



CODE

Google Cloud



CODE



Google Cloud

CODE



Google Cloud

ENERGY DRINK  
TNT

SOUL  
CODE



Google Cloud

SOUL  
CODE

ENERGY DRINK  
TNT



Google Cloud

SOUL

CODE

SOUL  
CODE



Google Cloud

ENERGY DRINK  
TNT

SOUL  
CODE

ENERGY DRINK  
TNT

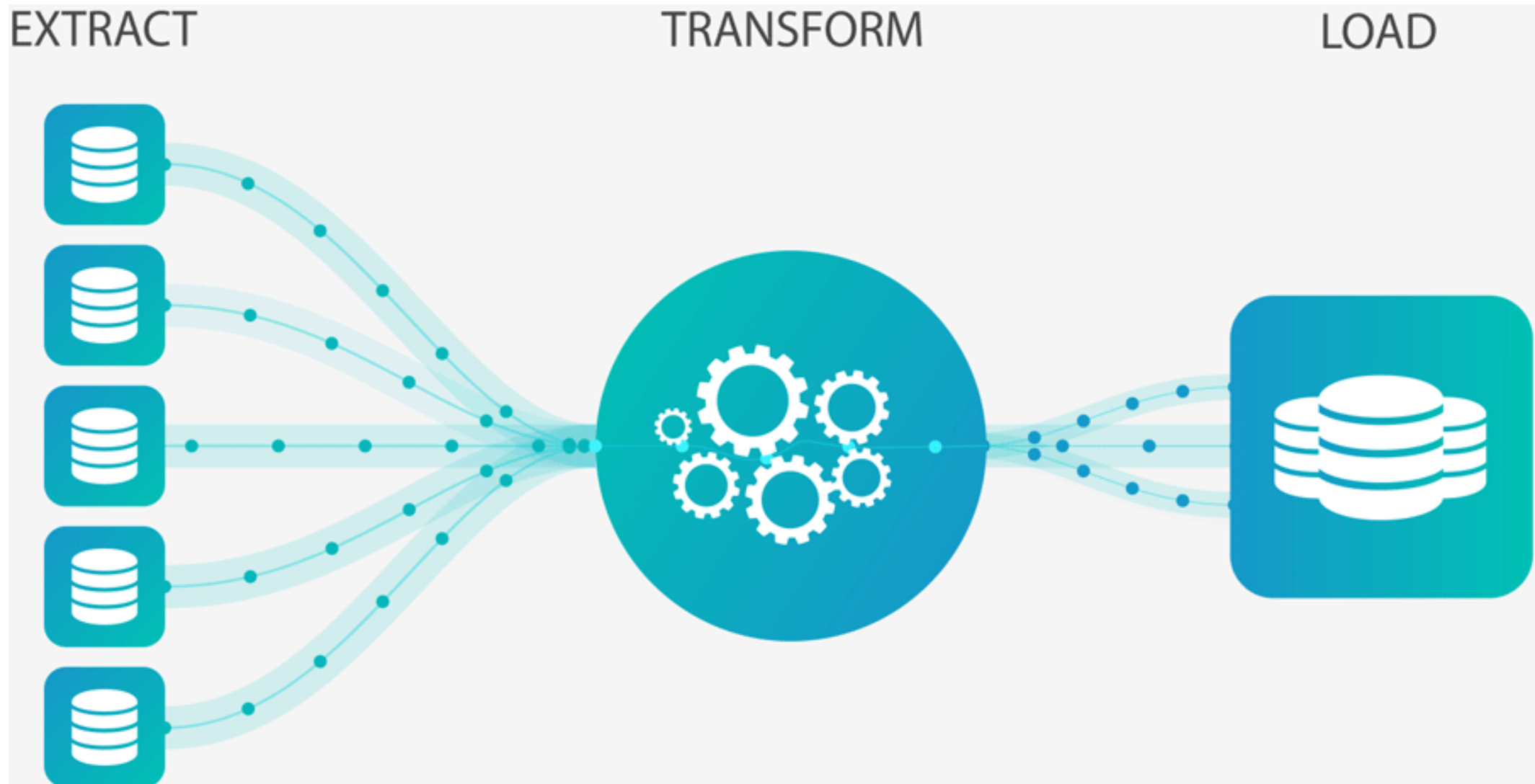


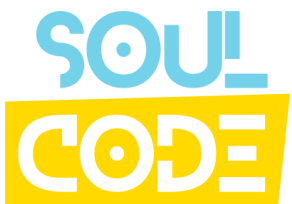
Google Cloud

SOUL  
CODE



# ETL - Engenharia de Dados



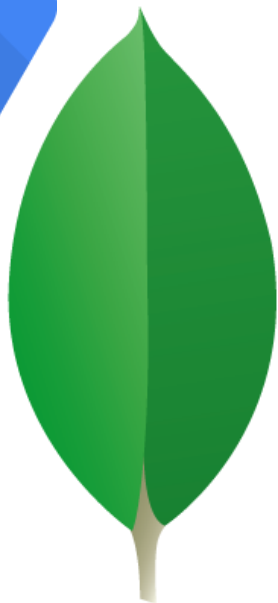


PANDAS

# Engenharia de Dados

PROJETO INDIVIDUAL - NÍVEIS

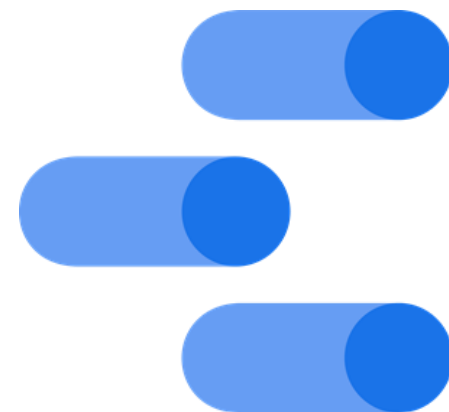
INFRA



SPARKSQL

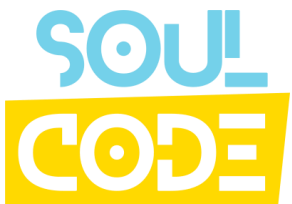


DATA STUDIO



PYSPARK



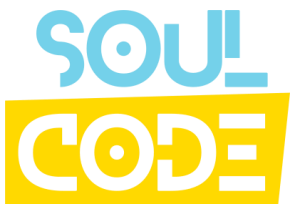


# Engenharia de Dados

## INFRA

- **Dataset** deve ser **salvo** em ambiente cloud
- Arquivo original e tratado deve ser salvo em MongoDB Atlas em coleções diferentes
- **DataFrames** devem ser obrigatoriamente **salvos** em uma **bucket** do CloudStorage.

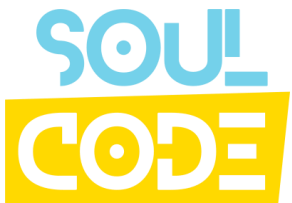




# Engenharia de Dados

## PANDAS

- Dados traduzidos para Português - BR
- Extração corretamente para um dataframe.
- Dados inconsistentes e realizar a limpeza para NaN ou NA, drops e exclusões
- **ETL** – Todos comentados

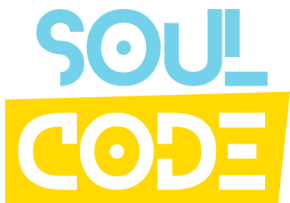


# Engenharia de Dados

## PYSPARK

- Estrutura do DataFrame utilizando o **StructType**.
- ETL
- Mudança de nome de pelo menos 2 colunas
- Duas novas colunas contendo alguma informação relevante sobre as outras colunas já existentes
- Utilizar filtros, ordenação e agrupamento, trazendo dados relevantes para o negócio
- Utilizar pelo menos duas Window Functions





