Teste Engenheiro de Dados

# Ingestão

* + Script organizacao.py descompacta a pasta zip e move os arquivos CSV para camada stage
  + Script ingestão.py lê os arquivos CSV adiciona a coluna DT\_CARGA, faz a carga da camada raw/tabela Pacientes e Exames com arquivos em formato *parquet*, compressão *snappy* e codificação UTF-8 e move os arquivos CSV para stage/historico com a data da movimentação

Normalização e aplicação de regras de qualidade

* Script nmlExames.py normalizar e aplicar regras de qualidade na tabela Exames, faz a carga da camada curated/hot/tb\_exames para os dados que passaram nas regras de qualidade e curated/rejected/tb\_exames para os que não passaram nas regras de qualidade.
* Script nmlPacientes.py normalizar e aplicar regras de qualidade na tabela Pacientes, faz a carga da camada curated/hot/tb\_pacientes para os dados que passaram nas regras de qualidade e curated/rejected/tb\_pacientes para os que não passaram nas regras de qualidade.

Transformação e processamento

* Script exames\_pacientes.py gera um conjunto de dados cruzando dados de exames e pacientes que estão na camada curated/hot, gerando desse conjunto um filtro com UF de São Paulo e fazendo a carga para camada service

Uso dos dados transformados

* Script consumo.py gera um histograma da coluna DE\_ANALITO