FUNDAÇÃO ESCOLA DE COMÉRCIO ÁLVARES PENTEADO

Entrega 1 - Inteligência Artificial e Aprendizado de Máquina

Estimador de preços UbaRun

Beatriz de Castilho Ferreira

Giulia Nogueira Lopes de Sá

Guilherme Alves de Oliveira

João Pedro Lima Paulo

Lara Marina de Oliveira

São Paulo

2025

**Sumário**

[Descrição do problema 3](#_je6oob9qrjoj)

[Coleta e limpeza de dados 3](#_eihzbqsrenhd)

[Instalando bibliotecas 3](#_b3gbskr41zw7)

[Importação das bibliotecas 4](#_91gxpcoplztf)

[Pegando os dados 4](#_202vo23r66mq)

[Padronizando nomes e campos 4](#_o0e3dywd84q4)

[Transformando, higienizando e padronizando os dados 4](#_pr0xlhputrm7)

[Tratamento do campo ‘State’ 6](#_5xypr7sin4ie)

[Visualização dos dados tratados 7](#_bdekqvpxcqsa)

[Exploração e análise inicial 9](#_fji7yzmobxj)

# 

# 

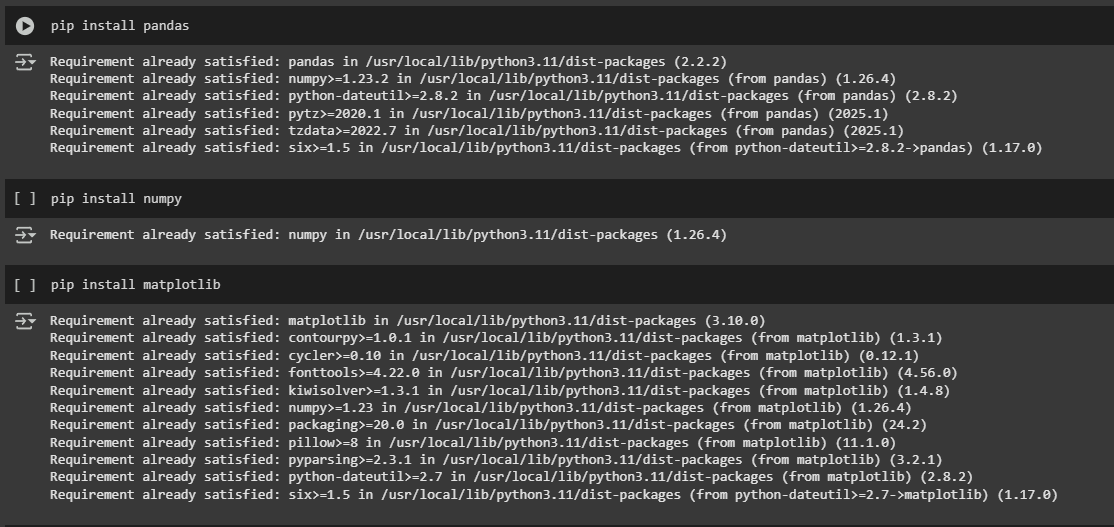
# **Descrição do problema**

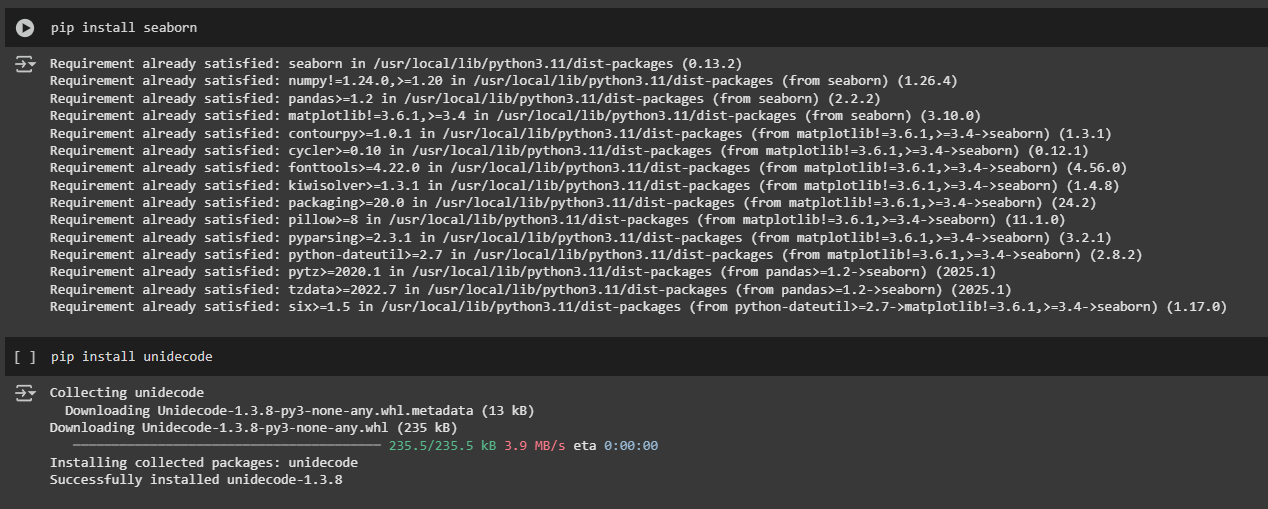
O objetivo é criar modelos de Machine Learning capazes de prever o preço estimado de corridas por aplicativo para as três categorias: UberX, Uber Confort e Uber Black. Também iremos utilizar dados de outros aplicativos de corrida ou transporte público para estimar a forma mais barata de chegar ao local desejado, gerando assim um estimador de vários serviços. Outra maneira é maximizar a acurácia das previsões por meio da escolha adequada de features, técnicas de dados e modelos.

