Introdução

Estamos em uma era tecnológica em que geramos milhões de dados a todo momento. Por conta desse grande volume, muitas vezes se torna desafiador conseguir extrair informações úteis com tantos dados a disposição. Por isso, o ato de lidar adequadamente com esse montante de fatos apenas aumentará pelos próximos anos.  
  
Enxergando esse cenário, fica clara a importância de se ter a melhor gestão desses dados. Desde as formas como eles são coletados e tratados, até como eles gerarão insights valiosos por meio de visualizações, quanto mais informações forem extraídas dessas bases maior será a clareza na tomada de decisão pelos colaboradores.

Com isso em mente, esse presente relatório tem como finalidade principal demonstrar quais as possibilidades podem ser geradas quando se há uma tratativa e exposição dos dados clara e trazer certas tomadas de decisão com base nas tabelas obtidas.

Processo de Análise

- Definição da Análise:

A primeira etapa da análise está ligada aos interesses que a Análise deseja alcançar e isso é feito pelas perguntas. Pensando na BanVic, algumas perguntas são importantes para o entendimento da eficiência do negócio, tais como:

* Quais ações podem ser tomadas para a aquisição de novos clientes?
* O banco digital traz mais ou menos retorno do que o banco físico?
* Quais produtos do banco são mais rentáveis?
* Pode ser aplicados indicadores para cada uma dessas e outras perguntas.

Sobre novos clientes, é possível trazer a visualização de quais estados possuem clientes mais ativos para focar o marketing de atração. Sobre as agências, é possível trazer a taxa de clientes ativos pelo total, assim como quantidade de colaboradores por tipo de agência. Por fim, sobre a rentabilidade, é possível comparar se os empréstimos geram mais ou menos retorno do que as transações.

Ainda é viável fazer mais perguntas. Quanto melhores elas forem elaboradas, mais fácil será os processos de limpeza, análise e visualização, assim como alcançar um melhor resultado no geral.

- Transformação e tratamento de dados

Obtendo o direcionamento por meios das perguntas, a limpeza ajudará tanto em ver se os dados estão consolidados e precisos, assim como garantir que apenas os dados de interesse ao projeto serão utilizados.

Primeiramente, foi checada se as tabelas estão concisas, sem elementos faltantes e bem estruturadas. Tendo aval positivo, foram delimitados os atributos de cada uma das tabelas que precisavam ser realocados.

Para tratar os dados, foi utilizado o Jupyter Notebook junto com o Python/Pandas por conta da facilidade do uso dessas ferramentas, tanto pela versatilidade quanto pela recorrência delas no processo de tratamento.

Mas, essa etapa pode ser feita também por SQL ao trabalhar com banco de dados ou Excel/Sheets caso seja melhor. Também é possível trabalhar com um pipeline de dados e data Warehouse no sistema em nuvem Azure ou AWS caso tenha interesse em segurança extra.

