

# PUC – CIÊNCIA DE DADOS E ANALYTICS

## OBJETIVO

A partir do histórico de preços das criptomoedas da corretora Binance, deseja-se consultar:

1. Qual período histórico contemplado pela tabela?
2. Quais são os últimos preços de todos os ativos listados?
3. Qual foi o recorde do preço do Bitcoin em 2023?
4. Quantos porcentos o Bitcoin rendeu em 2023?

## MODELAGEM

### Busca pelos dados

Em <https://www.binance.com/en/binance-api> é possível integrar a corretora à sistemas customizados via linguagem de programação.

```
1 from binance.client import Client
2 import pandas as pd
   Executed at 2023.09.27 11:34:38 in 1ms

1 client = Client()
2 info = client.get_exchange_info()
3 symbols = [x['symbol'] for x in info['symbols'] if x['symbol'].endswith('USDT')]

1 def get_daily_price_from(symbol):
2     frame = pd.DataFrame(client.get_historical_klines(symbol, '1d', '2023-01-01'))
3
4     if len(frame) > 0:
5         frame = frame.iloc[:, [0, 4]]
6         frame.columns = ['Time', symbol]
7         frame = frame.set_index('Time')
8         frame.index = pd.to_datetime(frame.index, unit='ms')
9         frame = frame.astype(float)
10    return frame
11
12 all_prices = pd.concat([get_daily_price_from(symbol) for symbol in symbols], axis=1)

1 relevant_prices = all_prices.dropna(axis=1)
2 relevant_prices.to_csv('daily_crypto_prices.csv')
3
   Executed at 2023.09.27 17:14:29 in 245ms
```

Figura 1 - Capturando Dados da Binance via API

Na figura 1, foi utilizada a linguagem Python para capturar o histórico de preços de todas a criptomoedas disponíveis na corretora desde o início de 2023. Os dados capturados foram salvos localmente em formato .csv.

## Coleta

Uma vez salvo localmente o próximo passo é inseri-lo em um data storage. Como neste projeto optou-se pelo uso do Google Cloud, foi feito o envio do arquivo para o Google Bucket.

