



Analytics para área da saúde

Organização de Dados

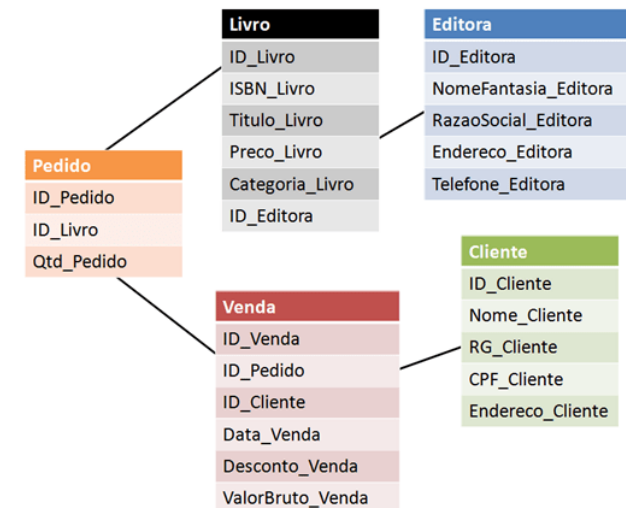
A organização dos dados em formato de tabelas é crucial para que os analistas e cientistas de dados possam ter um acesso facilitado aos dados e para que possam realizar suas análises com maior facilidade, podendo criar relações entre seus dados, enriquecendo assim seus objetos de estudo e análise.



Nome	Idade	Departamento		Salário	Telefone
Alberto da Silva	25	Vendas	RS	850,00	555-1902
Antônio dos Santos	32	Administração	RS	1.200,00	555-1117
Fabiana Rossi	40	Administração	RS	2.000,00	555-8929
Horácio Almeida	31	Recursos Humanos	RS	1.350,00	555-8907
João Pereira	35	Vendas	RS	1.500,00	555-7814
Márcia Souza	26	Vendas	RS	600,00	555-9800
Maria José Costa	22	Vendas	RS	600,00	555-6629
Mário Oliveira	54	Diretoria	RS	4.500,00	555-1237
Roberto Albuquerque	29	Administração	RS	1.200,00	555-8273
Sílvia Pires	23	Vendas	RS	600,00	555-8664

Relação de Dados

Um dos modelos mais utilizados na organização de dados em tabelas é o **modelo de dados relacional**, no qual as informações são pulverizadas em tabelas individuais que podem ser relacionadas através de chaves de cruzamento em comum. As chaves de cruzamento são campos nas tabelas que devem ser únicos em representar as linhas das tabelas.



Tabelas Cruzadas

Profissional	Compreensão do Prontuário		
	Não entendeu	Parcialmente	Totalmente
Enfermeiro	4	47	125
Fisio	7	50	119
Fono	32	38	106
Médico	14	60	102
Nutri	6	80	90
Psicologia	62	52	62
TO	39	28	109

<http://www.portaaction.com.br/tabela-de-contingencia/introducao-tabelas-de-contingencia>

A **tabela cruzada** é um exemplo de organização dos dados e de demonstração de dados como relatórios, também são chamados de **relatórios de matriz**, e também de **matriz de contingência**. Este tipo de organização de dados é utilizado para visualizar as relações entre três ou mais itens de consulta.

Os relatórios de tabela cruzada mostram os dados em linhas e colunas com informações resumidas nos pontos de interseção, ou seja as linhas e colunas são relacionadas. São comumente utilizadas em resumos estatísticos para quantificar populações de testes e experimentos. O número de observações podem ser dispostos em total absoluto ou em relativo (percentual) pelo total das linhas ou pelo total das colunas.

