

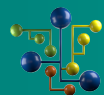
Apresentação

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

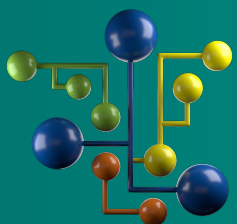




Data Science
Academy

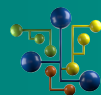
Data Science Academy santana.tavares.d@gmail.com 6176a538e32fc3fc3d50675a

Análise de Dados com Linguagem Python



Data Science
Academy

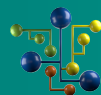




Análise de Dados com Linguagem Python

Engenharia de Atributos

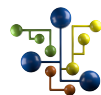




Análise de Dados com Linguagem Python

O Que é Engenharia de Atributos?



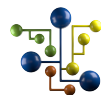


O Que é Engenharia de Atributos?

Idade	Valor Gasto na Compra de Roupas e Calçados Por Mês
25	R\$ 120
27	R\$ 150
38	R\$ 180
46	R\$ 90
52	R\$ 85

Qual o valor médio gasto por mês em roupas e calçados, por faixa etária?



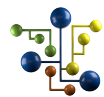


O Que é Engenharia de Atributos?

Faixa Etária	Valor Médio Gasto na Compra de Roupas e Calçados Por Mês
De 20 a 29 anos	R\$ 135
De 30 a 39 anos	R\$ 180
De 40 a 49 anos	R\$ 90
De 50 a 59 anos	R\$ 85

A criação de uma nova variável a partir dos dados originais é o que fazemos na Engenharia de Atributos.

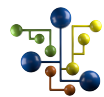




O Que é Engenharia de Atributos?

A Engenharia de Atributos (também chamada Engenharia de Recursos) é uma técnica de análise de dados que aproveita os dados para criar novas variáveis que não estão no conjunto de dados original.

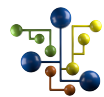




O Que é Engenharia de Atributos?

A Engenharia de Atributos permite produzir novos recursos para análise de dados, com o objetivo de simplificar e acelerar as transformações de dados.

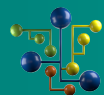




O Que é Engenharia de Atributos?

A Engenharia de Atributos permite compreender os dados por diferentes perspectivas.





Análise de Dados com Linguagem Python

Feature Selection
x
Feature Extraction





Feature Selection x Feature Extraction

Idade	Peso (Kg)	Altura (cm)	Hora de Dormir	Ingestão de Bebida Alcoólica	Nota nas Aulas de Matemática	Diabetes
34	78	167	23:00	Sim	10	Sim
52	83	175	23:30	Não	9	Sim
19	105	190	01:30	Sim	7.5	Não

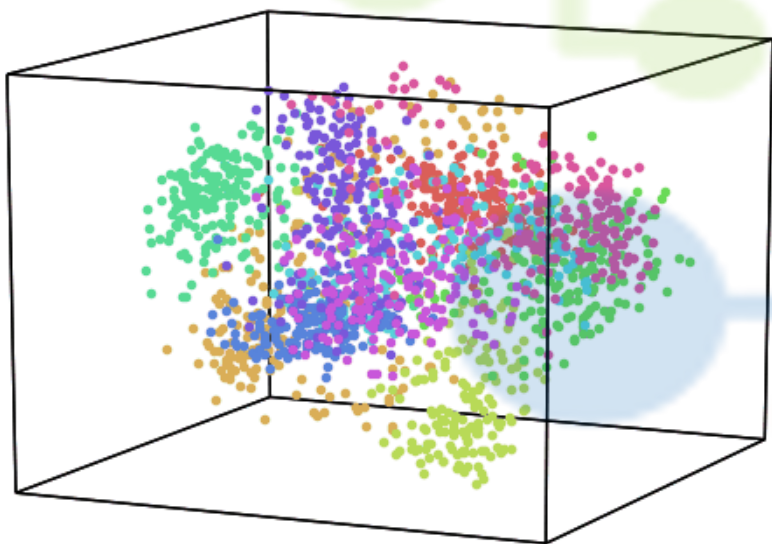
Será que todas as variáveis são relevantes para analisar a ocorrência de diabetes?

Para responder a pergunta precisamos aplicar técnicas de seleção de atributos (feature selection).





Feature Selection x Feature Extraction



Imagine um dataset com 1.000 colunas (muito comum em datasets de Processamento de Linguagem Natural). Torna-se inviável trabalhar com este dataset devido à chamada “maldição da dimensionalidade”.

