```python
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```python
import pandas as pd
import numpy as np
```

```python
df = pd.read_csv('/content/drive/MyDrive/touse.csv')
```

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 937 entries, 0 to 936
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   date    937 non-null    object
 1   stock   937 non-null    int64
 2   sales   937 non-null    int64
 3   price   937 non-null    float64
dtypes: float64(1), int64(2), object(1)
memory usage: 29.4+ KB
```

```python
df.head()
```

	date	stock	sales	price
0	01-01-2014	4972	0	1.29
1	02-01-2014	4902	70	1.29
2	03-01-2014	4843	59	1.29
3	04-01-2014	4750	93	1.29
4	05-01-2014	4654	96	1.29

```python
df['date']=pd.to_datetime(df['date'])
```

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 937 entries, 0 to 936
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   date    937 non-null    datetime64[ns]
 1   stock   937 non-null    int64
 2   sales   937 non-null    int64
 3   price   937 non-null    float64
```

```
dtypes: datetime64[ns](1), float64(1), int64(2)
memory usage: 29.4 KB
```

```
df['year']=df['date'].apply(lambda x: x.year)
```

```
df['month']=df['date'].apply(lambda x: x.month)
```
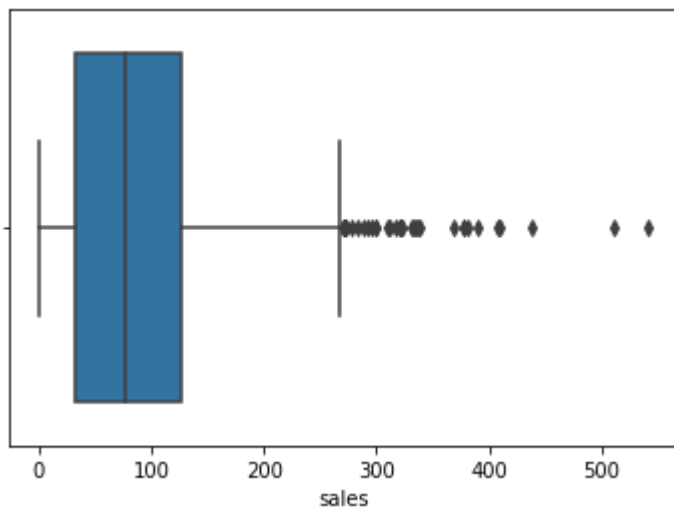
```
df = df.drop('date',axis=1)
```

```
df.head()
```

	stock	sales	price	year	month
0	4972	0	1.29	2014	1
1	4902	70	1.29	2014	2
2	4843	59	1.29	2014	3
3	4750	93	1.29	2014	4
4	4654	96	1.29	2014	5

```
import seaborn as sns
```

```
sns.boxplot(data=df,x='sales')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3365911110>
```



```
sns.histplot(data=df,x='sales')
```
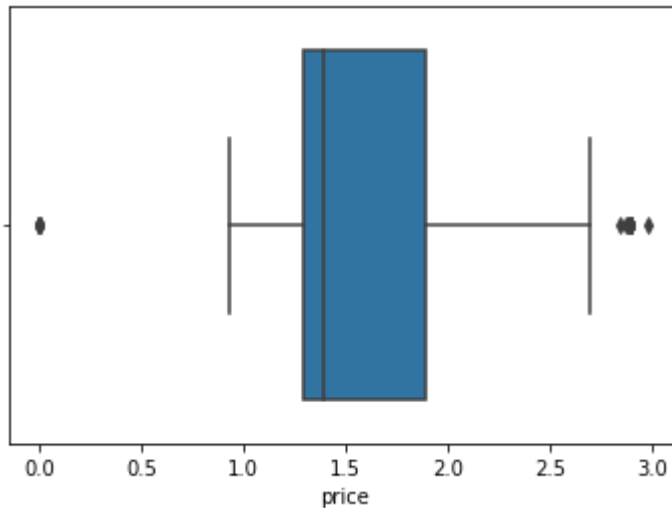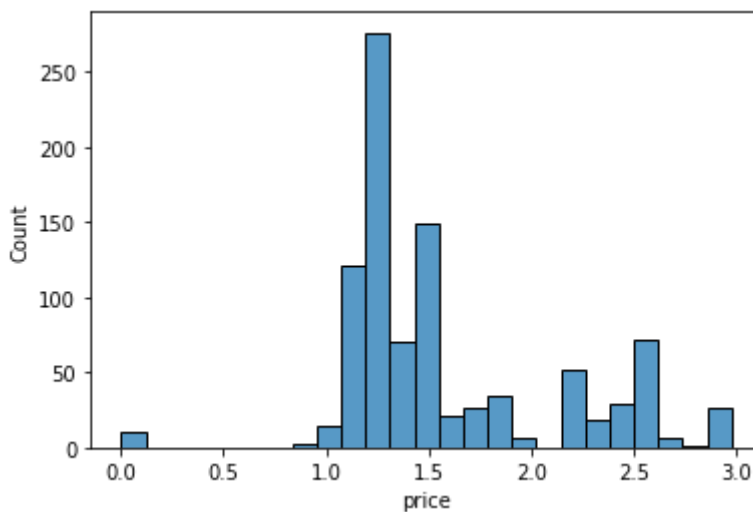
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f336580c610>
```
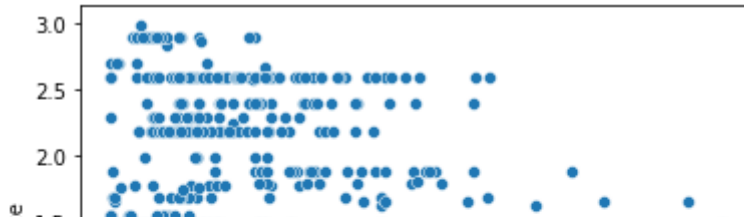