Tabela plana

ID colaborador	Data Nascimento	Nome Colaborador	Área	Tipo Função	data inicio empresa	Sexo	altura (m)	peso (kg)	IMC (kg/m2)
0001	22/05/1996	colaborador_0001	Produção	Operacional	13/01/2019	F	1,71	52,0	17,8
0002	07/06/1989	colaborador_0002	Produção	Operacional	21/05/2015	F	1,69	62,0	21,7
0003	06/12/1992	colaborador_0003	Produção	Operacional	18/04/2019	F	1,71	59,0	20,2
0004	22/05/1986	colaborador_0004	Produção	Operacional	16/02/2019	F	1,69	57,0	20,0
0005	07/09/1988	colaborador_0005	Produção	Operacional	22/03/2020	F	1,65	46,0	16,9
0006	25/06/1984	colaborador_0006	Produção	Operacional	02/12/2015	F	1,71	46,0	15,7
0007	30/07/1990	colaborador_0007	Produção	Operacional	13/03/2017	F	1,66	50,0	18,1
0008	29/01/1992	colaborador_0008	Produção	Operacional	31/01/2016	F	1,66	47,0	17,1
0009	17/11/1996	colaborador_0009	Produção	Operacional	09/02/2015	F	1,70	59,0	20,4
0010	24/11/1997	colaborador_0010	Produção	Operacional	26/01/2018	F	1,72	49,0	16,6
0011	26/09/1990	colaborador_0011	Produção	Operacional	21/01/2019	F	1,66	46,0	16,7
0012	20/08/1996	colaborador_0012	Produção	Operacional	23/08/2019	F	1,66	65,0	23,6
0013	26/06/1986	colaborador_0013	Produção	Operacional	02/01/2016	F	1,68	55,0	19,5
0014	20/04/1985	colaborador_0014	Produção	Operacional	13/06/2018	F	1,65	61,0	22,4
0015	08/06/1989	colaborador_0015	Produção	Operacional	01/08/2019	F	1,73	65,0	21,7
0016	14/06/1992	colaborador_0016	Produção	Operacional	19/01/2015	F	1,68	52,0	18,4
0017	11/12/1985	colaborador_0017	Produção	Operacional	23/05/2015	F	1,74	64,0	21,1
0018	01/01/1995	colaborador_0018	Produção	Operacional	29/01/2015	F	1,70	65,0	22,5
0019	03/03/1998	colaborador_0019	Produção	Operacional	12/08/2020	F	1,71	65,0	22,2
0020	20/01/1989	colaborador_0020	Produção	Operacional	05/02/2018	F	1,69	63,0	22,1
0021	25/07/1995	colaborador_0021	Produção	Operacional	14/11/2019	F	1,75	57,0	18,6
0022	11/01/1993	colaborador_0022	Produção	Operacional	28/03/2017	F	1,70	57,0	19,7
0023	04/10/1992	colaborador_0023	Produção	Operacional	20/01/2017	F	1,67	54,0	19,4
0024	05/09/1984	colaborador_0024	Produção	Operacional	08/05/2019	F	1,65	63,0	23,1
0025	24/09/1984	colaborador_0025	Produção	Operacional	20/12/2015	F	1,71	50,0	17,1
0026	08/01/1997	colaborador_0026	Produção	Operacional	31/01/2016	F	1,75	62,0	20,2
0027	14/01/1996	colaborador_0027	Produção	Operacional	18/04/2019	F	1,73	50,0	16,7

A **tabela plana** é um tipo de organização de dados verticalizado, ou seja, cada linha da tabela representa uma observação ou indivíduo que são identificados por chaves únicas como por exemplo ID, RG, CPF, entre outros. As colunas das tabelas planas são informações relativas aos indivíduos das linhas. Este tipo de organização de dados é bastante utilizado em cadastros, e sua característica vertical permite que a tabela seja atualizada linha a linha no final da tabela.

Tabela pivô

Altura média dos funcionários por sexo e função			
Função	F	M	Média Geral
cozinheiro	1,69	1,80	1,72
técnico	1,70	1,79	1,78
auxiliar	1,71	1,79	1,73
gerente	1,71	1,79	1,76
ajudante	1,70	1,80	1,75
aprendiz	1,72	1,80	1,76
estagiário	1,68	1,79	1,72
operador	1,69	1,79	1,74
encarregado	1,70	1,76	1,72
coordenador	1,69	1,80	1,76
supervisor	1,69	1,79	1,75
analista	1,71	1,78	1,76
engenheiro	1,68	1,79	1,75
Média Geral	1,70	1,79	1,75

A **tabela pivô** é uma forma de organizar dados já trabalhados a partir de uma tabela plana. Os campos são agrupados em categorias que são demonstrados nas primeiras colunas da tabela pivô.

As colunas agrupadas são em geral resumos estatísticos dos dados das categorias agrupadas, como por exemplo cálculos de média, contagem, desvio padrão, percentual entre outros.

Pode ser confundida com a tabela cruzada, porém a tabela pivô é geralmente dinâmica, podendo apresentar filtros em dashboards e planilhas Excel, enquanto que tabelas cruzadas são utilizadas em relatórios e como visões que também podem alimentar tabelas pivô.

