In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
df=pd.read_csv("Desktop/car sales.csv")
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 558837 entries, 0 to 558836
Data columns (total 16 columns):
 #   Column        Non-Null Count   Dtype  
---  ------        --------------   -----  
 0   year          558837 non-null  int64  
 1   make          548536 non-null  object 
 2   model         548438 non-null  object 
 3   trim          548186 non-null  object 
 4   body          545642 non-null  object 
 5   transmission  493485 non-null  object 
 6   vin           558833 non-null  object 
 7   state         558837 non-null  object 
 8   condition     547017 non-null  float64
 9   odometer      558743 non-null  float64
 10  color         558088 non-null  object 
 11  interior      558088 non-null  object 
 12  seller        558837 non-null  object 
 13  mmr           558799 non-null  float64
 14  sellingprice  558825 non-null  float64
 15  saledate      558825 non-null  object 
dtypes: float64(4), int64(1), object(11)
memory usage: 68.2+ MB
```

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

Out[4]:

	year	make	model	trim	body	transmission	vin	state	condition	odometer
0	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg566472	ca	5.0	16639.0
1	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg561319	ca	5.0	9393.0
2	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3c1c51ek116351	ca	45.0	1331.0
3	2015	Volvo	S60	T5	Sedan	automatic	yv1612tb4f1310987	ca	41.0	14282.0
4	2014	BMW	6 Series Gran Coupe	650i	Sedan	automatic	wba6b2c57ed129731	ca	43.0	2641.0

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

Out[5]:

	year	make	model	trim	body	transmission	vin	state	condition
558832	2015	Kia	K900	Luxury	Sedan	NaN	knalw4d4xf6019304	in	45.0
558833	2012	Ram	2500	Power Wagon	Crew Cab	automatic	3c6td5et6cg112407	wa	5.0
558834	2012	BMW	X5	xDrive35d	SUV	automatic	5uxzw0c58cl668465	ca	48.0
558835	2015	Nissan	Altima	2.5 S	sedan	automatic	1n4al3ap0fc216050	ga	38.0
558836	2014	Ford	F-150	XLT	SuperCrew	automatic	1ftfw1et2eke87277	ca	34.0

In [6]:
```
df.shape
```

Out[6]:
```
(558837, 16)
```

In [7]:
```
df.columns
```

Out[7]:
```
Index(['year', 'make', 'model', 'trim', 'body', 'transmission', 'vin', 'state',
       'condition', 'odometer', 'color', 'interior', 'seller', 'mmr',
       'sellingprice', 'saledate'],
      dtype='object')
```

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
year              0
make          10301
model         10399
trim          10651
body          13195
transmission  65352
vin               4
state             0
condition     11820
odometer         94
color           749
interior        749
seller            0
mmr              38
sellingprice     12
saledate         12
dtype: int64
```

In [9]:
```
df.dropna()
```

Out[9]:

	year	make	model	trim	body	transmission	vin	state	conditi
0	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg566472	ca	
1	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg561319	ca	
2	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3c1c51ek116351	ca	4
3	2015	Volvo	S60	T5	Sedan	automatic	yv1612tb4f1310987	ca	4
4	2014	BMW	6 Series Gran Coupe	650i	Sedan	automatic	wba6b2c57ed129731	ca	4
...	...	...	...	...	...	...	...	...	
558831	2011	BMW	5 Series	528i	Sedan	automatic	wbafr1c53bc744672	fl	3
558833	2012	Ram	2500	Power Wagon	Crew Cab	automatic	3c6td5et6cg112407	wa	
558834	2012	BMW	X5	xDrive35d	SUV	automatic	5uxzw0c58cl668465	ca	4
558835	2015	Nissan	Altima	2.5 S	sedan	automatic	1n4al3ap0fc216050	ga	3

	year	make	model	trim	body	transmission	vin	state	conditi
558836	2014	Ford	F-150	XLT	SuperCrew	automatic	1ftfw1et2eke87277	ca	3

