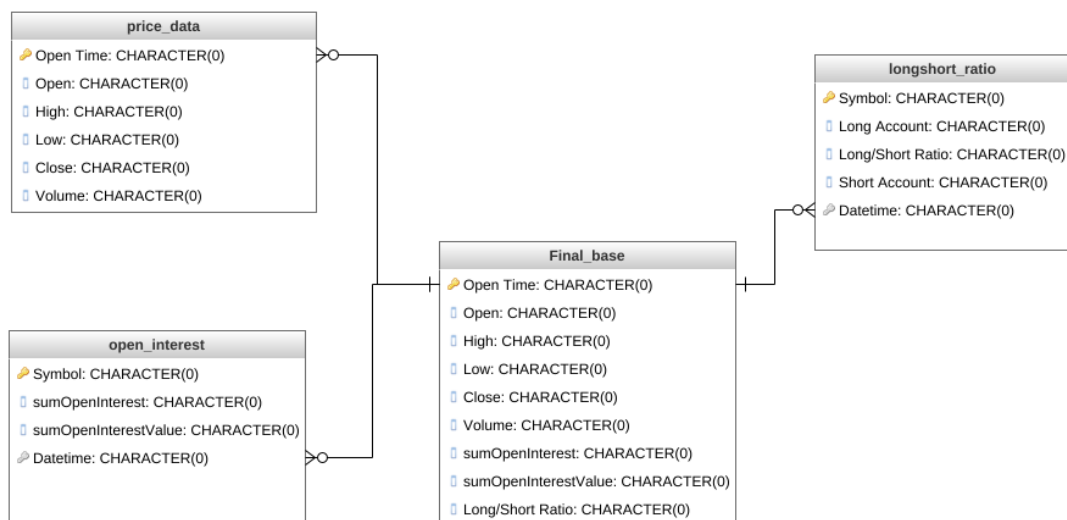


Figura 6: Esquema relacional das bases

5. Tratamento, Análise dos Dados e Resultados obtidos:

5.1 Qualidade dos dados:

Para a obtenção e tratamento dos dados, não havia problema com dados faltantes (seja nulos/vazios), visto que negociação de criptomoedas, diferentemente das bolsas de valores, funcionam 24hrs. Os tipos de dados foram os quesitos mais desafiadores, visto que os dados obtidos de preço, open interest e long/short ratio são separados por “:”, enquanto no csv havia a necessidade de trabalhar com “;”. Após essa conversão, houve nova necessidade de conversão para “.” Afim de transformar o tipo de dado de string para double.

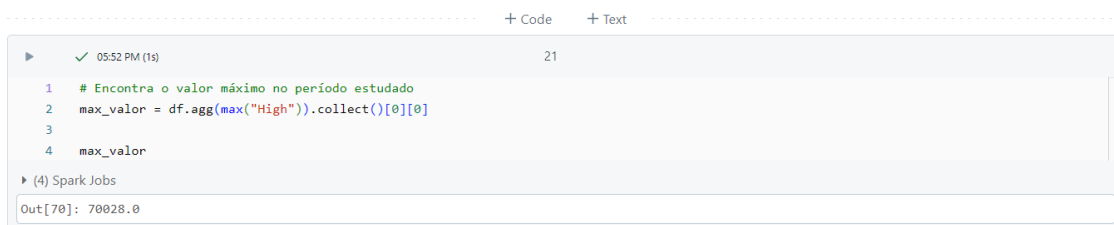
Preferi subir os dados todos como tipo string para que as alterações fossem realizadas dentro do próprio estudo. Além disso, houve a necessidade de renomear algumas colunas com objetivo de facilitar os “joins”.

5.2 Solução do problema:

Pergunta 1: Qual o topo do preço no período estudado, ou seja, qual o seu valor máximo?

R: U\$70.028,00

Verificando o topo do período estudado



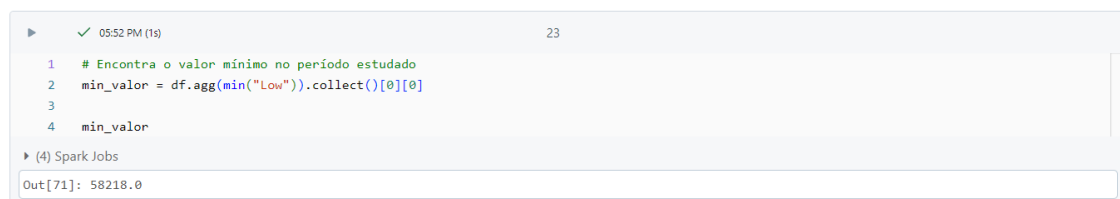
```
+ Code + Text
05:52 PM (1s) 21
1 # Encontra o valor máximo no período estudado
2 max_valor = df.agg(max("High")).collect()[0][0]
3
4 max_valor
(4) Spark Jobs
Out[70]: 70028.0
```

Figura 7: Valor máximo obtido

Pergunta 2: Qual o fundo do preço no período estudado, ou seja, qual o seu valor mínimo?

