

**Entendimento Dados Khipo**  
**Ciência de Dados e Big Data**

**INTEGRANTES DO PROJETO e RA'S**

Daniel Baptista Acioli Vanderlei	23025608
Fábio Oliveira Spíndola	22086131
Fabício Cândido do Nascimento	23025273
Leonardo de Souza Mouzinho	23025627

**Sumário**

<b>1.</b>	<b>Scripts .....</b>	<b>3</b>
<b>2.</b>	<b>Tabela product.....</b>	<b>4</b>
<b>3.</b>	<b>Tabela ride_v2.....</b>	<b>5</b>
<b>4.</b>	<b>Tabela rideaddress_v1 .....</b>	<b>6</b>
<b>5.</b>	<b>Tabela rideestimative_v3 .....</b>	<b>7</b>

## 1. Scripts

### **Link Google Colab:**

<https://colab.research.google.com/drive/1ux3P8wp4LcX6OhJ0KQ8xZvZQF8RhMCeE?usp=sharing>

## 2. Tabela product

Coluna	Descrição do Dado	Contém Nulos?	Qualidade do Dado
ProductID	Nome do produto Ex: 99pop, Uber X, Uber Comfort e etc	Não	Bom
ProviderID	Números sortidos sem ordem	Não	Ruim
CategoryID	Números sortidos sem ordem	Não	Ruim
Description	Repete o ProductID	Não	Bom

### 3. Tabela ride\_v2

Coluna	Descrição do Dado	Contém Nulos?	Qualidade do Dado
RideID	ID da corrida usada para correlacionar um Driver a um User, ela é gerada depois de uma consulta do user e usada como chave primaria recebendo chaves estrangeiras user/driver	Não	Bom
UserID	Código do usuário	Não	Médio
Schedule	Informa o momento que o user deseja a viagem Obs: Ele pode agendar viagem	Não	Bom
Create	Informa o momento que o user cria a viagem	Sim	Bom
RideStatusID	Status da corrida	Não	Ruim
CompanyID	Indica se é uma corrida corporativa	Não	Ruim
ProviderID		Sim	Ruim
RideProviderID		Sim	Ruim
price	Preço da corrida	Não	Bom
Updated	Momento que a corrida encontra um driver	Não	Bom
CategoryID	Sem dados para análise	Sim	Ruim
TotalUsers	Quantidade de driver alocados para a corrida Ex:driver aceita e cancela e direciona para o segundo	Não	Bom
Car	Veiculo do driver	Sim	Ruim
RideDriverLocationID	Localização da corrida	Sim	Médio
ScheduledRide	Sem dados para análise	Não	Ruim

#### 4. Tabela rideaddress\_v1

Coluna	Descrição do Dado	Contém Nulos?	Qualidade do Dado
RideAddressID	ID do endereço gera dois a cada consulta local de embarque e desembarque	Não	Bom
Address	Endereço completo concatenado rua, numero e bairro	Não	Bom
Street	Nome da rua de destino	Não	Bom
Number	Número da rua de destino	Não	Bom
Neighborhood	Bairro de destino	Não	Bom
City	Cidade de destino	Não	Bom
State	Estado de destino	Não	Bom
Lat	Latitude	Não	Médio
Lng	Longitude	Não	Médio
RideAddressTypeID	Tipo de endereço	Não	Ruim
RideID	ID da corrida	Não	Bom

## 5. Tabela rideestimative\_v3

Coluna	Descrição do Dado	Contém Nulos?	Qualidade do Dado
RideEstimativeID	Este ID é criado sequencialmente para cada produto	Não	Bom
RideID	Este ID é criado unitariamente para cada busca de preço e ele se repete em todos os produtos	Não	Bom
ProductID	Este ID é criado para cada produto da consulta Ex: Flash, Uber X, Comfort, 99Pop	Não	Bom
WaitingTime	Tempo aguardando pela corrida conforme o produto	Não	Bom
Price	Preço de cada produto	Não	Bom
FareID	Código de tarifa criado aleatoriamente na base essa informação não existe no produto 99	Sim	Ruim
Selected	Dados inexistente sem identificação	Não	Ruim
RideReasonSelectedEstimativeID	Razão pela qual o passageiro pesquisou a corrida, sem dados	Sim	Ruim
Fee	Representa a taxa que a empresa cobra pelo serviço	Sim	Ruim