

 Hi, I'm Ashish Mishra

 Internship in AI/ML

 at ShadowFox

Car selling price prediction and analysis

In [1]: `import pandas as pd`

In [2]: `df = pd.read_csv(r"C:\Users\Ashish Mishra\OneDrive\Desktop\ShadowFox\Shadowfox_AI-M`

In [3]: `df.head()`

Out[3]:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Trans
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	

In [4]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 301 entries, 0 to 300
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Car_Name        301 non-null    object
1   Year            301 non-null    int64
2   Selling_Price   301 non-null    float64
3   Present_Price   301 non-null    float64
4   Kms_Driven      301 non-null    int64
5   Fuel_Type       301 non-null    object
6   Seller_Type     301 non-null    object
7   Transmission    301 non-null    object
8   Owner           301 non-null    int64
dtypes: float64(2), int64(3), object(4)
memory usage: 21.3+ KB
```

In [5]: `df.describe()`

Out[5]:

	Year	Selling_Price	Present_Price	Kms_Driven	Owner
count	301.000000	301.000000	301.000000	301.000000	301.000000
mean	2013.627907	4.661296	7.628472	36947.205980	0.043189
std	2.891554	5.082812	8.644115	38886.883882	0.247915
min	2003.000000	0.100000	0.320000	500.000000	0.000000
25%	2012.000000	0.900000	1.200000	15000.000000	0.000000
50%	2014.000000	3.600000	6.400000	32000.000000	0.000000
75%	2016.000000	6.000000	9.900000	48767.000000	0.000000
max	2018.000000	35.000000	92.600000	500000.000000	3.000000

In [6]:

```
from datetime import datetime

current_year = datetime.now().year

df['Years_Used'] = current_year - df['Year']
df = df.drop(['Car_Name', 'Year'], axis=1)

df.head()
```

Out[6]:

	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	4.60	6.87	42450	Diesel	Dealer	Manual	0

In []:

```
import seaborn as sns

import matplotlib.pyplot as plt

sns.set(style="whitegrid")

fig, axes = plt.subplots(2, 2, figsize=(14, 10))

sns.histplot(df['Selling_Price'], kde=True, ax=axes[0, 0], color="blue")
axes[0, 0].set_title("Distribution of Selling Price")

sns.histplot(df['Present_Price'], kde=True, ax=axes[0, 1], color="green")
axes[0, 1].set_title("Distribution of Present Price")

sns.histplot(df['Kms_Driven'], kde=True, ax=axes[1, 0], color="purple")
axes[1, 0].set_title("Distribution of Kms Driven")
```

```

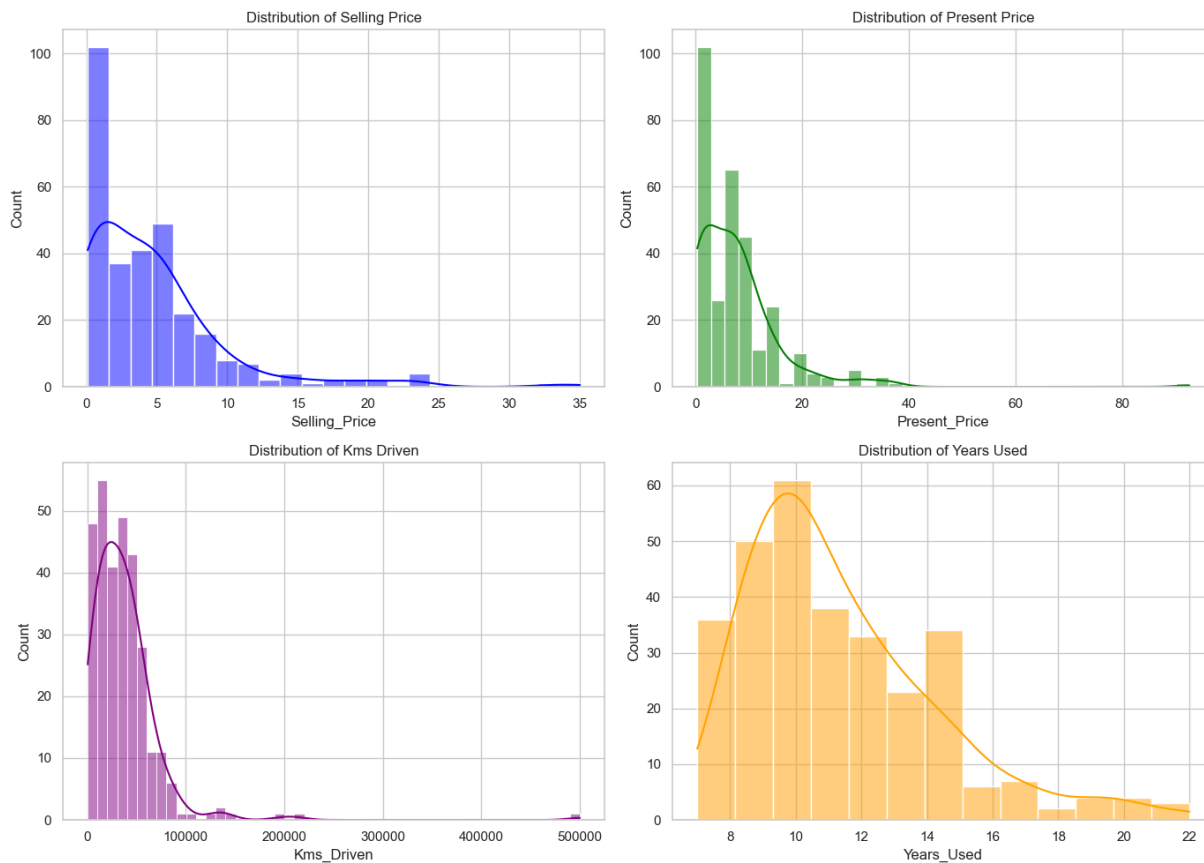
sns.histplot(df['Years_Used'], kde=True, ax=axes[1, 1], color="orange")
axes[1, 1].set_title("Distribution of Years Used")

