

Treinamento em Dados Ambientais 2022

XI SIMBIOMA - Desacelerando o Antropoceno: desafios e possibilidades na Mata Atlântica

Etapa 2 -Preparar

Banco de dados, conjunto de dados (Dataset) e software

Parte 01 - Etapas de Análise de Dados Ambientais

Parte 02 - Mais ferramentas e aprofundamento

Parte 01 - Etapas de Análise de Dados Ambientais

1. **Perguntar**
Fundamentos e recursos de dados ambientais
 2. **Preparar**
Banco de Dados, Dataset e Software
 3. **Limpar**
Configuração do dataset, limpeza e LOG
 4. **Processar**
Coleta e seleção de dados
 5. **Analisar**
Hipótese e ferramentas
 6. **Compartilhar**
Apresentar e mostrar
 7. **Agir**
Com base em dados
-

Parte 01 - Etapas de Análise de Dados Ambientais

Etapa 2 Preparar

Banco de dados, conjunto de dados (Dataset) e software

- A. Compreensão e estrutura do banco de dados
- B. definição do conjunto de dados
- C. Software para conjunto de dados e banco de dados.
- D. Relação dados, tempo e espaço
- E. Planilha - Excel/Google
- F. R-CRAN/R-Studio - Bibliotecas
- G. Banco de dados Estruturado (SQL)
- H. Sistemas de Informações
 - Geográficas (SIG)

Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

- A. Compreensão e estrutura do banco de dados
- B. Definição do conjunto de dados
- C. Software para conjunto de dados e banco de dados.
- D. Relação dados, tempo e espaço
- E. Planilha - Excel/Google
- F. R-CRAN/R-Studio - Bibliotecas
- G. Banco de dados Estruturado (SQL)
- H. Sistemas de Informações Geográficas (SIG)

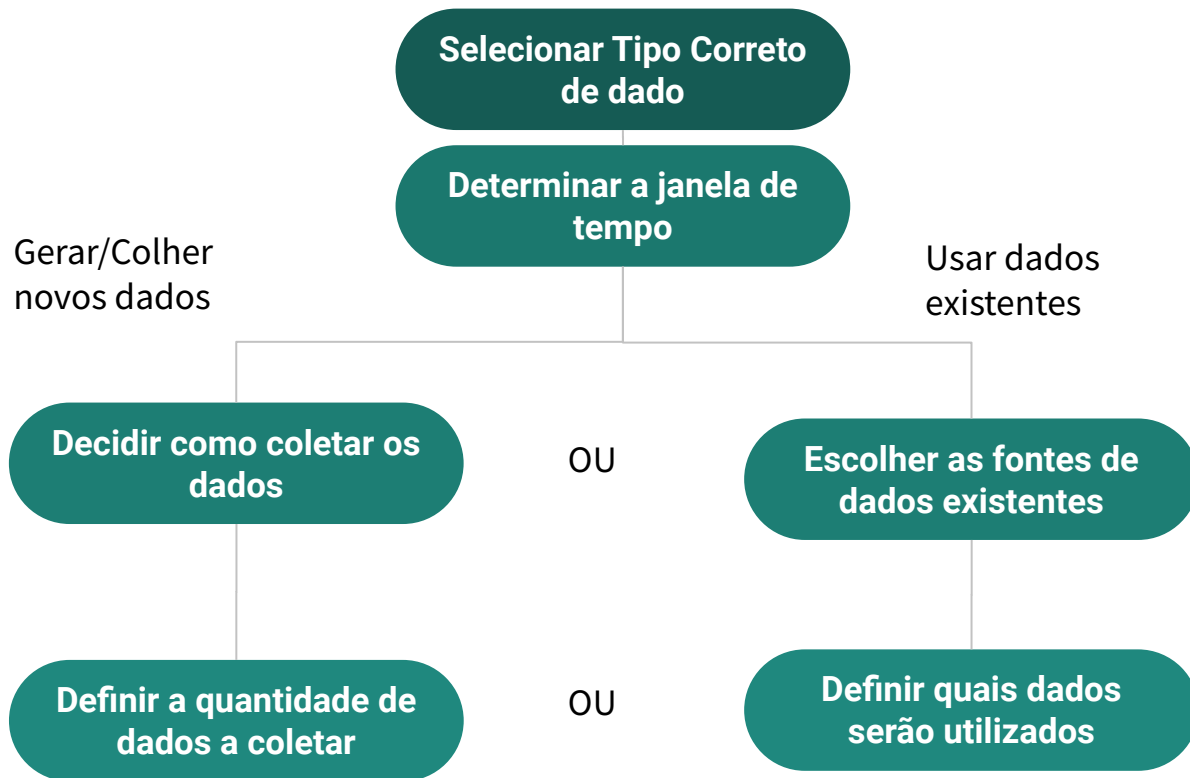
Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

