Projeto: Engenharia e Análise de Dados com T-SQL

​1. Visão Geral e Objetivo do Projeto

​Este projeto acadêmico foi concebido para simular o ciclo completo de um pipeline de dados, do levantamento de requisitos à entrega de insights acionáveis. Com o uso exclusivo do T-SQL, o objetivo é demonstrar proficiência em engenharia de dados, cobrindo aspectos de modelagem dimensional, processos de ETL, otimização de consultas e implementação de segurança. O tema escolhido, a análise dos gastos parlamentares, proporciona um cenário real e desafiador para a aplicação das técnicas.

​2. Tema e Problema de Negócio

​Tema: O impacto econômico dos gastos parlamentares: fornecedores, setores e oportunidades de mercado.

​A cota para o exercício da atividade parlamentar (CEAP) representa uma parte significativa do orçamento público. A transparência na sua utilização é fundamental, mas a análise dos dados brutos é complexa. O desafio deste projeto é transformar os dados fragmentados de diversas fontes públicas em um modelo de banco de dados analítico, capaz de responder a perguntas de negócio estratégicas e revelar padrões de gastos, concentração de mercado e correlações de comportamento.

​3. Plano de Análise: Perguntas de Negócio

​As seguintes perguntas guiam a modelagem, a transformação e a análise de dados do projeto:

​Quem são os maiores fornecedores dos deputados?

​Quais setores econômicos mais lucram com a cota parlamentar (via CNAE)?

​Existe concentração de mercado (poucas empresas dominam)?

​Quais partidos/estados concentram mais gasto em determinados setores?

​Há sazonalidade nos gastos (ex.: anos eleitorais, períodos específicos)?

​Deputados mais gastadores em publicidade têm comportamento diferente em votações/presenças?

​Empresas fornecedoras atuam em vários estados ou concentram contratos em regiões específicas?

​4. Arquitetura e Fontes de Dados

​O projeto utiliza um conjunto de dados robusto, com mais de 200.000 registros na tabela fato principal. A solução foi arquitetada em um esquema em estrela (star schema) para otimizar as consultas analíticas.

​Fontes de Dados Utilizadas:

​Cota Parlamentar - Brasil.io: Dados detalhados de despesas parlamentares por deputado e fornecedor.

​Câmara dos Deputados - Dados Abertos: Informações sobre deputados, partidos e legislaturas.

​Receita Federal - CNPJs: Classificação dos fornecedores por setor econômico (CNAE), essencial para a análise setorial.

​5. Modelagem de Dados: Abordagem e Soluções

​O modelo de dados lógico foi traduzido para um modelo físico otimizado para um ambiente de data warehouse (OLAP). As principais decisões de modelagem incluem:

​Esquema em Estrela: O modelo é centrado nas tabelas de fato FatoGastoParlamentar e FatoVotacaoPresenca, que se conectam diretamente a dimensões como DimDeputado, DimFornecedor, DimTempo, etc.

​Tabela de Ponte: A complexidade da afiliação partidária, que pode mudar ao longo do tempo, foi resolvida com uma tabela de ponte (BridgeDeputadoLegislaturaPartido). Esta tabela garante a precisão histórica ao associar cada gasto à afiliação partidária correta no momento da despesa.

​Dimensões Conformadas: As dimensões DimDeputado e DimTempo são compartilhadas entre as duas tabelas de fato, permitindo a análise de correlação entre os gastos e o comportamento de votação e presença dos deputados.

​6. Requisitos Técnicos e Implementação

​O projeto adere estritamente aos requisitos acadêmicos, com todos os artefatos técnicos desenvolvidos em T-SQL.

​Processo de ETL: Implementado via Stored Procedures e scripts T-SQL para a ingestão, limpeza e carga dos dados das fontes públicas para as tabelas de fato e dimensão.

​Objetos Programáticos: Implementação de 5 Views para análises pré-agregadas, 5 Stored Procedures (3 de ETL/CRUD e 2 analíticas), 2 Functions para lógicas de negócio específicas e 2 Triggers para validação e auditoria.

​Segurança (DCL): Criação de 3 perfis de acesso distintos (Roles) para simular um ambiente de produção.

​Documentação: Dicionário de dados completo e manual de uso das rotinas.

​Versionamento: Todo o código SQL e a documentação serão versionados no GitHub, com o acompanhamento de tarefas via quadro Trello.