

Descripción del proyecto

Eres un analista en Crankshaft List. Cientos de anuncios gratuitos de vehículos se publican en tu sitio web cada día. Necesitas estudiar los datos recopilados durante los últimos años y determinar qué factores influyen en el precio de un vehículo.

Instrucciones para completar el proyecto

Paso 1. Archivo de datos e información general

Ruta de archivo: /datasets/vehicles_us.csv

Paso 2. Preprocesamiento de datos

- **Identificar y estudiar los valores ausentes:**
 - En algunos casos hay una forma obvia de reemplazar valores ausentes. Por ejemplo, si un campo booleano contiene solo valores `True`, es razonable asumir que los valores ausentes son `False`. Para otros tipos de datos no hay arreglos tan obvios, y hay casos en los que el hecho de que haya un valor ausente es significativo. En tales casos, no sustituyas los valores.
 - Cuando sea apropiado, sustituye los valores. Explica por qué has decidido hacerlo así y cómo has seleccionado los valores sustitutos.
 - Describe los factores que puedan haber resultado en valores ausentes.
- **Convierte los datos en los tipos necesarios:**
 - Indica las columnas donde los tipos de datos necesitan ser cambiados y explica por qué.

Paso 3. Calcula y añade a la tabla lo siguiente:

- Día de la semana, mes y año en el que el anuncio se colocó
- Los años del vehículo cuando el anuncio se colocó

- La media de millaje del vehículo por año

En la columna `condition`, reemplaza los valores de cadena con una escala numérica:

- nuevo = 5
- como nuevo = 4
- excelente = 3
- bien = 2
- bastante = 1
- para rescate = 0

Paso 4. Lleva a cabo el análisis exploratorio de datos siguiendo las siguientes instrucciones:

- Estudia los siguientes parámetros: precio, años del vehículo cuando se colocó el anuncio, millaje, número de cilindrada y condición. Traza histogramas para cada uno de los parámetros. Estudia cómo los valores atípicos afectan a la forma y legitimidad de los histogramas.
- Determina los límites superiores de los valores atípicos, elimina dichos valores y almacénalos en un DataFrame apartado, y continúa tu trabajo con los datos filtrados.
- Utiliza los datos filtrados para plantear nuevos histogramas. Compáralos con los histogramas anteriores (aquellos con los valores atípicos incluidos). Obtén conclusiones de cada histograma.
- Estudia cuántos días los anuncios fueron mostrados (`days_listed`). Traza un histograma. Calcula la media y la mediana. Describe la vida útil habitual de un anuncio. Determina cuándo se eliminan rápidamente los anuncios y cuándo son publicados por un tiempo anormalmente largo.
- Analiza el número de anuncios y el precio medio para cada tipo de vehículo. Traza un gráfico mostrando la dependencia de los números de anuncios en cada tipo de vehículo. Selecciona los dos tipos con un mayor número de anuncios.
- ¿Qué factores impactan más sobre el precio? Toma cada uno de los tipos más populares que has detectado en la fase anterior y estudia si el precio depende de la edad, millaje, condición, tipo de transmisión y color. Para las variables categóricas (tipo de transmisión y color), traza

gráficos de caja y bigotes, y crea gráficos de dispersión para el resto. Cuando analices variables categóricas, observa que las categorías deben tener al menos 50 anuncios; si no, sus parámetros no serán válidos para el análisis.

Paso 5. Escribe una conclusión general

Formato: Completa la tarea en un Jupyter notebook. Escribe tu código en las celdas de código y tus explicaciones en las celdas markdown, después aplica los formatos y encabezados.

Descripción de los datos

El conjunto de datos contiene los siguientes datos:

- `price`
- `model_year`
- `model`
- `condition`
- `cylinders`
- `fuel` — gasolina, diesel, etc.
- `odometer` — el millaje del vehículo cuando el anuncio fue publicado
- `transmission`
- `paint_color`
- `is_4wd` — si el vehículo tiene tracción a las 4 ruedas (tipo Booleano)
- `date_posted` — la fecha en la que el anuncio fue publicado
- `days_listed` — desde la publicación hasta que se elimina

¿Cómo será evaluado mi proyecto?

Hemos recopilado los criterios de evaluación del proyecto. Léelos cuidadosamente antes de empezar a resolver tu caso.

Esto es lo que buscan los revisores de proyecto cuando evalúan tu proyecto:

- Cómo explicas los problemas identificados en los datos
- Qué métodos utilizaste para procesar valores ausentes
- Cómo utilizas segmentos de datos
- Si tus gráficos resuelven o no tareas específicas
- Qué métodos utilizas para trazar gráficos

- Si has automatizado o no el trazado de gráficos
- Si calculas las cifras de correlación de los datos y cómo las explicas
- Si sigues la estructura del proyecto y mantienes el código ordenado
- Tus conclusiones
- Si dejas comentarios en cada paso

Las hojas informativas y resúmenes de las lecciones anteriores tienen todo lo que necesitas para completar el proyecto.

¡Buena suerte!