Explicação dos arquivos:

* .git:

pasta padrão para subir o github.

* .ipynb\_checkpoints:

Pasta onde ficou salvo os checkpoint do Jupternotebook como uso de backup caso precisa-se voltar algum histórico.

* Codigos:

Nesta pasta estão todos os códigos feitos para o teste os códigos foram feitos em modelo Python seguindo o seguinte:

1. App.py -> seria o Main do código, com o processo de puxar os históricos de preços de uma forma padrão para os testes posteriores, esse código gera as querys e salva elas em csv dentro da pasta main do meu usuário;
2. Historical.py -> arquivo em que puxa a cotação de uma data especifica para ver como o código está se comportando.
3. Histórico\_3\_anos.py -> código já completo fazendo o push dos 3 anos de histórico contabilizando o dia 16/03/2020 até o dia 16/03/2023.
4. Merge\_das\_tabelas.py -> Esse código faz uma analise entre o CSV enviado pela Pismo (caffee.csv) e compara com o CSV gerado pelo meu código (base de dados SQLite) assim ele faz o merge na data, colocando as cotações de outras moedas e volumes na data em que está sendo apresentada nas duas tabelas.
5. Querys.sql -> Um código inicial para analise da sqlite, fiz as consultas em pyspark e mongodb então acabei deixando ela obsoleta dentro do código. Optei por deixa-la dentro do esqueleto caso precisasse utilizar algo dentro dela.
6. TempCodeRunner.py -> Deixei o código verdadeiro dentro mas como boa pratica e segurança eu gerei uma criptografia fazendo o código virar um modelo cripto: 6\_6ieae3S04IAkHxchFcXy4y7cgjtnVI5UhvYmg1cyU
7. transcript\_to\_csv\_pismodate.py -> Código que faz o transcript das moedas em todas as cotações solicitadas.
8. transcript\_to\_csv -> Código que faz o transcript para um CSV especifico.

* CSV: Aqui estão todos os CSV gerados e usados nos códigos.
* Databases: onde estão localizados todos os databases gerados, atenção para o uso do MongoDB.
* Evidencias: aqui estão as evidencias de código, detalhes e o double check em HTML ( usei o databricks ) mas deixei em html para qualquer pessoa ver.
* Pismo: dados adicionais.
* Pismo pdf: arquivo PDF que está o BI.
* Report link: links para acesso da tabela final:

<https://lookerstudio.google.com/s/q5f5iu-H8xQ>

<https://lookerstudio.google.com/reporting/8a700f52-8bf7-456c-8644-5865e24b7a58/page/p_2bqexdta4c>

Explicação do trabalho:

Primeiramente foi analisado o arquivo enviado direto na fonte, após isso fiz o read dele dentro do jupternotebook para ter mais informações do arquivo, após essa primeira etapa de análise, fiz o salvamento da planilha como Bronze, fiz uma tabela provisória de visualização em python com pyspark, e comecei a arrumar os dados, verifiquei se possui colunas faltantes, se possuía valores nulos ou erros de padrão e digitação.

Fiz o change table nas tabelas necessárias, transformei os valores que precisavam ser transformados, converti o volume para (int) e os valores em (float), a data segui o padrão datetime do csv e para análises fiz em timestamp64.

Após o tratamento fiz um novo salvamento com o nome gold.

Após isso a planilha estava com o ETL pronto e podia ser analisada, fiz as análises solicitadas utilizando o pandas no jupternotebook e depois para um double check fiz em pyspark. Os resultados podem ser vistos tanto no arquivo csv com os respectivos nomes quanto no histórico de evidência.

Para ter uma visão de tratamento diferenciada, subi o arquivo em databricks e refiz as análises em spark. Após isso salvei em formato HTML para poder ser visto, vendo que o databrics é um serviço pago.

Quanto ao python, eu fiz em VS code, utilizei python versão 3 para alguns tratamentos e análises, quanto ao API deixei a chave no histórico do código caso precise de alguma análise extra, os dados eu salvei em sqlite3 e mongoDB ( prefiro fazer database documental no mongodb compass), subia as currency das datas, acabei fazendo algo um pouco mais diferenciado que achei interessante de programar, fiz uma análise das datas que estavam em formato do csv do café, fiz um código em python para buscar essas informações e atualizar no database.

Após isso salvei o arquivo currencyusd em formato raw.

Utilizando o python fiz a transformação em csv e parquet.

Abri o arquivo de BI solicitado no documento da Pismo, criei duas folhas, uma com a análise de café X usd e uma como solicitado da cotação das moedas em 3 anos de histórico.