

DOCUMENTAÇÃO DA DATABASE GRUPO 1

Version Control

Version Number	Date Issued	Changes Included	Changed By:	Review and Approved By:
0.1	04/11/2021	Desenvolvimento do template inicial		N/A
0.2				
0.3				
0.4				
0.5				

1) Definição do Objetivo

Objetivo
Análise do cenário brasileiro com relação aos índices de taxa SELIC e IPCA nacionais para avaliar a abertura ou compra de um banco. Posteriormente realizar a comparação com o cenário americano para, através dos indicadores, definir qual país é mais favorável ao investimento.

2) Responsáveis

Nome	Cargo
Kevin Liu	Scrum Master (SM)
Guilherme Montes	Data Engineering
Karina Fernandes	Data Analyst & UX
Marcus Carceles	Data Engineering
William Leamari	Data Analyst & UX

3) Fontes de Dados

Fonte de Dados

Foram utilizados dois tipos de taxas para realizar o comparativo, Taxa IPCA e Taxa Selic. As fontes de dados consultadas foram coletadas do FED e BLS para os EUA e BACEN para o Brasil para a implementação do código em Python.

Banco Central do Brasil (BACEN): tem o objetivo fundamental de assegurar a estabilidade de preços e zelar pela estabilidade do sistema financeiro. Site: <https://www.bcb.gov.br/>.

Federal Reserve (FED): site oficial do governo americano com os demonstrativos das taxas históricas econômicas do país. Site: <https://www.federalreserve.gov>.

U.S. Bureau of Labor Statistics (BLS): plataforma oficial de estatísticas nacionais do governo americano. Site: <https://www.bls.gov/>.

APIs: Chave fornecida pela fonte oficial para acesso ao banco de dados de interesse da organização. Para o projeto em questão foi utilizada uma API do Banco Central do Brasil para extrair a informação da Selic e IPCA brasileiras, através do link abaixo. É possível substituir no código o número '11' para obter Selic e '433' para obter IPCA diretamente no endereço de rede para extrair outra informação do site, através da tabela específica disponibilizada online. <https://api.bcb.gov.br/dados/serie/bcdata.sgs.4390/dados?formato=json>.

- Significados das taxas utilizadas para o comparativo:

Taxa Selic - Sistema Especial de Liquidação e Custódia. A SELIC é a taxa média ajustada dos financiamentos diários apurados no Sistema Especial de Liquidação e de Custódia (SELIC) para títulos públicos federais, sendo definida como a taxa básica de juros da economia.

Equivalente americano: Interest Rate.

Taxa IPCA - Índice Nacional de Preços ao Consumidor Amplo. O IPCA mede a variação dos preços de produtos e serviços para o consumidor final, o IPCA é também considerado como o principal indicador para a taxa de inflação do Brasil. Já inflação é o nome dado ao aumento dos preços de **produtos** e serviços, sendo calculada pelos índices de preços, ou comumente chamados, índices de inflação.

