

Documentação da Tabela: trusted-zone.tb\_corrida\_taxi\_verde

Descrição

Esta tabela contém os dados das corridas de táxi verde, processados a partir de arquivos Parquet brutos encontrados no diretório /Volumes/workspace/raw-zone/taxi\_verde/. Os dados são tratados para padronização, com nomes das colunas traduzidos para português, tipos de dados ajustados e valores numéricos corrigidos para o padrão decimal com ponto.

A tabela serve para análise de desempenho, faturamento, e comportamento das corridas, permitindo cruzamentos com a base de táxi amarelo ou outras fontes.

Processo de Ingestão e Transformação

- Arquivos Parquet são lidos do diretório raw.
- Todas as colunas são convertidas para lowercase.
- Renomeação das colunas conforme o dicionário renomear\_colunas, garantindo nomes descritivos em português.
- Valores numéricos com vírgulas são convertidos para ponto decimal.
- Aplicação de cast seguro conforme tipos definidos no dicionário tipos\_colunas, convertendo valores inválidos para null.
- Inclusão da coluna origem\_taxi com valor fixo "taxi\_verde" para identificação da fonte dos dados.
- Salvamento da tabela no catálogo trusted-zone com opção mergeSchema habilitada para facilitar alterações futuras.

Esquema da Tabela

Nome da Coluna	Tipo de Dados	Descrição
----- ----- -----		
-----		
cod_motorista	string	Código identificador do motorista (originado do campo vendorid).
dt_hr_inicio	timestamp	Data e hora do início da corrida (lpep_pickup_datetime).
dt_hr_fim	timestamp	Data e hora do término da corrida (lpep_dropoff_datetime).

qtd_pessoas	integer	Quantidade de passageiros na corrida.
dist_percorrida	decimal(10,2)	Distância percorrida durante a corrida, em milhas.
cod_taxa	integer	Código do tipo de taxa aplicada (ratecodeid).
ind_armazenamento	string	Indicador se os dados foram armazenados e encaminhados (store_and_fwd_flag).
cod_bairro_origem	string	Código do bairro ou local de origem da corrida (pulocationid).
cod_bairro_destino	string	Código do bairro ou local de destino da corrida (dolocationid).
tipo_pagamento	string	Forma de pagamento utilizada (exemplo: cartão, dinheiro).
tipo_viagem	integer	Tipo de viagem (campo trip_type, com código específico da modalidade da corrida).
vlr_taxa_corrida	decimal(10,2)	Valor base da corrida (fare amount).
vlr_taxa_extra	decimal(10,2)	Valor adicional cobrado na corrida (extra).
vlr_taxa_mta	decimal(10,2)	Taxa MTA cobrada (Metropolitan Transportation Authority).
vlr_troco	decimal(10,2)	Valor do troco recebido (tip amount).
vlr_pedagio	decimal(10,2)	Valor pago em pedágios durante a corrida (tolls amount).
vlr_gorjeta	decimal(10,2)	Valor da gorjeta cobrada (ehail_fee).
cod_taxa_melhoria	integer	Código da taxa de melhoria aplicada (improvement surcharge).
vlr_total	decimal(10,2)	Valor total pago na corrida, somando todas as taxas.
vlr_taxa_congestao	decimal(10,2)	Valor da taxa de congestionamento aplicada (congestion surcharge).
vlr_taxa_aeroporto	decimal(10,2)	Valor da taxa cobrada para corridas com destino ou origem em aeroporto.
origem_taxi	string	Origem da base, sempre com valor fixo: "taxi_verde".

### Observações

- Valores numéricos fora do padrão decimal esperado são convertidos para null para manter a integridade dos dados.
- Campos de data/hora são convertidos para o tipo timestamp para facilitar análise temporal.
- A opção mergeSchema no salvamento da tabela permite alterações evolutivas no esquema.
- A coluna origem\_taxi permite combinar e diferenciar dados entre táxi verde e amarelo em análises

futuras.

#### Uso Sugerido

- Análises temporais para identificar períodos de maior demanda.
- Avaliação do comportamento do passageiro e preferências de pagamento.
- Cálculo de receita total por motorista, bairro, ou modalidade de viagem.
- Cruzamento com outras bases para análise integrada de mobilidade urbana.