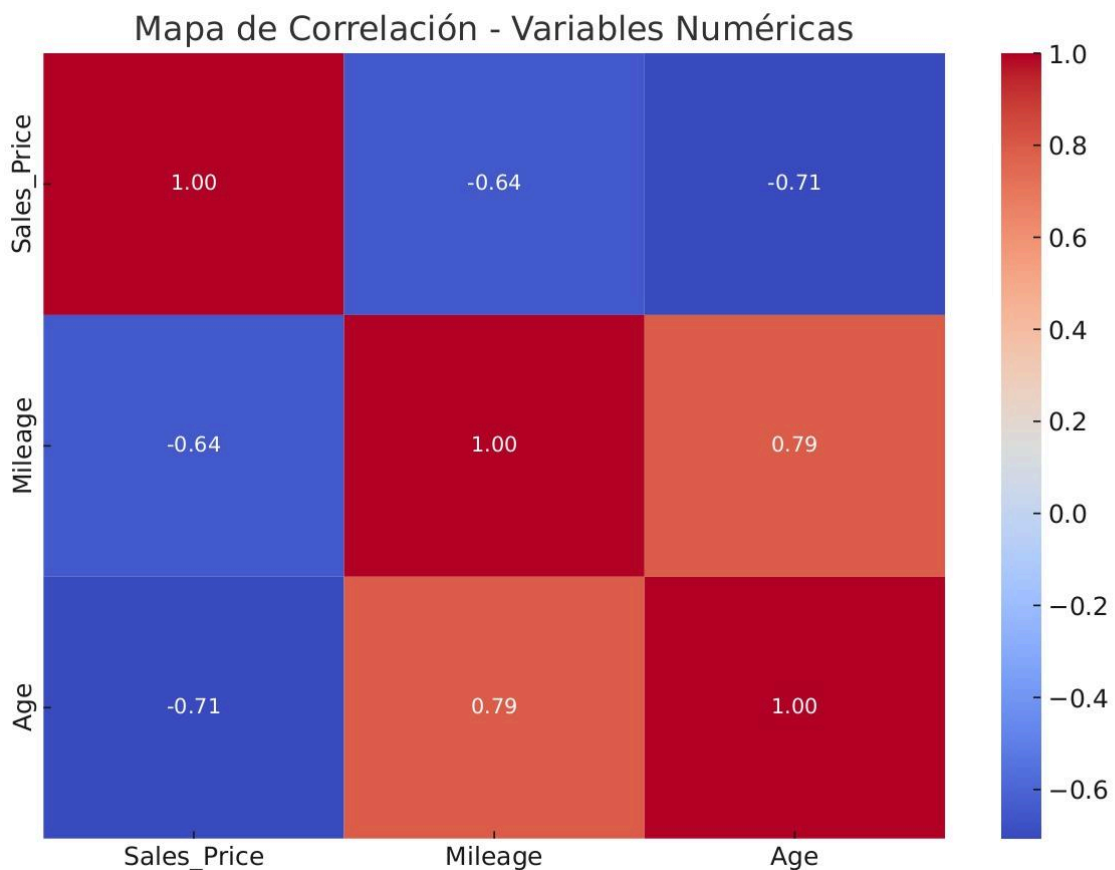


Descripción de Gráficas del EDA

1. Mapa de Correlación - Variables Numéricas

Este gráfico muestra la matriz de correlaciones entre tres variables clave: Sales_Price (precio de venta), Mileage (kilometraje) y Age (antigüedad del vehículo).

- Existe fuerte correlación negativa entre Sales_Price y Age (-0.71) y entre Sales_Price y Mileage (-0.64), lo cual es coherente:
→ A mayor antigüedad o uso, menor precio de mercado.
- Mileage y Age también están fuertemente correlacionados (+0.79), lo que indica redundancia parcial.

