**Arquitetura de Dados**

Nossa documentação de arquitetura de dados, contém:

* DER - Diagrama Entidade Relacionamento
* Planilha de dicionarização
* MER - Modelo Entidade Relacionamento
* Estrutura das tabelas (SQL)
* Mapa do fluxo de Dados

**DER - Diagrama Entidade Relacionamento**

Um modelo de dados conceitual das relações entre as entidades.

Permite visualizar como as entidades se relacionam entre si, como funcionam seus relacionamentos e onde estes relacionamentos podem ser melhorados.

Nosso banco se refere a um projeto de criação de dashboard para analisar as vendas dos combustíveis da empresa, considerando os posto e funcionários neste processo.

**Planilha de dicionarização**

É uma ferramenta de extrema importância na criação da estrutura dos dados.

Está diretamente relacionada a:

* Boas práticas de coleta: objetivos claros definidos, que ajudam a determinar quais dados são realmente necessários.
* Padronização de Dados – padrões consistentes facilitando a análise e comparação entre conjuntos de dados diferentes.
* Padrões de nomenclatura consistentes para facilitar a identificação e recuperação de dados.
* Verificações de validação implementados para garantir que os dados coletados atendam aos critérios definidos.

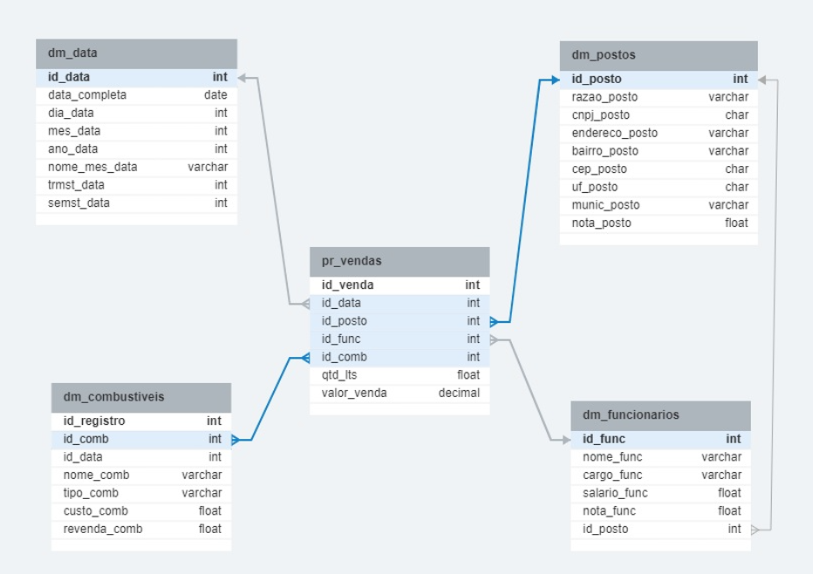
Após as definições e preenchimento concluídos, a planilha de dicionarização será importada na ferramenta de modelagem, onde a macro espera esse formato.

Uma estrutura organizada e com padrões bem definidos traz muitas vantagens:

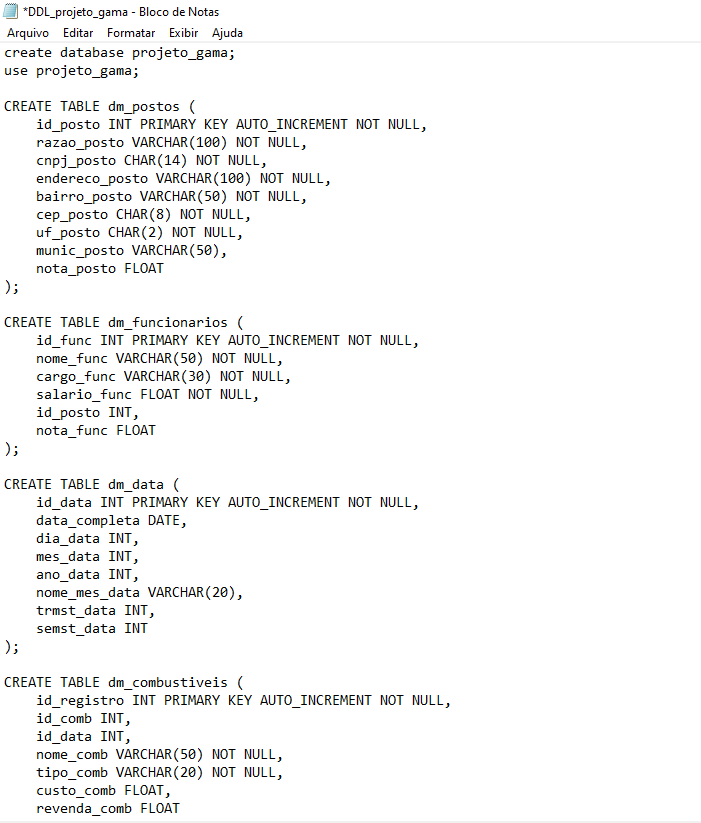
* Permite dar visibilidade para a companhia sobre a existência dos dados
* Democratização na consulta e utilização gerando insights importantes na tomada de decisão.
* Evita retrabalho