# Engenharia de Dados

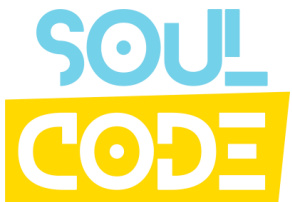
## SPARKSQL

### 5 consultas diferentes utilizando o SparkSQL

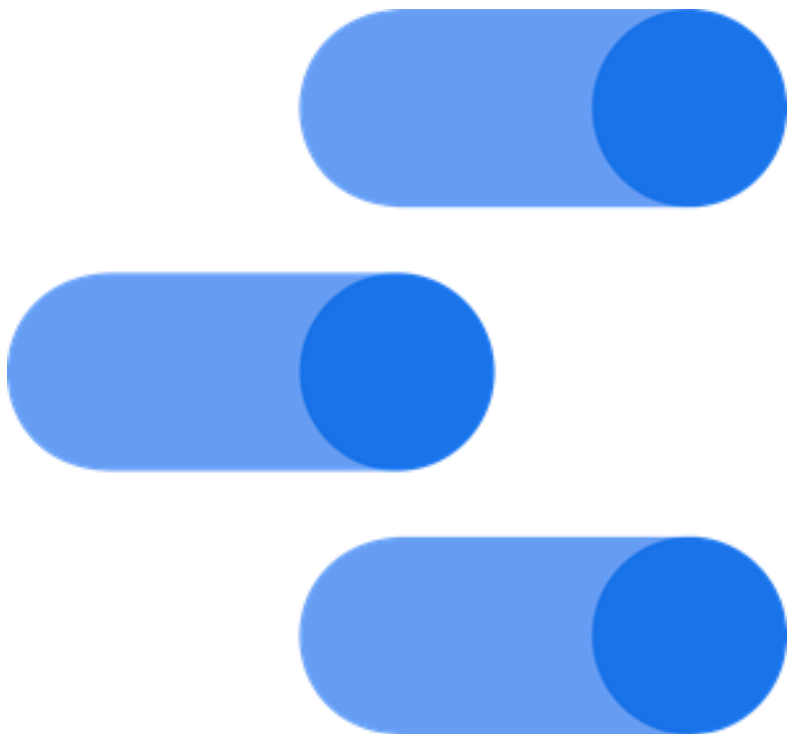
- 1 - Visualização do Dataframe
- 2 - Visualização os indexes da tabela.
- 3 - Seleção da renda anual e perfil de compras com base nos dados socio-econômicos
- 4 - Estudo com critério de quantidade de compras, renda anual e dias desde da ultima compra.
- 5 - Estudo com critério do maior numero de dias sem comprar.







# Engenharia de Dados



**DATA STUDIO**

# Engenharia de Dados

## Campanha de Marketing - Soulcode

COMPRAS NA LOJA

12.970

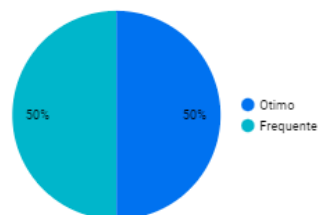
COMPRAS ONLINE

9.150

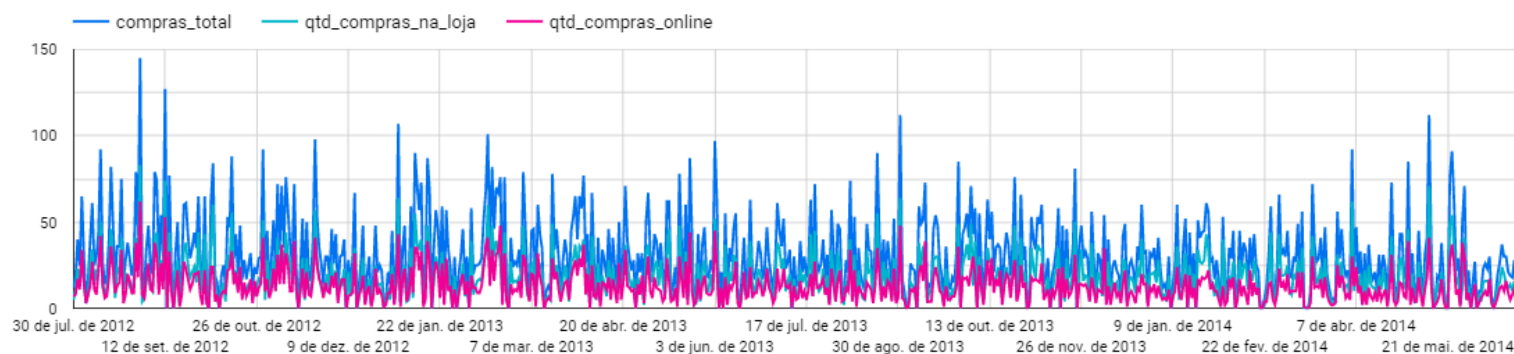
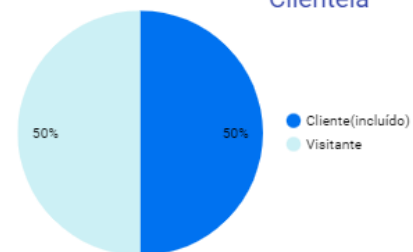
TOTAL DE COMPRAS