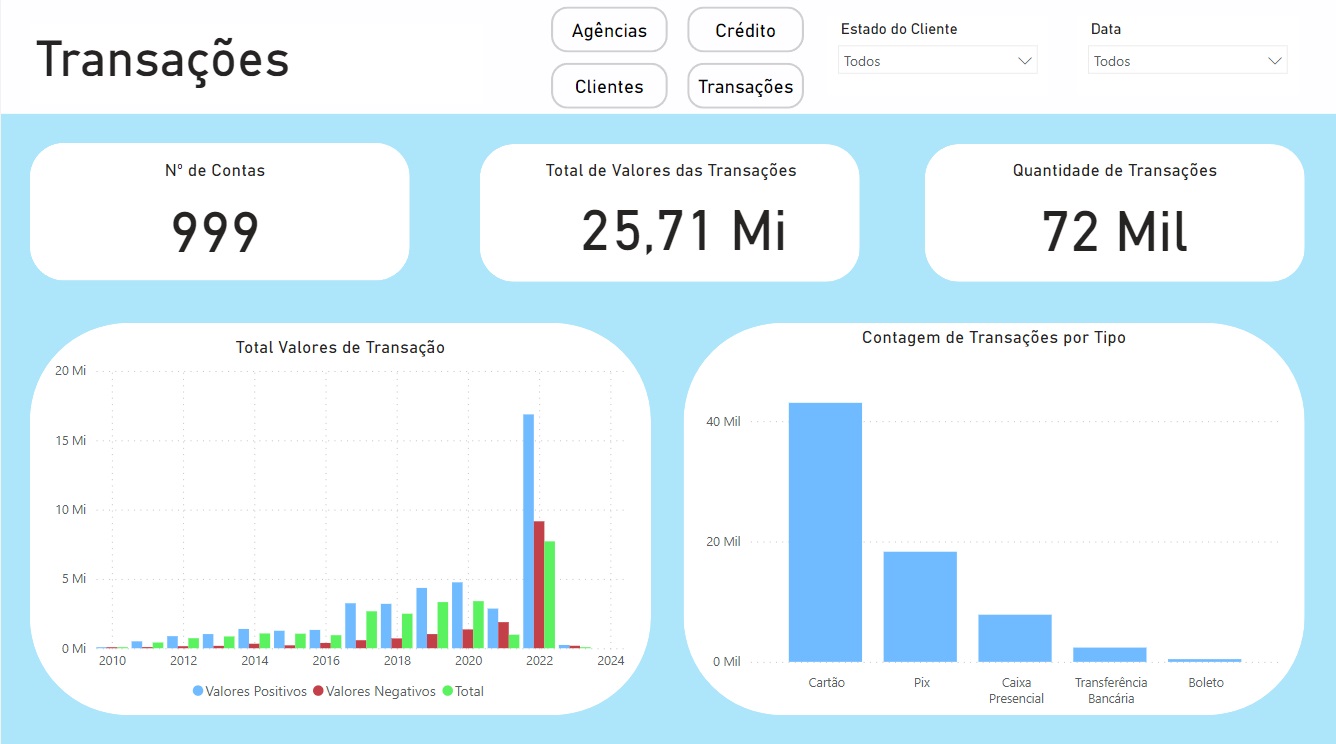
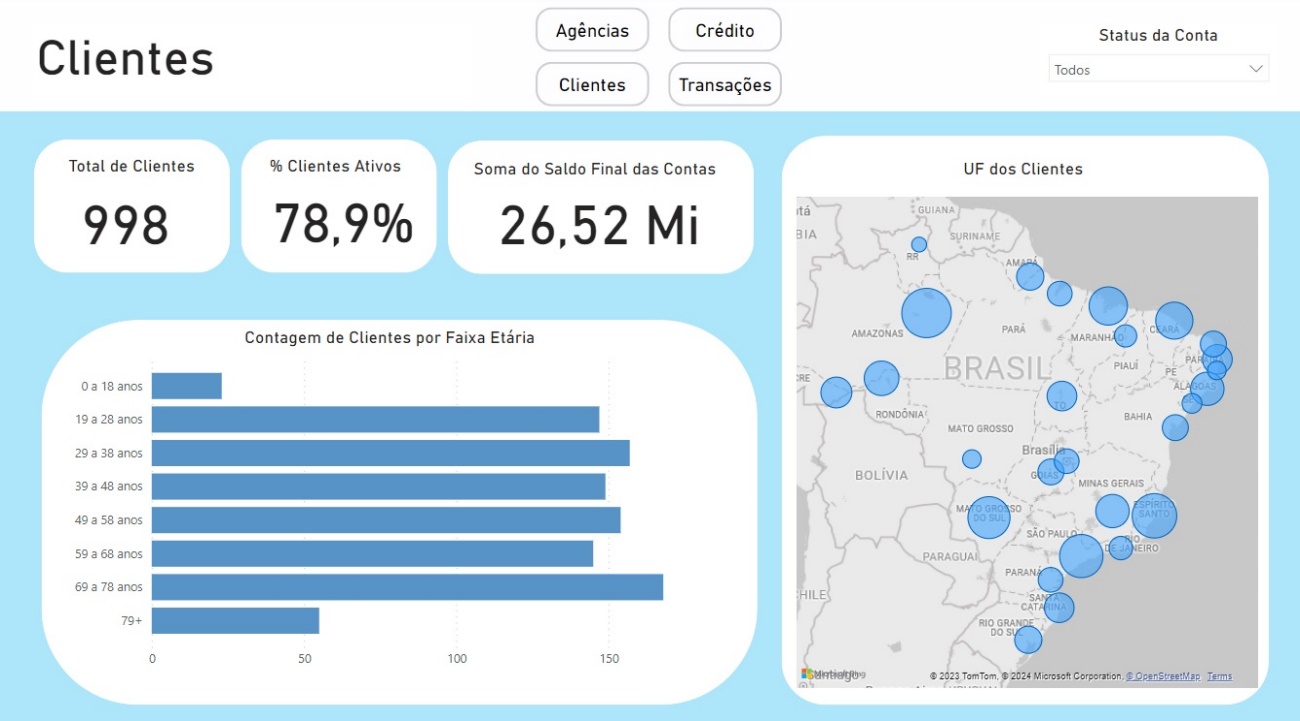
Sobre esse processo, algumas alterações formam feitas nas planilhas, dentre elas:

* Separação do UF dos clientes e colaboradores do endereço para análise por estado;
* Normalização das datas em um formado possível de ser compreendido pelo Power Bi;
* Normalização do CEP, deixando todos sem o hífen;
* Criação de uma coluna de tipo de Transação, para facilitar o filtro no Dashboard;
* Seleção de colunas de interesse, para melhorar a performance do Dashboard;

O arquivo do notebook com todas essas informações está presente neste [link](https://github.com/caiombribeiro/lighthouse_estudo/blob/main/consulta_planilhas.IPYNB).

- Visualização dos dados

Após o tratamento adequado, entra a fase de dispor esses dados de uma forma clara e objetiva, sempre respeitando os interesses levantados já na primeira etapa das perguntas.



D.

Algumas análises

E.

Recomendações ou Sugestões de negócio