A. Compreensão e estrutura do banco de dados



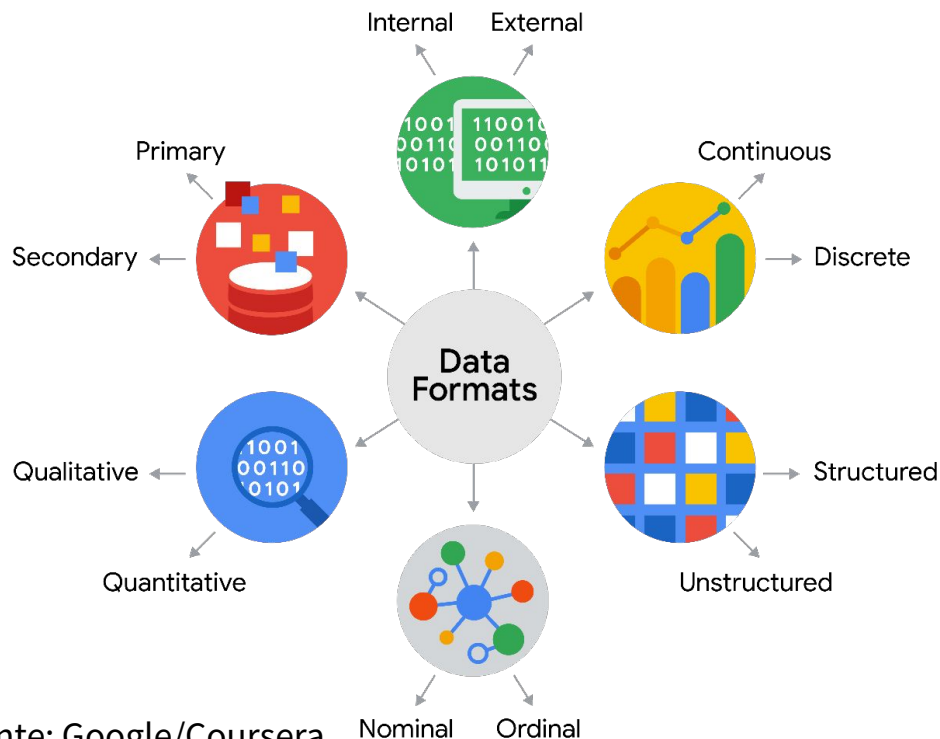
Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

A. Compreensão e estrutura do banco de dados



Formato de dados

1. Interno x Externo
2. Contínuo x Discreto
3. Estruturado x Não Est.
4. Nominal x Ordinal
5. Quantitativo x Qualitativo
6. Primário x Secundário

Fonte: Google/Coursera

[Prepare Data for Exploration](#)

Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Formato de dados

Dados primários

Coletado por um pesquisador de fontes em primeira mão

- Dados de uma entrevista que você realizou
- Dados de uma pesquisa retornada de 20 participantes
- Dados de questionários que você recebeu de um grupo de trabalhadores

Dados secundários

Recolhidos por outras pessoas ou de outras pesquisas

- Dados que você comprou dos perfis de clientes de uma empresa de análise de dados local
- Dados demográficos coletados por uma universidade
- Dados censitários coletados pelo governo federal

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

Fonte: Google/Coursera

[Prepare Data for Exploration](#)

Dados internos

Dados que vivem dentro dos próprios sistemas de uma empresa

- Salários dos funcionários em diferentes unidades de negócios monitorados pelo RH
- Dados de vendas por localização da loja
- Níveis de estoque de produtos nos centros de distribuição

Dados externos

Dados que residem fora de uma empresa ou organização

- Salários médios nacionais para os vários cargos em toda a sua organização
- Relatórios de crédito para clientes de uma concessionária de automóveis

Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Formato de dados

Dados contínuos

Dados que são medidos e podem ter praticamente qualquer valor numérico

- Altura das crianças nas classes da terceira série (52,5 polegadas, 65,7 polegadas)
- Marcadores de tempo de execução em um vídeo
- Temperatura

Dados discretos

Dados contados e com um número limitado de valores

- Número de pessoas que visitam um hospital diariamente (10, 20, 200)
- Capacidade máxima do quarto permitida
- Ingressos vendidos no mês atual

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

Fonte: Google/Coursera

[Prepare Data for Exploration](#)

Qualitativo

Medidas subjetivas e explicativas de qualidades e características

- Atividade de exercício mais apreciada
- Marcas favoritas dos clientes mais fiéis
- Preferências de moda de jovens adultos