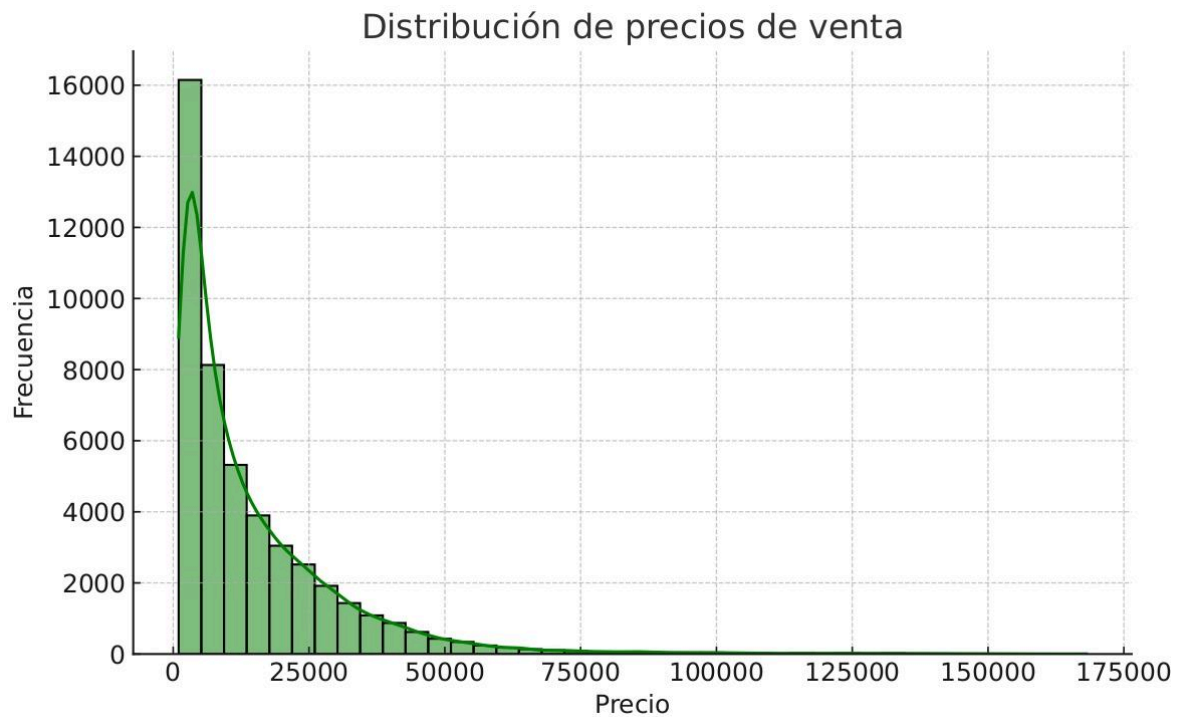


2. Distribución de Precios de Venta

Histograma con KDE (estimación de densidad) que muestra cómo se distribuyen los precios de los autos usados.

- Distribución sesgada a la derecha (right-skewed), lo que implica presencia de valores atípicos (vehículos de lujo o errores).

- Sugerencia: aplicar logaritmo a Sales_Price en etapas futuras del modelo.



3. Boxplot del Precio de Venta

Visualización de los valores mínimos, máximos, cuartiles y outliers del precio.

- Muestra claramente que hay outliers hacia la derecha (vehículos costosos).
- Es útil para hacer limpieza previa al modelado y decidir estrategias de normalización.

