Documentação da Tabela: refined-zone.tb_corrida_taxi_unificado

Descrição

Tabela consolidada que unifica os dados das corridas de táxi verde e amarelo, filtradas para o período de 1º de janeiro a 31 de maio de 2023. Esta tabela faz parte da camada refined do pipeline, trazendo dados essenciais para análises simplificadas e otimizadas para consulta.

Processo de Construção

- Leitura das tabelas base trusted-zone.tb_corrida_taxi_verde e trusted-zone.tb_corrida_taxi_amarelo.
- Aplicação de filtro temporal para manter somente corridas cujo início e fim estejam entre "2023-01-01 00:00:00" e "2023-05-31 23:59:59".
- Seleção apenas das colunas relevantes:
 - cod_motorista: identificador do motorista
 - dt_hr_inicio: data e hora de início da corrida
 - dt_hr_fim: data e hora de fim da corrida
 - qtd_pessoas: quantidade de passageiros
 - vlr_total: valor total da corrida
 - origem_taxi: origem da corrida (taxi_verde ou taxi_amarelo)
- União dos dois dataframes via unionByName para garantir alinhamento correto das colunas.
- Salvamento do resultado na camada refined, na tabela refined-zone.tb_corrida_taxi_unificado, com sobrescrição da tabela existente.

Esquema da Tabela

string

Observações

- Os filtros garantem que apenas corridas dentro do período definido são incluídas.
- A camada refined visa otimizar consultas, fornecendo um dataset limpo e reduzido.
- A coluna origem_taxi facilita análises segmentadas por tipo de táxi.
- A tabela é atualizada via overwrite a cada execução do pipeline.

Uso Sugerido

- Análises de receita e volume de corridas por tipo de táxi.
- Estudos temporais para avaliar demanda e sazonalidade.
- Dashboards executivos e relatórios simplificados.
- Base para modelos preditivos com dados limpos e padronizados.