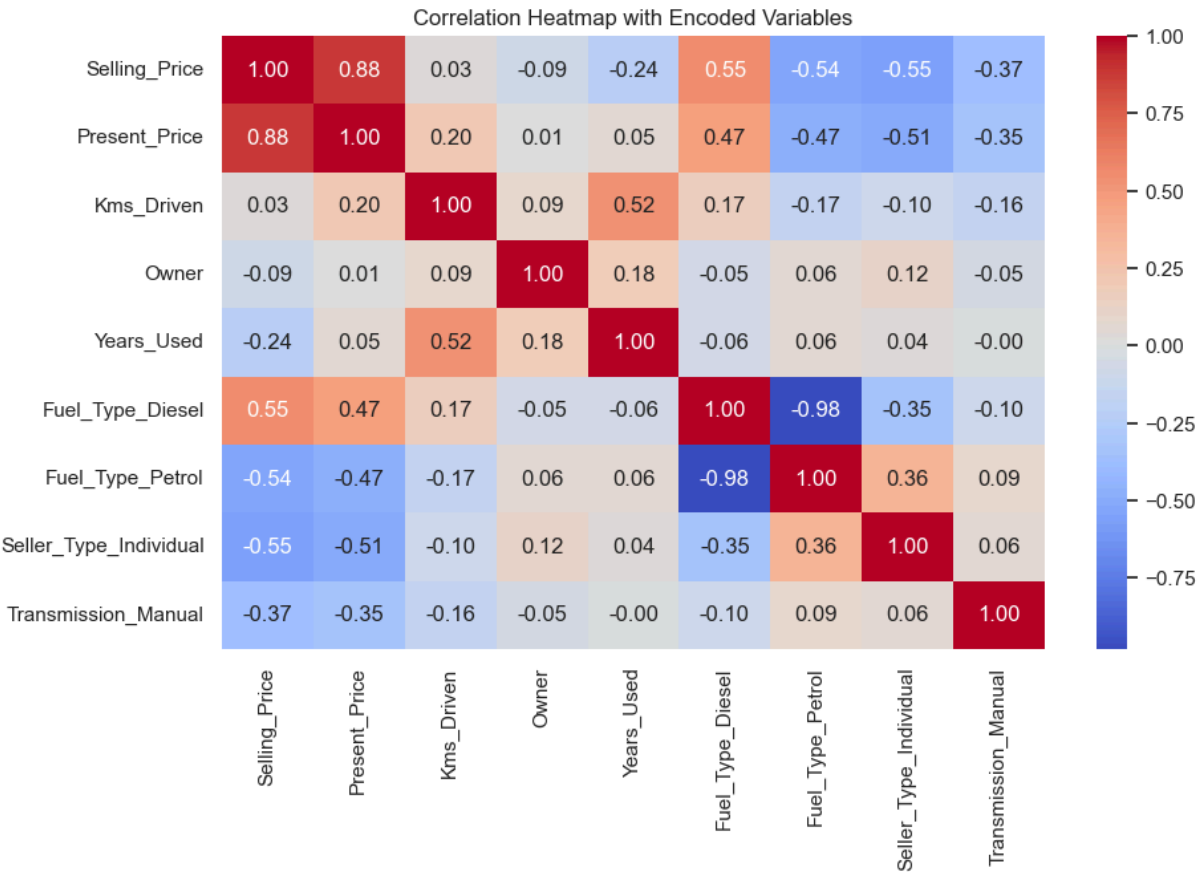
plt.tight_layout()
plt.show()

df_encoded = pd.get_dummies(df, columns=['Fuel_Type', 'Seller_Type', 'Transmission'])

plt.figure(figsize=(10, 6))
sns.heatmap(df_encoded.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap with Encoded Variables")
plt.show()

```





```
In [ ]: encoded_data = pd.get_dummies(df, columns=['Fuel_Type', 'Seller_Type', 'Transmission_Manual'])
encoded_data.head()
```

Out[]:

	Selling_Price	Present_Price	Kms_Driven	Owner	Years_Used	Fuel_Type_Diesel	Fuel_Type_Petrol
0	3.35	5.59	27000	0	11	False	True
1	4.75	9.54	43000	0	12	True	False
2	7.25	9.85	6900	0	8	False	True
3	2.85	4.15	5200	0	14	False	True
4	4.60	6.87	42450	0	11	True	False