22.018

Relacionamento



Clientela



## Campanha de Marketing - Soulcode

COMPRAS NA LOJA

12.970

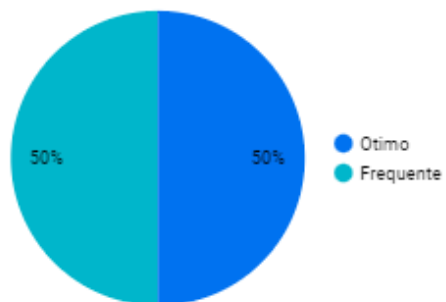
COMPRAS ONLINE

9.150

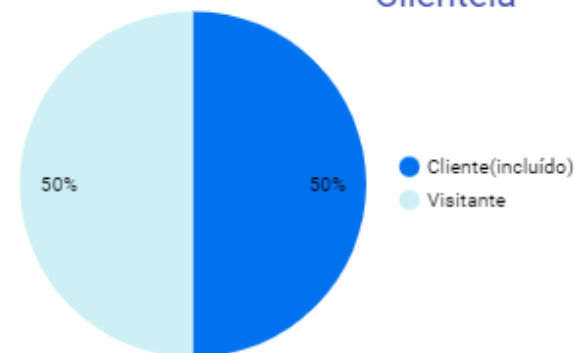
TOTAL DE COMPRAS

22.018

Relacionamento

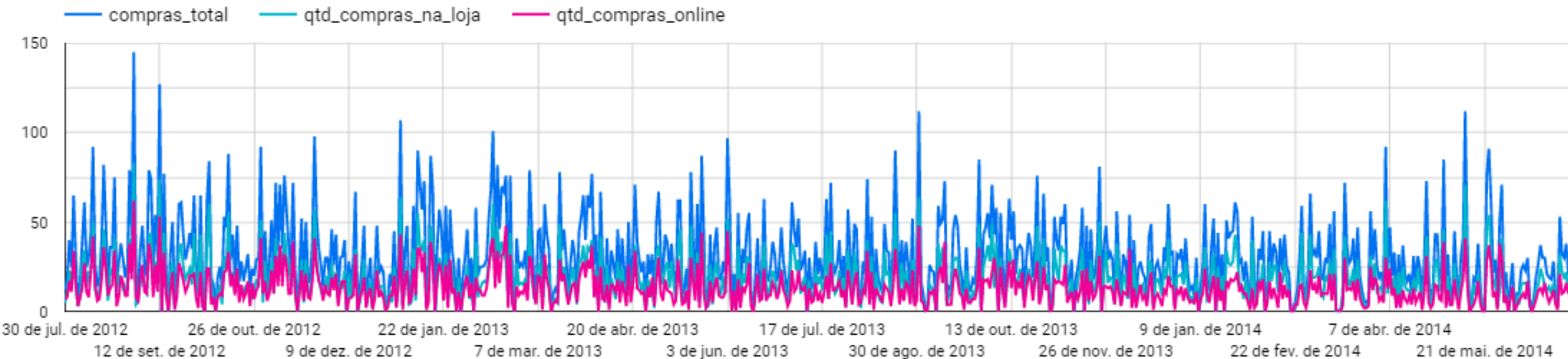


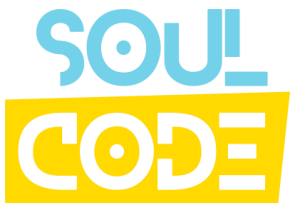
Clientela



# Engenharia de Dados

## Análise Temporal $\approx$ 2 anos

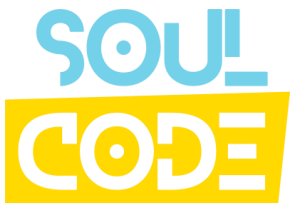




# Engenharia de Dados

## Caracterização - Campanha de marketing

- Um loja de produtos de alto custo.
- Dados de 2012 a 2014.
- Público seletivo - USA.
- Itens de consumo alimentos e joias.



# Engenharia de Dados

## Avaliação - Campanha de marketing

- **Problema x Solução**

Falta de informações estratégicas; telefone, email e canais de relacionamento e baixa escala.

- **Product x Marketing**

50% dos clientes prestou alguma reclamação.

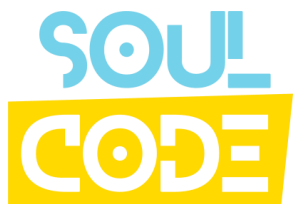
50% dos clientes em acesso online e da loja física não compram

- **Scale**

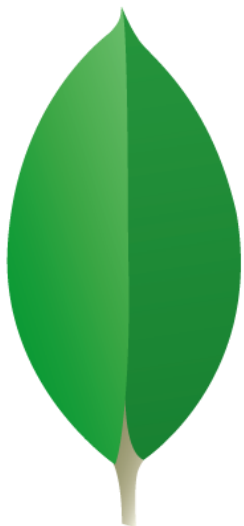
Aumentar a escala de 50%

Aproximar o relacionamento com o cliente

Aumentar os itens específicos para cada público. (Educação, sexo ...)



