

# EDA - Desafio Lighthouse Ciência de Dados

## Hipóteses de Negócio

### Primeira Hipótese: O valor do hodômetro

A pergunta é: o aumento da quilometragem marcada no hodômetro é inversamente proporcional à média de valores? Para responder, comparei a média de preços dos carros na mesma faixa de quilometragem. Segue o código:

```
valores_hodometro = [0, 20000, 40000, 60000, 80000, 100000, 120000, 140000]

for i in range(len(valores_hodometro)):
    if i < (len(valores_hodometro) - 1):
        nome = 'carros_' + str(round(valores_hodometro[i] / 1000)) + '_' +
str(round(valores_hodometro[i+1] / 1000)) + 'K'
        carros = df_cars_train[(df_cars_train['hodometro'] > valores_hodometro[i])
& (df_cars_train['hodometro'] <=
(valores_hodometro[i+1]))]
    else:
        nome = 'carros_' + str(round(valores_hodometro[i] / 1000)) + 'K_'
        carros = df_cars_train[(df_cars_train['hodometro'] >
valores_hodometro[i])]

    print(nome, ': R$' + str(round(carros['preco'].mean())))
```

O resultado, conforme esperado, mostra que, na medida em que o valor do hodômetro aumenta, o preço do veículo diminui. O mais interessante é que o resultado mostra que a queda do preço é maior quando se trata de valores 'cheios' de quilometragem como, por exemplo, 50.000 e 100.000. Creio que aqui há mais coisas a se pensar. De todo modo, a hipótese foi **confirmada**: carros menos rodados tem preço médio maior que carros muito rodados.

### Segunda Hipótese: Revisões em concessionárias

A pergunta é: o fato de as revisões de um veículo terem sido feitas apenas em concessionárias influencia positivamente seu preço? Para responder, em primeiro lugar, dividi as entradas pelos intervalos de quilometragem no hodômetro e, em segundo, dividi pelo fato ter as revisões na concessionária ou não. Comparei então a média de preços de cada uma das categorias. Segue o código:

```
diferencas_percentuais = []
valores_hodometro = []

for i in range(0, 100000, 20000):
    j = i + 20000
    valores_hodometro.append(i)
    df_revisados = df_cars_train[(df_cars_train['hodometro'] > i) &
```

```

(df_cars_train['hodometro'] <= (j)) &
    (df_cars_train['revisoes_concessionaria'] ==
'Sim'])
    df_nao_revisados = df_cars_train[(df_cars_train['hodometro'] > i) &
(df_cars_train['hodometro'] <= (j)) &
    (df_cars_train['revisoes_concessionaria'] ==
'Não'')]

    diferencas_percentuais.append(round((1 - (df_revisados['preco'].mean() /
df_nao_revisados['preco'].mean())) * -100))

print('### Diferença percentual de preço dos carros revisados APENAS na
concessionários e carros NÃO revisados APENAS na concessionária ###')

for i in range(len(diferencas_percentuais)):
    print('Entre ', valores_hodometro[i], 'e', (valores_hodometro[i] + 20000),
'km: ', str(diferencas_percentuais[i]) + '%')

```

Os resultados mostram que o valor de carros que foram revisados apenas em concessionárias chega a ser 25% maior que carros na mesma faixa de kilometragem mas que não foram revisados apenas em concessionárias. A hipótese foi, poranto, **confirmada**: revisões em concessionária aumentam o preço do carro.

### Terceira Hipótese: Único dono

A pergunta é: o fato de ter sido possuído apenas por um único dono influencia positivamente o preço de um veículo? Para responder, utilizei o 'tipo' do carro como critério e comparei, portanto, o valor médio de preço dos carros de um mesmo tipo que a) possuíam apenas um dono e que b) não possuíam apenas um dono. Segue o código:

```

tipos = df_cars_train['tipo'].unique().tolist()

for tipo in tipos:
    df_unico_dono = df_cars_train[(df_cars_train['veiculo_unico_dono'] == 'Único
dono') & (df_cars_train['tipo'] == tipo)]
    df_multiplos_donos = df_cars_train[(df_cars_train['veiculo_unico_dono'] ==
'Não') & (df_cars_train['tipo'] == tipo)]
    diferenca_percentual = (df_unico_dono['preco'].mean() /
df_multiplos_donos['preco'].mean()) * 100

    if not np.isnan(diferenca_percentual):
        print('Para ' + tipo + ', a média de preço dos carros de um único dono
equivale a '
            + str(round(diferenca_percentual, 2)) + '% do valor médio de preço
dos carros de múltiplos donos.')