Mineração de Dados



N. Trans.	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	não	sim	não	sim	sim	não	não
2	sim	não	sim	sim	sim	não	não
3	não	sim	não	sim	sim	não	não
4	sim	sim	não	sim	sim	não	não
5	não	não	sim	não	não	não	não
6	não	não	não	não	sim	não	não
7	não	não	não	sim	não	não	não
8	não	não	não	não	não	não	sim
9	não	não	não	não	não	sim	sim
10	não	não	não	não	não	sim	não

<i>Dor de cabeça</i>	<i>Dor nas articulações</i>	<i>Febre alta</i>	<i>Moleza e dor no corpo</i>	<i>Está com dengue?</i>
<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Não</i>	<i>Sim</i>
<i>Sim</i>	<i>Não</i>	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
<i>Não</i>	<i>Sim</i>	<i>Não</i>	<i>Sim</i>	<i>Sim</i>
<i>Sim</i>	<i>Sim</i>	<i>Sim</i>	<i>Não</i>	<i>Não</i>
<i>Não</i>	<i>Sim</i>	<i>Sim</i>	<i>Não</i>	<i>Não</i>

Mineração de dados (tradução livre do termo data mining) é o processo de explorar dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis, detectando assim novos subconjuntos de dados para extrair insights e confirmações de hipóteses. Exemplos de mineração de dados são detecção de outliers, aglomeração, associação, classificação, regressão.

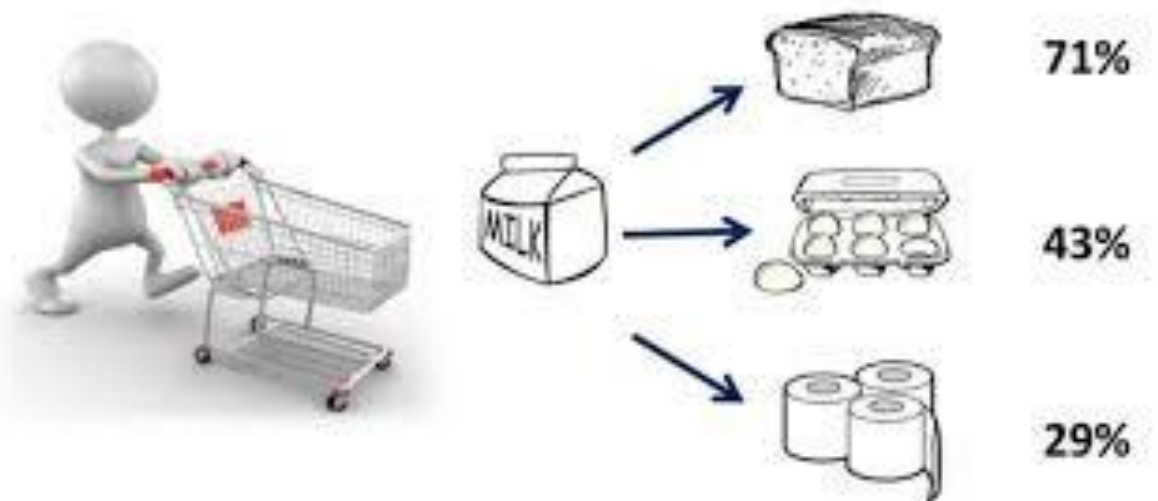
Associações: São ocorrências ligadas a um único evento. Por exemplo: um estudos de modelos de compra em supermercados pode revelar que na compra de fraldas para bebês compra-se também cerveja em 75% das vezes.

Sequências e séries: os eventos são conectados ao longo do tempo. Por exemplo pode-se descobrir que quando se compra uma casa, em 65% as vezes se adquire uma nova geladeira no período de duas semanas; e que em 45% das vezes, um fogão também é comprado três semanas depois da compra da casa.

