

Cheat Sheet: Manipulação de Dados com KNIME Analytics Platform

ACESSO AOS DADOS

**CSV Reader**  
Lê um arquivo CSV de seu sistema de arquivos local ou de outro sistema de arquivos remoto. Clique no canto esquerdo inferior para adicionar uma porta de entrada de conexão dinâmica para se conectar a um sistema de arquivos externo, como Amazon S3, Azure Blob Storage, etc.

**Excel Reader**  
Lê planilha(s) de um ou mais arquivos Excel. Uma folha de cada arquivo Excel. Um loop pode ser usado para ler várias planilhas de um arquivo Excel.

**Table Reader**  
Lê dados do arquivo .table Os arquivos .table são organizados usando formato proprietário do KNIME, incluindo estrutura completa do arquivo, fornecendo máximo desempenho e velocidade com configuração mínima.

**SAP Reader (Theobald Software)**  
Carrega vários dados do Sistema SAP (ex: SAP S/4HANA, SAP BW, SAP R/3).



**Amazon S3 Connector**  
Conecta-se ao Amazon S3 e aponta para um diretório de trabalho (com uma sintaxe semelhante ao UNIX, por exemplo, / mybucket / myfolder / myfile). Permite que nodes de leitura downstream acessem dados do Amazon S3 como um sistema de arquivos.

Configurações comuns de nodes Reader e Writer

**Caminho do arquivo:** todos os nodes - Reader e Writer precisam de um caminho de arquivo. O caminho do arquivo pode ser inserido como um caminho absoluto no sistema de arquivos local, um caminho relativo para uma chave local na instalação KNIME atual ou um caminho definido em um sistema de arquivos externo se tal conexão for usada.

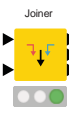
**Múltiplos arquivos:** os nodes Reader e Writer podem ler e concatenar vários arquivos, de acordo com uma extensão de arquivo ou padrão de nome de arquivo selecionado.

**Guia Transformação:** os nodes Reader e Writer incluem guia de Transformação para renomear, filtrar, reordenar e alterar o tipo das colunas

COMBINAR DADOS



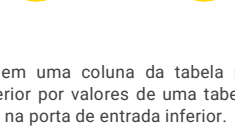
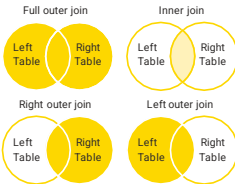
**Concatenate**  
Concatena as linhas de todas as tabelas de entrada, gravando-as uma abaixo da outra. Isso é especialmente útil para tabelas com cabeçalhos de coluna compartilhados.



**Joiner**  
Une as colunas das duas tabelas de entrada com base em uma ou várias colunas de união. Permite selecionar entre diferentes modos de junção e usar várias colunas de junção.



**Cell Replacer**  
Substitui os valores em uma coluna da tabela na porta de entrada superior por valores de uma tabela de pesquisa fornecida na porta de entrada inferior.



FILTRO DE DADOS



**Row Filter**  
Filtrar linhas dentro ou fora da tabela de entrada de acordo com uma regra de filtragem. A regra de filtragem pode corresponder a um valor em uma coluna selecionada ou números em um intervalo numérico.



**Rule-based Row Filter**  
Filtrar as linhas de entrada ou saída de acordo com um conjunto de regras, definido em sua janela de configuração. As regras são avaliadas de cima para baixo. Usar TRUE como antecedente aplica a regra a todas as linhas sem correspondência.



**Reference Row Filter**  
Filtrar as linhas da tabela de entrada superior de acordo com os valores correspondentes na coluna selecionada da tabela de entrada inferior.



**Column Filter**  
Filtrar colunas para dentro ou para fora da tabela de entrada de acordo com uma regra de filtragem. As colunas a serem retidas podem ser escolhidas manualmente ou selecionadas de acordo com seu tipo, ou com base em uma expressão regex correspondente ao seu nome.



**DB Row Filter**  
Expandir uma consulta SQL de entrada para incluir os critérios de filtro de linha definidos na janela de configuração. O agrupamento de várias condições com conjunções AND ou OR também são suportados. Nenhuma codificação SQL necessária.



**DB Query**  
Modifica a consulta SQL de entrada usando SQL customizado. A consulta SQL de entrada é representada pelo marcador de posição #table#.

GRAVAR DADOS



**Excel Writer**  
Grava as tabelas de entrada em planilhas em um arquivo Excel (XLS ou XLSX). Clique nos três pontos no canto esquerdo inferior para adicionar uma porta de entrada de conexão dinâmica para gravar várias tabelas de dados em várias folhas.



**CSV Writer**  
Grava a tabela de dados de entrada em um arquivo CSV. Clique nos três pontos no canto esquerdo inferior para adicionar uma porta de entrada de conexão dinâmica para gravar em um sistema de arquivos externo, como Amazon S3, Armazenamento de Blob do Azure, etc.



**Send to Tableau Server**  
Carrega a tabela de entrada em um servidor Tableau para relatórios.



**Send to Power BI**  
Carrega a tabela de entrada para o Microsoft Power BI para relatórios.

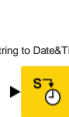


**DB Writer**  
Insere as linhas de dados da porta de entrada superior em uma tabela no banco de dados especificado pela porta de conexão de entrada. Se a tabela do banco de dados não existir, ela será criada.



**DB Connection Table Writer**  
Grava as linhas resultantes da consulta SQL de entrada em uma nova tabela dentro do banco de dados.

DATA & HORA



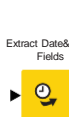
**String to Date&Time**  
As strings são analisadas nas colunas selecionadas de acordo com um formato de data / hora e convertidas em células de Data e Hora. Quatro formatos de data e hora são suportados: apenas data, apenas hora, data e hora e data e hora mais fuso horário.



**Date&Time-based Row Filter**  
Extraí linhas em que o valor de tempo na coluna selecionada está dentro de uma determinada janela de tempo. A janela de tempo é especificada por uma data de início e / ou término ou por uma data de início e uma duração.



**Date&Time Difference**  
Calcula a diferença entre dois objetos de data e hora. Por exemplo, de duas colunas selecionadas, de uma coluna selecionada e um valor fixo, de uma coluna selecionada e o tempo de execução atual, ou uma célula e a célula na linha anterior para uma coluna selecionada.



**Extract Date&Time Fields**  
Extraí campos de hora e data selecionados de uma coluna selecionada do tipo data e hora e acrescenta seus valores em novas colunas.

LIMPEZA DE DADOS



**Missing Value**  
Define e aplica uma estratégia para substituir os valores ausentes na tabela de entrada - globalmente em todas as colunas ou individualmente para cada coluna.



**Duplicate Row Filter**  
Detecta linhas duplicadas e aplica a operação selecionada, por exemplo, remove linhas duplicadas. Duplicados são linhas que possuem o mesmo valor em todas as colunas selecionadas.



**Numeric Outliers**  
Detecta e trata valores discrepantes numéricos para cada uma das colunas selecionadas individualmente usando o intervalo interquartil (IQR).

RECURSOS

**E-Books:** KNIME Advanced Luck cobre recursos avançados e muito mais. Praticando ciência de dados é uma coleção de estudos de caso de ciência de dados de projetos anteriores. Ambos disponíveis em [knime.com/knimepress](https://www.knime.com/knimepress)

**KNIME Blog:** Tópicos envolventes, desafios, notícias do setor e pilulas de conhecimento em [knime.com/blog](https://www.knime.com/blog)

**E-Learning Courses:** Faça nossos cursos on-line e gratuitos para aprender sobre as diferentes etapas de um projeto de Data Science, (com exercícios e soluções para testar seus conhecimentos) em [www.knime.com/knime-self-paced-courses](https://www.knime.com/knime-self-paced-courses)

**KNIME Hub:** Navegue e compartilhe fluxos de trabalho, nodes e componentes. Adicione classificações ou comentários a outros fluxos de trabalho em [hub.knime.com](https://hub.knime.com)

**KNIME Forum:** Junte-se à nossa comunidade global e participe de conversas em [forum.knime.com](https://forum.knime.com)

**KNIME Server:** Para colaboração baseada em equipe, automação, gerenciamento e implantação, verifique o servidor KNIME em [www.knime.com/knime-server](https://www.knime.com/knime-server)

REMODELAR E AGREGAR DADOS



**GroupBy**  
Agrupar as linhas de uma tabela pelos valores exclusivos nas colunas selecionadas e calcula agregação e medidas estatísticas para os grupos definidos. Apesar de seu nome simples, ele oferece uma funcionalidade poderosa e tem muitos usos confiáveis.



**Pivoting**  
Estende a funcionalidade de agregação do node GroupBy criando tabela de saída com colunas e linhas para os valores exclusivos nas colunas de entrada selecionadas. Os valores exclusivos das colunas de agrupamento tornam-se linhas e os valores exclusivos das colunas dinâmicas tornam-se colunas.



**Category to Number**  
Mapeia os valores categóricos nas colunas selecionadas para valores inteiros e exporta as regras de mapeamento para a porta de saída do modelo. Os nodes Category to Nubber (Aplicar) e Número para Categoria (Aplicar) aplicam a regra de mapeamento em ambas as direções.



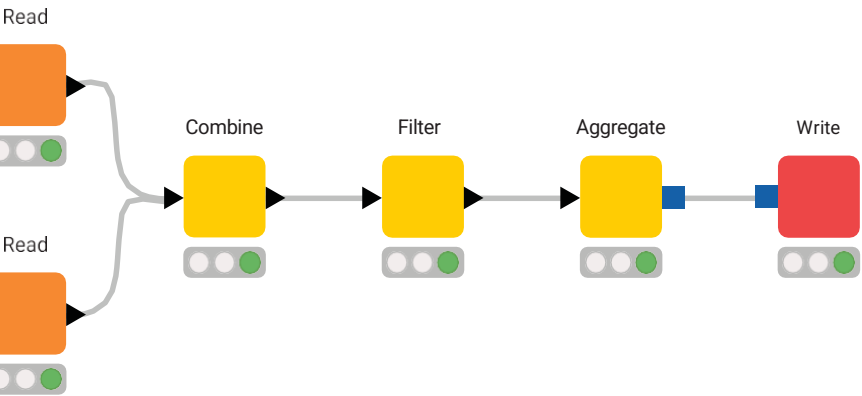
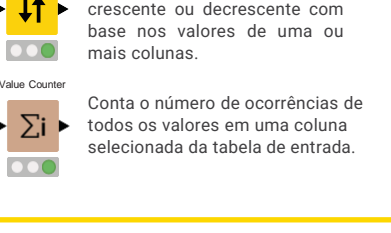
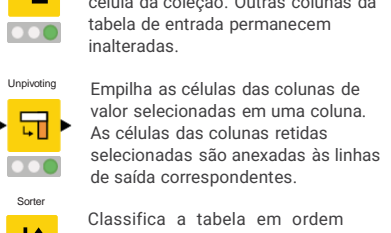
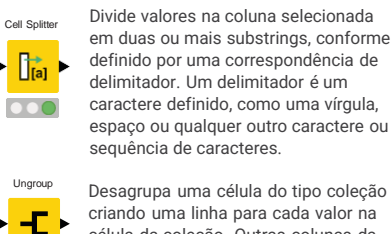
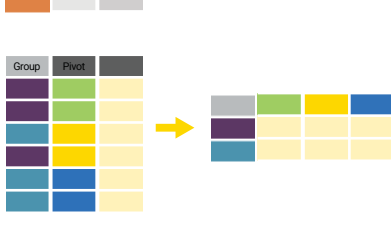
**One to Many**  
Cria uma nova coluna para cada valor na coluna de entrada selecionada. Esses valores se tornam os cabeçalhos das colunas. As células nas colunas recém-criadas são definidas como 0 se o valor não estiver presente, caso contrário, 1. Esse tipo de codificação é chamado de vetor one-hot.



**Transpose**  
Converte as linhas em colunas e as colunas em linhas.



**Table Manipulator**  
Executa várias transformações de uma vez, como renomear, filtrar, reordenar e alterar o tipo nas colunas de entrada. Ao adicionar portas dinâmicas, ele pode substituir o node concatenate.



TIPOS DE DADOS & CONVERSÕES

**String:** Sequência de caracteres, por ex. "This is a string" [.] **Collection Cell:** Coleção de vários valores do mesmo tipo ou de tipos diferentes, por exemplo, pode ser uma lista de valores ou um conjunto de valores. Em um conjunto, cada valor ocorre apenas uma vez.

**Integer:** Número inteiro com valor real, por ex. -100 ou 345

**Double:** Número com valor real, por ex. -0,432 ou 45,39

**Date&Time:** Um formato de dados para data, hora, data e hora ou data e hora mais fuso horário.

**Boolean:** Apenas dois valores possíveis (ex. V e F)

**Document/Image:** O KNIME Analytics Platform oferece suporte a muitos outros tipos de dados, com documentos de texto, imagens, impressões digitais, etc.

**String to Number**  
Converte uma entrada de dados do tipo numérica para string por exemplo int ou double, para string.

**String to Number**  
Converte uma entrada de dados do tipo string para numérica, por exemplo string para int ou double

CRIAR COLUNAS

**Math Formula**  
Implementa várias operações matemáticas em várias colunas de entrada. As operações matemáticas podem ser aplicadas a várias colunas com o node Math Formula (Multi Column).

**String Manipulation**  
Executa operações em valores de string em colunas, como combinar duas ou mais strings, extrair uma ou mais substrings, cortar espaços em branco e assim por diante.

**Rule Engine**  
Aplica um conjunto de regras a cada linha da tabela de entrada. As regras são aplicadas de cima para baixo. A primeira regra correspondente é usada.

**String Replacer**  
Substitui valores em uma coluna de string selecionada se eles corresponderem a um padrão definido.

**Counter Generation**  
Cria uma nova coluna com um contador. O valor inicial e o tamanho do passo são definidos na janela de configuração.

**Column Expressions**  
Combina a funcionalidade dos nodes de fórmula matemática, mecanismo de regra e manipulação de string. Mais de uma expressão pode ser definida para modificar ou adicionar várias colunas ao mesmo tempo.

PORTAS DINÂMICAS



**Portas dinâmicas:** Portas de entrada adicionais podem ser adicionadas clicando nos três pontos no canto esquerdo inferior de um node.

FORMATAR PLANILHAS DE EXCEL

A extensão Continental Nodes for KNIME permite que você formate automaticamente uma planilha Excel existente. A chave é uma tabela de dados adicional do mesmo tamanho da planilha original do Excel, onde cada célula contém um ou mais valores de tag separados por vírgula, por exemplo, cabeçalho, borda, etc. Com base nessas tags, os nodes do Formataador XLS adicionam novas instruções de formatação para as instruções existentes, conforme disponíveis na porta de entrada inferior (opcional).



**XLS Control Table Generator**  
Transforma a tabela de entrada em uma Tabela de Controle XLS, o que significa que troca os nomes das colunas para A, B, C, ... e os IDs das linhas para 1, 2, 3, ... É o node inicial para coletar instruções de formatação para uma planilha do Excel e alimenta todos os nodes do formataador XLS.



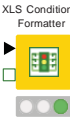
**XLS Background Colorizer**  
Adiciona cor de fundo e / ou instruções de formatação de preenchimento de padrão a todas as células com uma tag especificada na Tabela de controle XLS na porta de entrada superior.



**XLS Border Formatter**  
Adiciona instruções de formatação de borda para um determinado intervalo especificado por uma tag na tabela de controle XLS na porta de entrada superior.



**XLS Cell Merger**  
Adiciona instruções de formatação para mesclar todas as células com uma tag especificada na tabela de controle XLS na porta de entrada superior.



**XLS Conditional Formatter**  
Adiciona instruções de formatação para colorir fundos de células de acordo com seu valor numérico para todas as células especificadas por uma tag na tabela de controle XLS na parte superior



**XLS Formatter (apply)**  
Aplica todas as instruções de formatação a uma planilha Excel existente.

Amplie seu conhecimento sobre KNIME com nossa coleção de livros da KNIME Press. Para usuários iniciantes e avançados, até aqueles interessados em tópicos especializados, como detecção de tópicos, combinação de dados e soluções clássicas para casos de uso comuns usando KNIME Analytics Platform - há materiais para todos. Disponível para download em [www.knime.com/knimepress](http://www.knime.com/knimepress).

