

Análise de Dados com Linguagem Python

O Ciclo de Vida da Análise de Dados



Ciclo de Vida da Análise de Dados



Nesta etapa definimos o problema a ser resolvido.



Ciclo de Vida da Análise de Dados



Nesta etapa realizamos limpeza, modelagem, checagem de qualidade dos dados e transformações quando necessárias.



Ciclo de Vida da Análise de Dados



Aqui o objetivo é explorar os dados, extrair estatísticas descritivas, segmentar os dados, incluir ou remover variáveis.



Ciclo de Vida da Análise de Dados



Se os dados estiverem prontos para uso podemos agora visualizá-los e compreendê-los.



Ciclo de Vida da Análise de Dados



Por fim, extraímos insights e recomendamos ações aos tomadores de decisão. Ou simplesmente passamos os dados adiante para modelagem preditiva.





Análise de Dados X Ciência de Dados

A grande diferença aqui está no fato que a Análise de Dados não aborda Machine Learning.

Ciência de Dados engloba Análise de Dados mais aprendizado de máquina ou mesmo IA.





Data Analytics X Business Analytics

Data Analytics tem o foco maior nas técnicas de análise de dados. Aqui o perfil é técnico.

Business Analytics tem foco maior nas características do negócio e como a análise de dados pode ser empregada para resolver os problemas. Normalmente o perfil é menos técnico.



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.