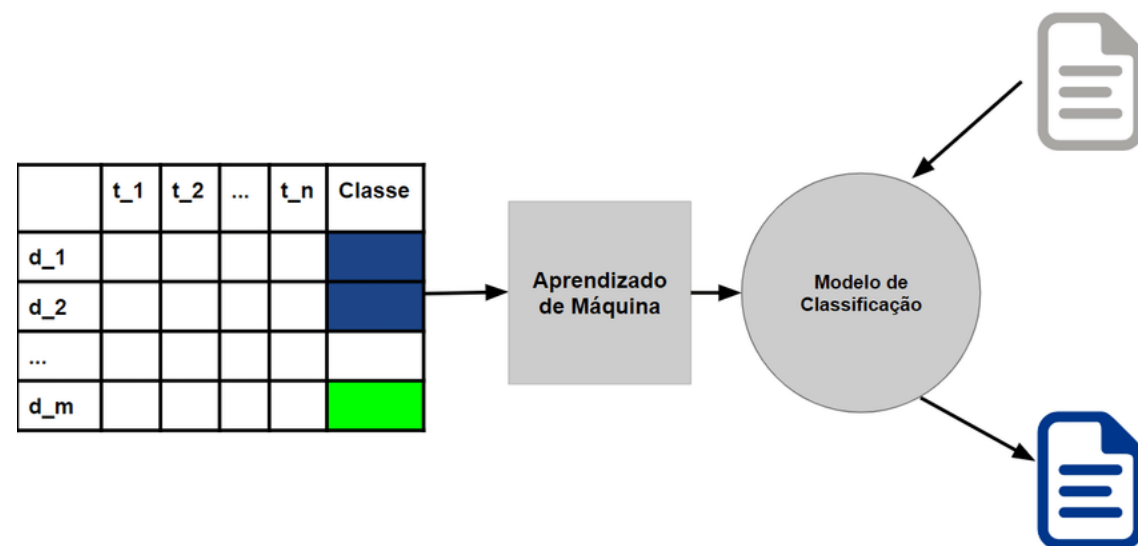
Classificação: Reconhecer indivíduos e classifica-los em grupos previamente classificados através de regras específicas. Por exemplo bancos e outros tipos de instituições financeiras que ofertam cartões ou linhas de crédito classificam seus clientes em grupos de risco para a oferta de linhas de crédito e limite de cartões.

Aglomerção (clustering): Funciona de maneira semelhante a classificação, porém não existe classificação prévia. Por exemplo classificar tipos de lojas de uma rede de varejo de acordo com seu tipo de operação.

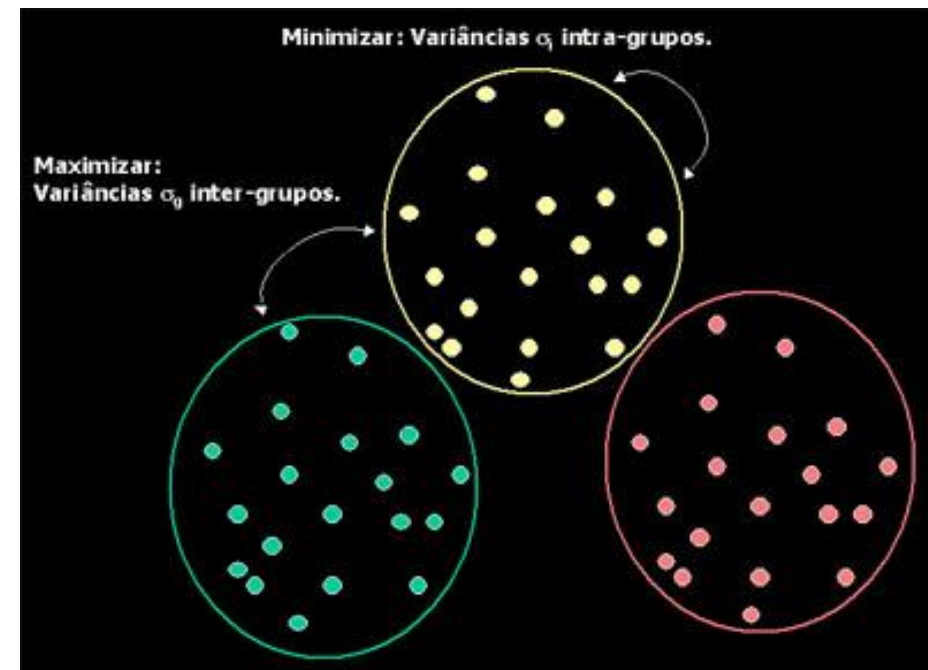
Modelo de Associação



Modelo de Classificação



Modelo de Aglomeração



Séries Temporais



Open	25.38	Mkt cap	168.00B	Prev close	25.85
High	26.09	P/E ratio	499.90	52-wk high	28.31
Low	25.38	Div yield	0.18%	52-wk low	6.25