

Relatório Final de Análise de Performance de Estudantes

1. Contexto e Objetivo Executivo

O objetivo principal deste projeto foi identificar e quantificar a relação entre os hábitos de estudo/vida do aluno e seu desempenho final em avaliações. Utilizou-se um conjunto de dados simulado de performance de estudantes para demonstrar proficiência em **Modelagem SQL**, **Otimização de Consultas**, e **Visualização Interativa de Dados** (Power BI).

Insight Principal: A análise de correlação demonstrou que o **Tempo de Estudo (hours_studied)** é o fator mais forte e previsível da performance, com uma correlação positiva de **+0,78** com a Nota Final (exam_score).

2. Metodologia de Dados e Estruturação SQL

A fase inicial do projeto focou na criação de um esquema de banco de dados otimizado para a análise.

2.1. Modelagem

Foi criado o banco de dados estudantes_db e a tabela estudantes com a seguinte estrutura, garantindo a integridade dos dados e o tipo de dado correto para cálculos (DECIMAL):

SQL

```
CREATE TABLE estudantes (  
    student_id INT PRIMARY KEY AUTO_INCREMENT,  
    hours_studied DECIMAL(5, 2),  
    sleep_hours DECIMAL(5, 2),  
    attendance_percent DECIMAL(5, 2),  
    previous_scores DECIMAL(5, 2),  
    exam_score DECIMAL(5, 2)  
);  
  
-- Criação de Índices para Otimização de Performance  
CREATE INDEX idx_hours_studied ON estudantes (hours_studied);
```

CREATE INDEX idx_exam_score ON estudantes (exam_score);

-- Índices adicionais para consultas de agrupamento e segmentação.

A criação de índices (idx...) visa acelerar as consultas de agrupamento e filtragem no SQL, melhorando o tempo de resposta em bancos de dados maiores.

2.2. Análises e Agrupamentos

As consultas foram desenvolvidas em SQL para realizar a limpeza de dados e criar os agrupamentos (faixas) utilizados posteriormente no Power BI.

Objetivo da Consulta	Função Principal do SQL
Resumo Descritivo	AVG() para média de todas as variáveis.
Distribuição de Faixas	CASE WHEN para criar intervalos de notas (Ruim, Médio, Excelente).
Impacto de Variáveis	Uso de CASE WHEN e AVG(exam_score) agrupados por faixas (Ex: Baixo Estudo, Alto Estudo) para identificar o impacto médio de cada hábito.
Rendimento Combinado	Lógica avançada de CASE WHEN para classificar alunos em grupos (Alto Rendimento, Baixo Rendimento, Misto/Médio) com base na combinação de todos os três hábitos.

3. Visualização Interativa no Power BI

O Power BI foi utilizado para conectar-se aos dados tratados, criar a modelagem DAX para correlação e desenvolver a visualização interativa do dashboard.

3.1. Modelagem e Medidas DAX

- Medida Principal:** Foi criada uma medida DAX para calcular o Coeficiente de Correlação de Pearson entre as Horas Estudadas e a Nota Final.
- KPIs Dinâmicos:** As medidas de Média (AVERAGE) e rendimento foram calculadas e dispostas em visuais de cartão, tabela e gráfico para que mudassem dinamicamente conforme os filtros eram aplicados.

3.2. Análise do Dashboard

O dashboard foi dividido em áreas lógicas para guiar a narrativa:

Área do Dashboard	Elementos e Insights
KPI Principal	Cartão de Destaque (Fator Preditor): O valor de +0,78 de correlação entre <i>Horas de Estudo</i> e <i>Nota Final</i> é o destaque, servindo como o título da análise.
Filtros	Slicers Horizontais: Filtros de Horas Estudadas , Horas de Sono e Frequência permitem que o usuário teste o impacto dos hábitos nas notas em tempo real.
Análise de Distribuição	Gráfico de Barras que mostra a Contagem de Alunos em diferentes faixas de notas. O visual se adapta ao filtro, mostrando a distribuição dos alunos dentro da faixa selecionada (ex: como os alunos de 'Alto Estudo' se distribuem entre as notas).

4. Conclusão e Insights Finais

O projeto validou a premissa de que a métrica de esforço (horas dedicadas) é o principal motor do desempenho acadêmico neste conjunto de dados.

- **Fator Mais Forte:** O coeficiente de **+0,78** é estatisticamente significativo e forte, comprovando que o tempo de estudo é o preditor mais confiável da nota final.
- **Fatores Secundários:** As baixas correlações com Horas de Sono e Frequência sugerem que, isoladamente, estes hábitos têm um impacto menor na nota do que o esforço dedicado.

Este projeto demonstra domínio na manipulação de dados em SQL, na criação de lógica analítica complexa (funções CASE WHEN), e na apresentação de *insights* acionáveis em um dashboard interativo.