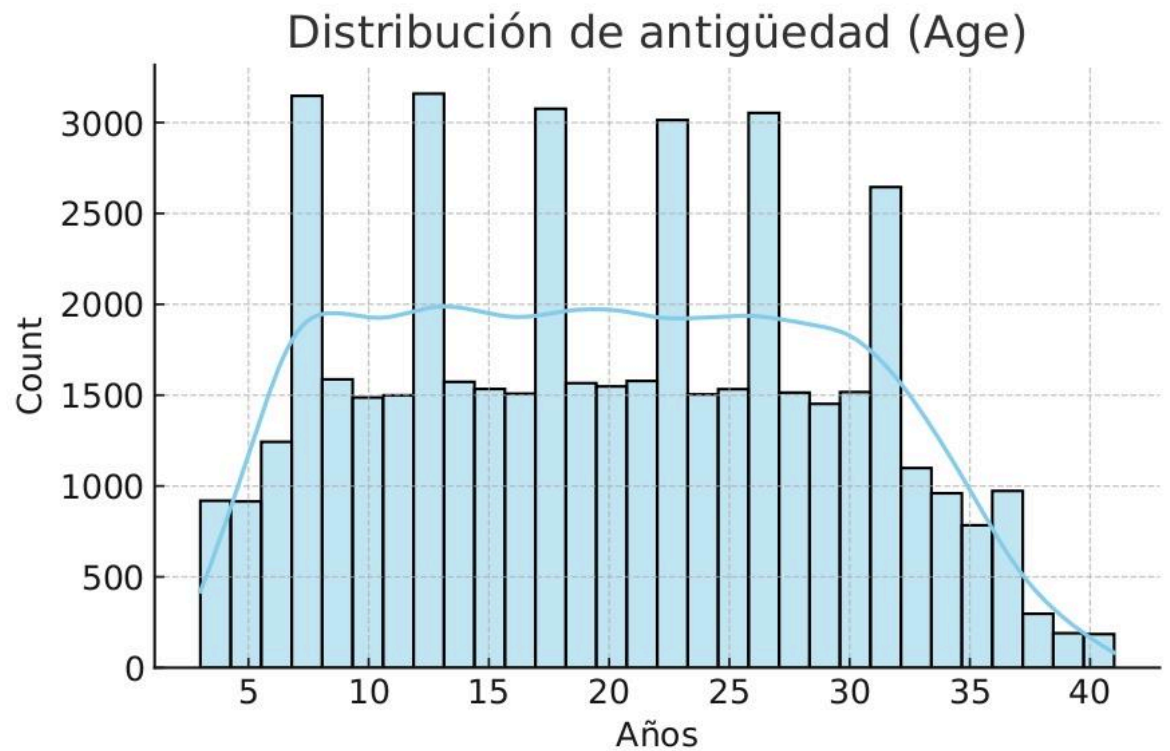


4. Distribución de Antigüedad (Age)

Histograma de la variable Age, calculada como 2025 - Year of manufacture.

- La mayoría de los autos tienen entre 5 y 15 años, con muy pocos extremos (0 o más de 30).

- Apoya la decisión de incluir Age como feature numérica clave.