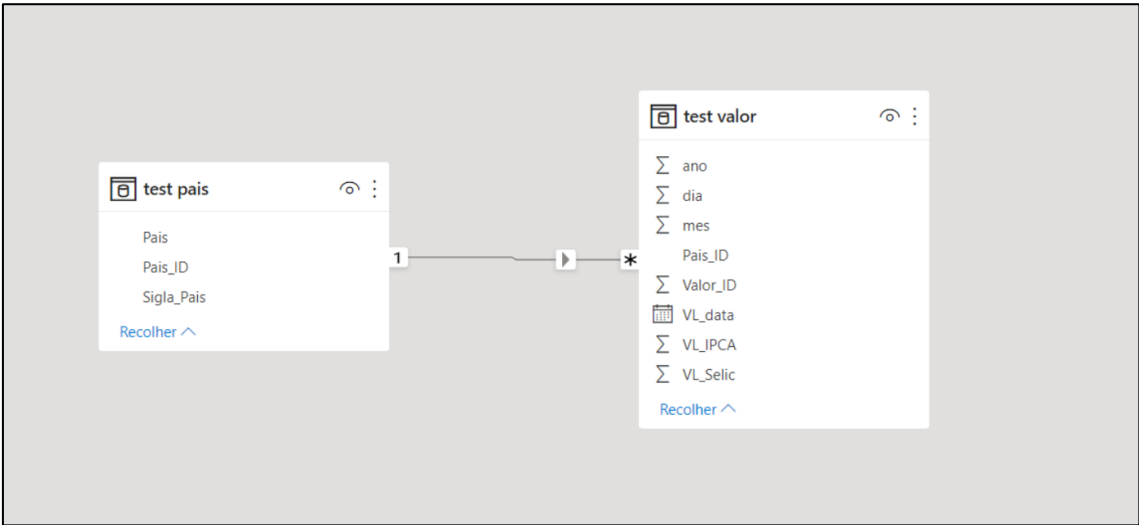
Equivalente americano: Consumer Price Index (CPI).

4) Dicionário de Dados:

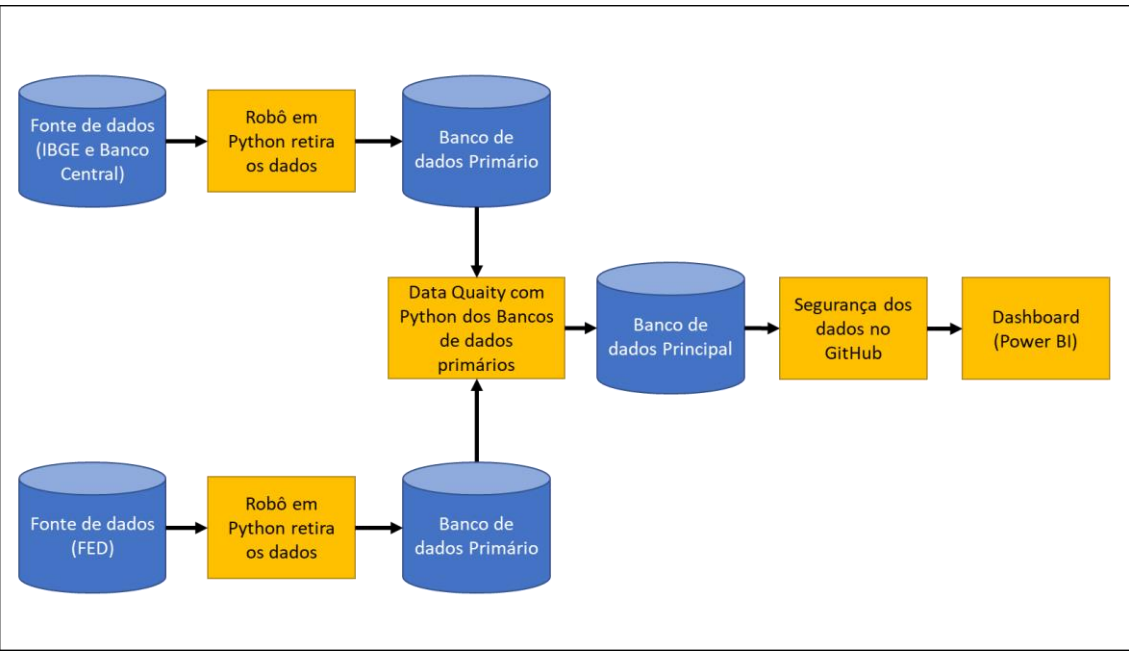
Tabela	Valores			
Descrição	Armazenará as taxas IPCA e SELIC relacionadas a país e data.			
Observações	A tabela possui chave estrangeira da tabela país.			
Campos				
Nome(valores)	Descrição	Tipo de Dado	Tamanho	Restrições de Domínio
Taxa IPCA	Índice Nacional de Preços do Consumidor Amplo - representatividade da inflação da moeda nacional.	Varchar	5	Default
Taxa SELIC	Sistema Especial de Liquidação de Custódia - Taxa básica de juros da economia.	Varchar	5	Default
País_ID	Nome do país de referência da taxa em estudo.	Varchar	15	FK
Dia	Dia da data de consolidação das taxas	Int	2	Default
Mês	Mês da data de consolidação das taxas	Int	2	Default
Ano	Ano da data de consolidação das taxas	Int	4	Default
Data Completa	Data completa de consolidação das taxas	Varchar	15	Default
Instância	Número de referência do segmento de dados (Index)	Int		PK/Unique