* Acesse o Colab da entrega pelo link: [Entrega 1 - Colab](https://drive.google.com/file/d/1A3r66tfPtEYheG3mXxRKX3ZkADPCNS3G/view?usp=sharing)

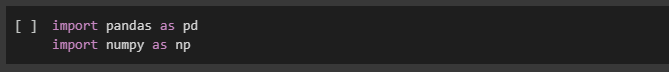
# **Coleta e limpeza de dados**

## Instalando bibliotecas

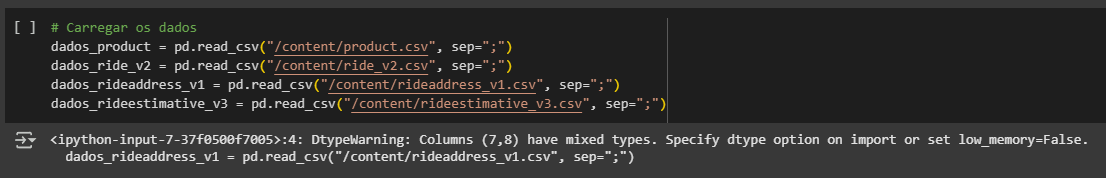




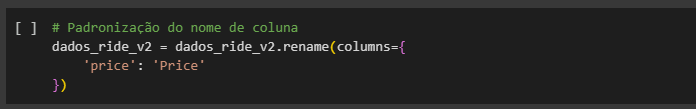
## Importação das bibliotecas



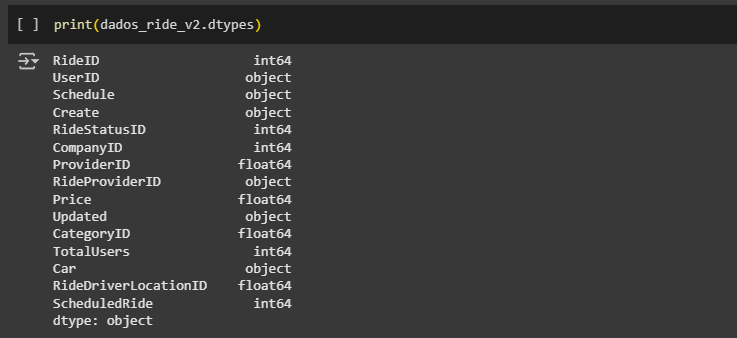
