

In [1]: *#Import the necessary libraries: You'll need to import libraries such as pandas, numpy, seaborn, and matplotlib to load and visualize the data.*

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]: *#Load the data: Load the dataset using pandas read_csv() function.*

```
df = pd.read_csv('sports_cars_dataset.csv')
```

In [3]: *#Check the shape of the data: Use the shape attribute to check the number of rows and columns in the dataset.*

```
print(df.shape)
```

```
(713, 8)
```

In [4]: *#Preview the data: Use the head() function to preview the first few rows of the dataset.*

```
print(df.head())
```

	Car Make	Car Model	Year	Engine Size (L)	Horsepower	Torque (lb-ft)	\
0	Porsche	911	2022	3	379	331	
1	Lamborghini	Huracan	2021	5.2	630	443	
2	Ferrari	488 GTB	2022	3.9	661	561	
3	Mercedes-Benz	AMG GT	2021	4	523	494	
4	Chevrolet	Corvette	2021	6.2	490	465	

	0-60 MPH Time (seconds)	Price (in USD)
0	4	101,200
1	2.8	274,390
2	3	333,750
3	3.8	118,500
4	2.8	59,900

In [5]: *#Check for missing values: Use the isnull() function to check for missing values in the dataset.*

```
print(df.isnull().sum())
```

```
Car Make      0
Car Model     0
Year          0
Engine Size (L)  10
Horsepower    0
Torque (lb-ft)  3
0-60 MPH Time (seconds)  0
Price (in USD)  0
dtype: int64
```

In [6]: *#Check data types: Use the dtypes attribute to check the data types of each column.*

```
print(df.dtypes)
```

Car Make	object
Car Model	object
Year	int64
Engine Size (L)	object
Horsepower	object
Torque (lb-ft)	object
0-60 MPH Time (seconds)	object
Price (in USD)	object
dtype:	object

In [7]: *#Check for duplicates: Use the duplicated() function to check for duplicate rows in the dataset.*

```
print(df.duplicated().sum())
```

180

In [8]: *#Descriptive statistics: Use the describe() function to get descriptive statistics for the numeric columns in the dataset.*

```
print(df.describe())
```

	Year
count	713.000000
mean	2021.082749
std	2.363141
min	1965.000000
25%	2021.000000
50%	2021.000000
75%	2022.000000
max	2023.000000

In [9]: *# Replace commas with empty strings in Price (in USD) column*

```
df['Price (in USD)'] = df['Price (in USD)'].str.replace(',', '')
```

In [10]: *#The errors='coerce' parameter tells the function to convert any non-numeric values to NaN (Not a Number) values, which can be safely ignored or removed later on.*

```
df['Price (in USD)'] = pd.to_numeric(df['Price (in USD)'], errors='coerce')
```

In [11]: *# Use visualization libraries such as seaborn and matplotlib to create various visualizations to explore the relationships between variables and identify patterns and trends in the data.*

```
sns.pairplot(df)
plt.show()

sns.histplot(df['Price (in USD)'], bins=20)
plt.show()

sns.boxplot(x='Car Make', y='Price (in USD)', data=df)
plt.show()

sns.scatterplot(x='Horsepower', y='Price (in USD)', data=df)
plt.show()
```