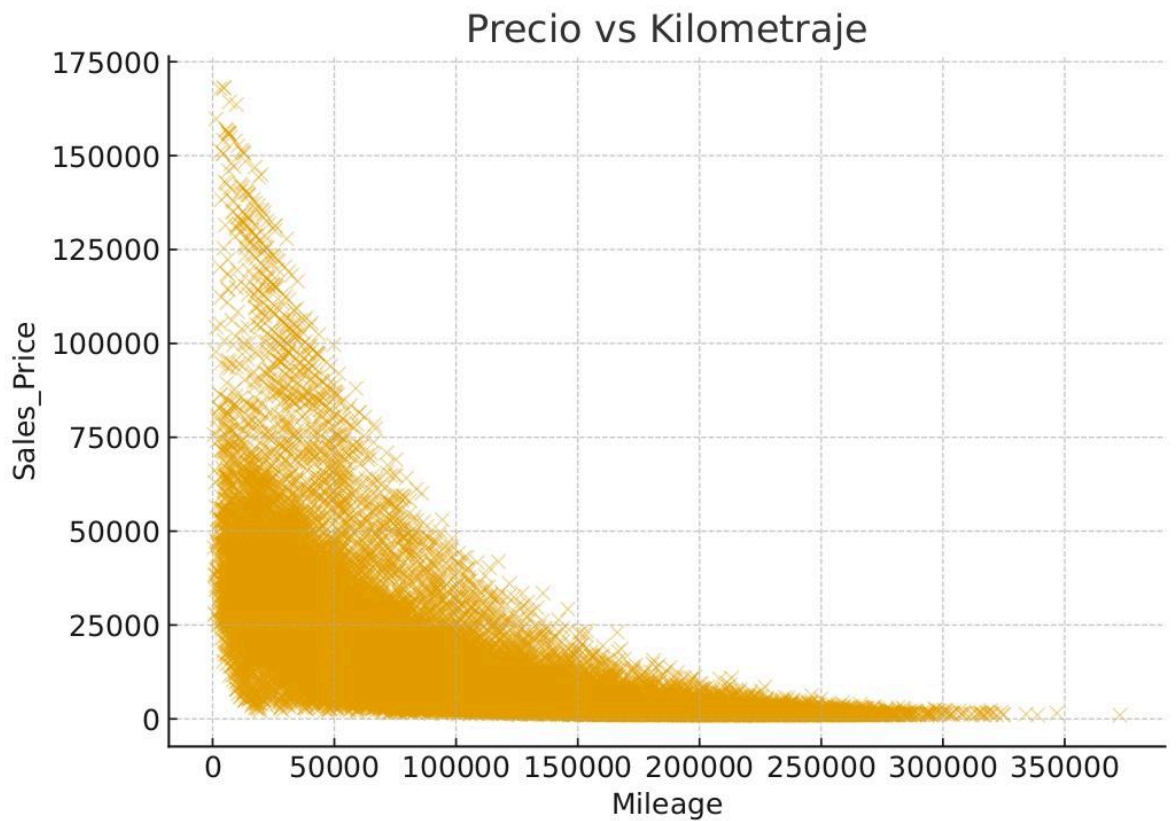


5. Precio vs Kilometraje (Scatterplot)

Gráfico de dispersión entre Mileage y Sales_Price.

- Muestra tendencia decreciente clara: más kilometraje → menor valor.

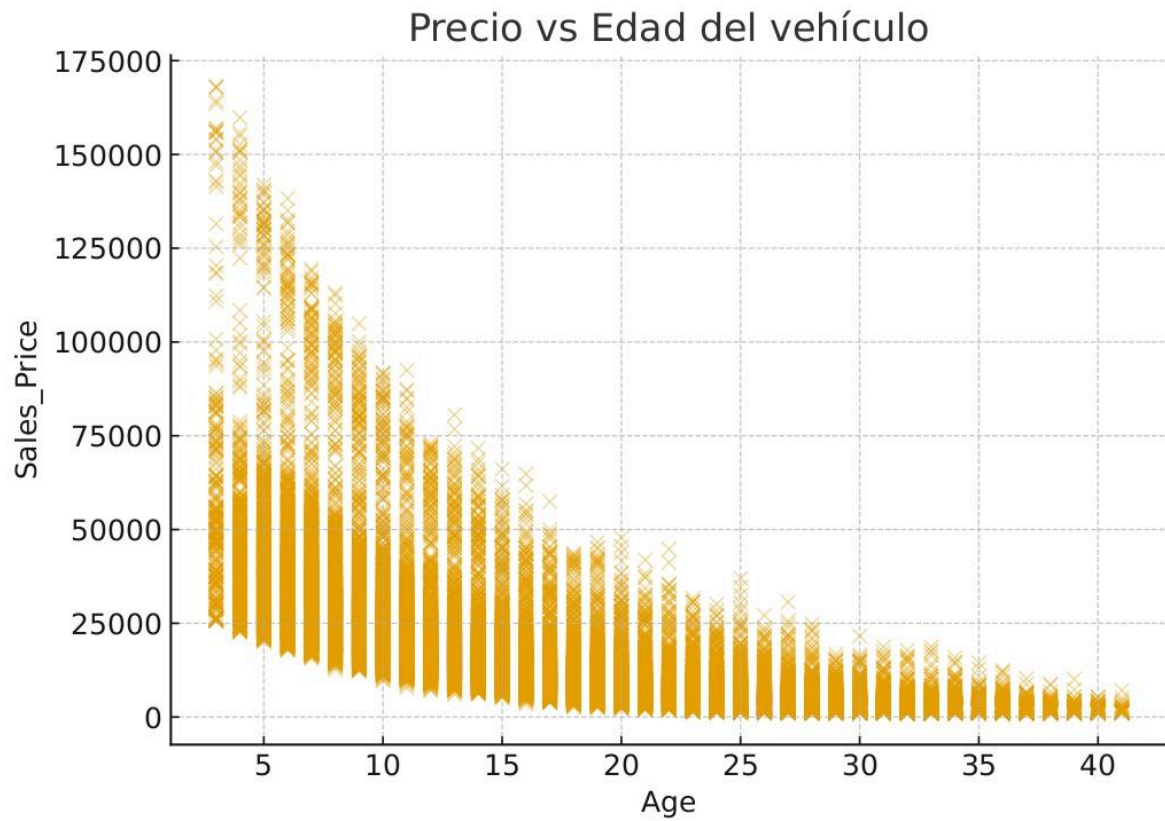
- La dispersión sugiere que esta relación no es perfectamente lineal (ideal para modelos tipo árbol).



