

Entrega 1 – Inteligência Artificial e Aprendizado de Máquina

Descrição do Problema

O problema a ser resolvido foca no âmbito de mobilidade urbana, em que, se busca centralizar e catalogar simulações e preços de corridas de aplicativos que oferecem o serviço de carros particulares, como, Uber e 99.

Dessa forma, o grupo pretende, por meio de Inteligência Artificial, produzir uma solução que seja capaz de prever preços em simulações de corridas, dos mais diversos aplicativos e categorias de corrida.

Dicionário de Dados

- **ProductID:** ID que referencia a categoria de corrida
- **Description:** Descrição da categoria de corrida
- **RideID:** ID que referencia a corrida que foi realizada
- **Schedule:** data que a corrida realizada aconteceu
- **Create:** data que a corrida realizada foi criada
- **Price:** preço que a corrida realizada teve
- **TotalUsers:** total de passageiros que a corrida teve
- **RideAddressID:** ID que referencia o endereço da corrida realizada
- **Address:** endereço completo da corrida realizada
- **Street:** rua derivada do endereço da corrida
- **Number:** número derivado do endereço da corrida
- **Neighborhood:** bairro derivado do endereço da corrida
- **City:** cidade derivada do endereço da corrida
- **State:** estado derivado do endereço da corrida
- **Lat:** latitude do endereço de origem ou destino
- **Lng:** longitude do endereço de origem ou destino
- **RideEstimativeID:** ID que referencia a simulação de uma corrida
- **WaitingTime:** tempo que o cliente esperou até a corrida iniciar
- **Selected:** indica qual simulação o cliente escolheu para realizar
- **Fee:** taxa utilizada para calcular o preço da corrida