In [10]:
```
df.drop_duplicates()
```

Out[10]:

	year	make	model	trim	body	transmission	vin	state	conditi
0	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg566472	ca	
1	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg561319	ca	
2	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3c1c51ek116351	ca	4
3	2015	Volvo	S60	T5	Sedan	automatic	yv1612tb4f1310987	ca	4
4	2014	BMW	6 Series Gran Coupe	650i	Sedan	automatic	wba6b2c57ed129731	ca	4
...	...	...	...	...	...	...	...	...	
558832	2015	Kia	K900	Luxury	Sedan	NaN	knalw4d4xf6019304	in	4
558833	2012	Ram	2500	Power Wagon	Crew Cab	automatic	3c6td5et6cg112407	wa	
558834	2012	BMW	X5	xDrive35d	SUV	automatic	5uxzw0c58cl668465	ca	4
558835	2015	Nissan	Altima	2.5 S	sedan	automatic	1n4al3ap0fc216050	ga	3

	year	make	model	trim	body	transmission	vin	state	conditi
558836	2014	Ford	F-150	XLT	SuperCrew	automatic	1ftfw1et2eke87277	ca	3

In [11]:
```
df.describe()
```

Out[11]:

	year	condition	odometer	mmr	sellingprice
count	558837.000000	547017.000000	558743.000000	558799.000000	558825.000000
mean	2010.038927	30.672365	68320.017767	13769.377495	13611.358810
std	3.966864	13.402832	53398.542821	9679.967174	9749.501628
min	1982.000000	1.000000	1.000000	25.000000	1.000000
25%	2007.000000	23.000000	28371.000000	7100.000000	6900.000000
50%	2012.000000	35.000000	52254.000000	12250.000000	12100.000000
75%	2013.000000	42.000000	99109.000000	18300.000000	18200.000000
max	2015.000000	49.000000	999999.000000	182000.000000	230000.000000

In [12]:
```
df.columns
```

Out[12]:
```
Index(['year', 'make', 'model', 'trim', 'body', 'transmission', 'vin', 'state',
       'condition', 'odometer', 'color', 'interior', 'seller', 'mmr',
       'sellingprice', 'saledate'],
      dtype='object')
```

In [23]:
```
df.clip()
```

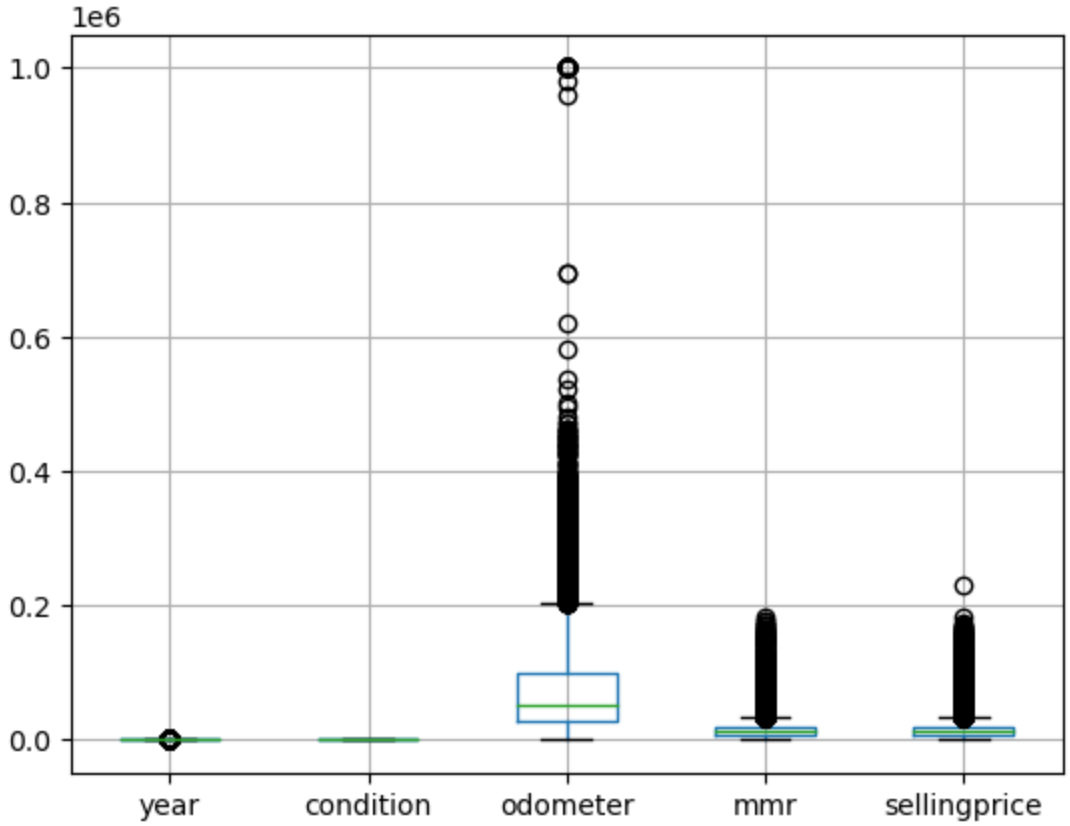
Out[23]:

	year	make	model	trim	body	transmission	vin	state	conditi
0	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg566472	ca	
1	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg561319	ca	
2	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3c1c51ek116351	ca	4
3	2015	Volvo	S60	T5	Sedan	automatic	yv1612tb4f1310987	ca	4
4	2014	BMW	6 Series Gran Coupe	650i	Sedan	automatic	wba6b2c57ed129731	ca	4
...	...	...	...	...	...	...	...	...	
558832	2015	Kia	K900	Luxury	Sedan	NaN	knalw4d4xf6019304	in	4
558833	2012	Ram	2500	Power Wagon	Crew Cab	automatic	3c6td5et6cg112407	wa	
558834	2012	BMW	X5	xDrive35d	SUV	automatic	5uxzw0c58cl668465	ca	4
558835	2015	Nissan	Altima	2.5 S	sedan	automatic	1n4al3ap0fc216050	ga	3

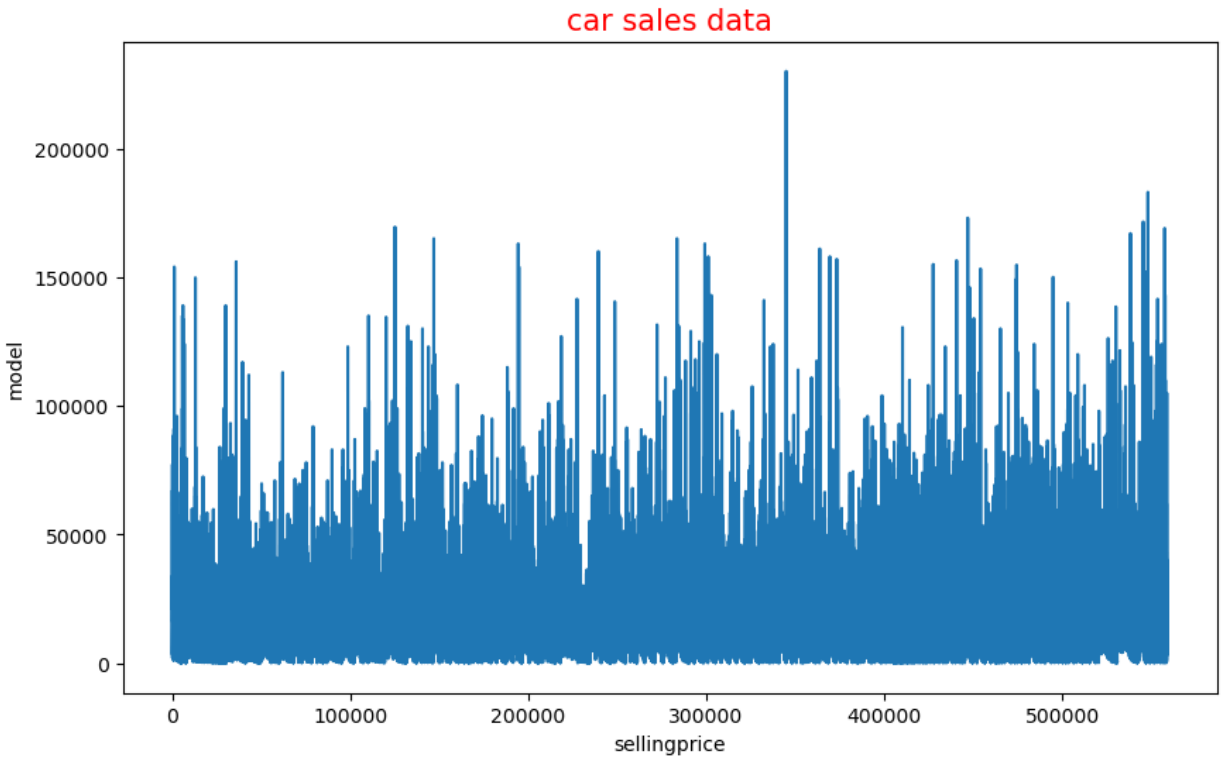
	year	make	model	trim	body	transmission	vin	state	conditi
558836	2014	Ford	F-150	XLT	SuperCrew	automatic	1ftfw1et2eke87277	ca	3

