

STONE_DATA_CHALLENGE_2022

NOME: MURILO-FARIAS

1. Preenchendo o framework de avaliação

2.1 Validações iniciais

- a. Quantos contratos distintos existem em sua tabela de observação final, após todos os merges e etapas de ETL?

Na tabela final temos número total de: 14.756 contratos

- b. Você considerou ou aplicou algum filtro de exclusão? Caso afirmativo, descreva-o e informe quantos contratos foram filtrados.

Mesmo que considerando um filtro para os contratos devedores que seria filtrar a coluna DSPP com números maior do que zero, ainda assim restam 14.756 contratos.

2.2 Fluxo de etapas realizadas

1.1.1 Disponibilização dos arquivos no Google Drive

- Descompactar o arquivo #DC2022.zip
- Upload no Google Drive e disponibilização da pasta em modo público para acesso dos arquivos tratados e realizados durante o desafio:
https://drive.google.com/drive/folders/1szW3yjqo6OF27udVJHj_3fN7vZtQ4H2L?usp=sharing

1.1.2 Criar novo arquivo do Google Colab

- Arquivo criado: STONE_DATA_CHALLENGE_2022_MURILO-FARIAS.ipynb

1.1.3 Análise Exploratória com Python do arquivo: PORTFOLIO_GERAL.csv

- Importar biblioteca: pandas
- Carregar o arquivo: portfolio_geral.csv
- Visualizar as 10 primeiras linhas para confirmar carregamento com sucesso.
- Verificar os nomes das colunas.
- Verificar os tipos das colunas.
- Verificar as quantidades de linhas e colunas.
- Verificar se existe dados nulos.
- Importar bibliotecas matplotlib e seaborn.
- Gerar tabela correlação para colunas numéricas do arquivo portfolio_geral.csv
- Criar uma cópia do arquivo original para realizar tratamento de dados.
- Exclusão de 3 colunas do novo arquivo com nome: portfolio_geral_tratado
- Confirmar quantidade de linhas e colunas para validar exclusão das colunas.
- Verificar as estruturas de todas as colunas do arquivo tratado.
- Contagem de valores para 3 colunas escolhidas utilizando uma lista personalizada.
- Validando os tipos de contratos existentes.
- Criar uma compactação para Zip.
- Fazer o download do arquivo compactado: PORTFOLIO_GERAL_TRATADO_COMPACTADO

1.1.4 Análise Exploratória com Python do arquivo: PORTFOLIO_TPV.csv

- Carregar o arquivo: portfolio_tpv.csv
- Visualizar as 10 primeiras linhas para confirmar carregamento com sucesso.
- Verificar os nomes das colunas.
- Verificar os tipos das colunas.
- Verificar as quantidades de linhas e colunas.
- Verificar se existe dados nulos.
- Gerar tabela correlação para colunas numéricas do arquivo: portfolio_tpv.csv
- Verificar as estruturas de todas as colunas do arquivo.

Após a análise para este arquivo foi considerado manter nas próximas etapas o arquivo original sem modificação.

1.1.5 Análise Exploratória com Python do arquivo: PORTFOLIO_COMUNICADOS.csv

- Carregar o arquivo: portfolio_comunicados.csv
- Visualizar as 10 primeiras linhas para confirmar carregamento com sucesso.
- Verificar os nomes das colunas.
- Verificar os tipos das colunas.
- Verificar as quantidades de linhas e colunas.
- Verificar se existe dados nulos.
- Verificar as estruturas de todas as colunas do arquivo.
- Contagem de valores para 3 colunas escolhidas utilizando uma lista personalizada.

Após a análise para este arquivo foi considerado manter nas próximas etapas o arquivo original sem modificação.

1.1.6 Análise Exploratória com Python do arquivo: PORTFOLIO_CLIENTES.csv

- Carregar o arquivo: portfolio_clientes.csv
- Visualizar as 10 primeiras linhas para confirmar carregamento com sucesso.
- Verificar os nomes das colunas.
- Verificar os tipos das colunas.
- Verificar as quantidades de linhas e colunas.
- Verificar se existe dados nulos.
- Verificar as estruturas de todas as colunas do arquivo.
- Contagem de valores para 5 colunas escolhidas utilizando uma lista personalizada.
- Validando as cidades existentes utilizando lista personalizada.

Após a análise para este arquivo foi considerado manter nas próximas etapas o arquivo original sem modificação.