Sendo assim, devemos aplicar técnicas de extração de atributos (feature extraction), reduzindo de 1.000 variáveis para 5 (por exemplo) e então realizar o processo de análise.





Feature Selection x Feature Extraction

Feature Selection e Feature Extraction são técnicas avançadas de Engenharia de Atributos que normalmente envolvem Machine Learning ou técnicas matemáticas/estatísticas avançadas.

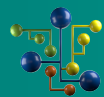




Feature Selection x Feature Extraction

Neste curso trabalharemos com técnicas fundamentais de Engenharia de Atributos e aplicaremos fundamentos estatísticos.

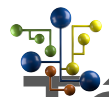




Análise de Dados com Linguagem Python

Técnicas Mais Comuns
de Engenharia de
Atributos





Técnicas Mais Comuns de Engenharia de Atributos

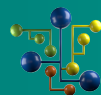
Técnicas Fundamentais:

1. Imputação
2. Tratamento de Outliers
3. Binarização
4. Transformação de Log
5. One-Hot Encoding
6. Operações de Agrupamento
7. Split (Divisão) de Atributos
8. Scaling (Padronização)
9. Extração de Data
10. Categorização

Técnicas Avançadas:

1. Seleção Univariada de Atributos
2. Principle Components Analysis (PCA)
3. Independent Component Analysis (ICA)
4. Linear Discriminant Analysis (LDA)
5. Locally Linear Embedding (LLE)
6. t-distributed Stochastic Neighbor Embedding (t-SNE)
7. Autoencoders





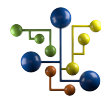
Data Science
Academy

Data Science Academy santana.tavares.d@gmail.com 6176a538e32fc3fc3d50675a

Análise de Dados com Linguagem Python

Engenharia de Atributos Manual
x
Engenharia de Atributos
Automática

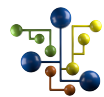




Engenharia de Atributos Manual x Engenharia de Atributos Automática

A Engenharia de Atributos Manual requer conhecimento em programação, essencialmente como fizemos em todo o curso até aqui.

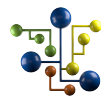




Engenharia de Atributos Manual x Engenharia de Atributos Automática

Mas o processo pode ser automatizado via AutoML ou mesmo usando modelos avançados de Deep Learning.

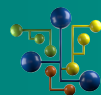




Engenharia de Atributos Manual x Engenharia de Atributos Automática

Com a automação podemos perder a capacidade de interpretação dos dados, prejudicando a entrega do resultado de análise. Devemos usar a Engenharia de Atributos Automática com parcimônia.

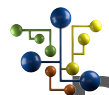




Análise de Dados com Linguagem Python

Processo Iterativo de Engenharia de Atributos

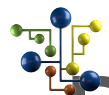




Processo Iterativo de Engenharia de Atributos

Iteração é o processo na programação de repetição de uma ou mais ações. É importante salientar que cada iteração se refere a apenas uma instância da ação, ou seja, cada repetição possui uma ou mais iterações.

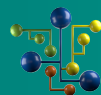




Processo Iterativo de Engenharia de Atributos

Engenharia de Atributos é um processo iterativo e cada transformação que fazemos pode gerar a necessidade de novas transformações.

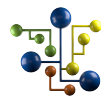




Análise de Dados com Linguagem Python

Decomposição Categórica de Atributos

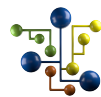




Decomposição Categórica de Atributos

Trabalhar com variáveis categóricas é sempre um desafio e na maioria das vezes queremos converter uma variável categórica para uma representação numérica sem perder a informação, ou realizar agrupamentos a fim de simplificar o processo de análise de variáveis categóricas. Para isso empregamos a decomposição categórica de atributos durante a engenharia de atributos.

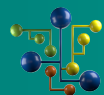




Decomposição Categórica de Atributos

Em Machine Learning não usamos variáveis categóricas como variáveis preditoras e sempre convertemos as variáveis categóricas para sua representação numérica através das técnicas de encoding.

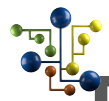




Análise de Dados com Linguagem Python

Decomposição Baseada em Tempo

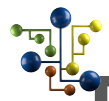




Decomposição Baseada em Tempo

Em muitos casos uma variável do tipo data deve ser decomposta em sua hierarquia com ano, mês, dia, hora, minuto, segundo, etc..., para atender aos requisitos do que precisamos entregar no processo de análise.

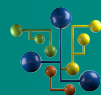




Decomposição Baseada em Tempo

Podemos fazer a decomposição em tempo de análise ou decompor e gerar novos atributos em nosso conjunto de dados.