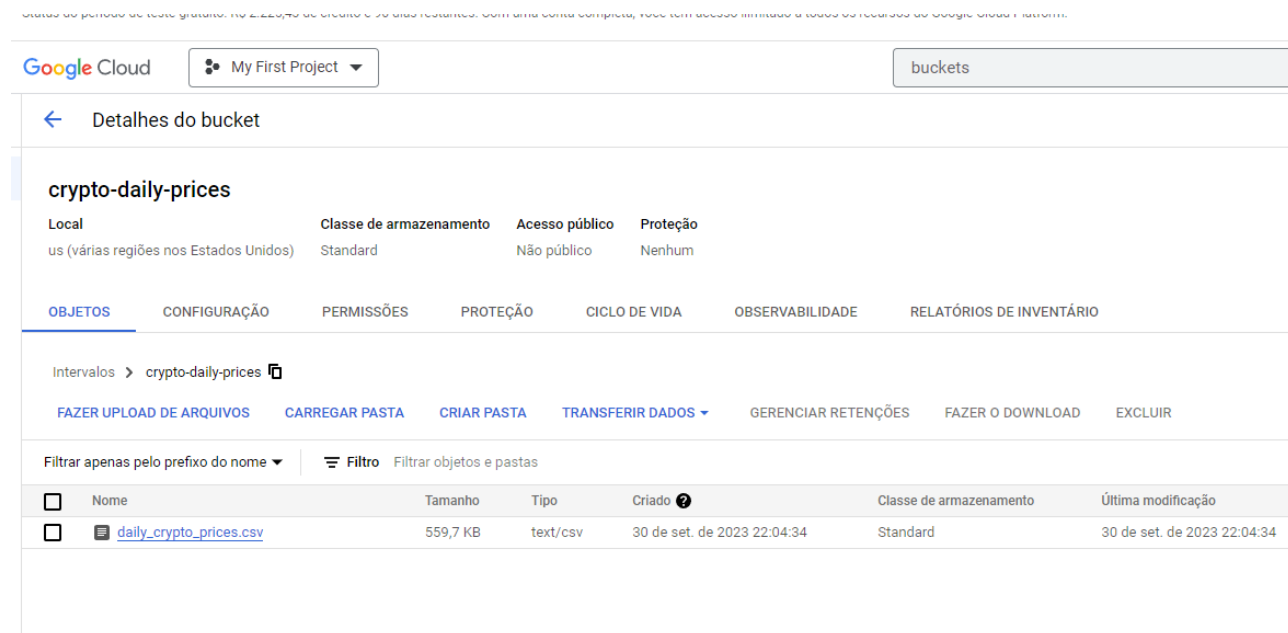


Figura 2 - Arquivo local salvo no Bucket do Google Cloud

## Carga (ETL)

Por definição, ETL – Extract, Transform, Load é o processo de carga que ocorre na sequência descrita. A ferramenta ETL da Google é o Data Fusion.

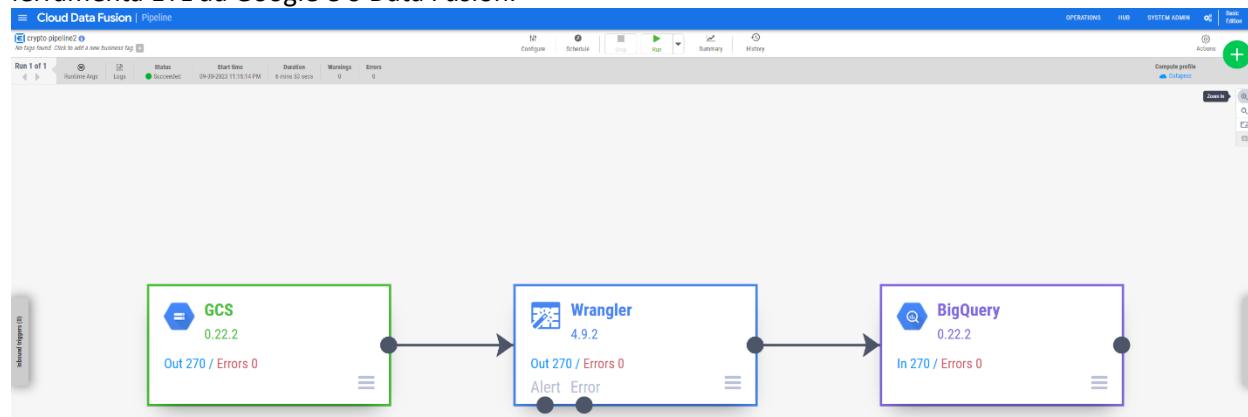


Figura 3 - Ferramenta ETL Gráfica (Data Fusion)

No Data Fusion, foi feito esse processo ETL de integração entre:

1. GCS (Google Cloud Storage): Onde o .csv foi armazenado.

2. Wrangler: Onde foi alterado o tipo da coluna Time, que relaciona o tempo histórico dos preços. Neste caso foi alterado para o tipo Datetime.
3. Big Query: Armazenamento de dados, onde será criada a tabela para as consultas seguintes.

## Consultas

1. Qual período histórico contemplado pela tabela?

```
1 WITH STARTING_TIME AS
2 (
3   SELECT Time
4   FROM `fluted-reporter-400701.cryptoprices.pricetable1` ORDER BY Time LIMIT 1
5 ),
6 ENDING_TIME AS
7 (
8   SELECT Time
9   FROM `fluted-reporter-400701.cryptoprices.pricetable1` ORDER BY Time DESC LIMIT 1
10 )
11 (
12   SELECT *
13   FROM ENDING_TIME
14 )
15 UNION ALL
16 (
17   SELECT *
18   FROM STARTING_TIME
19 )
```

Resultados da consulta

Informações do Job	Resultados	JSON	Detalhes da Execução	Gráfico	Pré-visualização	Gráfico de Execução
Linha	Time					
1	2023-01-01					
2	2023-09-27					