6. Precio vs Edad del Vehículo (Scatterplot)

Gráfico de dispersión entre Age y Sales_Price.

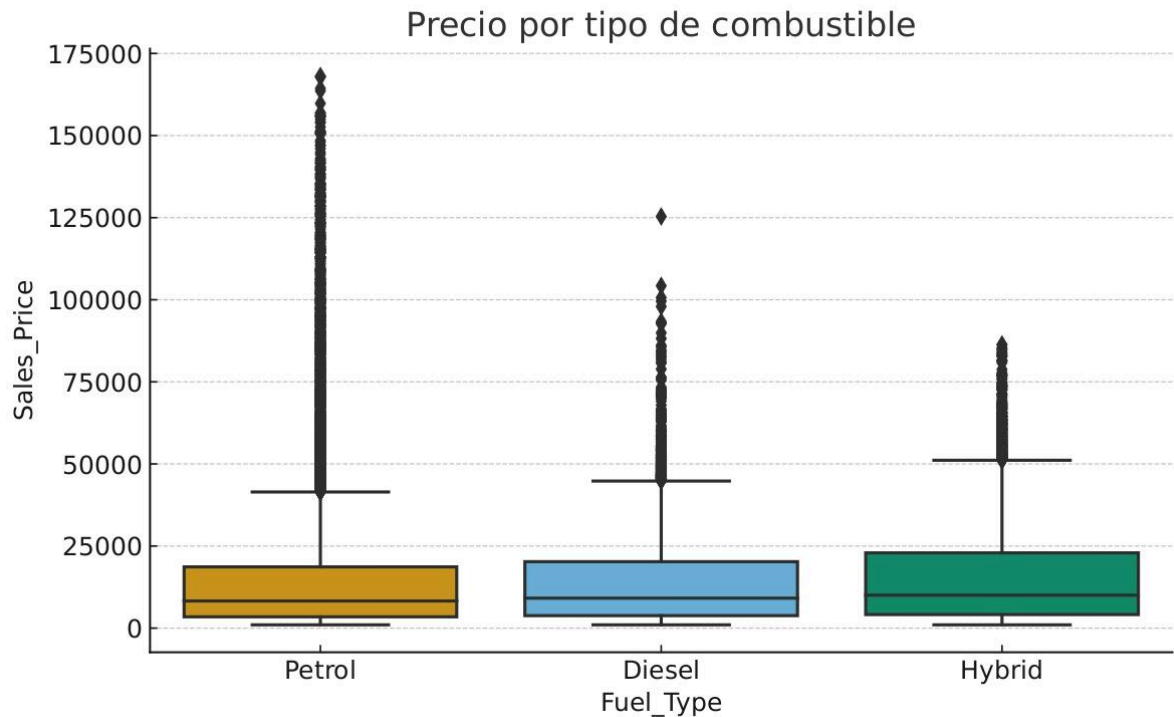
- Relación negativa esperada: autos más viejos tienen menor precio.
- También se observa alta variabilidad (autos de la misma edad con precios muy diferentes).



7. Precio por Tipo de Combustible (Boxplot)

Comparación del precio entre autos según su tipo de combustible (Fuel_Type).

- Se observa que los vehículos diésel e híbridos tienden a tener un precio promedio más alto que los de gasolina (petrol).
- Es una variable categórica importante para el modelo.



8. Precio por Marca - Top 10 (Boxplot)

Boxplot para las 10 marcas con mayor número de registros (Make).

- Se aprecian diferencias notables en los rangos de precios entre marcas.
- Por ejemplo: Porsche y BMW tienen precios mucho más altos que Ford o VW.
- Es útil para codificar correctamente esta variable al modelar (e.g. OneHotEncoding o Target Encoding).