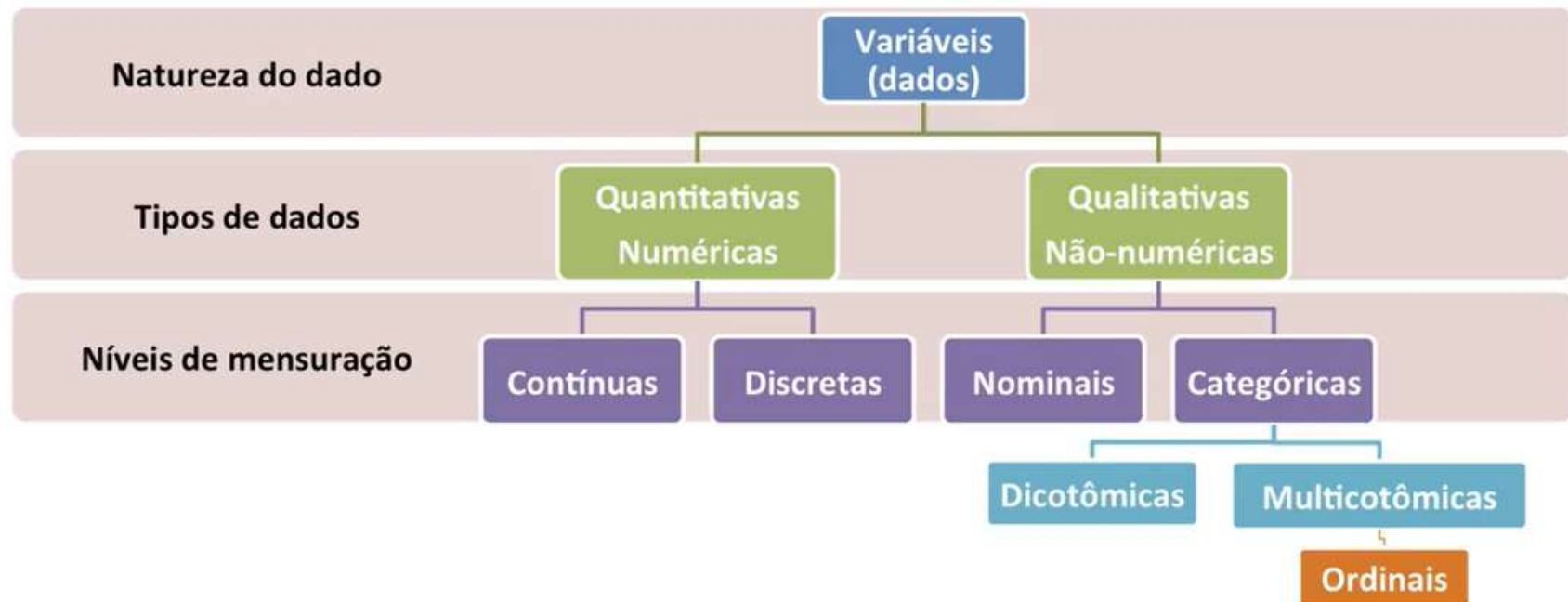


Análise de Dados com Linguagem Python

Numéricos
X
Categóricos

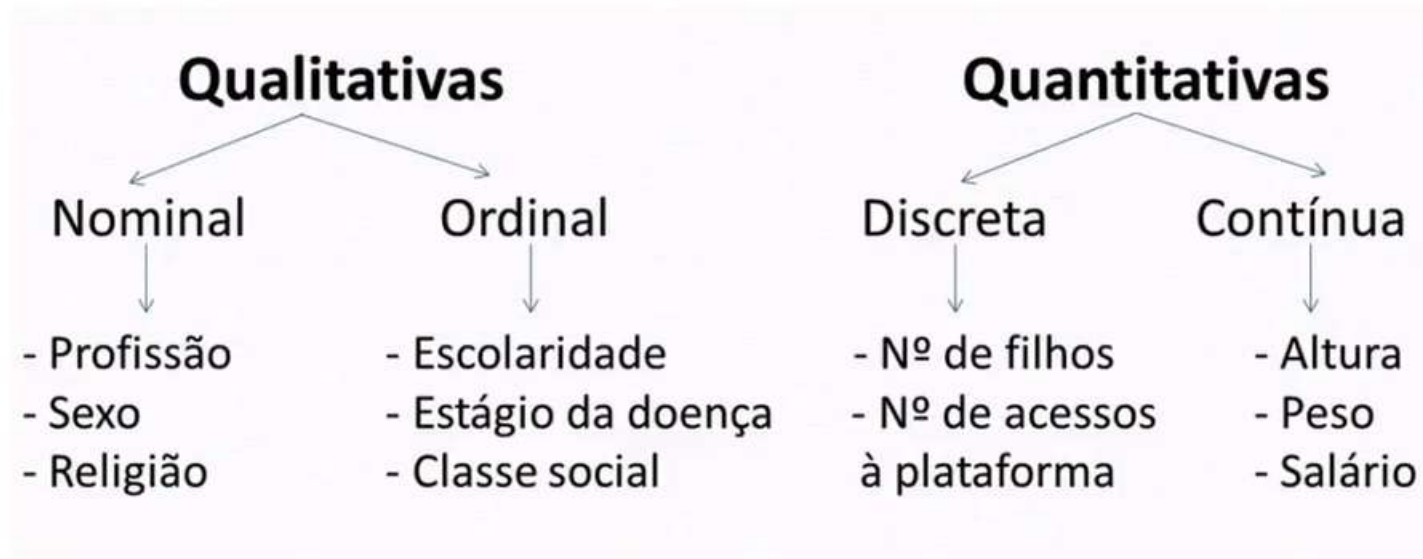


Numéricos x Categóricos



Numéricos x Categóricos

Tipos de Variáveis (Tipos de Dados)



Análise de Dados com Linguagem Python

Primários
X
Secundários



Primários x Secundários

Dados primários

Dados obtidos de geração própria, da própria empresa ou da própria pesquisa. São dados captados no CRM, nas redes sociais, no site da própria empresa, por exemplo. Tudo aquilo que é de domínio próprio da organização.



Primários x Secundários

Dados secundários

São dados não próprios, de uma segunda fonte externa, mas que tem uma característica importante do dado não ser público. O dado que se obtém de forma secundária não pode ser adquirido livremente por qualquer empresa. É o caso de parcerias de troca de dados entre empresas, que muitas vezes estão em setores complementares, mas não concorrentes.



Primários x Secundários

Dados terciários

São também dados externos, não próprios, mas esses sim públicos. Qualquer outra empresa pode ter acesso, de forma paga ou gratuita. Hoje muitas empresas fazem coleta, armazenamento e enriquecimento de dados públicos para serem comercializados livremente no mercado.