```python
sns.boxplot(data=df,x='price')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33657fe290>
```



```python
sns.histplot(data=df,x='price')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33652ad650>
```



```python
sns.scatterplot(data=df,x='sales',y='price')
```
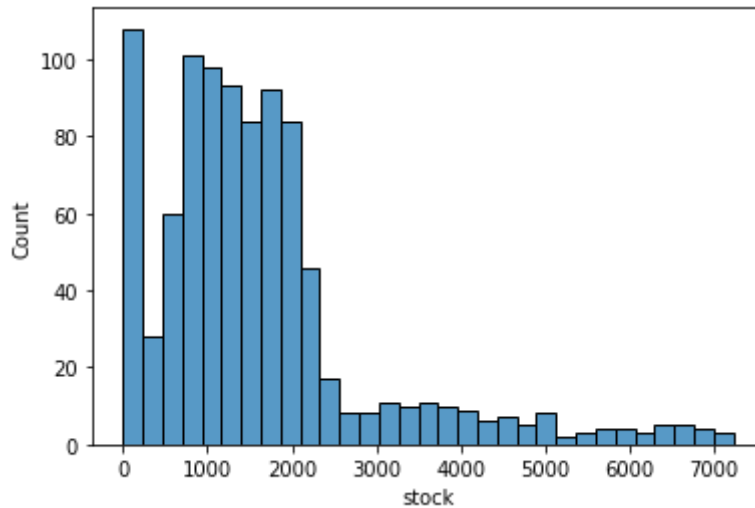
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3365174550>
```



```
sns.histplot(data=df,x='stock')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3365168f10>
```



```
df[df['stock'] == 0]
```

	stock	sales	price	year	month

```
import matplotlib.pyplot as plt
```

| **307** | 0 | 51 | 1.49 | 2014 | 10 |

```
plt.figure(figsize=(12,10),dpi=150)
sns.scatterplot(data=df,x='sales',y='price',hue='stock')
```
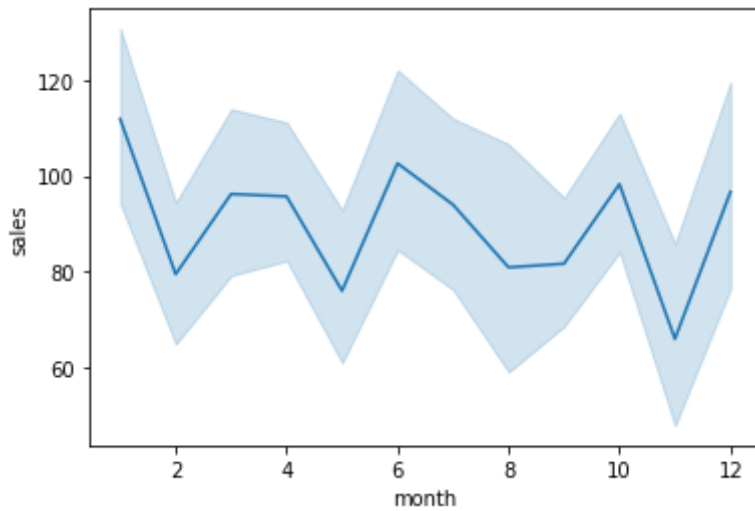
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3364fa6f90>
```

3.0

```
sns.lineplot(data=df,x='month',y='sales')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33630382d0>
```



```
zero_price_sales = df[df['price'] == 0].index
```

```
df[df['price'] == 0]
```

	stock	sales	price	year	month
157	2136	0	0.0	2014	12
190	794	0	0.0	2014	7
191	794	0	0.0	2014	7
192	794	0	0.0	2014	7
193	794	0	0.0	2014	7
233	1186	0	0.0	2014	8
487	15	0	0.0	2015	9
732	584	0	0.0	2016	9
740	107	0	0.0	2016	1
859	80	0	0.0	2016	5

```
df = df.drop(zero_price_sales)
```

```
df.to_csv('real_stock_prediction_dataset.csv')
```

```
pwd
```

```
'/content'
```

Colab paid products - Cancel contracts here