



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3



Apresentação

A Formação Analista de Dados é um programa de aperfeiçoamento profissional 100% online e 100% em português, com certificado de conclusão.

Seja Muito Bem-vindo(a)!

Formação Analista de Dados

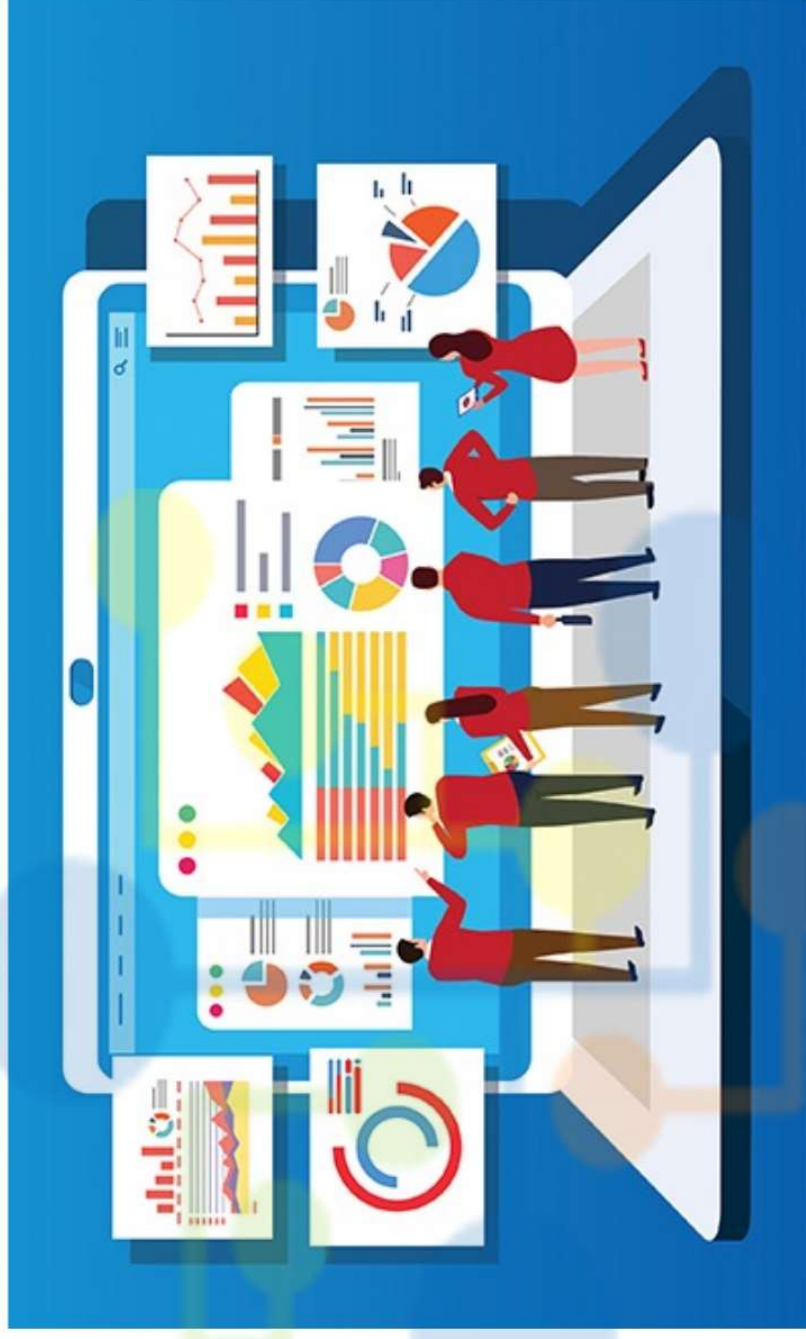


Copyright © Data Science Academy



SQL Para Data Science

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3



Copyright © Data Science Academy



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3

Categorização, Codificação e Binarização de Variáveis com SQL



Copyright © Data Science Academy