Quantitativo

Medidas específicas e objetivas de fatos numéricos

- Porcentagem de médicos certificados pelo conselho que são mulheres
- População de elefantes na África
- Distância da Terra a Marte

Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Formato de dados

Nominal

Um tipo de dados qualitativos que não são categorizados com uma ordem definida

- Cliente pela primeira vez, cliente recorrente, cliente regular
- Novo candidato a emprego, candidato existente, candidato interno
- Nova listagem, listagem de preço reduzido, encerramento

Ordinal

Um tipo de dados qualitativos com uma ordem ou escala definida

- Classificações de filmes (número de estrelas: 1 estrela, 2 estrelas, 3 estrelas)
- Seleções de votação de escolha ranqueada (1º, 2º, 3º)
- Nível de renda (baixa renda, renda média, alta renda)

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

Fonte: Google/Coursera

[Prepare Data for Exploration](#)

Dados estruturados

Dados organizados em um determinado formato, como linhas e colunas

- Relatório de despesas
- Declarações fiscais
- Estoque da loja

Dados não estruturados

Dados que não são organizados de maneira facilmente identificável

- Postagens nas redes sociais
- E-mails
- Vídeos

Etapa 2 - Preparar

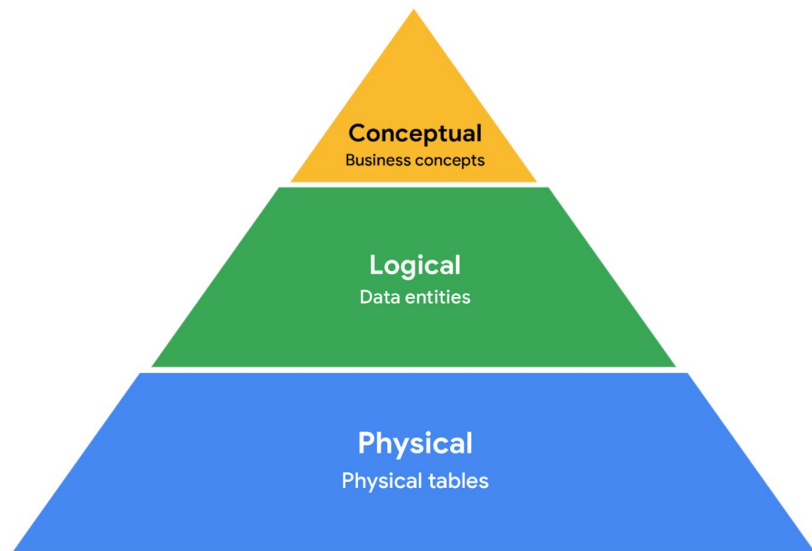
Banco de dados, conjunto de dados (Dataset) e software

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

B. Definição do conjunto de dados - Modelagem

The three most common types of data modeling



1. A modelagem de dados conceitual fornece uma visão de alto nível da estrutura de dados, como como os dados interagem em uma organização. Um modelo de dados conceitual não contém detalhes técnicos.
2. A modelagem de dados lógicos se concentra nos detalhes técnicos de um banco de dados, como relacionamentos, atributos e entidades. Não explica os nomes reais das tabelas do banco de dados.
3. A modelagem de dados físicos descreve como um banco de dados opera. Um modelo de dados físico define todas as entidades e atributos usados; por exemplo, inclui nomes de tabelas, nomes de colunas e tipos de dados para o banco de dados.

Fonte: Google/Coursera

[Prepare Data for Exploration](#)

B. Definição do conjunto de dados - Transformações

A transformação de dados é o processo de alteração do formato, estrutura ou valores dos dados para facilitar sua análise.

- Adicionando, copiando ou replicando dados
- Excluindo campos ou registros
- Padronizando os nomes das variáveis
- Renomear, mover ou combinar colunas em um banco de dados
- Unindo um conjunto de dados com outro
- Salvando um arquivo em um formato diferente.

C. Software para conjunto de dados e banco de dados.

- **LAI / LGPD / Dados Abertos**

Lei de Acesso à Informação

Lei Geral de Proteção de Dados

Dados abertos no Brasil e no Mundo

C. Software para conjunto de dados e banco de dados.

O que são dados abertos? Parte da ética dos dados. A abertura refere-se ao livre acesso, uso e compartilhamento de dados. Devem:

1. Estar disponível e acessível ao público como um conjunto de dados completo
2. Ser fornecido em termos que permitam a sua reutilização e redistribuição
3. Permitir a participação universal para que qualquer pessoa possa usar, reutilizar e redistribuir os dados

Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

C. Software para conjunto de dados e banco de dados.

Elementos de **metadados**

Título e descrição

Qual é o nome do arquivo ou site que você está examinando? Que tipo de conteúdo contém?