Tabela	País			
Descrição	Armazenará os nomes dos países em estudo.			
Observações	A tabela possui chave primária para a tabela valores.			
Campos				
Nome(país)	Descrição	Tipo de Dado	Tamanho	Restrições de Domínio
País_ID	ID atribuído ao país.	Varchar	15	PK
Nome do País	Nome do país	String	20	Not null
Abreviação do País	Sigla do país.	String	3	Not null

4.1) Relação entre as Tabelas



5) Fluxograma de Dados:



Tipos de Dados	
Estruturados	MS Excel/ CSV, My SQL
Semi	JSON
Não Estruturados	NA

Explicação do Fluxograma de Dados

- 1- O código automatizado em Python extrai o Banco de Dados Primário das fontes pré-determinadas, sejam elas IBGE e Banco Central no caso das taxas brasileiras e FED para o caso americano. Os dados históricos extraídos são armazenados em Excel CSV, com as variáveis respectivas às taxas e datas correlacionadas;
- 2- Um tratamento em Python é aplicado à base primária para atribuir cada variável a sua respectiva tabela, definida na arquitetura do projeto (tabela 'valores', tabela 'país' e tabela 'data'). Consecutivamente à separação entre as três tabelas, são atribuídos os IDs de interesse para as Primary e Foreign Key e as tabelas relacionais, obtendo-se o Banco de Dados Principal;
- 3- O banco de dados e suas atualizações são armazenados no GitHub, com a devida segurança de acesso e histórico de versões e edições;
- 4- O Banco de Dados Principal do Excel é importado para o Power BI, onde os dashboards de interesse do cliente são gerados para melhor visualização e dinamismo dos dados, possibilitando insights e conclusões referentes aos cenários analisados.

6) Domínios e Diretrizes

Domínios de Dados

Administrador	Os membros responsáveis, citados no início do documento, têm acesso livre para administrar o banco de dados.
Stewards	A equipe de desenvolvimento realizará manutenções e atualizações periódicas da base de dados.
Consumidores	Os consumidores da base de dados se referem à equipe de DnA da EY em sua completude.

Diretrizes de Confidencialidade

Privacidade	Os dados utilizados no projeto são de domínio público, retirados de fontes oficiais, portanto são gratuitos e abertos a todos.
Classificação de Dados	Os dados são públicos, confiáveis e gratuitos, não necessitam de autorização para serem retirados da fonte.
Acesso a dados	As fontes de dados são de acesso público. Os Bancos de Dados produzidos no projeto terão acesso limitado aos papéis relacionados diretamente ao projeto (sponsor, administrador, stewards e consumidores) e à equipe DnA da EY.

Diretrizes de Integridade	
Qualidade de Dados	
Controles	Os dados devem ser tratados para obedecer a conformidade, padronização e integridade.
Ética no Uso	Os dados devem ser usados para os propósitos relacionados ao objetivo deste projeto. Assim como obedecer às normas da LGPD (Lei Geral de Proteção de Dados).

Diretrizes de Disponibilidade	
Retenção e Expurgo	A retenção e armazenamento da base de dados estão disponíveis no GitHub. O banco de dados e o Power BI com as análises estão armazenados no seguinte link do GitHub: https://github.com/MarcusCarceles/Grupo1_DnA .
Ingestão	Através de aplicativo de análise de dados que importe Excel, no caso foi utilizado o Power BI.