In [13]:
```
df.boxplot()
```

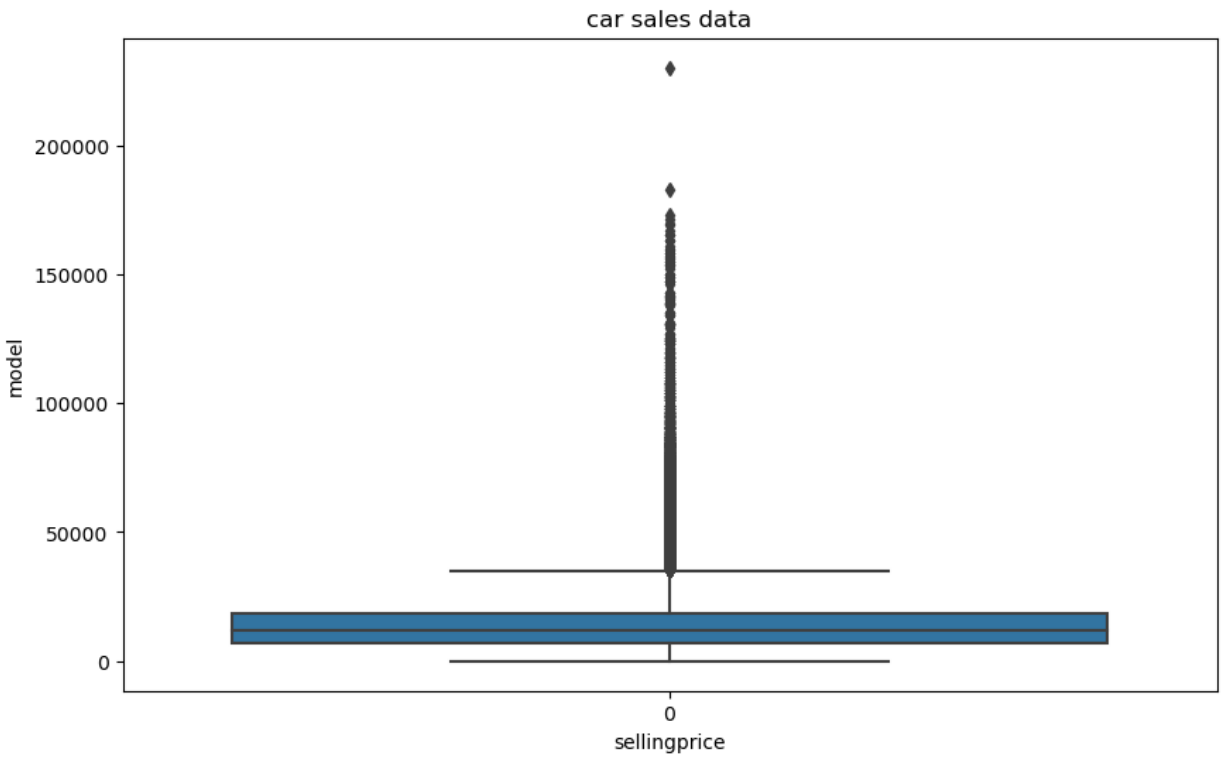
Out[13]: <Axes: >



In [14]:
```
plt.figure(figsize=(10,6))
sns.lineplot(df['sellingprice'])
plt.title('car sales data',fontsize=15,color='red')
plt.xlabel('sellingprice')
