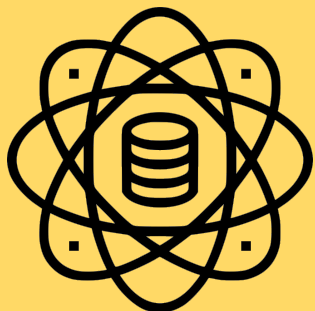




Curso:

Introdução à Análise de dados

Visão geral sobre análise de dados



28 de Janeiro de 2020
São José dos Campos - SP

Felipe Carvalho
Felipe Menino

Quem somos

FELIPE CARVALHO

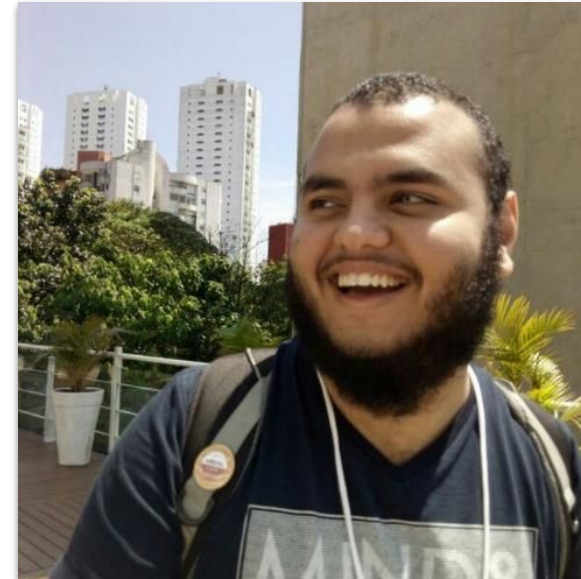


@oldlipe



lipecaso@gmail.com

FELIPE MENINO



@M3nino



efelipecarlos@gmail.com

Material

Disponível no Github:

- Ambiente: <https://github.com/dataAt/inpe-ambiente>
- Apresentação: <https://introducao-ds-inpe-2020.netlify.com>
- Códigos: <https://github.com/dataAt/introducao-analise-de-dados>
- Livro-Texto: <https://dataat.github.io/introducao-analise-de-dados/>

Objetivo

Objetiva-se apresentar ao aluno o processo básico de análise de dados, utilizando as linguagens de programação R e Python, de acordo com a tecnologias mais usadas no mercado e na academia.

Sobre os dados



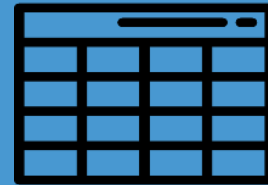
Dado

É tudo aquilo que é registrado, para posteriormente, ser analisada



Formatos

Cada formato pode apresentar etapas diferentes de organizações



Data Frame

Tabelas utilizadas para armazenar, manipular e representar dados



Peso

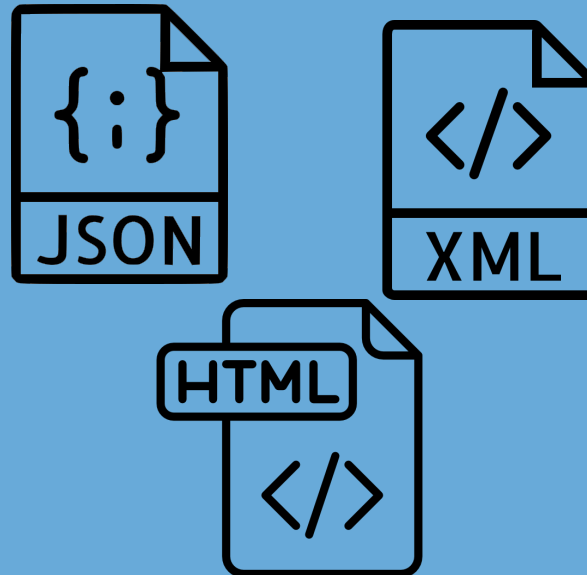
Entenda que todo dado tem um peso, contexto e impacto ao ser analisado

Formato dos dados

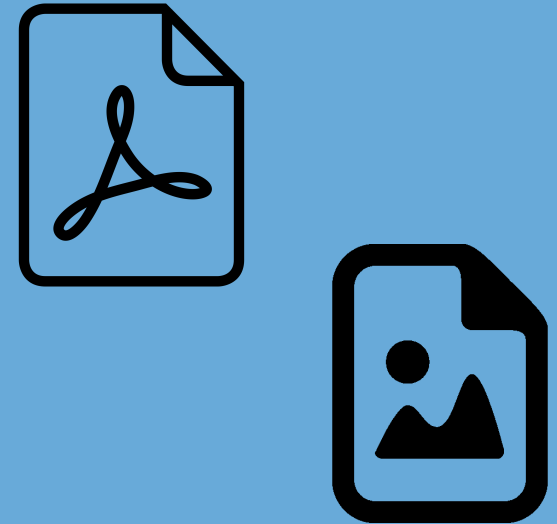
Dados Estruturados



Dados Semiestruturados



Dados não Estruturados



Análise de dados

Para John W. Tukey, grande parte da análise de dados é inferencial, ou seja, o ato de extrair informações de uma amostra em relação ao conjunto todo.

→ Não é um conceito novo!

- Tukey fez esta definição em 1961! No artigo *The Future of Data Analysis*

O analista de dados

De acordo com artigo publicado pela Dataquest, o analista de dados é um:

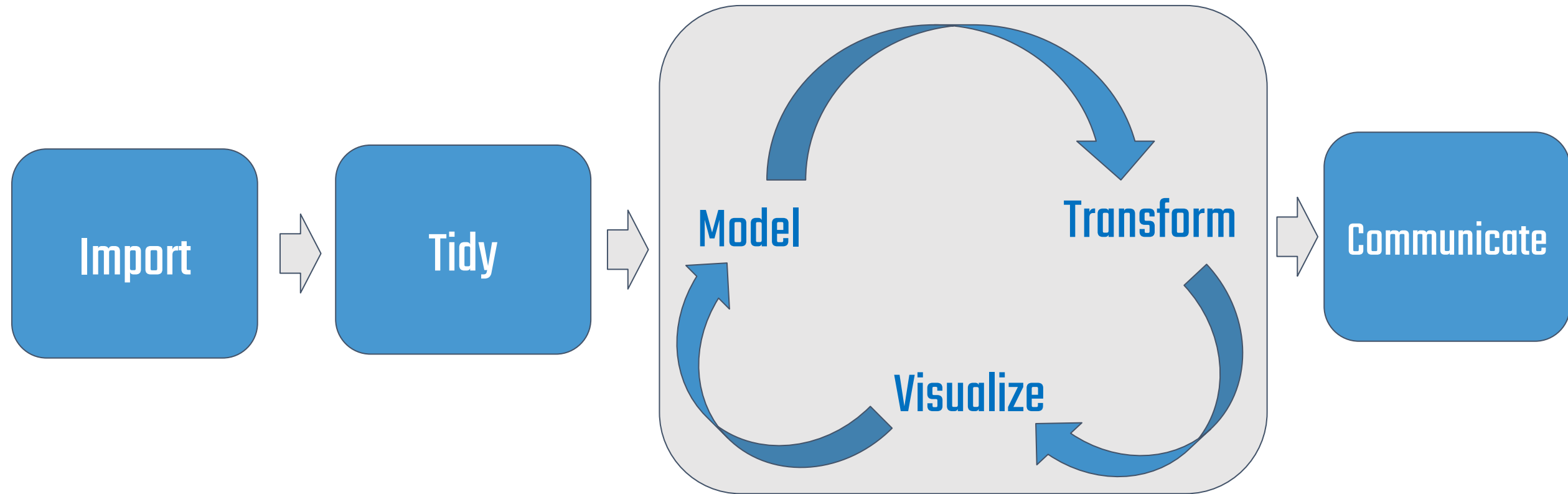
- Agregador de valor para a companhia através de respostas obtidas com dados, dessa forma, ajudando na tomada de decisão da empresa

O analista de dados

Possui como principais atividades:

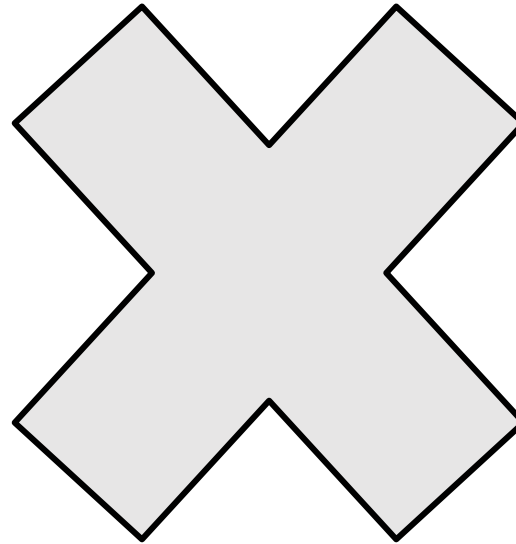
- Limpeza e organização de dados;
- Uso de estatística para ter uma visão geral dos dados;
- Análise de tendências encontradas nos dados;
- Criação de visualização e dashboards para ajudar na interpretação e tomada de decisão da empresa;
- Apresentação dos resultados obtidos pela análise para os clientes.

O processo de Análise de dados



Fazendo Análise de Dados

R VS Python ?



Ambiente

Jupyter Notebook

- Documentos interativos;
- Apresentação bonita e amigável;
- Fácil distribuição dos resultados das análises.



Plataforma

Kaggle

- Repositório de dados;
- Ambiente simples e fácil de utilizar;
- Recursos computacionais disponibilizados gratuitamente.