## Pegando os dados

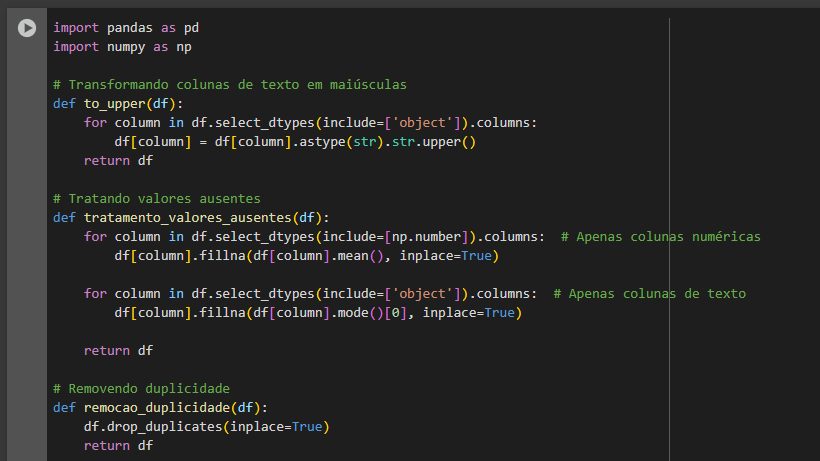


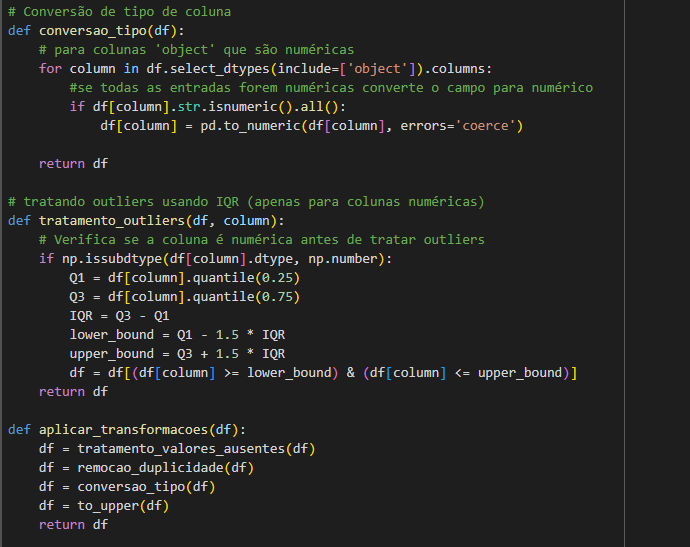
## Padronizando nomes e campos

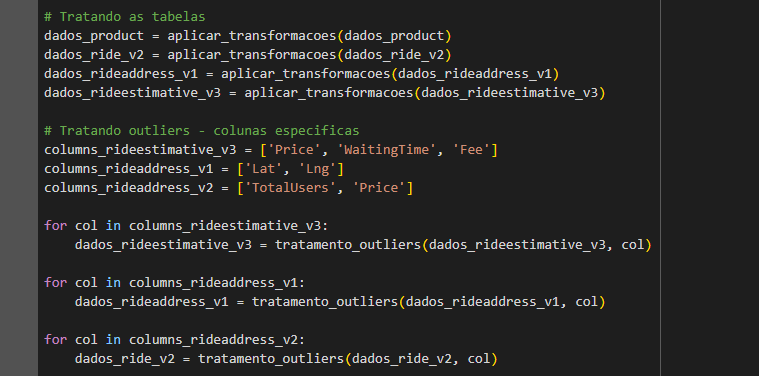


## Transformando, higienizando e padronizando os dados

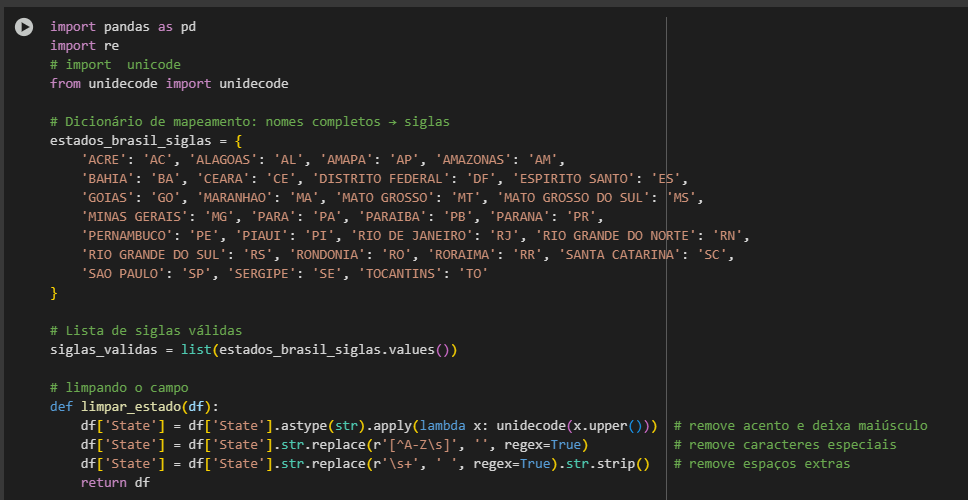


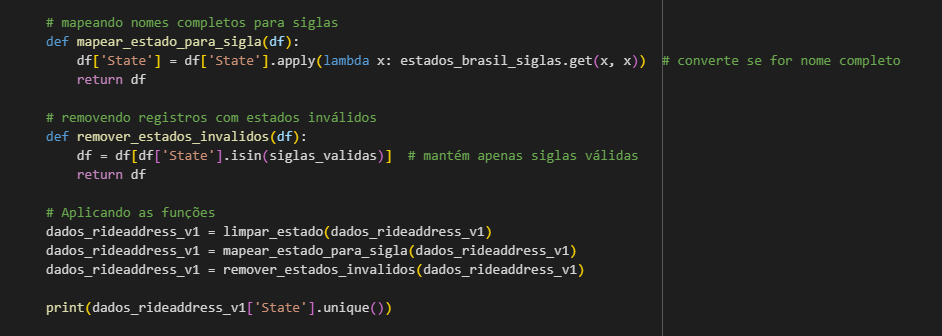






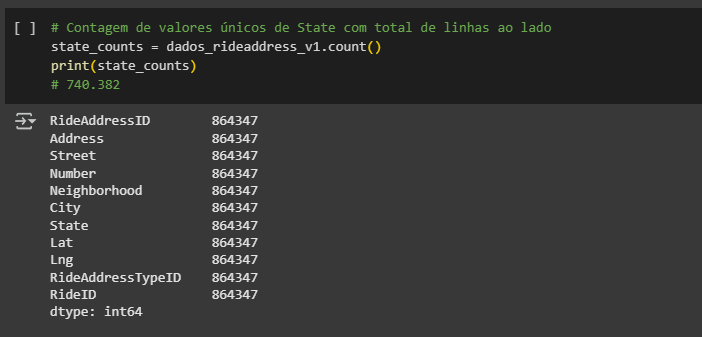
## Tratamento do campo ‘State’