Marcas e categorias

Qual é a visão geral dos dados que você tem? Os dados são indexados ou descritos de uma forma específica?

Quem o criou e quando

De onde vieram os dados e quando foram criados? É recente ou existe há muito tempo?

Quem o modificou pela última vez e quando

Houve alguma alteração nos dados? Se sim, as modificações foram recentes?

Quem pode acessá-lo ou atualizá-lo

Este conjunto de dados é público? São necessárias permissões especiais para personalizar ou modificar o conjunto de dados?

Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

C. Software para conjunto de dados e banco de dados.

1. Planilha e editor de texto local
2. Documentos compartilhados online - <https://docs.google.com/>
3. Software especializado em análise de dados - <https://cran.r-project.org/> e <https://rstudio.cloud/> ou <https://qgis.org/>
4. Conexão com banco de dados online - <https://console.cloud.google.com/bigquery>
5. Banco de dados local sincronizado - <https://dbeaver.io/>

Instruções BigQuery:

https://scribhow.com/shared/Selecting_a_BigQuery_public_dataset_Qr3pplQVTOC7rn6HXRTZJw

D. Relação dados, tempo e espaço

Considerando que seu problema está delineado e você definiu quais dados precisa, o formato, tipo e temporalidade das informações iniciais, é necessário:

- 1. Ter uma estratégia comum de uso dos dados;**
- 2. Prever o período disponível para preparar os dados;**
- 3. Prever infraestrutura e tecnologias a serem usadas.**
- 4. Se houver dados pessoais ou restritos, abordar.**

D. Relação dados, tempo e espaço

Práticas recomendadas para convenções de nomenclatura de arquivos

1. Trabalhe e concorde com as convenções de nomenclatura de arquivos no início de um projeto para evitar renomear arquivos repetidamente.
2. Alinhe a nomenclatura de seus arquivos com as convenções de nomenclatura de arquivos existentes de sua equipe ou empresa.
3. Certifique-se de que seus nomes de arquivo sejam significativos; considere incluir informações como o nome do projeto e qualquer outra coisa que o ajude a identificar (e usar) rapidamente o arquivo para a finalidade correta.
4. Inclua a data e o número da versão nos nomes dos arquivos; formatos comuns são YYYYMMDD para datas e v### para versões (ou revisões).
5. Crie um arquivo de texto como um arquivo de amostra com conteúdo que descreva (decomponha) a convenção de nomenclatura de arquivo e um nome de arquivo que a aplique.
6. Evite espaços e caracteres especiais em nomes de arquivos. Em vez disso, use traços, sublinhados ou letras maiúsculas. Espaços e caracteres especiais podem causar erros em alguns aplicativos.

D. Relação dados, tempo e espaço

Práticas recomendadas para manter os arquivos organizados

1. Crie pastas e subpastas em uma hierarquia lógica para que os arquivos relacionados sejam armazenados juntos.
2. Separe o trabalho em andamento do concluído para que seus arquivos de projeto atuais sejam mais fáceis de encontrar.
3. Se o backup dos seus arquivos não for feito automaticamente, faça backup deles manualmente com frequência para evitar a perda de trabalhos importantes

Etapa 2 - Preparar

Banco de dados, conjunto de dados (Dataset) e software

Treinamento em Dados Ambientais 2022

Parte 01 - Etapas de Análise de Dados Ambientais

E. Planilha - Excel/Google

F. R-CRAN/R-Studio - Bibliotecas

G. Banco de dados Estruturado (SQL)

H. Sistemas de Informações Geográficas (SIG)

Parte 01 - Etapas de Análise de Dados Ambientais

Etapa 2 Preparar

Banco de dados, conjunto de dados (Dataset) e software

- A. Compreensão e estrutura do banco de dados
- B. definição do conjunto de dados
- C. Software para conjunto de dados e banco de dados.
- D. Relação dados, tempo e espaço
- E. Planilha - Excel/Google
- F. R-CRAN/R-Studio - Bibliotecas
- G. Banco de dados Estruturado (SQL)
- H. Sistemas de Informações
 - Geográficas (SIG)

Parte 01 - Etapas de Análise de Dados Ambientais

1. **Perguntar**
Fundamentos e recursos de dados ambientais
 2. **Preparar**
Banco de Dados, Dataset e Software
 3. **Limpar**
Configuração do dataset, limpeza e LOG
 4. **Processar**
Coleta e seleção de dados
 5. **Analisar**
Hipótese e ferramentas
 6. **Compartilhar**
Apresentar e mostrar
 7. **Agir**
Com base em dados
-