

ESTRATÉGIA GERAL DE ANÁLISE DE DADOS

- Cabeçalho;
- Requisitos;
- Informações da base de dados com dicionários;
- Informar fontes de pesquisas (se houver);
- Problema de negócio: aplicação dos 5w;
- Instalação de pacotes/bibliotecas;
- Importar esses pacotes/bibliotecas;
 - Pandas;
 - Numpy;
 - Matplotlib;
 - Pandera
- Extração
 - Google Drive (bruto);
 - GCP;
 - MongoDB;
 - Backup1;
- Pré-Análise;
 - Visualização dos dados;
 - Avaliação das possibilidades do que faremos;
- Transformação
 - Escolher as possíveis colunas importantes para a análise;
 - Verificar se há colunas iguais;
 - Verificação e tratamento de duplicadas;
 - Verificação e tratamento de valores nulos;
 - Verificação e tratamento de valores (ou observações) únicos;
 - Verificação e tratamento de inconsistência;
 - Tradução/Renomeação;
 - Nome das colunas: minúsculo;
 - Nome das categorias: maiúsculo;
 - Transformação de tipos adequados dos dados;
 - Float;

- Int;
 - String;
 - Bool;
 - Datetime;
- Qualidade e integridade dos dados via Pandera;
- Backup2;
- Carregamento;
 - Google Drive (tratado);
 - GCP;
 - MongoDB;
- Análise
 - Busca ou Localização;
 - Filtro;
 - GroupBy;
 - Combinando filtro e agrupamento;
 - Resumo de estatística;
 - Média
 - Mediana
 - Moda
 - Quartis
 - Desvio-padrão
 - Visualização pelo Pandas e Matplotlib;
 - Barras
 - Setores
 - Boxplot
 - Linhas
 - Dispersão
 - Histograma

Sugestão: Fluxo de trabalho

Link: <https://colab.research.google.com/drive/19NS6grsbQ2WOINOInHtyREyG0MR5wmyX?usp=sharing>
<https://github.com/sahtcarneiro>