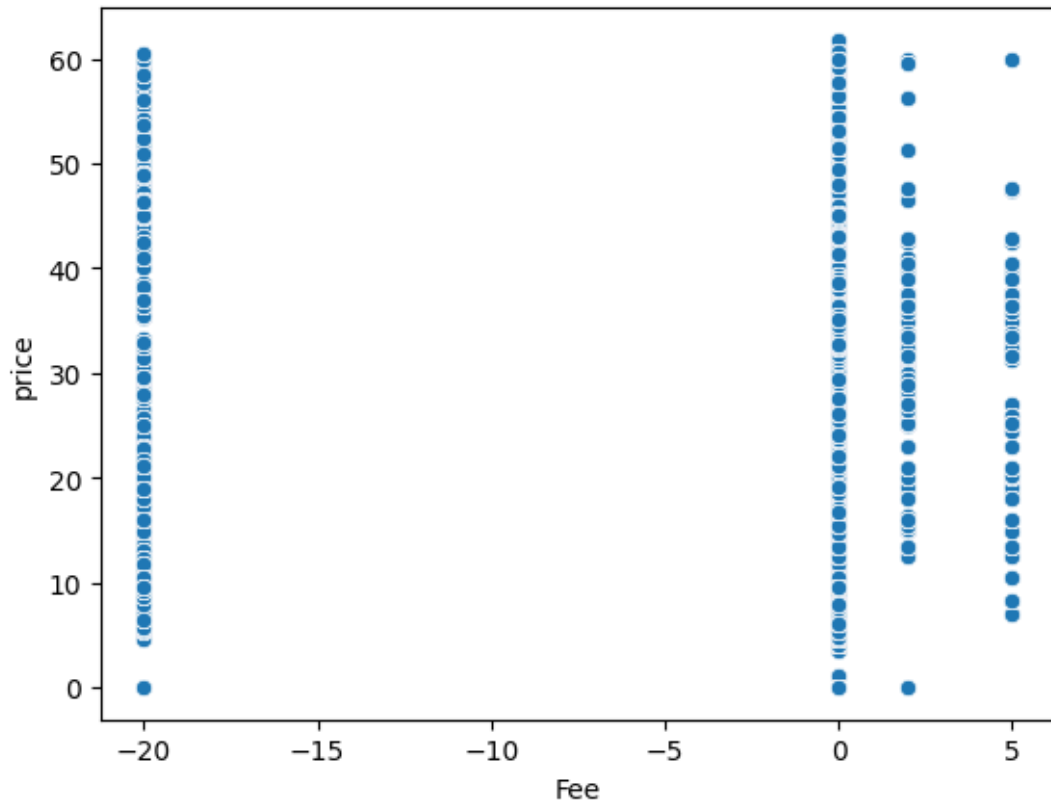
Transformações nos dados

Para fins de limpeza de base, fizemos a exclusão de algumas colunas, seja por inutilização dos dados ou por falta de um dicionário para fazer sentido aos dados. As colunas excluídas foram: UserID, CategoryID, RideProviderID, ProviderID, RideStatusID, CompanyID, Car, RideDriverLocationID, RideAddressTypeID, FareID e RideReasonSelectedEstimativeID.

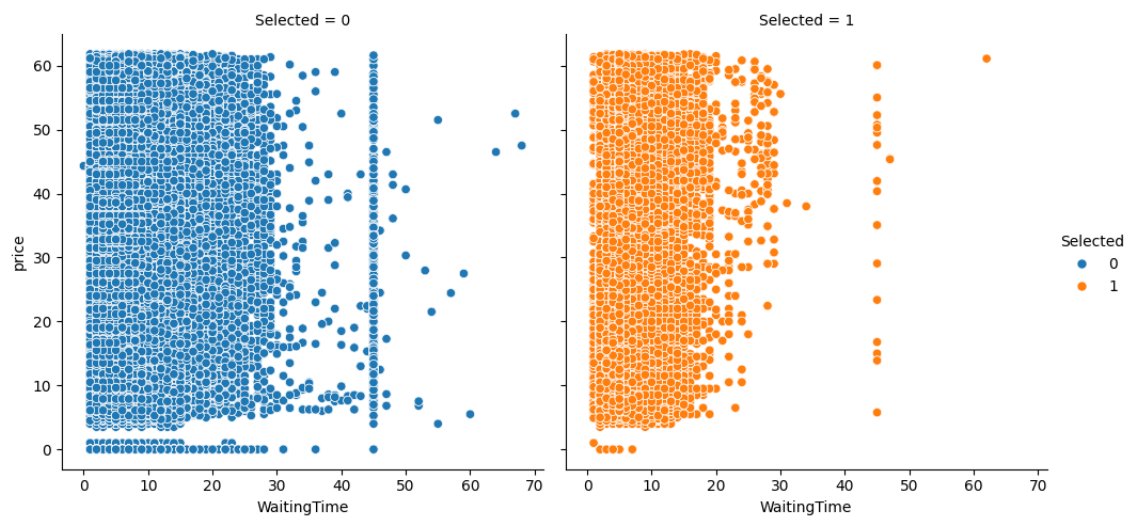
Fizemos também, a remoção de alguns outliers, principalmente, pela coluna de preço, em que havia alguns registros de corridas acima dos R\$ 2000,00, o que resultaria num desvio muito grande dos dados.

