

## Nomes:

Bruno Cardoso da Silveira  
Christian Aguiar Plentz

### 1. Breve descrição da base:

Esse conjunto de dados traz informações de 150.000 corridas que foram solicitadas por meio da Uber. Ele abrange o ciclo completo da reserva, desde o pedido até a conclusão, cancelamento ou status de incompleta (corridas que não foram concluídas e tiveram que ser canceladas no durante o trajeto). Os dados são da Região da Capital Nacional (NCR) da Índia, uma espécie de região metropolitana, que engloba a capital Nova Delhi e cidades satélites. Neste dataset temos informações como:

- **Informações de Reserva (Solicitação de corrida):** Data, Hora, Status.
- **Detalhes da corrida:** Tipo de veículo, Locais de partida e destino, Distância da viagem, Valor.
- **Dados de cancelamento:** Registros de quem cancelou (motorista ou passageiro), Motivos do cancelamento.
- **Detalhes de Incompletos:** Motivo que impediu a corrida de ser concluída.
- **Avaliações:** Notas de avaliação fornecidas por clientes e motoristas.
- **Pagamento:** O método de pagamento utilizado para a corrida.

### 2. Análise Descritiva com principais estatísticas:

Nossa base conta com 150.000 corridas na Índia ao longo de quase um ano. Nossa análise nos revela um serviço de transporte com ampla cobertura geográfica, com 176 localidades, e uma grande versatilidade, atendendo tanto a viagens curtas quanto longas (1 a 50 km).

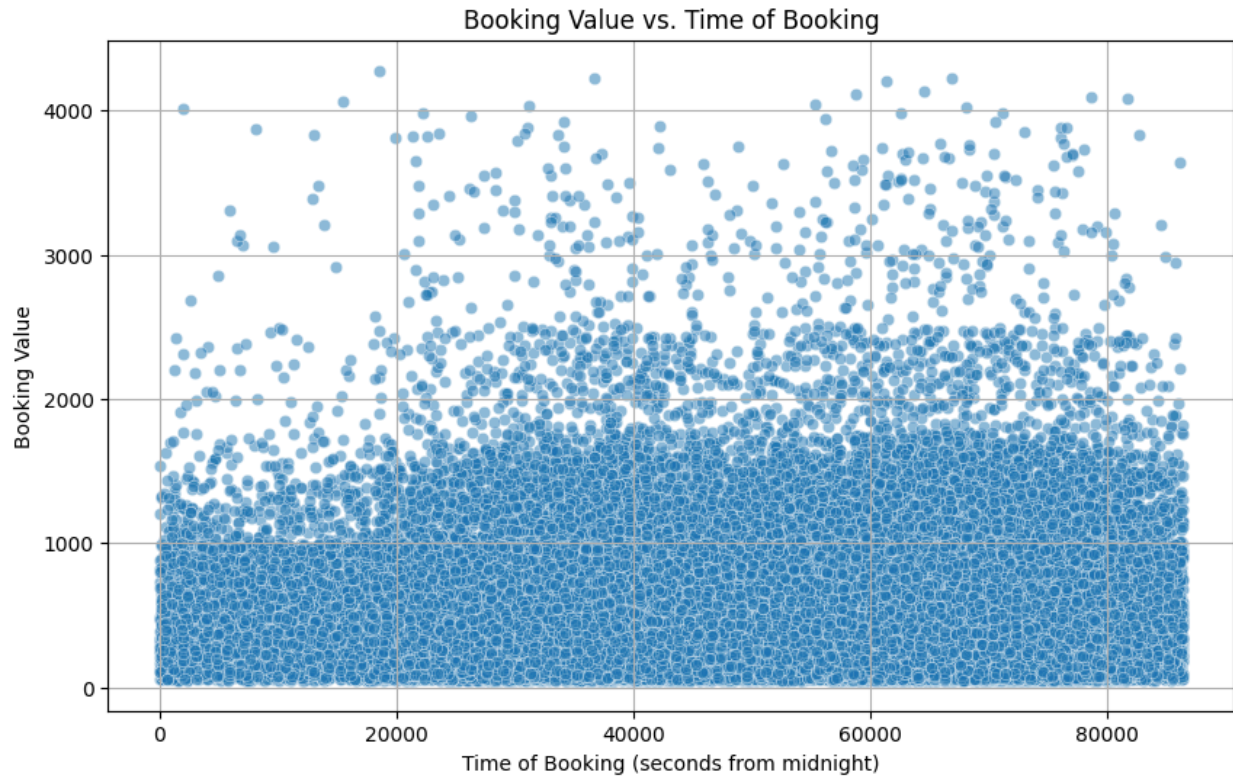
Como nosso dataset representa corridas em Delhi, fizemos nossa análise de medidas descritivas centrada no conceito de corrida.

Métrica	Valor
Contagem Corridas	150.000
Pontos de partidas distintos	176
Destinos distintos	176
Menor distância (km)	1.00

Maior distância (km)	50.00
Preço mínimo da corrida (INR)	50.00
Preço máximo da corrida (INR)	4277.00
Média de custo das corridas (INR)	508.30
Mediana de preço das corridas (INR)	414.00
Desvio padrão de preço (INR)	395.81
Amplitude em dias	364
Amplitude em meses	11

## 2.1 Análise de dinâmica:

Como usuários da Uber, sabemos que em horários de pico é certo que vamos sofrer com os preços mais caros. Será que nosso dataset reflete esse comportamento



A seguir vamos o gráfico de valor de corrida comparado aos segundos até meia noite, no caso quando começa a chegar aos 20000 segundos ( $\sim$  5,5 horas em torno das 19:30) pode se ver uma ausência de uma população grande no valor de corrida na ala entre 2000 e 3000. então pode se sustentar essa hipótese do preço da corrida está correlacionado ao movimento do horário.,

## 3. Identificação do tipo de cada variável:

Classificação das Variáveis:

- **Date:** Categórica (Temporal)
- **Time:** Categórica (Temporal)
- **Booking ID:** Categórica (Nominal)
- **Booking Status:** Categórica (Nominal)
- **Customer ID:** Categórica (Nominal)

- **Vehicle Type:** Categórica (Nominal)
- **Pickup Location:** Categórica (Nominal)
- **Drop Location:** Categórica (Nominal)
- **Avg VTAT:** Contínua
- **Avg CTAT:** Contínua
- **Cancelled Rides by Customer:** Discreta
- **Reason for cancelling by Customer:** Categórica (Nominal)
- **Cancelled Rides by Driver:** Discreta
- **Driver Cancellation Reason:** Categórica (Nominal)
- **Incomplete Rides:** Discreta
- **Incomplete Rides Reason:** Categórica (Nominal)
- **Booking Value:** Contínua
- **Ride Distance:** Contínua
- **Driver Ratings:** Ordinal
- **Customer Rating:** Ordinal
- **Payment Method:** Categórica (Nominal)

#### **4. Consistência nos dados:**

A base em um geral é bem consistente, mas para verificar a consistência demos ênfase nos pontos principais:

- Consistência nas Corridas Completas: Não temos nenhuma corrida com status 'Completed' que esteja sem valor (Booking Value) ou distância (Ride Distance).
- Não há nenhuma corrida registrada com distância igual a 0 e que tenha um valor cobrado.
- Com a análise descritiva dos dados de valores de corrida, os valores máximos, embora altos (4277.00 INR  $\approx$  R\$ 265,00), são esperados para corridas longas ou com veículos "premium".