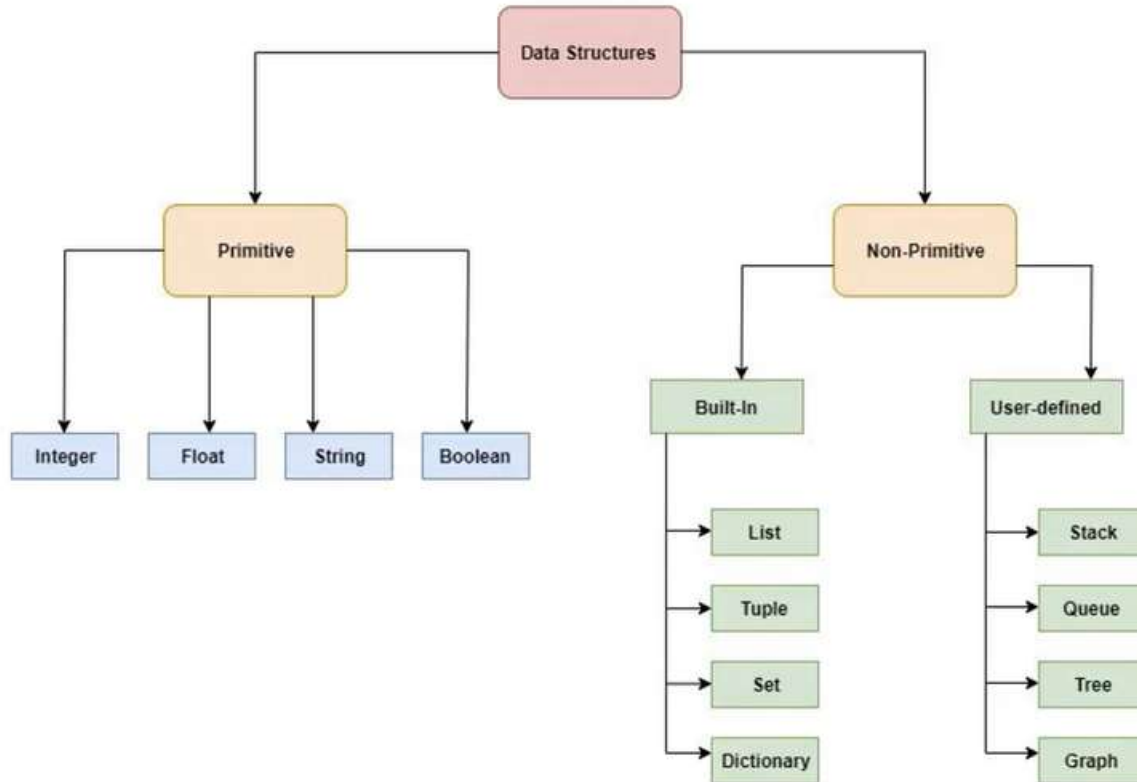


Análise de Dados com Linguagem Python

Primitivos
X
Não Primitivos



Primitivos x Não Primitivos



Análise de Dados com Linguagem Python

Estruturados
X
Não Estruturados
X
Semi Estruturados



Estruturados x Não Estruturados x Semi Estruturados

Dados estruturados

São aqueles que têm uma classificação e uma lógica formal na maneira como são esquematizados. Aqui estamos falando de dados em um banco de dados, com linhas e colunas ou mesmo em uma planilha do Excel. Dados estruturados são dados tabulares com estrutura bem definida.



Estruturados x Não Estruturados x Semi Estruturados

Dados não-estruturados

São os dados estruturalmente mais brutos, capturados sem classificação ou esquema formal. São dados como posts em redes sociais, imagens, vídeos, e-mails, etc. Apesar de serem, sem dúvida, os dados mais abundantes que existem, eles requerem um grande trabalho – estratégico e operacional – para poderem ser utilizados de maneira útil no ciclo de análise de dados.



Estruturados x Não Estruturados x Semi Estruturados

Dados semi-estruturados

São dados igualmente sem uma organização esquemática, mas aqui já classificados de alguma forma. Isso significa que há como saber como os dados estão separados, mas algum trabalho de estruturação deverá ser feito. Um bom exemplo são arquivos XML e JSON muito usados em aplicações web.



Análise de Dados com Linguagem Python

Definido os
Metadados



Definido os Metadados

Metadados

São aqueles dados que fornecem informações sobre outros dados. Um exemplo fácil do dia a dia está no nosso computador. Quando vemos o tamanho de um arquivo, sua extensão, a data de criação ou de modificação, são todos metadados que dão informação sobre um documento que em si já carrega uma série de dados (desestruturados nesse caso).

Metadados são dados sobre os dados.