Figura 4 - Entre 01-01-2023 e 27-09-2023

2. Quais são os últimos preços de todos os ativos listados?

```
1 SELECT *
2 FROM `fluted-reporter-400701.cryptoprices.pricetable1` ORDER BY Time DESC LIMIT 1
3
```

Resultados da consulta

Informações do Job	Resultados	JSON	Detalhes da Execução	Gráfico	Pré-visualização	Gráfico de Execução					
Linha	Time	BTCUSD	ETHUSD	BNBUSDT	NEOUSDT	LTCUSD	QTUMUSD	ADAUSD	XRPUSD	EOSUSD	IOTAI
1	2023-09-27	26162.19	1587.99	211.4	7.14	63.35	2.146	0.2431	0.5017	0.557	

Figura 5 - 270 criptomoedas exibindo seu último preço registrado

### 3. Qual foi o recorde do preço do Bitcoin em 2023?

Q LAST-PRICES ▾ ×

Q HIGHEST\_BTC\_PRICE ▾ ×

Q BTC\_RETURNS ▾ ×

Q TIME\_PERIOD ▾ ×

+

Q HIGHEST\_BTC\_PRICE

EXECUTAR

SALVAR CONSULTA ▾

COMPARTILHAR ▾

PROGRAMAÇÃO

MAIS ▾

1 SELECT MAX(BTCUSDT)

2 FROM 'fluted-reporter-400701.cryptoprices.pricetable1'

Resultados da consulta

INFORMAÇÕES DO JOB

RESULTADOS

JSON

DETALHES DA EXECUÇÃO

GRÁFICO

PRÉ-VISUALIZAÇÃO

GRÁFICO DE EXECUÇÃO

Linha //

f0\_ ▾

//

1

31454.23

Figura 6 - U\$31.454,23 - preço recorde do Bitcoin no período

### 4. Quantos porcentos o Bitcoin rendeu em 2023?

Q

LAST-PRICES

×

Q

HIGHEST\_BTC\_PRICE

×

Q

BTC\_RETURNS

×

Q

TIME\_PERIOD

×

+

Q

BTC\_RETURNS

EXECUTAR

MAIS

SALVAR CONSULTA

COMPARTILHAR

PROGRAMAÇÃO

```

1 WITH LAST_PRICE AS
2 (
3   SELECT BTCUSDT
4   FROM `fluted-reporter-400701.cryptoprices.pricetable1` ORDER BY Time DESC LIMIT 1
5 ),
6
7 INITIAL_PRICE AS
8 (
9   SELECT BTCUSDT
10  FROM `fluted-reporter-400701.cryptoprices.pricetable1` ORDER BY Time LIMIT 1
11 ),
12
13 BOTH_PRICES AS
14 (
15   (
16     SELECT *
17     FROM LAST_PRICE
18   )
19   UNION ALL
20   (
21     SELECT *
22     FROM INITIAL_PRICE
23   )
24 )
25
26 select *,
27 round(ifnull(100 * (BTCUSDT/lag(BTCUSDT) over(order by BTCUSDT) - 1), 0), 2) || '%'
28 from BOTH_PRICES
29
30

```

Resultados da consulta

INFORMAÇÕES DO JOB

RESULTADOS

JSON

DETALHES DA EXECUÇÃO

GRÁFICO

PRÉ-VISUALIZAÇÃO

GRÁFICO DE EXECUÇÃO

Linha	BTCUSDT	f0_
1	26162.19	57.44%
2	16616.75	0%

Figura 7 - 57,44% de Janeiro à Setembro de 2023

## **Conclusão**

A Google Cloud atendeu os requisitos para este projeto. Foram encontradas algumas dificuldades quanto a limitação de uso gratuito, que foi resolvido em seguida.

A Big Query mostrou-se mais eficiente para datasets colunares, para dataframes do tipo preço x tempo (time series), que foi este caso, as consultas de tornam demasiadamente complexas para níveis de requisições mais avançados.

Como sugestão para próximo projeto, recomenda-se analisar uma ferramenta de consulta (query) que seja mais otimizada para time series.