

Vinícius Miranda Andrade Piovesan RA: 23025544

Felipe Ribeiro Almeida

RA: 23024683

Sérgio Ricardo Pedote Junior

RA:23747441

MATHEUS DE MEDEIROS TAKAKI

RA: 23025143

1. Limpar / Uniformizar os Dados

- **Seleção de colunas:** filtragem de colunas desnecessárias, mantendo apenas RideID, RideAddressID, RideAddressTypeID, Lat, Lng.
- **Tratamento de coordenadas:** criação da coluna Coordenadas no formato padronizado "lat,lng" para padronizar entrada de dados.
- **Filtragem de registros:** separação de registros de origem (RideAddressTypeID == 1) e destino (RideAddressTypeID == 2), removendo registros sem par.

2. Derivar Dados

- **Cálculo de métricas de rota:** uso da API Google Maps para derivar duas novas variáveis:
 - **Distancia_km:** distância de trajeto em quilômetros (arredondado a duas casas decimais).
 - **Duracao_min:** tempo de viagem em minutos (arredondado a duas casas decimais).

- **Limitação para entrega:** processamento de apenas 5.000 primeiras linhas para cumprir o escopo de entrega.
-

3. Integrar os Dados

- **Unificação de origens e destinos:** merge interno (inner join) por RideID, associando coordenadas de início e fim em uma única linha.
 - **Enriquecimento com estimativas:** merge à esquerda com rideestimative_v3.csv, adicionando as colunas:
 - ProductID
 - Price Baseado no campo comum RideID.
-

4. Formatar os Dados

- **Delimitador CSV:** uso de ponto e vírgula (;) para compatibilidade regional.
 - **Estrutura de colunas:** ordenação lógica das colunas em cada arquivo de saída:
 1. RideID
 2. RideAddressID
 3. Coordenadas Inicio
 4. Coordenadas Fim
 5. Distancia_km
 6. Duracao_min
 7. ProductID e Price (após integração)
 - **Precisão numérica:** valores de distância e tempo com duas casas decimais.
 - **Modularidade:** geração de arquivos intermediários (tabela_unificada_entrega2.csv, tabela_derivada_entrega2.csv) e final (tabela_completa_entrega2.csv).
-

<https://colab.research.google.com/drive/1MqWSidfS5KsR8JcEMbFXfAKSGW-hg5zY?usp=sharing>

