



Linhagem, Descrição e Proveniência dos Dados


Fonte dos Dados: Student Performance Element

Nome: Dados de Performance Estudantil

Versão: Única

Descrição: O Dataset escolhido é composto por apenas uma planilha "StudentPerformanceFactors.csv" extraída do Kaggle. Essa planilha contém vinte atributos que correspondem as colunas da tabela. O conteúdo detém dados categóricos e numéricos característicos de estudantes. Informações que podem ser utilizadas para avaliação dos impactos de aspecto no seu desempenho através da análise dos dados.

Estrutura:

Os dados foram extraídos em uma só tabela, mas quando modelados irão se dividir em cinco no esquema Estrela que pode ser consultado na seção .

Modelagem dos Dados.

Tabelas:

 Catálogo de Dados/Metadados dos Campos

Chaves Identificadoras e de Ligação:

- As Chaves Identificadoras nesse modelo são os campos "Codigo" de cada tabela Dimensão (Dim_Estudo; Dim_Escola; Dim_Comportamento; Dim_Aluno).
- Todos os campos de identificação (Id_Estudo; Id_Escola; Id_Comportamento; Id_Aluno) da Tabela Fato (Fato_Desempenho) são Chaves Estrangeiras desse modelo.

Granularidade:

Os dados são relativos a performance de um estudante. De maneira que cada linha represente um aluno, contendo dados pertencentes ao mesmo.

Temporalidade:

Atemporal. Não há especificidade de datas nem nos dados e nem na página de onde foi extraído o Dataset.

Origem: Dados criados para fins educacionais e de pesquisa que não contém dados reais mas simulam situações/fatores possíveis de influenciar o resultado de um aluno.

Não houve qualquer outro arquivo para compor essa análise e somatizar aos dados extraídos do Kaggle.

Licença: Dados de domínio público.

Transformações Realizadas:

- As colunas foram renomeadas para português para facilitar a compreensão e equalizar a linguagem, uma vez que toda a documentação do projeto se encontra em português também.
- Todas as linhas que continham algum valor NULL foram removidas e as que estavam fora do intervalo de domínio também.
- Todos os espaços extras que existissem nos dados categóricos, os mesmo foram removidos e os dados foram uniformizados para todas as categorias serem armazenadas em Caps Lock (caixa alta).
- TRANSFORMAR NOME DE COLUNAS EM PT (SE TIVER TEMPO)