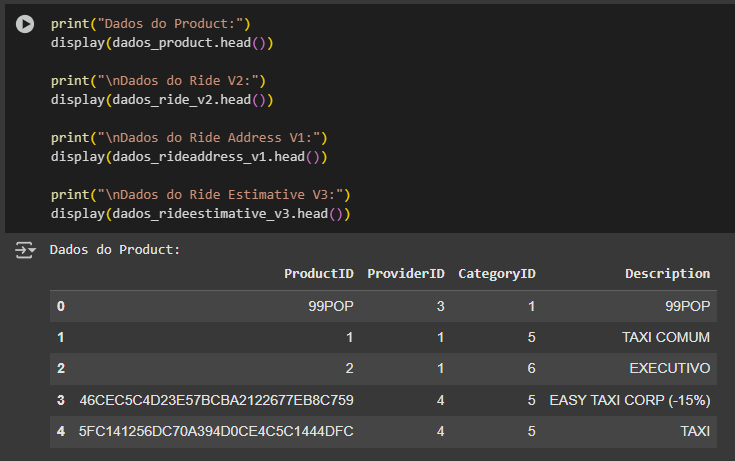


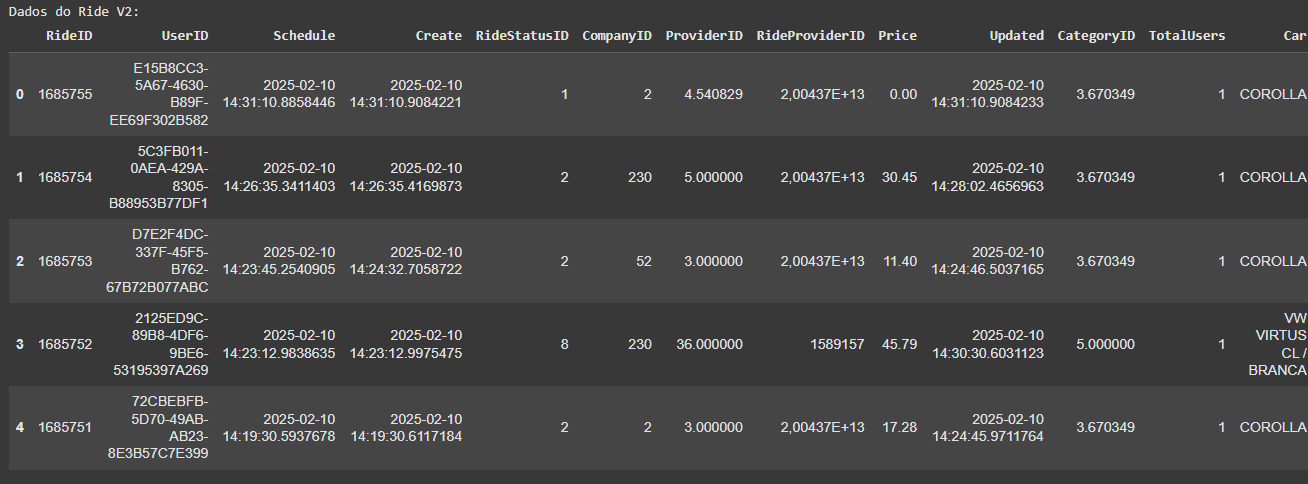


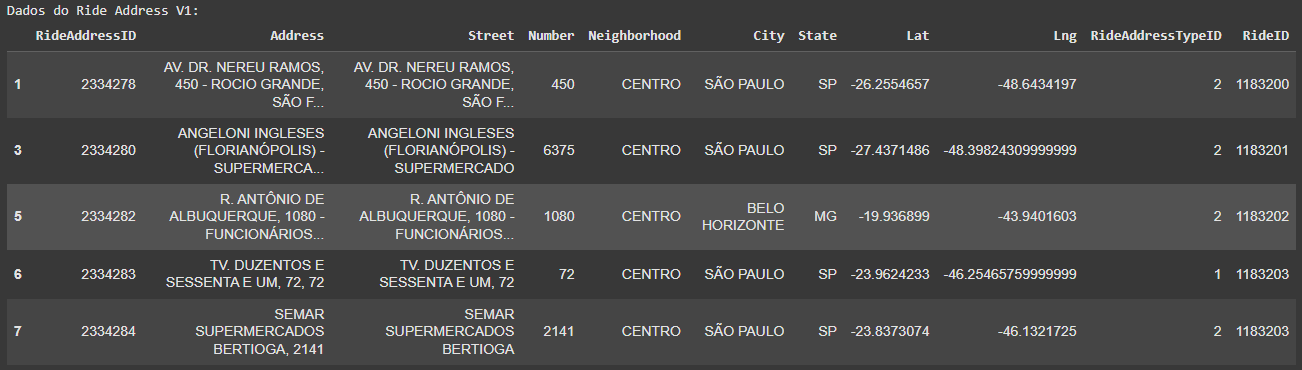


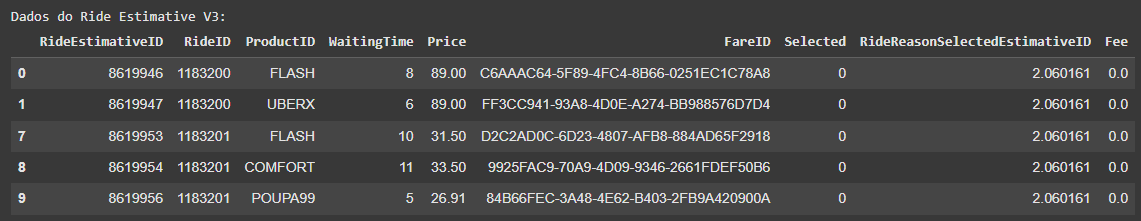
## Visualização dos dados tratados





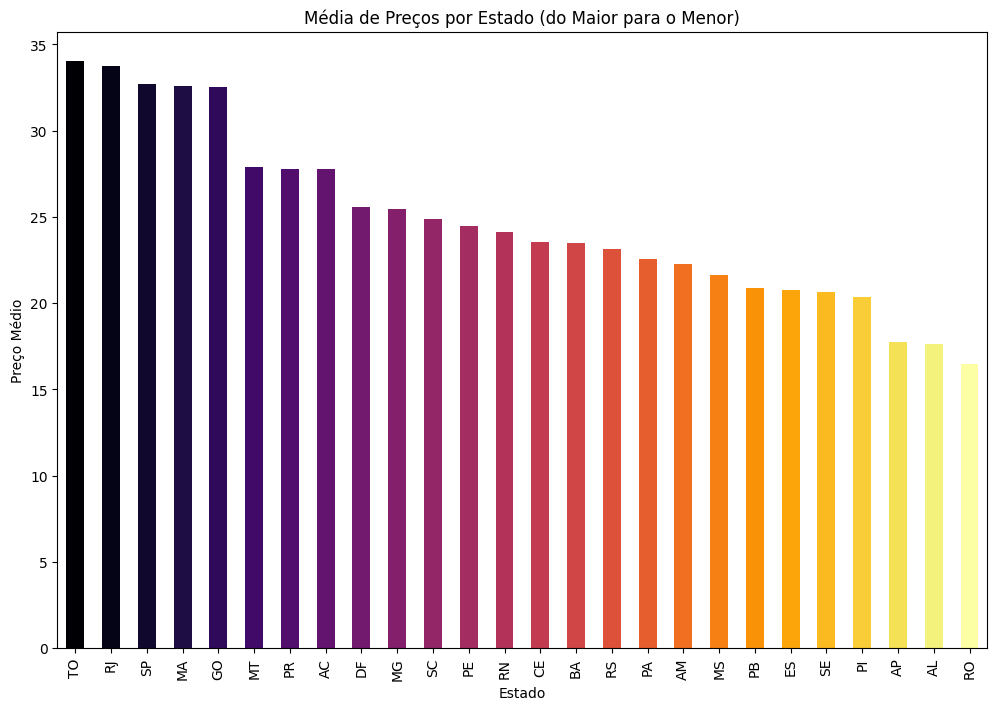






# **Exploração e análise inicial**

1. Gráfico de distribuição - Média de preços por estado



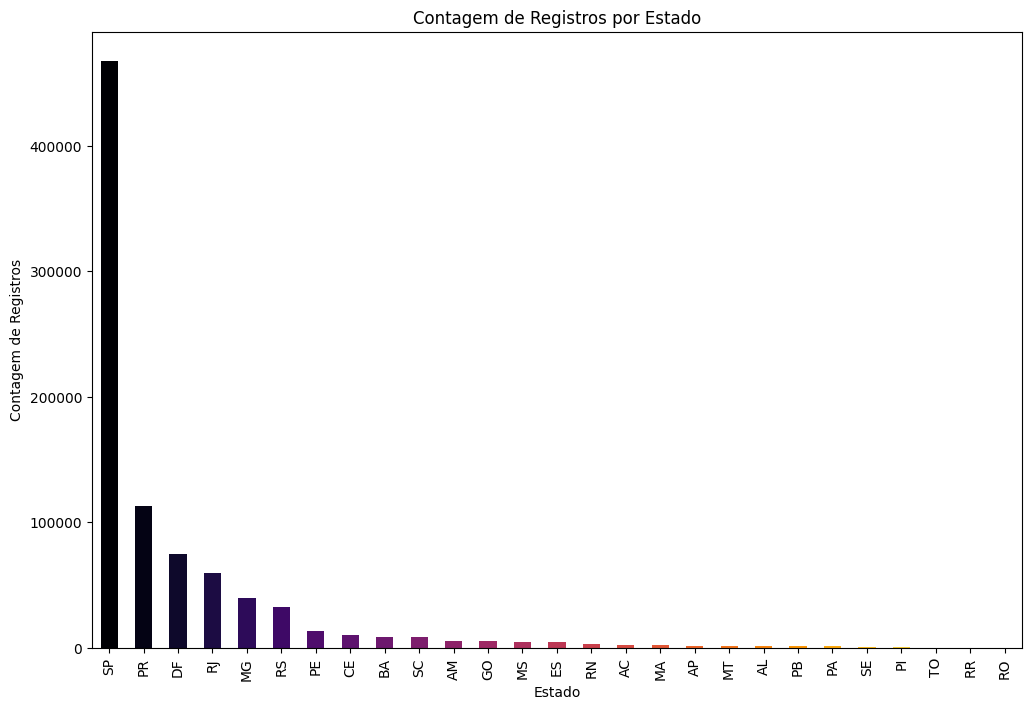
No gráfico de distribuição acima, temos um comparativo entre média de preços por cada estado brasileiro. As maiores médias de preço no Uber em Tocantins, Rio de Janeiro, São Paulo e Maranhão podem ser explicadas por uma combinação de fatores como o custo de vida, densidade populacional, infraestrutura local, concorrência de motoristas, demanda por serviços de transporte e características econômicas regionais.

Em relação ao custo de vida, SP e RJ são as maiores cidades do Brasil, com preços elevados de combustíveis, manutenção de veículos e salários, impactando no custo da tarifa dos aplicativos. Já nos estados de Maranhão e Tocantins, mesmo tendo o custo de vida mais baixo, a concorrência pode fazer com que esses valores aumentem significativamente.

Diferente de SP e RJ, que possuem uma densidade populacional muito grande, TO e MA podem ser impactados pelo turismo, alta demanda em locais sem motoristas e entre outros.