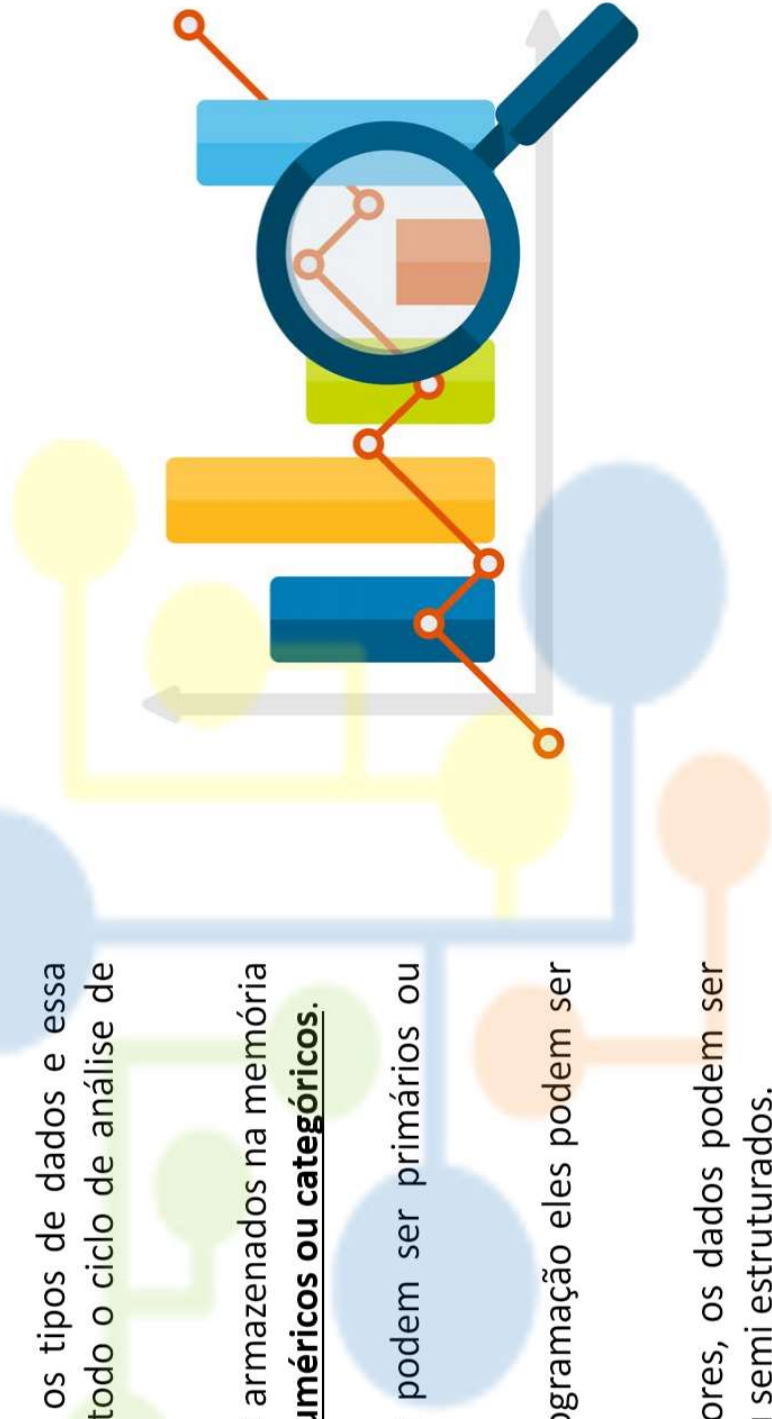


Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3

Tipos de Dados

Temos quatro classificações para os tipos de dados e essa compreensão é importante para todo o ciclo de análise de dados.

- 1- Com relação aos tipos de dados armazenados na memória do computador, eles podem ser numéricos ou categóricos.
- 2- Em relação à fonte os dados podem ser primários ou secundários.
- 3- Em relação à linguagem de programação eles podem ser primitivos ou não primitivos.
- 4- Independente dos itens anteriores, os dados podem ser estruturados, não estruturados ou semi estruturados.