R: U\$58.218,00

Verificando o fundo do período estudado



```
05:52 PM (1s) 23
1 # Encontra o valor mínimo no período estudado
2 min_valor = df.agg(min("Low")).collect()[0][0]
3
4 min_valor
(4) Spark Jobs
Out[71]: 58218.0
```

Figura 8: Valor mínimo obtido

Pergunta 3: Há algum tipo de relação entre o indicador do long/short ratio e o preço de fechamento do ativo?

R: De acordo com o gráfico de correlação a seguir, o valor de r é negativo e muito próximo a 1, portanto, indica que, quanto maior o valor de long/short ratio, menor é o preço de fechamento. Com isso, obtemos que, quanto maior o número de posições em aberto compradas no mercado futuro de Bitcoin em relação a quantidade de posições vendidas, maior a chance do preço cair e liquidar essas posições, indicando que deve-se tomar cuidado com o aumento do preço e do indicador long/short ratio, pois a tendência é o preço desabar em seguida.

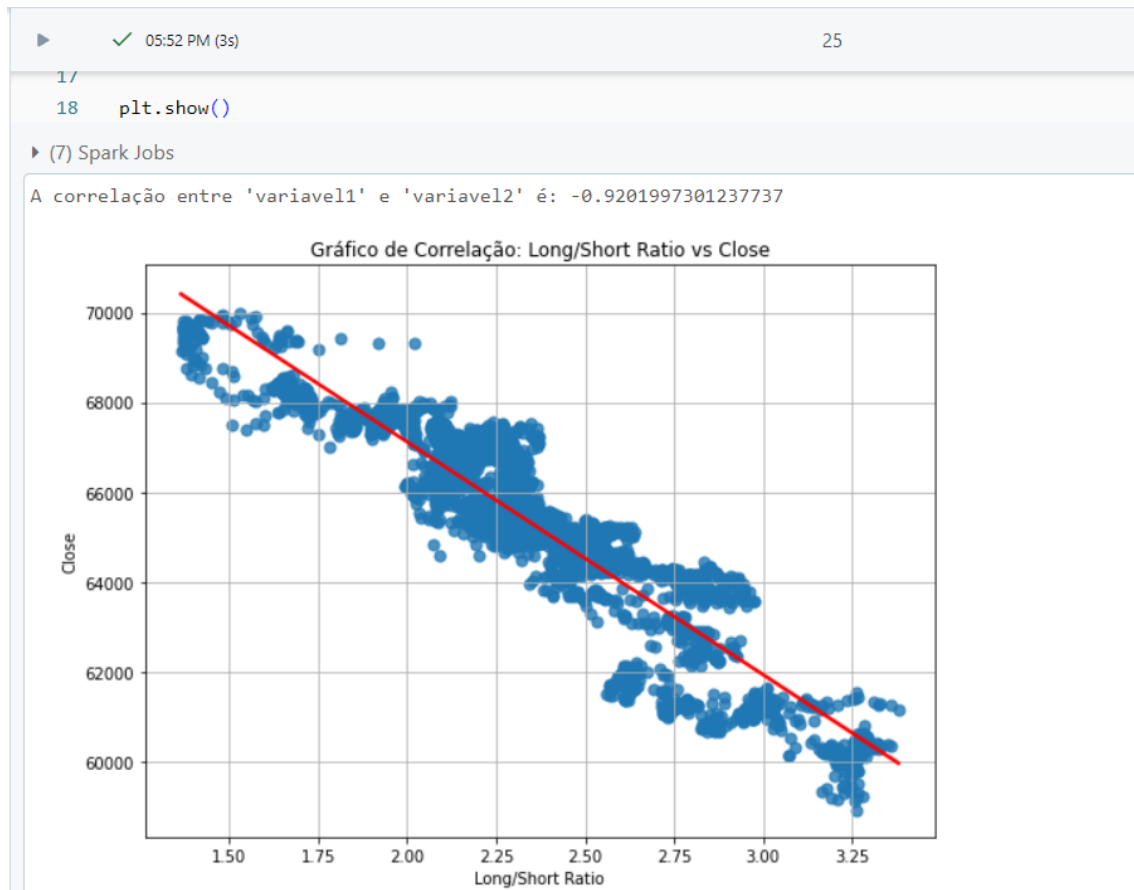


Figura 9: Gráfico de correlação do preço de fechamento e do indicador de long/short ratio

Pergunta 4: Há algum tipo de relação entre o indicador do open interest e o preço de fechamento do ativo?