plt.ylabel('model')
plt.show()
```

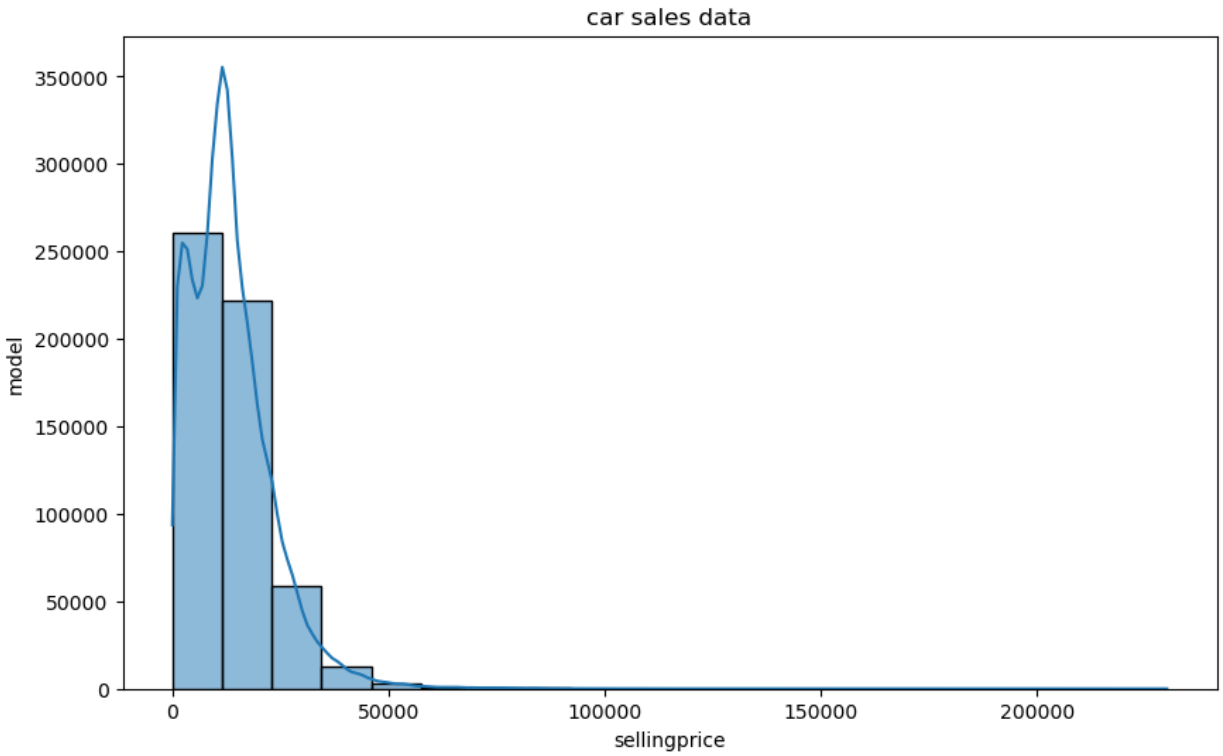


In [15]:
```
plt.figure(figsize=(10,6))
sns.boxplot(df['sellingprice'])
plt.title('car sales data')
plt.xlabel('sellingprice')
plt.ylabel('model')
plt.show()
```