Análise Descritiva dos Dados

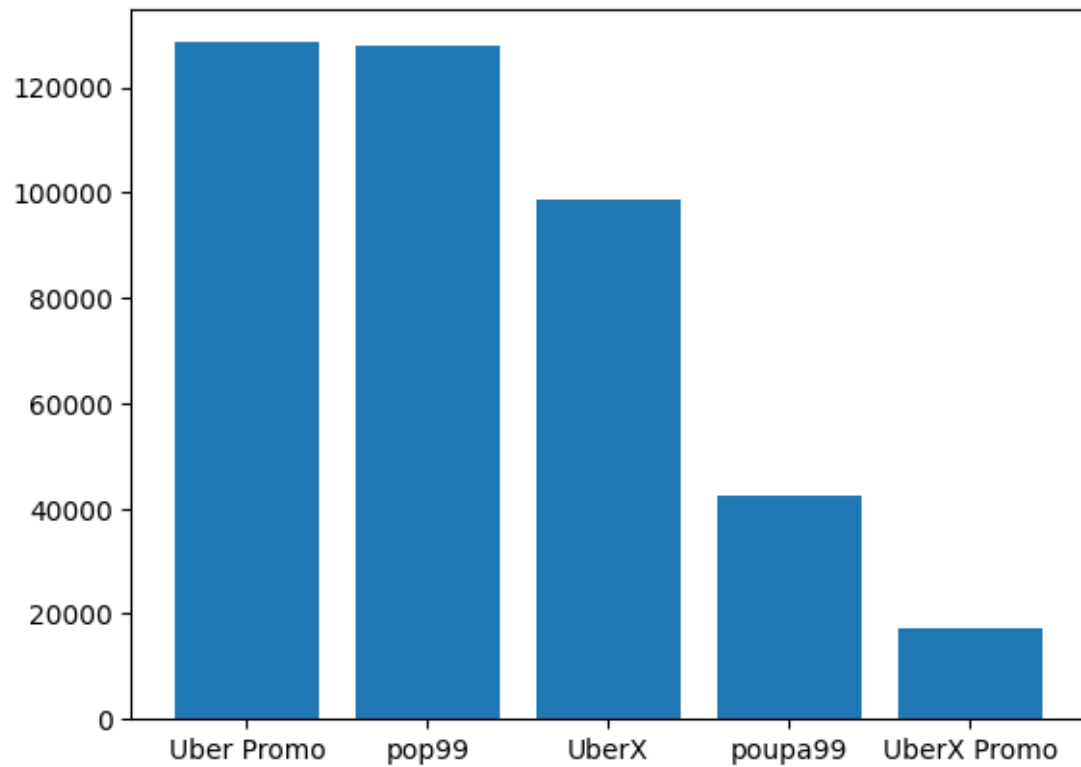
- Gráfico que mostra a concentração dos preços de corridas a partir das tarifas aplicadas



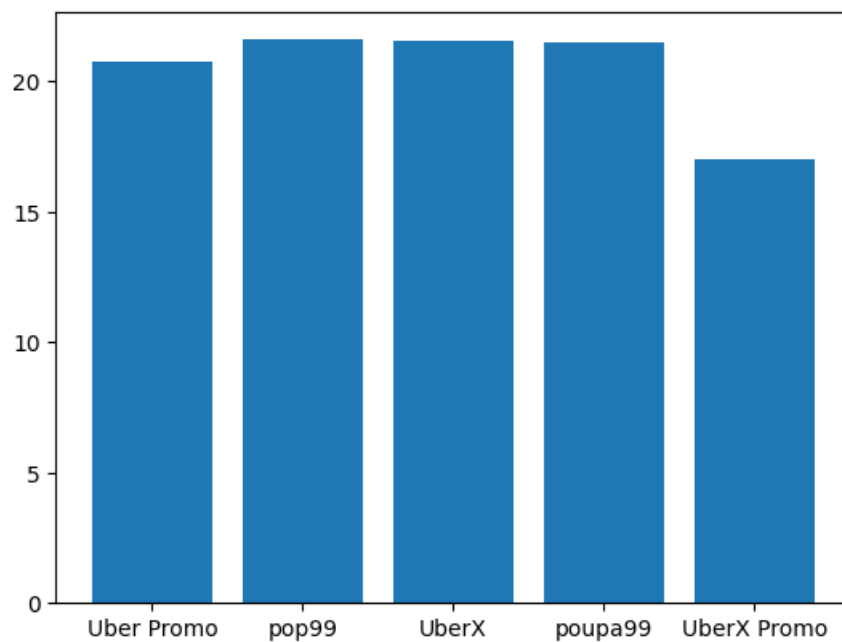
- Gráfico que mostra a relação de tempo de espera e preço das corridas, diferenciando, por cores, se a simulação foi aceita ou não



- Gráfico que mostra as 5 categorias com mais corridas realizadas



- Gráfico que mostra a média de preço das 5 categorias mais selecionadas



- Gráfico que mostra a distribuição do tempo de espera e das categorias de corrida

