


PROJETO APLICADO I



Professor Fabio Lopes





PROJETO APLICADO I PROPOSTA E AED

PROPOSTA DE SOLUÇÃO ANALÍTICA

Quanto ao formato, deve apresentar:

- Sumário.
 - Glossário.
 - Objetivo de estudo.
 - Apresentação da empresa e problema de pesquisa.
 - Apresentação dos metadados e análise exploratória de dados.
-

APRESENTAÇÃO DA EMPRESA

Objetivo: buscar familiaridade com o contexto do estudo.

Utilize uma abordagem top/down para apresentar, indicando:

- Nome da empresa.
- Missão/visão/valores.
- Segmento de atuação (pode ser vários, mas destaque aquele que será o objeto do estudo).
- Market Share/posicionamento no mercado.
- Número de colaboradores.
- Iniciativas na área de Data Science.
- Trabalhos em destaque.

PROBLEMA DO ESTUDO

Objetivo: identificar perguntas que pretendemos responder.

O que falta?

O que incomoda?

O que pode ser melhorado?

Qual é o gap?

Há um padrão que pode ser observado?

Há uma afirmação que pode ser contestada?

Utilize a abordagem do pensamento computacional:

- Temos um problema que pode ser decomposto em problemas menores?
- Temos padrões que podem ser observáveis?
- Podemos abstrair (ignorar) detalhes e usar uma generalização para buscar uma solução?
- Podemos escrever um algoritmo para alcançar resultados?

METADADOS

Objetivo: caracterizar e registrar os datasets que temos para o estudo.

Itens que podemos registrar neste processo:

- Tipo de arquivo (csv, json, xml etc.).
- Origem dos dados (aberto ou privado).
- Sensibilidade (possui dados sensíveis).
- Validade (quando foi gerado, quando se torna obsoleto).
- Proprietário do dado (quem é responsável pelo dataset).
- Informações adicionais.
- Restrições de uso (LGPD, por exemplo).
- Descrição dos atributos (definição e tipo de dado).

ANÁLISE EXPLORATÓRIA DE DADOS

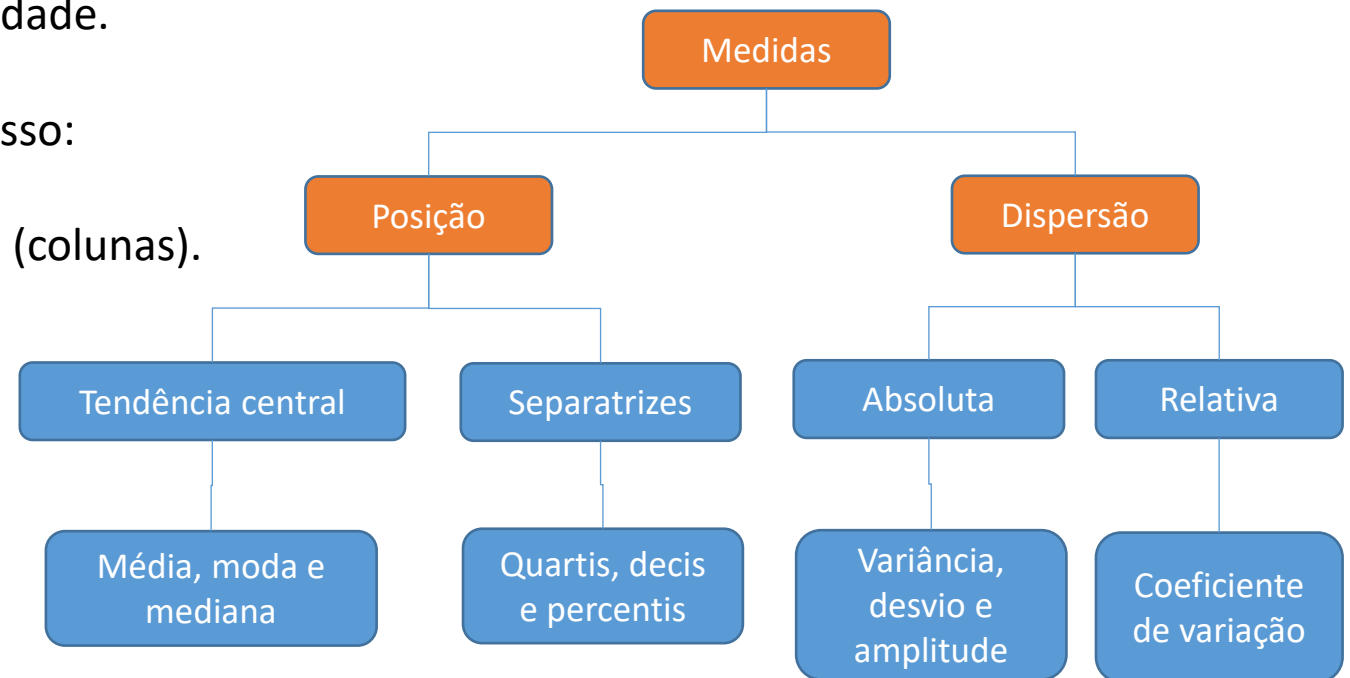
Objetivo: descobrir quais são os dados que temos para nosso estudo.

Trata-se de uma abordagem para analisar datasets e sumarizar suas principais características.

Utilizamos a estatística descritiva para esta atividade.

Perguntas que podemos responder neste processo:

- Número de exemplares (linhas) e dimensões (colunas).
- Tipos de dados.
- Medidas de posição e dispersão.
- Distribuição e frequência.
- Correlações.
- Valores perdidos ou incorretos.
- Anomalias e outliers.



Torturaremos os dados até que eles confessem...

