**ENTREGA 2.0 – CIÊNCIA DE DADOS**

**Objetivo da entrega:**  
Selecionar os Dados  
Limpar/ Uniformizar os dados  
Derivar dados  
Integrar os Dados  
Formatar os Dados

**Selecionar os Dados:**

Para iniciar a construção do projeto foi necessário instalar as bibliotecas **Pandas** e **BCB**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

Após isso, buscamos os buscamos os dados dos **indicadores** que utilizaremos durante o projeto, sendo eles: **Taxa Selic**, **Ipca**, **Igpm**, **Salário Mínimo** e o **Índice de inadimplência.** Todos com o inicio da captação dos dados sendo no início de 2020.

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Limpar/ Uniformizar os dados**

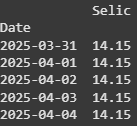
Nesta fase foram necessárias algumas mudanças nesses Dataframes. Primeiramente uma verificação de valores nulos, feita por uma função que verifica se contem valores nulos e soma a quantidade total, felizmente não registrou nenhum valor nulo nas tabelas.

Texto

O conteúdo gerado por IA pode estar incorreto.

Outro problema relatado foi que o index dos Dataframes eram a própria data do lançamento dos indicadores, desta forma tanto a formatação quanto eventualmente a criação de outros indicadores era prejudicada. Portanto foi necessário a criação de um index numérico, e as datas foram convertidas para uma coluna apropriada.

**Formatação antes:**



**Função:**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Formatação após a função:**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

Agora só restava ajustar a formatação das datas para o padrão brasileiro.

Texto

O conteúdo gerado por IA pode estar incorreto.

Desta forma, a formatação ficou assim:

Texto branco sobre fundo preto

O conteúdo gerado por IA pode estar incorreto.

E para finalizar, foi percebido que a tabela da Taxa Selic era **diária**, e não **mensal,** portanto, foi necessário ajustá-la para que futuramente seja corretamente integrada com as outras. Portanto setamos a coluna ‘Data’ de volta para index, com isso filtramos ela para receber apenas seus valores no inicio de cada mês, após isso retornamos a coluna ‘Data’ para seu local correto.

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Derivar Dados:**

Para iniciar a derivação dos dados foi criada uma coluna de variação em todos os Dataframes para comparar seu desempenho com o passar do tempo.

Texto

O conteúdo gerado por IA pode estar incorreto.

Após isso foi necessário alterar o nome das colunas, deixando cada uma com um nome único. Com essa alteração agora seria possível entrar na fase de **integração de dados.**

Texto

O conteúdo gerado por IA pode estar incorreto.

**Integração de dados:**

Para começar a fase de integração de dados, fizemos a união de todos os Dataframes pela coluna de **data** de maneira que os dados se entrelacem quando as datas em todos os indicadores sejam **iguais.**

Texto

O conteúdo gerado por IA pode estar incorreto.

Com isso foi possível criar **comparação entre indicadores**, sendo eles um que calcula o **salário real**, dividindo o salário mínimo pela inflação do Ipca. E o outro que calcula a **diferença entre inflações** do Ipca e do Igpm.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

**FORMATAR DADOS:**

Agora, com tudo já construído, só restou ajeitar as colunas de forma que o número de casas decimais seja reduzido.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Desta forma, a formatação das tabelas ficou assim:

Tabela

O conteúdo gerado por IA pode estar incorreto.