6



Parabéns! Você completou a sessão sobre a análise exploratória de dados. É hora de aplicar o conhecimento e as habilidades que você adquiriu em um projeto: um estudo de caso analítico da vida real que você concluirá por conta própria.

Quando você terminar o projeto, envie seu trabalho para o revisor do projeto para avaliação. Eles te darão feedback dentro de 24 horas. Use o feedback para fazer alterações e, em seguida, envie a nova versão de volta ao revisor do projeto.

Você pode obter mais feedback sobre a nova versão. Isso é completamente normal. Não é incomum passar por vários ciclos de feedback e revisão.

Seu projeto será considerado concluído assim que o revisor do projeto o aprovar.

Descrição do projeto

Você é analista na Lista de Eixo de Manivela. Centenas de propagandas gratuitas de veículos são publicadas no seu site todos os dias. Você precisa estudar os dados coletados nos últimos anos e determinar quais fatores influenciaram o preço de um veículo.

Instruções para completar o projeto

Passo1. Abra o arquivo de dados e estude a informação geral

Caminho do arquivo: /datasets/vehicles_us.csv. Baixar o conjunto de dados

Passo 2. Pré-processamento de dados

- Identifique e estude valores ausentes:
 - Em alguns casos há uma maneira óbvia de substituir valores ausentes. Por exemplo, se um campo booleano contém apenas valores True, é razoável presumir que valores ausentes são False . Não há correções tão óbvias para outros tipos de dados, e há casos quando o fato de que um valor está ausente é significativo. Em tais instâncias, não preencha os valores.
 - fazer isso e como você selecionou os valores substitutivos.

Quando for apropriado, preencha os valores. Explique porque você escolheu

- Descreva os fatores que podem ter resultado em valores ausentes.
- Indique as colunas nas quais os tipos de dados precisam ser mudados e

Converta os dados para os tipos necessários:

explique por quê.

Dia da semana, mês, e ano que a propaganda foi colocada

Passo 3. Calcule e adicione à tabela o seguinte:

- A idade do veículo (em anos) quando a propaganda foi colocada
- A média de quilometragem por ano
- Na coluna | condition |, substitua valores de string por uma escala numérica:

novo = 5

- como novo = 4
- excellente = 3 bom = 2
- razoável = 1
- sucata = 0

as instruções abaixo: Estude os seguintes parâmetros: preço, idade do veículo quando a

Passo 4. Realize uma análise exploratória de dados, seguindo

- propaganda foi colocada, quilometragem, número de cilindros, e condição. Construa histogramas para cada um desses parâmetros. Estude como valores atípicos afetam a forma e a legibilidade dos histogramas. Determine os limites superiores de valores atípicos, remova os valores
- atípicos, armazene-os em um DataFrame separado e continue seu trabalho com os dados filtrados. Use os dados filtrados para construir novos histogramas. Compare-os com os histogramas anteriores (aqueles que incluem valores atípicos). Tire conclusões
- para cada histograma. Estude quantos dias as propagandas foram exibidas (days_listed). Construa um histograma. Calcule a média e a mediana. Descreva o tempo de vida útil comum de uma propaganda. Determine quando as propagandas foram

removidas rapidamente, e quando elas foram listadas por um tempo

Construa um gráfico mostrando a dependência do número de propagandas em relação ao tipo de veículo. Selecione os dois tipos com o maior número de propagandas. Que fatores mais influenciam o preço? Pegue cada um dos tipos populares que você detectou no estágio anterior e estude se o preço depende da idade,

Analise o número de propagandas e o preço médio para cada tipo de veículo.

quilometragem, condição, tipo de transmissão e cor. Para variáveis categóricas (tipo de transmissão e cor), construa gráficos de extremos e quartis, e crie gráficos de dispersão para o restante. Ao analisar variáveis categóricas, note que as categorias precisam ter pelo menos 50 propagandas; de outro modo, seus parâmetros não serão válidos para análise. Passo 5. Escreva uma conclusão geral

Formato: Complete a tarefa em um notebook Jupyter. Coloque seu código nas células de código e suas explicações de texto em células Markdown, então aplique a formatação e os cabeçalhos.

O conjunto de dados contém os seguintes campos:

Descrição dos dados

price model_year

anormalmente longo.

- condition
- cylinders

model

- fuel gasolina, diesel etc. odometer – a quilometragem do veículo quando a propaganda foi publicada
- transmission paint_color

começar a resolver o caso.

is_4wd — Se o veículo é 4 por 4 (tipo Booleano) date_posted — a data que a propaganda foi publicada

days_listed — dias desde a publicação até a retirada

Como meu projeto será avaliado?

Isso é o que os revisores do projeto procuram ao avaliar seu projeto:

Como você explica os problemas identificados nos dados Que métodos você usou para processar valores ausentes

Nós reunimos alguns critérios de avaliação. Leia-os cuidadosamente antes de

Se seus gráficos resolvem ou não tipos específicos de tarefas

Como você usa fatias de dados

- Que métodos você utiliza para construir gráficos Se você automatizou ou não a construção de gráficos
- Se você calcula figuras de correlação para os dados, e como você as explica
- Se você segue a estrutura do projeto e mantém o código organizado Suas conclusões

Se você deixa comentários a cada passo As folhas de conclusões e resumos das lições anteriores têm tudo o que você

precisa para completar o projeto. Boa sorte!

Avançar