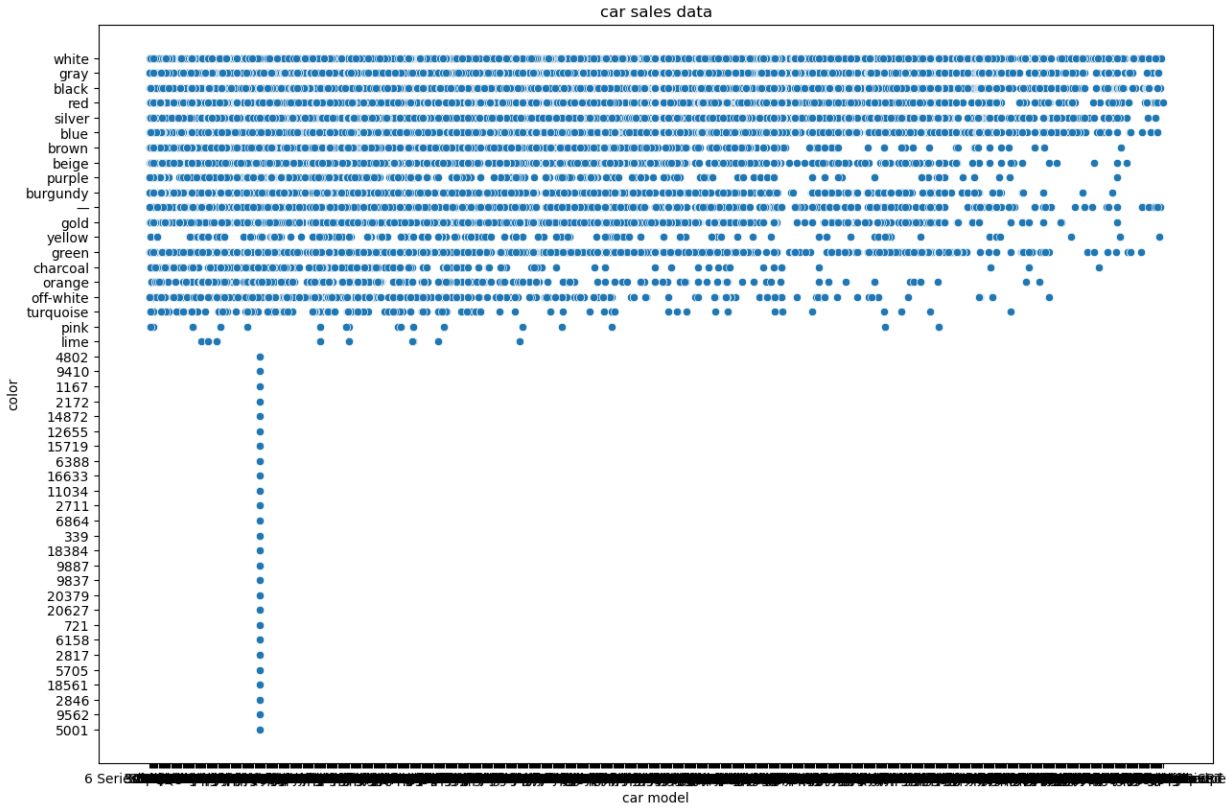


In [16]:
```
plt.figure(figsize=(10,6))
sns.histplot(df['sellingprice'],bins=20,kde=True)
plt.title('car sales data')
```

```
plt.xlabel('sellingprice')
plt.ylabel('model')
plt.show()
```



In [17]:
```
plt.figure(figsize=(15,10))
sns.scatterplot(x='model',y='color',data=df)
plt.title('car sales data')
plt.xlabel('car model')
plt.ylabel('color')
plt.show()
```