```

Entre os 6 tipos de carro, apenas as picapes que possuem um único dono aparecem com valor menor do que as picapes que não possuem um único dono. Para a maior parte dos casos, a hipótese foi **confirmada**: ter

apenas um único dono influencia positivamente o valor do carro.

## Perguntas de Negócio

### Primeira pergunta: Vender um carro de marca popular

Não entendi exatamente se a pergunta se referia a carros populares ou a carros de marcas populares. Resolvi explorar as duas possibilidades. No primeiro caso, escolhi os 10 modelos de carros mais frequentes no conjunto de dados. Somei as médias de preço de cada um dos modelo por estado e tirei novamente a média. Comparei então a média de cada estado. Segue o código:

```
modelos_populares = df_cars_train['modelo'].value_counts().keys().tolist()[0:10]
estados = df_cars_train['estado_vendedor'].value_counts().keys().tolist()

estados_media_populares = []
for estado in estados:
    media_estado = 0
    df_estado = df_cars_train[df_cars_train['estado_vendedor'] == estado]
    for modelo in modelos_populares:
        df_modelo = df_estado[df_estado['modelo'] == modelo]
        media_estado += df_modelo['preco'].mean()
    estados_media_populares.append(media_estado / len(modelos_populares))

print('Maior preço médio de um carro de modelo popular é o do estado de',
      estados[estados_media_populares.index(max(estados_media_populares))],
      ': R$', round(max(estados_media_populares), 2))
```

O estado de Goiás (GO) aparece com a média mais alta, R\$ 158176.07, e é, portanto, o estado mais indicado para se vender um carro popular. Em relação às marcas mais populares, fiz um processo análogo. A única diferença é que não escolhi os 10 modelos mais frequente no conjunto de dados mas as 5 marcas mais frequentes. Segue o código:

```
marcas_populares = df_cars_train['marca'].value_counts().keys().tolist()[0:5]
estados = df_cars_train['estado_vendedor'].value_counts().keys().tolist()

estados_media_populares = []
for estado in estados:
    media_estado = 0
    df_estado = df_cars_train[df_cars_train['estado_vendedor'] == estado]
    for marca in marcas_populares:
        df_marca = df_estado[df_estado['marca'] == marca]
        media_estado += df_marca['preco'].mean()
    estados_media_populares.append(media_estado / len(marcas_populares))

print('Maior preço médio de um carro de marca popular é o do estado de',
      estados[estados_media_populares.index(max(estados_media_populares))],
      ': R$', round(max(estados_media_populares), 2))
```

O estado de Goiás (GO) aparece aqui também com a média mais alta, R\$ 134465.97, e é, portanto, o estado mais indicado para se vender um carro de marca popular.

## Segunda Pergunta: Melhor estado para se comprar picape automática

Para responder essa questão, selecionei as linhas do conjunto de dados que continham 'picapes' na coluna 'tipo' e 'automática' na coluna 'cambio' e comparei a média de preços por estado. Segue o código:

```
picapes = df_cars_train[df_cars_train['tipo'] == 'Picape']
picapes_automaticas = picapes[picapes['cambio'] == 'Automática']

estados_media_picapes_automaticas = []
for estado in estados:
    media_estado = 0
    df_estado = picapes_automaticas[picapes_automaticas['estado_vendedor'] == estado]
    media_estado = df_estado['preco'].mean()
    estados_media_picapes_automaticas.append(media_estado)

print('Menor preço médio de uma picape com transmissão automática é o do estado de', estados[estados_media_picapes_automaticas.index(min(estados_media_picapes_automaticas))], ': R$', round(min(estados_media_picapes_automaticas), 2))
```

O estado da Paraíba (PB) aparece com a menor média de preços, R\$ 93157.04, e é, portanto, o estado mais indicado para se comprar uma picape com transmissão automática.

## Terceira Pergunta: Carros que ainda têm garantia de fábrica

Para responder essa questão, selecionei as linhas do conjunto de dados com os veículos que ainda tinham garantia de fábrica e comparei a média de preços por estado. Segue o código:

```
garantidos = df_cars_train[df_cars_train['garantia_de_fábrica'] == 'Sim']

estados_media_garantidos = []
for estado in estados:
    media_estado = 0
    df_estado = garantidos[garantidos['estado_vendedor'] == estado]
    media_estado = df_estado['preco'].mean()
    estados_media_garantidos.append(media_estado)

print('Menor preço médio de um carro com garantia de fábrica é o do estado de', estados[estados_media_garantidos.index(min(estados_media_garantidos))], ': R$', round(min(estados_media_garantidos), 2))
```

O estado da Paraíba (PB) aparece aqui também com a menor média de preços, R\$ 95762.75, e é, portanto, o estado mais indicado para se comprar uma picape com transmissão automática.