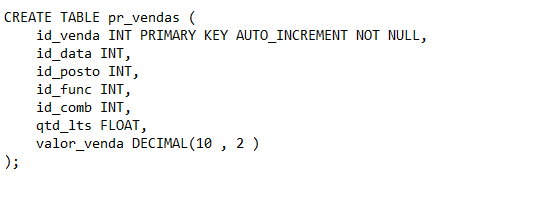
**MER - Modelo Entidade Relacionamento**

Aqui descrevemos os objetos (entidades) envolvidos no negócio, com suas características (atributos) e como elas se relacionam entre si (relacionamentos). Os relacionamentos entre as tabelas contidas no banco e cardinalidades.

****

**Estrutura das tabelas criadas (SQL)**

****

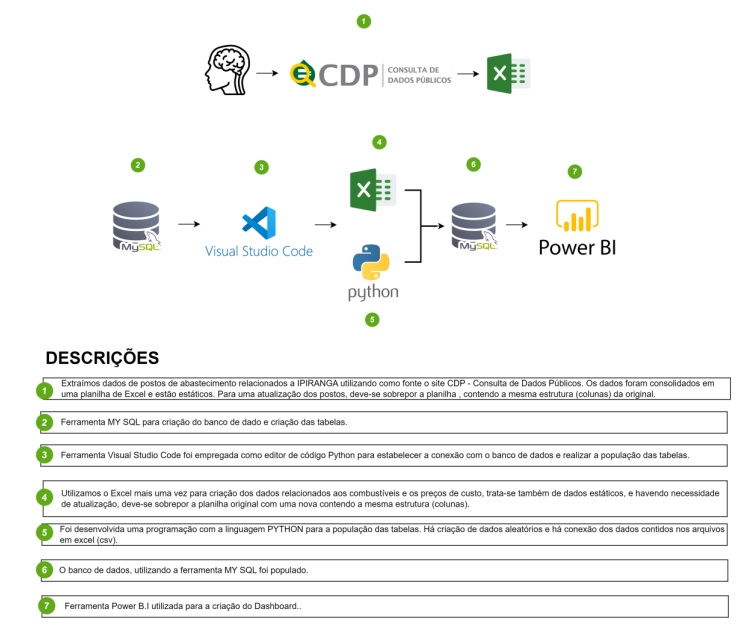
****

**Fluxo de Dados**

Consolidação das informações inerentes a estrutura do banco de dados.

Mapear e registrar as fontes dos dados, tecnologias utilizadas, garantindo a confiabilidade

e qualidade dos dados.



**Uma OBS importante é que nosso projeto é um protótipo e contém dados aleatórios e fictícios gerados em python.**

Em um cenário contendo dados reais, coletados de sistemas ou fontes operacionais da empresa, há a necessidade de maiores controles e ações, além das apresentadas acima.

* Avaliação relacionada a confiabilidade e a credibilidade das fontes de dados antes

de integrá-las ao conjunto.

* Segurança e Privacidade
* Backup e Recuperação
* Checagem/atualização de sistemas e softwares de armazenamento
* Cuidado com a limpeza de dados
* Monitoramento e Desempenho