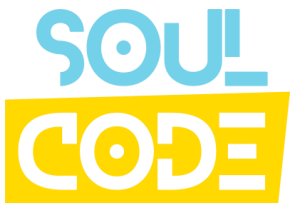
# Engenharia de Dados



PySpark

APACHE  
Spark™ SQL





# Engenharia de Dados

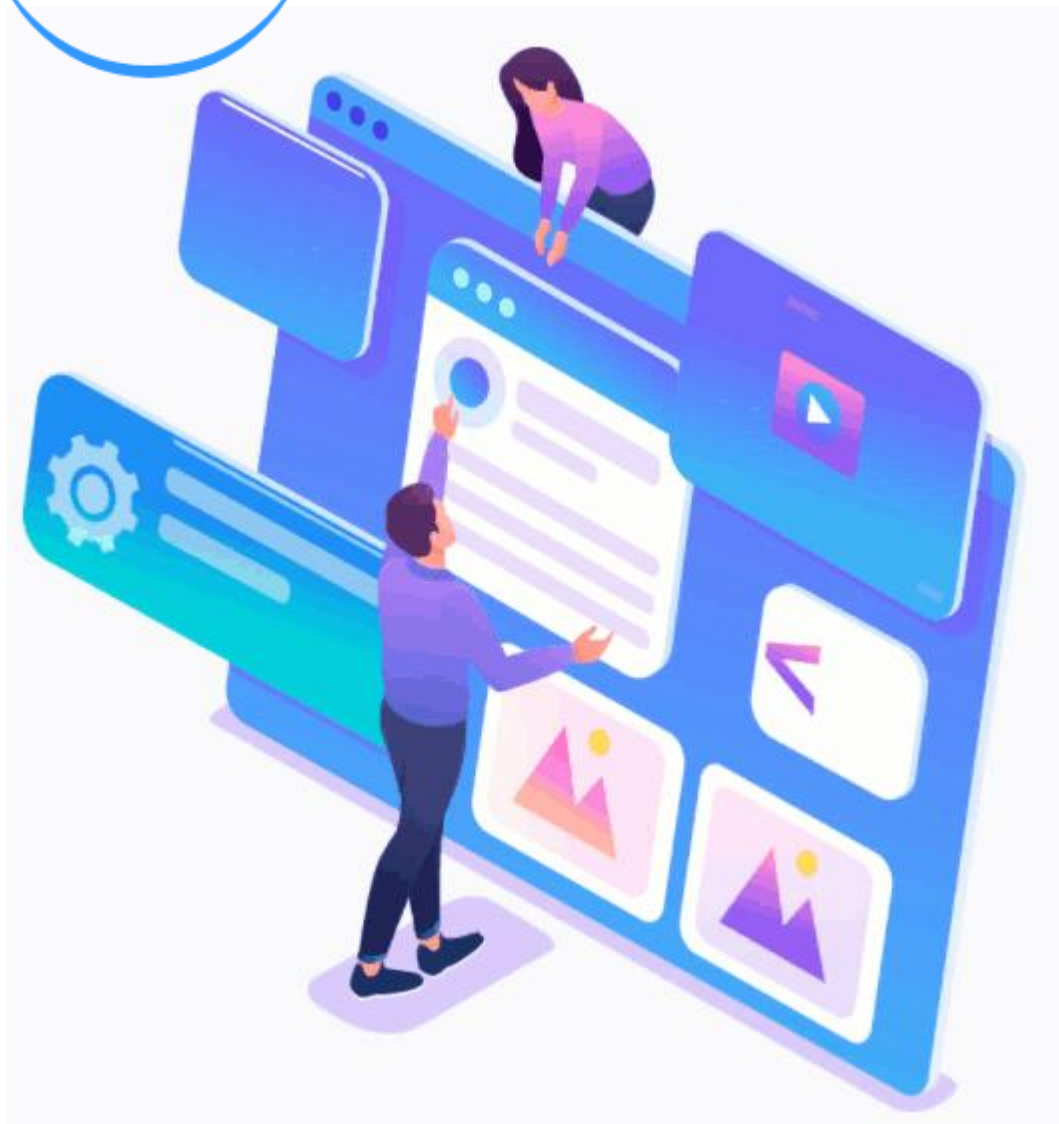
## Conclusões - Skills

- A finalidade do porque dos processos de ETL devem ser conhecido.
- O conhecimento de regras de negócio é um fator primordial nos processos de ETL.
- O Engenheiro de dados deve analisar os processos de ETL e sua finalidade.





# Engenharia de Dados



**Perguntas?**

**Respostas:**

**Grato!**

**Grato!**



CODE Google Cloud TNT CODE TNT Google Cloud CODE



SOUL

CODE