Copyright © Data Science Academy



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3

O Que é Categorização?

Estado do Cliente	Idade do Cliente	Faixa Etária do Cliente
Rio de Janeiro	28	De 21 a 30 Anos
São Paulo	34	De 31 a 40 Anos
Ceará	62	De 61 a 70 Anos
Santa Catarina	46	De 41 a 50 Anos
Amazonas	43	De 41 a 50 Anos
Goiás	67	De 61 a 70 Anos

Esta é uma variável categórica que representa classes ou categorias.

Esta é uma variável numérica (operações matemáticas fazem sentido).

Categorização é converter uma variável para o tipo categórico a fim de obter um novo tipo de informação.



Copyright © Data Science Academy



O Que é Codificação (Encoding)?

Carro	Cor	Cor_Label
Fusca	Branco	1
Passat	Verde	2
Uno	Preto	3
Kadet	Verde	2
Santana	Azul	4
Monza	Preto	3

Esta é uma variável categórica que representa classes ou categorias.

Esta é uma variável categórica que representa classes ou categorias.

Codificação é transformar uma variável para uma representação numérica, sem perder a essência da informação.





Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3

O Que é Codificação (Encoding)?

Existem 2 tipos principais de Encoding:
Label Encoding e One-Hot-Encoding.



Copyright © Data Science Academy



O Que é Codificação (Encoding)?

Carro	Cor	Cor_Label
Fusca	Branco	1
Passat	Verde	2
Uno	Preto	3
Kadet	Verde	2
Santana	Azul	4
Monza	Preto	3

O exemplo acima chamamos de Label Encoding e normalmente usamos quando a variável categórica tem até 5 categorias.





O Que é Codificação (Encoding)?

Carro	Cor	Cor_Label_ Branco	Cor_Label_ Verde	Cor_Label_ Preto	Cor_Label_ Azul	Cor_Label_ Vermelho	Cor_Label_ Prata
Fusca	Branco	1	0	0	0	0	0
Passat	Verde	0	1	0	0	0	0
Uno	Preto	0	0	1	0	0	0
Kadet	Verde	0	1	0	0	0	0
Santana	Azul	0	0	0	1	0	0
Monza	Preto	0	0	1	0	0	0
Chevette	Vermelho	0	0	0	0	1	0
Puma	Prata	0	0	0	0	0	1

O exemplo acima chamamos de One-Hot-Encoding, sendo usado para criar variáveis dummy (representações numéricas para cada categoria de uma variável categórica) e normalmente usamos quando a variável categórica tem mais de 5 categorias.





Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3

O Que é Codificação (Encoding)?

Aplicamos a Codificação às variáveis categóricas e raramente precisamos aplicar a variáveis numéricas.

O objetivo é exatamente converter a variável categórica para uma representação numérica sem perder a essência da informação contida na variável.



Copyright © Data Science Academy



O Que é Binarização?

Estado do Paciente	Paciente com Diabetes	Paciente com Diabetes Bin
Rio de Janeiro	Sim	1
São Paulo	Não	0
Ceará	Não	0
Santa Catarina	Não	0
Amazonas	Sim	1
Goiás	Não	0

Esta é uma variável categórica que representa classes ou categorias.

Esta é uma variável categórica que representa classes ou categorias.

Binarização é converter uma variável para sua representação binária sem perder a essência da informação.





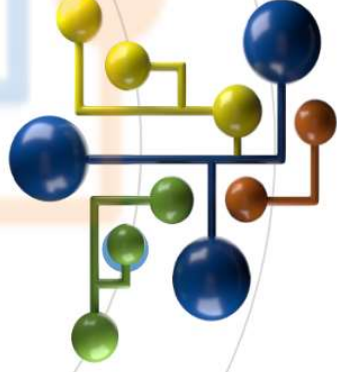
Data Science
Academy

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3



Muito Obrigado!

Continue Trilhando uma Excelente Jornada de Aprendizagem!



Copyright © Data Science Academy