Lembrando que o Uber usa uma política de preços dinâmicos, onde o valor da corrida varia conforme a oferta e a demanda. Isso pode resultar em variações significativas de preço nas cidades, dependendo das condições de tráfego, do número de motoristas disponíveis, do horário e de outros fatores que podem gerar picos nas tarifa

1. Gráfico de distribuição - Contagem de registros por estado

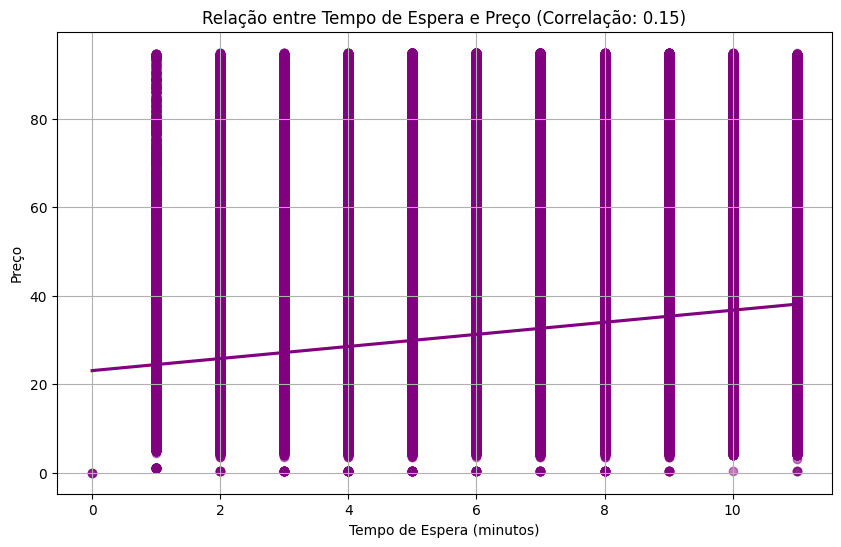


No gráfico de comparação entre contagem de registro de viagens por estados brasileiros, vemos que São Paulo lidera com mais do triplo de registros do segundo colocado, sendo Paraná.

São Paulo, sendo a maior cidade do Brasil, possui características que favorecem uma demanda maior e um maior volume de viagens, como alta densidade populacional, maior oferta de motoristas, maior infraestrutura e mobilidade, economia, turismo e eventos.

Esses elementos fazem com que São Paulo se destaque em termos de volume de viagens do Uber em comparação a outros estados brasileiros.

1. Gráfico de correlação - Tempo de espera e preço



Quando o tempo de espera nos aplicativos é maior, é comum que o preço da corrida aumente, e isso pode ser explicado por vários fatores dentro da política de preços dinâmicos e a estratégia de oferta e demanda do serviço.

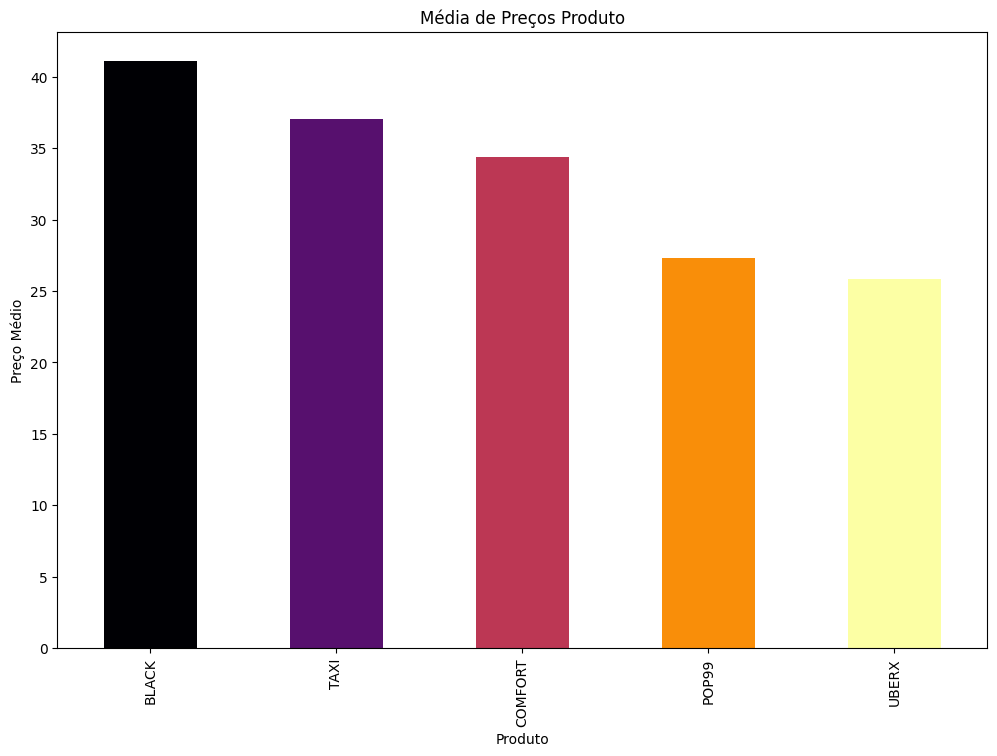
Motoristas ocupando mais tempo de espera acabam perdendo oportunidades de fazer outras corridas. Isso significa que eles estão dedicando mais tempo a uma única viagem, o que acaba impactando o custo da corrida.