R: De acordo com o gráfico de correlação entre o preço de fechamento do Bitcoin e o valor total de contratos ativos, obtivemos uma relação positiva, porém, não tão próxima a 1. Com isso, temos como indicativo de que quanto maior a quantidade de contratos ativos, maior a tendência do preço subir, e quanto menor a quantidade de contratos ativos, maior a tendência do preço cair, ressaltando que a relação não é tão próxima quanto a relação vista na pergunta anterior (mais próxima a 1).

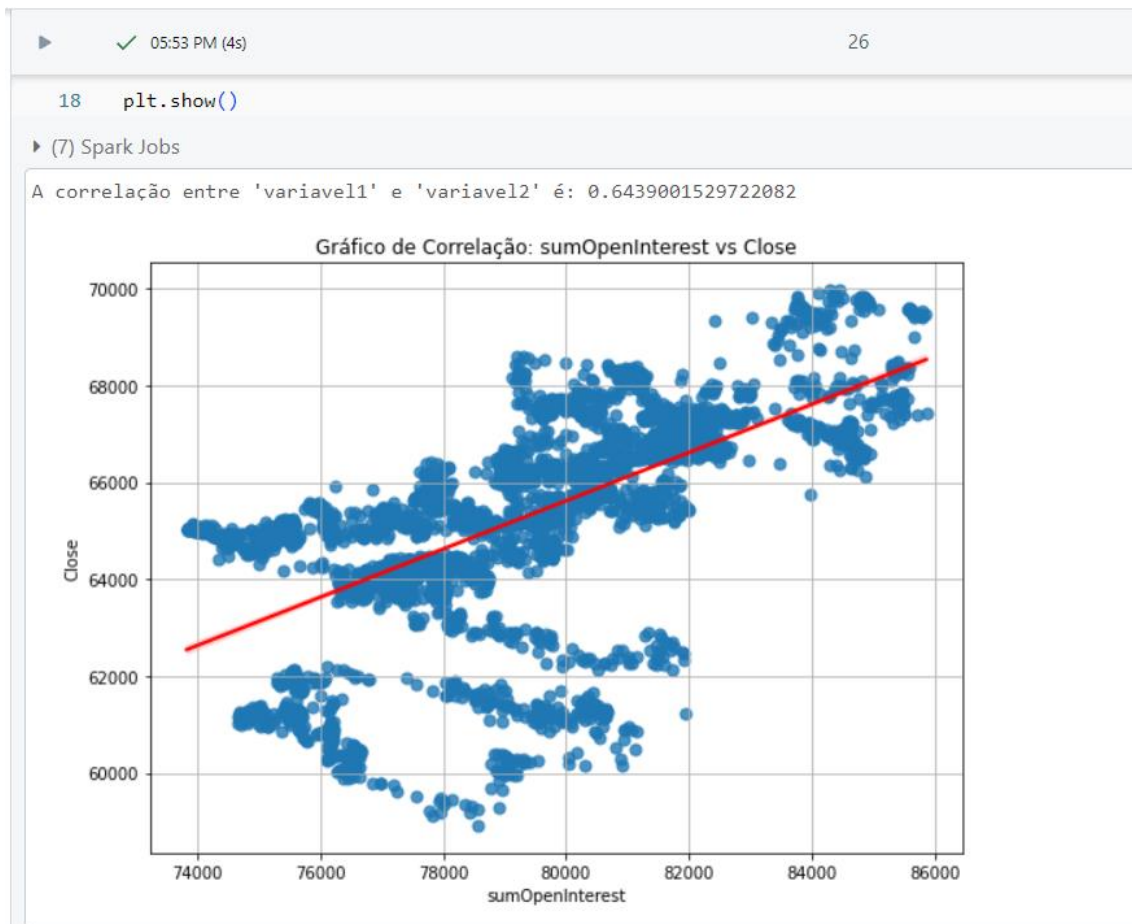


Figura 10: Gráfico de correlação do preço de fechamento e da quantidade total de contratos ativos

6. Auto-Avaliação e dificuldades encontradas:

Dentro do que foi proposto, entendo que atingi os resultados esperados, porém, devido ao tempo que estou sem praticar python, tive algumas dificuldades na parte de transformações de dados e também nunca havia utilizado a biblioteca do pyspark. Atualmente, utilizo como ferramenta de ETL o Alteryx, o que fez com que fosse necessário pesquisar muitas informações na internet. No início do projeto, havia pensado em outras possibilidades, tais como: quantidade de ordens enviadas, histórico de ordens – porém, devido ao alto volume de dados, preferi seguir somente com os dados de indicadores e os preços. A ideia é que, nas próximas 2 sprints, os mesmos dados sejam reutilizados afim de criar uma automação de negociação de criptomoedas utilizando de um modelo de série temporal.