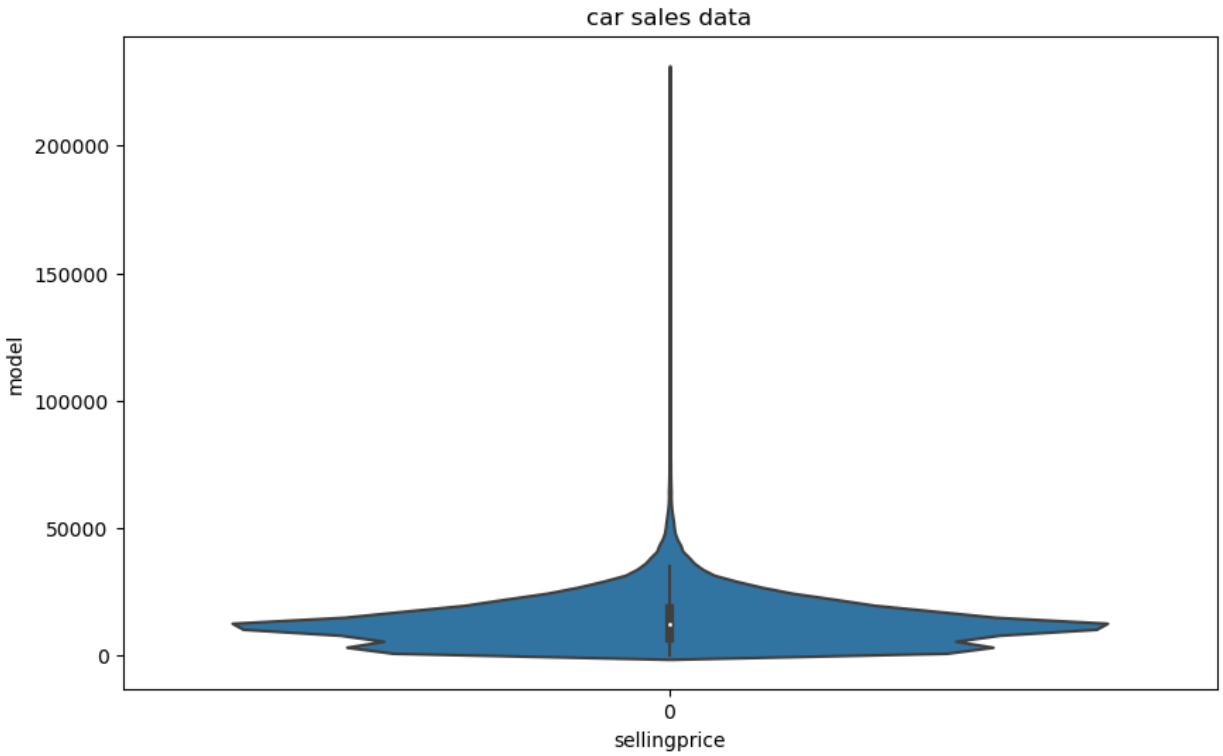
In [18]:
```
df.corr()
```

```
C:\Users\Administrator\AppData\Local\Temp\ipykernel_15888\1134722465.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  df.corr()
```