A empresa pode aumentar o preço para compensar o motorista por esse tempo adicional, de forma a compensar o motorista pela espera mais longa e pelo tempo que ele está deixando de usar para pegar outros passageiros.

Além disso, se o tempo de espera maior for causado por trânsito ou outros obstáculos, o tempo gasto no deslocamento até o ponto de embarque vai gerar um custo adicional que pode ser repassado ao passageiro.

O objetivo é ajustar o preço para manter o equilíbrio entre oferta e demanda, garantir a viabilidade da corrida e proporcionar uma experiência mais eficiente para todos os envolvidos.

1. Gráfico de distribuição - Média de preço por produto



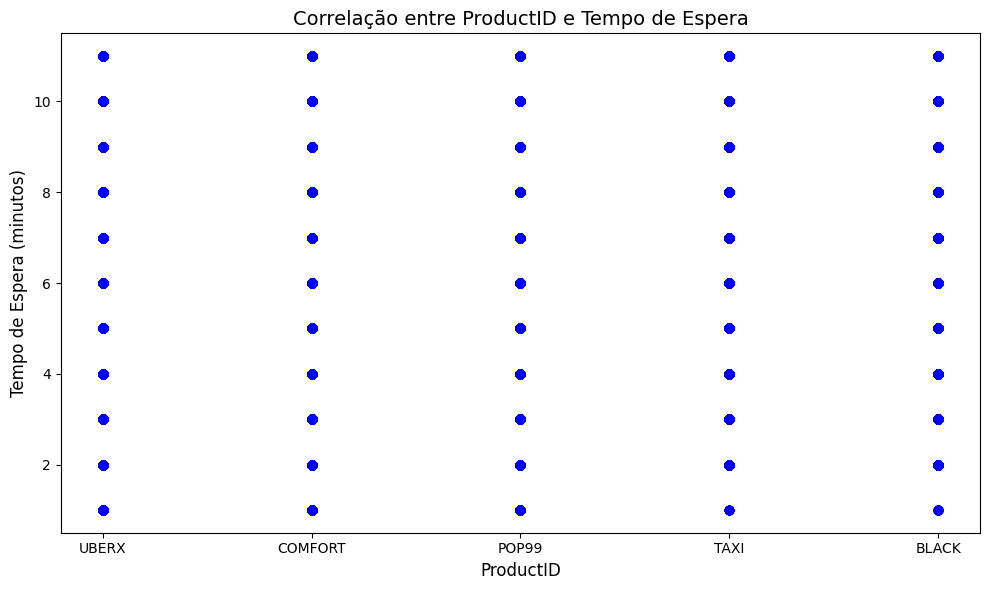
O Uber Black é mais caro do que outras categorias de serviço, como o Uber Confort ou o próprio Táxi convencional, por vários motivos. Esses fatores estão relacionados à qualidade do serviço, ao tipo de veículo utilizado e ao nível de exclusividade e conforto oferecido.

O Black exige que os motoristas utilizem carros de luxo ou premium, como sedans de marcas renomadas ou veículos de alto padrão. Esses carros são mais caros tanto para comprar quanto para manter, o que reflete no valor das corridas.A exclusividade também envolve um nível de privacidade maior, com carros mais sofisticados que atraem pessoas que buscam uma experiência diferenciada.

A categoria Uber Black também é percebida como mais segura devido aos tipos de veículos e motoristas mais qualificados. Além disso, a aparência do carro e do motorista transmite um nível de profissionalismo e cuidado que contribui para o custo mais elevado.

Esses fatores criam uma experiência de transporte premium, que naturalmente tem um custo mais elevado do que outras categorias mais acessíveis, como o Confort.

1. Gráfico de correlação - Produto e tempo de espera



O gráfico acima é um gráfico de dispersão que mostra a relação entre o produto (categorias de serviços de transporte, como UberX, Confort, Pop 99, Táxi e Black) e o tempo de espera (em minutos).  
 Cada ponto azul representa um tempo de espera para um determinado tipo de serviço. O gráfico mostra que há tempos de espera variados para cada categoria de serviço.  
 O tempo de espera para cada categoria está espalhado de forma semelhante, sugerindo que todas possuem tempos de espera entre 1 e 11 minutos. Não há uma tendência clara de que um serviço tenha tempos de espera consistentemente menores ou maiores que os outros, igualmente como não há uma correlação visível entre o tipo de serviço e o tempo de espera, pois os pontos estão distribuídos de forma semelhante para todas as categorias.