O Código Python está disponível no arquivo: STONE_DATA_CHALLENGE_2022_MURILO-FARIAS.ipynb
https://github.com/murilofarias10/STONE_DATA_CHALLENGE_2022-MURILO-FARIAS

Todos os arquivos originais e o tratado estão disponíveis no Google Drive:
https://drive.google.com/drive/folders/1szW3yjqo6OF27udVJHj_3fN7vZtQ4H2L?usp=sharing

1.1.7 Preparar e carregar dados no SGBD PostgreSQL

- Criar base de dados com o nome: Stone
- Criar um schema com o nome: Data-Challenge
- Criar tabela com nome: `portfolio_geral_tratado` utilizando comando SQL.
- Criar tabela com nome: `portfolio_tpv` utilizando comando SQL.
- Criar tabela com nome: `portfolio_comunicados` utilizando comando SQL.
- Criar tabela com nome: `portfolio_clientes` utilizando comando SQL.
- Alterar na tabela: `portfolio_tpv` o tipo da coluna: `dt_transacao` para data.
- Alterar a tabela: `portfolio_clientes` retirando a coluna: `nr_documento` como primary key uma vez que existem valores duplicados para esta coluna.
- Alterar a tabela: `portfolio_comunicados` retirando a chave primary key uma vez que existem valores duplicados para esta coluna.
- Alterar a tabela: `portfolio_tpv` alterando o tipo da coluna: `dt_transacao` para o tipo data e retirando a coluna: `nr_documento` como primary key uma vez que existem valores duplicados para esta coluna.
- Alterar a tabela `portfolio_geral_tratado` o tipo da coluna: `dt_contrato` para string
- Alterar a tabela `portfolio_geral_tratado` o tipo da coluna: `prazo` para: numérico
- Alterar a formatação de todas as colunas numéricas da tabela: `portfolio_geral_tratado` utilizando comando SQL.
- Alterar as configurações do PostgreSQL para que possa ser importados arquivos externos:
File > Preference > Paths > Binary Path
Adicionar o caminho: C:\Program Files\PostgreSQL\13\bin
- Importar dados para as tabelas utilizado:
Formato csv
Encoding: UTF8
Arquivo com cabeçalho (header)
Delimitador virgula (,)
- Importar dados do arquivo: `portfolio_clientes.csv` para a tabela: `portfolio_clientes`
- Importar dados do arquivo: `portfolio_comunicados.csv` para a tabela: `portfolio_comunicados`
- Importar dados do arquivo: `portfolio_tpv.csv` para a tabela: `portfolio_tpv`
- Importar dados do arquivo: `portfolio_geral_tratado.csv` para a tabela: `portfolio_geral_tratado`

O Scrip SQL está disponível no arquivo: SCRIP-SQL-MURILO-FARIAS-PARTE1.txt

https://github.com/murilofarias10/STONE_DATA_CHALLENGE_2022-MURILO-FARIAS

1.1.8 Conectando o Power Bi no PostgreSQL e importando as tabelas

- Conectar no banco de dados:
Servidor: Localhost
Banco de dados: Stone
Modelo de conectividade: DirectQuery

Instrução SQL: O script SQL está disponível no arquivo: SCRIP-SQL-MURILO-FARIAS-PARTE2.txt

https://github.com/murilofarias10/STONE_DATA_CHALLENGE_2022-MURILO-FARIAS

Nome do usuário: postgres

Senha: (definida quando instalei o SGDB)

1.1.9 Tratamento de dados para a tabela: PORTFOLIO_GERAL_TRATADO

- Para a tabela `portfolio_geral_tratado` foi escolhido o modelo de conectividade: importar
- Alterar o nome da tabela de: Consulta1 para: `portfolio_geral_tratado`
- Alterar o tipo de dados da coluna: `dt_ref_portfolio` de: data/hora para o tipo: data
- Alterar o tipo de dados da coluna: `dt_desembolso` de: data/hora para o tipo: data
- Fechar e aplicar.

1.1.10 Tratamento de dados para a tabela: PORTFOLIO_TPV

- Para a tabela `portfolioi_tpv` foi escolhido o modelo de conectividade: importar
- Alterar o tipo de dados da coluna: `dt_transacao` de: data/hora para o tipo: data
- Alterar o nome da tabela de: Consulta1 para: `portfolio_tpv`
- Alterar o tipo de dados da coluna: `qtd_trasacoes` de texto para o tipo: número inteiro
- Fechar e aplicar.

1.1.11 Tratamento de dados para a tabela: PORTFOLIO_COMUNICADOS

- Alterar o nome da tabela de: Consulta1 para: `portfolio_comunicados`
- Alterar o tipo de dados da coluna: `dt_ref_portfolio` de: data/hora para o tipo: data
- Fechar e aplicar.

1.1.12 Tratamento de dados para a tabela: PORTFOLIO_CLIENTES

- Alterar o nome da tabela de: Consulta1 para: `portfolio_clientes`
- Fechar e aplicar.

1.1.13 Relacionamento das tabelas no Power Bi

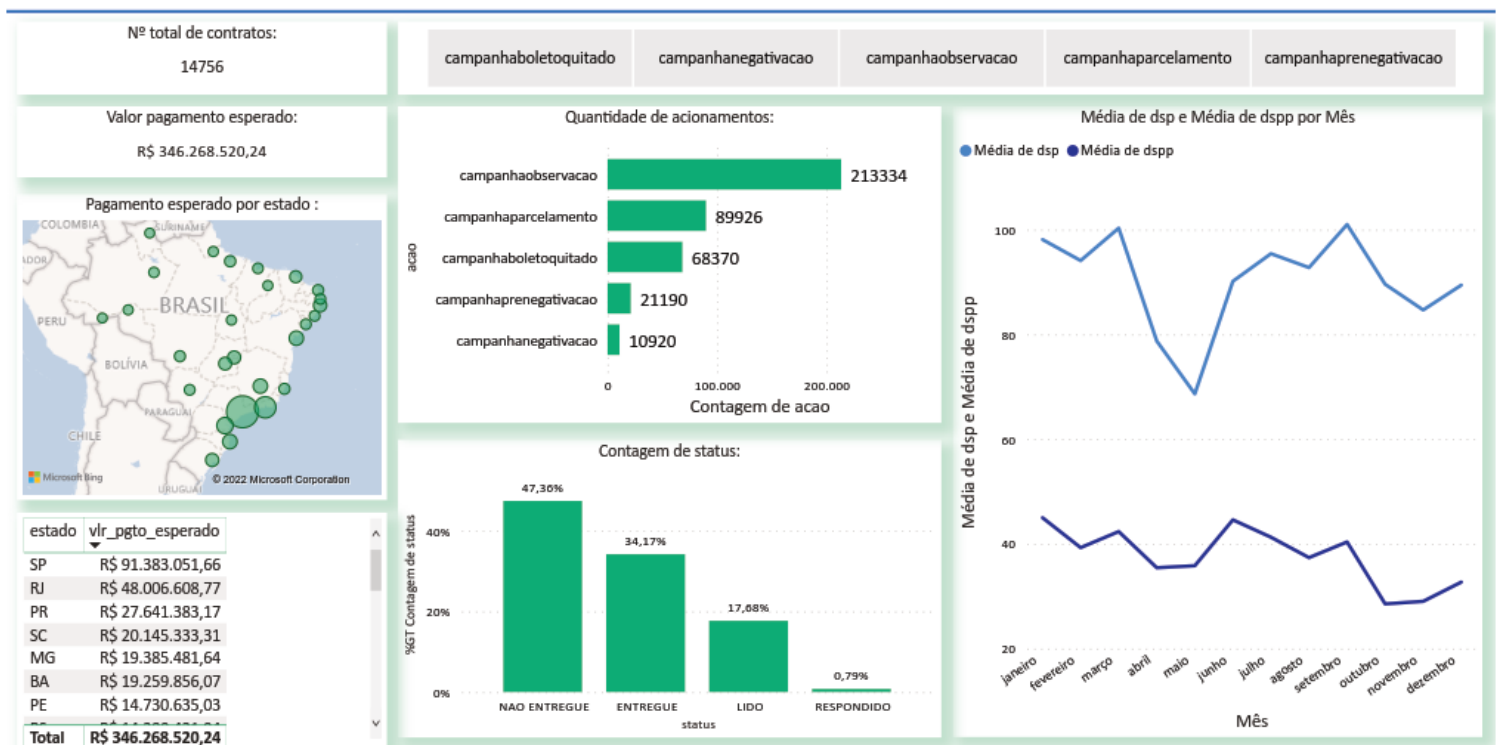
- Montando um esquema onde a tabela `portfolio_geral_tratado` será a tabela fato e as outras tabelas: `portfolio_clientes`, `portfolio_comunicados`, `portfolio_tpv` serão tabelas dimensão.
- Criar relacionamento muitos para muitos da tabela: `PORTFOLIO_GERAL_TRATADO` coluna: `contrato_id` com a tabela: `PORTFOLIO_COMUNICADOS` coluna: `contrato_id`
- Criar relacionamento muitos para muitos da tabela: `PORTFOLIO_GERAL_TRATADO` coluna: `nr_documento` com a tabela: `PORTFOLIO_TPV` coluna: `nr_documento`
- Criar relacionamento muitos para muitos da tabela: `PORTFOLIO_GERAL_TRATADO` coluna: `nr_documento` com a tabela: `PORTFOLIO_CLIENTES` coluna: `nr_documento`.

1.1.14 Elaboração do Dashboard no Power Bi

- Inserir um cartão com a contagem da coluna: contrato_id da tabela: portfolio_geral_tratado
- Inserir uma segmentação de dados utilizando a coluna: acao da tabela: portfolio_comunicados
- Inserir um cartão com a soma da coluna: vlr_pgto_esperado da tabela: portfolio_geral_tratado
- Inserir um gráfico de barra clusterizado com a contagem da coluna: acao da tabela: portfolio_comunicados por acao.
- Inserir um gráfico de linhas com as medias das colunas: dsp e dspp da tabela: portfolio_geral_tratado
- Inserir um mapa com a localização da tabela: portfolio_clientes coluna: estado e o valor da tabela: portfolio_geral_tratado coluna: vlr_pgto_esperado
- Inserir um gráfico de colunas clusterizado utilizando a coluna: status da tabela: portfolio_comunicados e a contagem de valor da coluna status da mesma tabela.
- Inserir uma tabela somando o valor da coluna: vlr_pgto_esperado da tabela: portfolio_geral_tratado pela coluna: estado da tabela: portfolio_clientes
- Inserir uma caixa de texto para o titulo: STONE_DATA_CHALLENGE_2022 Murilo Farias
- Renomear a pagina para: Análise_Campanhas
- Aplicar o filtro em toda pagina: valor da tabela: portfolio_geral_tratado coluna: dspp é maior do que 0 (zero).
- Formatação geral da pagina (Eixos, Titulos, Rótulos, Sombras, Cabeçalhos e outros).

STONE_DATA_CHALLENGE_2022

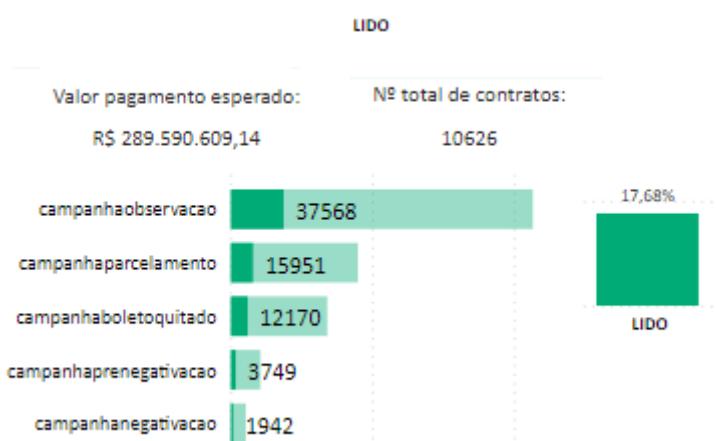
Murilo Farias



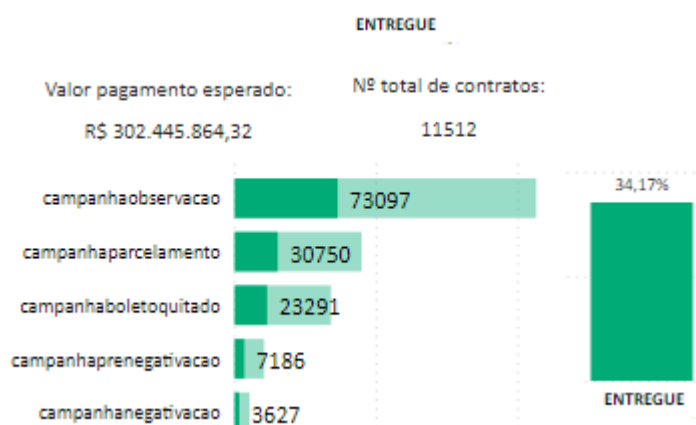
2.3 Conclusões e insights

Após a análise exploratória dos dados, entendimento dos problemas e elaboração do dashboard a análise para identificar qual é a melhor campanha foi baseada no percentual de leitura e percentual de entregas, também considerando somente os contratos devedores ou seja (DSPP maior do que zero).

Detalhando um pouco mais, temos uma taxa de leitura de aproximadamente 17% independente da campanha conforme foto abaixo retirado do dashboard:



E temos um percentual de entrega de aproximadamente 34% independente da campanha conforme foto abaixo retirado do dashboard:



Sendo assim eleito a melhor campanha logo é aquela que temos o maior número de alcance independente se a quantidade de disparo também for proporcional, para o marketing digital a campanha mais viral é que tende a trazer maior retorno financeiro por conta também de um maior número de contratos ativos seria a: **campanha observação**.