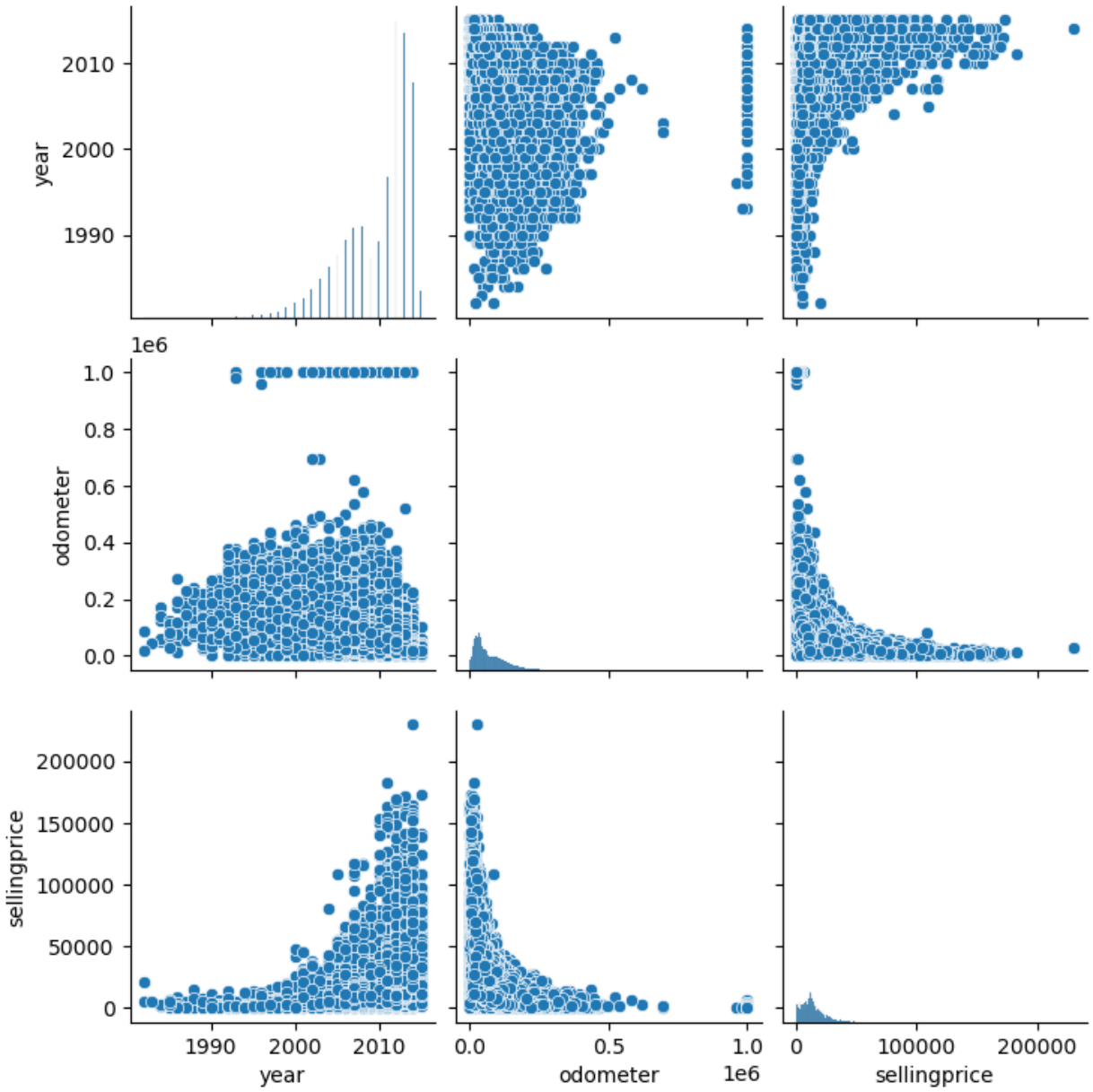
Out[18]:

	year	condition	odometer	mmr	sellingprice
year	1.000000	0.340320	-0.772737	0.596631	0.586480
condition	0.340320	1.000000	-0.318188	0.281415	0.321944
odometer	-0.772737	-0.318188	1.000000	-0.588113	-0.582393
mmr	0.596631	0.281415	-0.588113	1.000000	0.983635
sellingprice	0.586480	0.321944	-0.582393	0.983635	1.000000

In [19]:
```
plt.figure(figsize=(10,6))
sns.violinplot(df['sellingprice'],bins=20,kde=True)
plt.title('car sales data')
plt.xlabel('sellingprice')
plt.ylabel('model')
plt.show()
```



In [20]:
```
selected_columns=['year','odometer','sellingprice']
sns.pairplot(df[selected_columns])
plt.show()
```


year
odometer
1e6
sellingprice
2010
2000
1990
1.0
0.8
0.6
0.4
0.2
0.0
200000
150000
100000
50000
0
1990
2000
2010
0.0
0.5
1.0
1e6
0
100000
200000

In [ ]: