



Aula 18 – Parte 2

Introdução à Ciência de Dados

Desvendando o mundo dos dados, das profissões ao pipeline completo, com foco em Python.

Por que Ciência de Dados?



Decisões Orientadas por Dados

Transição da intuição para insights embasados em evidências concretas.



Explosão do Volume de Dados

O Big Data exige novas abordagens para extrair valor e conhecimento.



Ferramentas Acessíveis

Com Python, Pandas e BI, a análise de dados está ao alcance de todos.

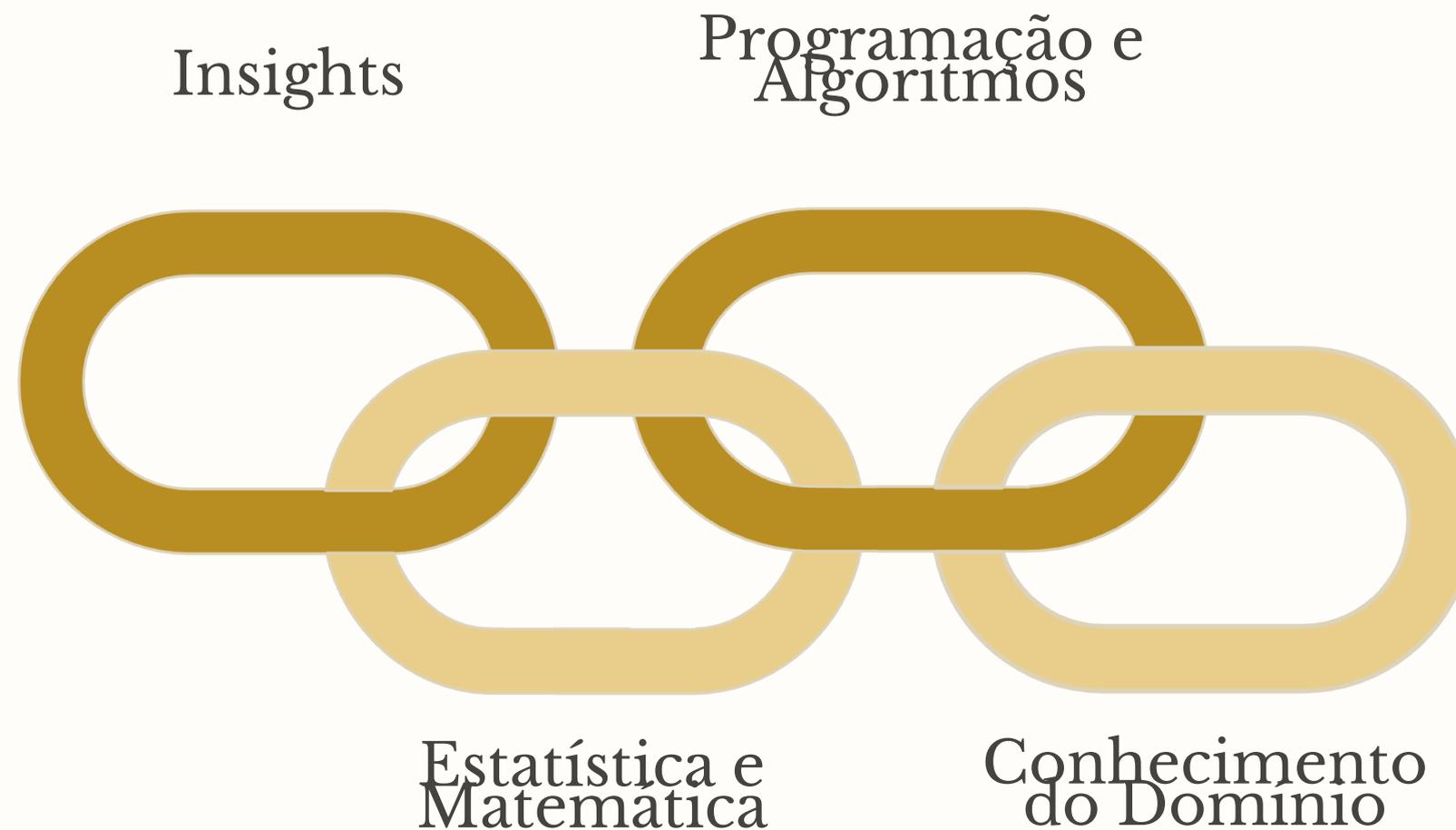


Aplicações Universais

Da saúde às finanças, a Ciência de Dados impulsiona a inovação em todos os setores.

🧠 O que é Ciência de Dados?

É a arte de transformar dados brutos em inteligência acionável, combinando:



O objetivo é gerar informações úteis, previsões e decisões inteligentes.



Profissões na Ciência de Dados

Cientista de Dados	Modelagem, previsão e análise profunda de dados.
Analista de Dados	Dashboards, relatórios e análise descritiva de dados.
Engenheiro de Dados	Infraestrutura e pipelines de dados para sistemas.
Engenheiro de IA / ML	Desenvolve e implanta modelos de aprendizado de máquina.
Arquiteto de Dados	Projeta a arquitetura de dados da organização.



O Pipeline da Ciência de Dados

1. Coleta de Dados

Aquisição de dados de diversas fontes.

2. Limpeza e Pré-processamento

Tratamento de dados ausentes, inconsistentes e formatos.

3. Análise Exploratória (EDA)

Compreensão inicial dos dados e identificação de padrões.

4. Modelagem e Algoritmos

Criação de modelos preditivos ou descritivos.

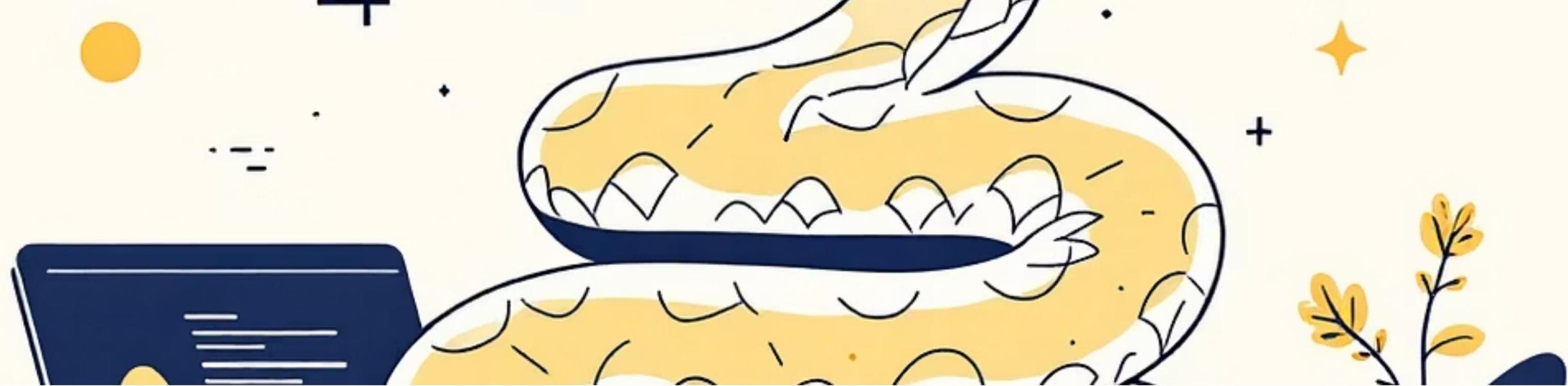
5. Avaliação de Modelos

Verificação da performance e robustez dos modelos.

6. Implantação e Monitoramento

Colocação do modelo em produção e acompanhamento contínuo.

Exemplo prático: Análise de sobrevivência no Dataset Titanic.



Pacotes Python Essenciais

- **Pandas:** Manipulação de dados tabulares (DataFrames).
- **NumPy:** Operações numéricas eficientes e arrays.
- **Matplotlib & Seaborn:** Visualização de dados de alta qualidade.
- **Scikit-learn:** Ferramentas abrangentes de Machine Learning.

Python é o ecossistema padrão e mais robusto da área!

🔍 Estudo de Caso: Titanic



Objetivo: Prever a sobrevivência no Titanic

Utilizamos as características dos passageiros (idade, classe, sexo) para construir um modelo preditivo.

- **Limpeza:** ``read_csv()`, `drop()`, `fillna()``.
- **Engenharia de Features:** ``get_dummies()``.
- **Modelagem:** ``train_test_split()`` e ``RandomForestClassifier()``.
- **Avaliação:** ``score()`` - Acurácia de ~79%.

Pensamento Computacional & Ciência de Dados

Decomposição

O pipeline se divide em pequenas e gerenciáveis etapas.

Abstração

DataFrames e funções simplificam a complexidade dos dados.

Reconhecimento de Padrões

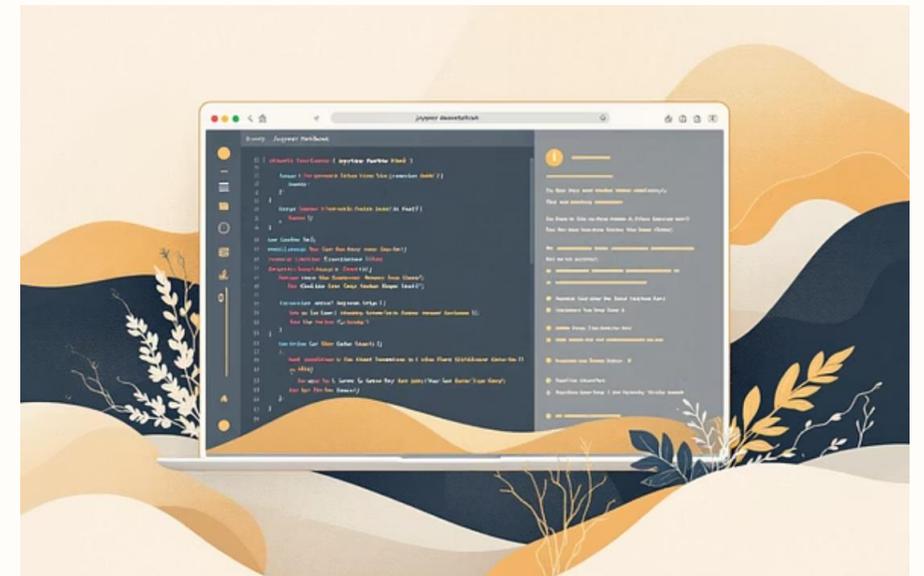
A lógica de agrupamento e filtragem revela insights.

Algoritmos

Bibliotecas de ML aplicam soluções avançadas.

🧠 Conexões com a Disciplina

Ciência de Dados integra e aplica conceitos fundamentais aprendidos:



É a prova de que tudo se conecta!

Para Explorar Mais

Recursos Adicionais:

- pypro.com.br: Livros e cursos completos.
- **Capítulo 11** – Do livro *Python Essencial*.

Próximos Passos:

- Projeto real com dados de **Credit Score**.
- Aplicação prática de todo o pipeline da Ciência de Dados.

 Ciência de Dados é o **futuro** e o **presente!**