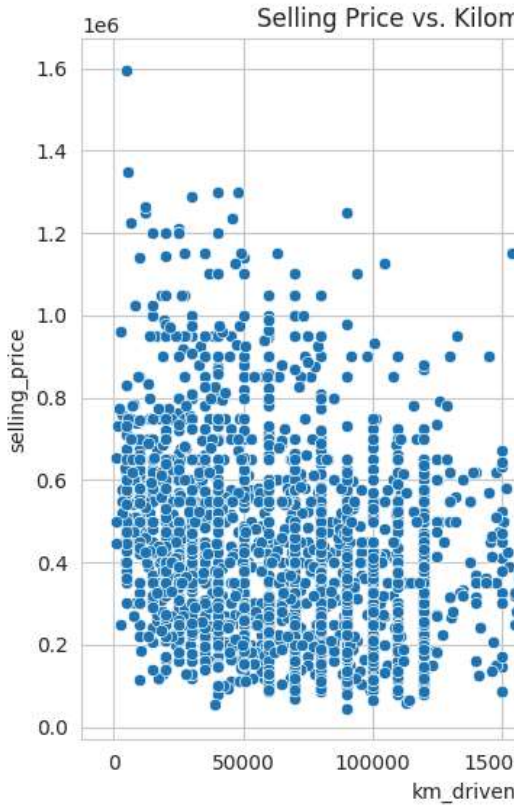
```
import pandas as pd

df_clean = pd.read_csv("processes2.csv")
```

```

# Calculate descriptive statistics for the dataset
desc_stats = data.describe()

# now to Visualize Data - We will select a few key variables for visualizationand
#focus on 'selling_price', 'km_driven', and 'year' as key variables of interest.

# Import  visualization scematics
import matplotlib.pyplot as plt
import seaborn as sns

# Seting the any  style of the plots
sns.set_style("whitegrid")

plt.figure(figsize=(15, 5))
fig, ax = plt.subplots(1, 3, figsize=(18, 6))

# Histogram of selling prices as states well focus on the main variables.
sns.histplot(df_clean['selling_price'], bins=30, kde=True, ax=ax[0])
ax[0].set_title('Distribution of Selling Prices')

# Scatter plot of km_driven vs. selling_price Scatter plot helps pin point outliers
sns.scatterplot(x='km_driven', y='selling_price', data=df_clean, ax=ax[1])
ax[1].set_title('Selling Price vs. Kilometers Driven')

# Box plot of selling prices by year
sns.boxplot(x='year', y='selling_price', data=data, ax=ax[2])
ax[2].set_title('Selling Prices by Year')
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability

# Display the plots
plt.tight_layout()
plt.show()

# Display descriptive statistics ( this ties into the first code)
desc_stats

```

```
<Figure size 1500x500 with 0 Axes>
```



	Unnamed: 0	year	selling_price	km_driven	seats	max_power (in bph)	Mileage	Engine (CC)
count	2095.000000	2095.000000	2.095000e+03	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000
mean	3155.537947	2014.052506	4.272799e+05	67413.593795	5.096897	77.410427	21.397928	1205.078282
std	1791.039215	3.273451	2.326698e+05	41580.846333	0.457000	19.884938	1.209234	245.058696
min	0.000000	2000.000000	4.595700e+04	1000.000000	4.000000	35.000000	19.600000	624.000000
25%	1589.500000	2012.000000	2.500000e+05	35000.000000	5.000000	67.040000	20.360000	998.000000
50%	3185.000000	2014.000000	3.900000e+05	60000.000000	5.000000	74.000000	21.100000	1248.000000