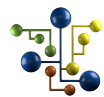




Análise de Dados com Linguagem Python

Reframe Numérico

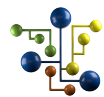




Reframe Numérico

O Reframe Numérico compreende um conjunto de técnicas para ajustar colunas numéricas. Podemos usar Padronização, Normalização, Transformação de Log e qualquer outra técnica de engenharia de atributos aplicável a variáveis numéricas.

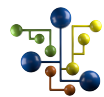




Reframe Numérico

Dependendo da variável técnicas matemáticas mais avançadas podem ser necessárias.

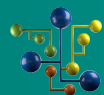




Reframe Numérico

Lembre-se: em qualquer transformação não podemos perder a essência da informação contida na variável.





Data Science
Academy

Data Science Academy santana.tavares.d@gmail.com 6176a538e32fc3fc3d50675a

Análise de Dados com Linguagem Python

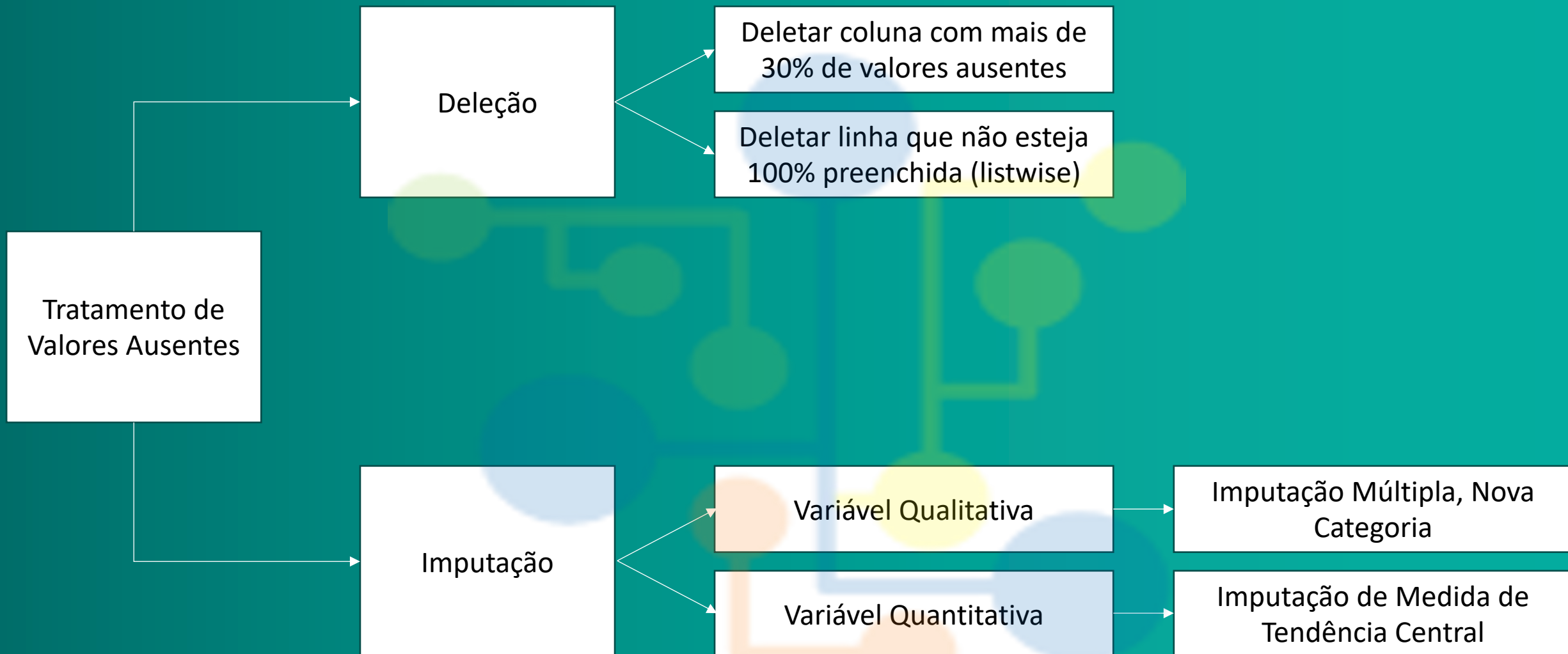
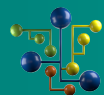
Projeto 8

Engenharia de Atributos em Dados Médicos



Data Science
Academy







Muito Obrigado!

Continue Trilhando uma Excelente Jornada de Aprendizagem!

