In [1]:

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

In [3]:

```
import os
os.chdir("C:/Users/mathava naresh/Desktop/naresh")
```

In [5]:

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

In [6]:

```
df
```

Out[6]:

	data	venda	estoque	preco
0	2014-01-01	0	4972	1.29
1	2014-01-02	70	4902	1.29
2	2014-01-03	59	4843	1.29
3	2014-01-04	93	4750	1.29
4	2014-01-05	96	4654	1.29
...	...	...	...	...
932	2016-07-27	98	3179	2.39
933	2016-07-28	108	3071	2.39
934	2016-07-29	128	4095	2.39
935	2016-07-30	270	3825	2.39
936	2016-07-31	183	3642	2.39

937 rows × 4 columns

In [7]:

```
df.info()
```

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

In [8]:

```
df.columns
```

Out[8]:

```
Index(['data', 'venda', 'estoque', 'preco'], dtype='object')
```

In [9]:

```
df.head()
```

Out[9]:

	data	venda	estoque	preco
0	2014-01-01	0	4972	1.29
1	2014-01-02	70	4902	1.29
2	2014-01-03	59	4843	1.29
3	2014-01-04	93	4750	1.29
4	2014-01-05	96	4654	1.29

In [10]:

```
df.tail()
```

Out[10]:

	data	venda	estoque	preco
932	2016-07-27	98	3179	2.39
933	2016-07-28	108	3071	2.39
934	2016-07-29	128	4095	2.39
935	2016-07-30	270	3825	2.39
936	2016-07-31	183	3642	2.39

In [11]:

```
df.describe()
```

Out[11]:

	venda	estoque	preco
count	937.000000	937.000000	937.000000
mean	90.533618	1608.258271	1.592572
std	80.682089	1356.691877	0.529502
min	0.000000	0.000000	0.000000
25%	33.000000	794.000000	1.290000
50%	76.000000	1348.000000	1.390000
75%	127.000000	1964.000000	1.890000
max	542.000000	7228.000000	2.980000

In [12]:

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

Out[12]:

```
data       0
venda      0
estoque    0
preco      0
dtype: int64
```

In [13]:

```
df.corr()
```

Out[13]:

	venda	estoque	preco
venda	1.000000	0.153659	0.094779
estoque	0.153659	1.000000	-0.032604
preco	0.094779	-0.032604	1.000000

In [14]:

```
df.cov()
```

Out[14]:

	venda	estoque	preco
venda	6509.599563	1.681963e+04	4.049096
estoque	16819.631265	1.840613e+06	-23.421562
preco	4.049096	-2.342156e+01	0.280372

In [15]:

```
df.dtypes
```

Out[15]:

```
data        object
venda        int64
estoque      int64
preco      float64
dtype: object
```

In [16]:

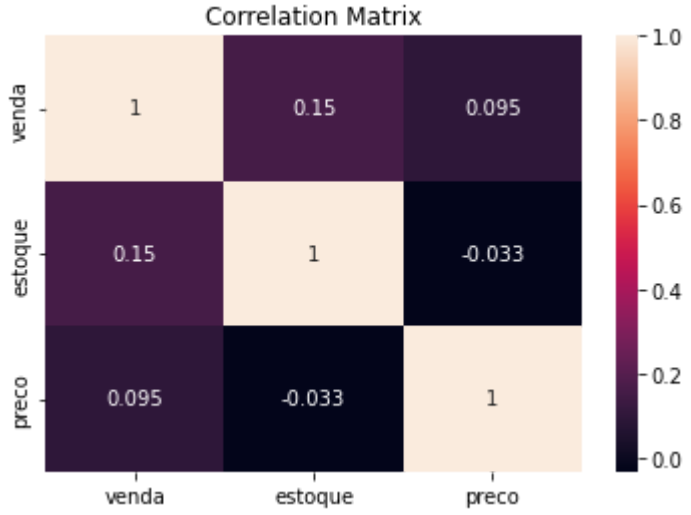
```
df.shape
```

Out[16]:

```
(937, 4)
```

In [17]:

```
sns.heatmap(df.corr(), annot = True)
plt.title("Correlation Matrix")
plt.show()
```



In [18]:

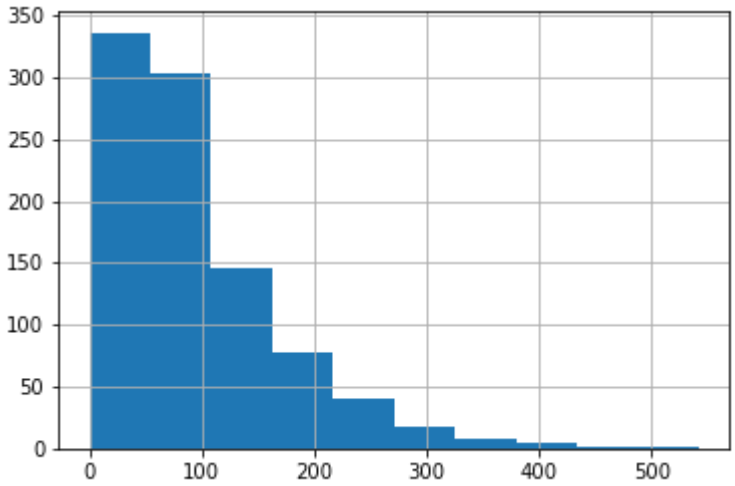
```
%matplotlib inline
```

In [20]:

```
df['venda'].hist(bins=10)
```

Out[20]:

<AxesSubplot:>

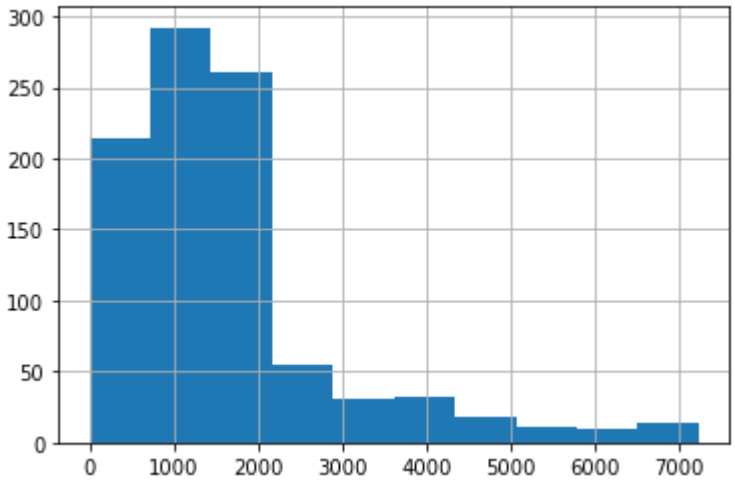


In [21]:

```
df['estoque'].hist(bins=10)
```

Out[21]:

<AxesSubplot:>

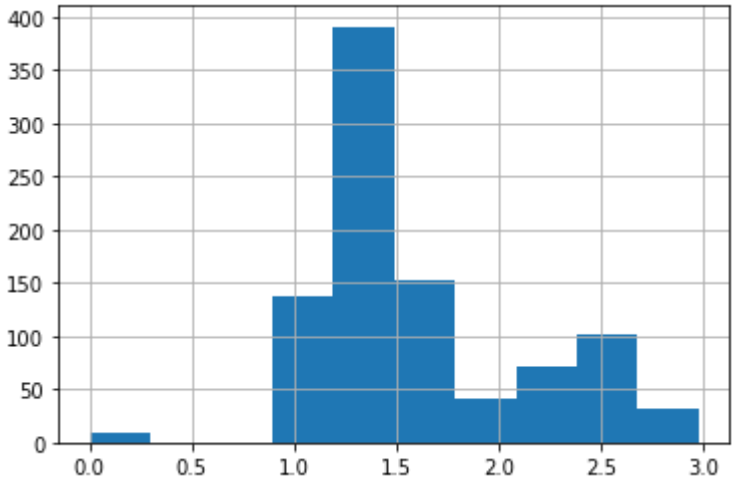


In [22]:

```
df['preco'].hist(bins=10)
```

Out[22]:

<AxesSubplot:>

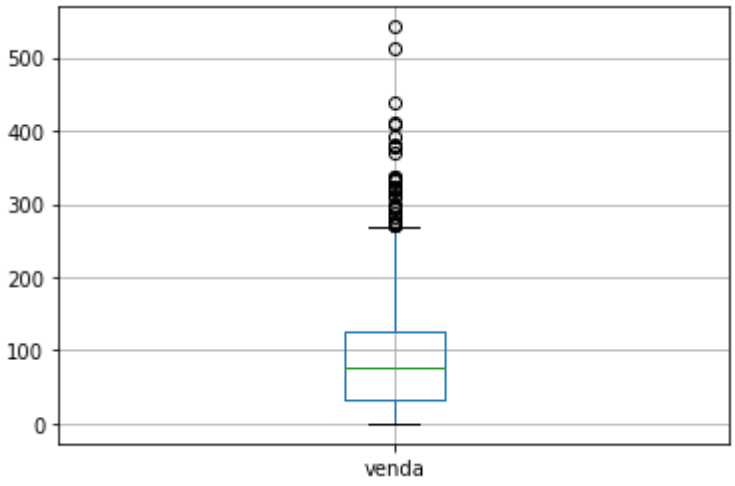


In [23]:

```
df.boxplot(column='venda')
```

Out[23]:

<AxesSubplot:>

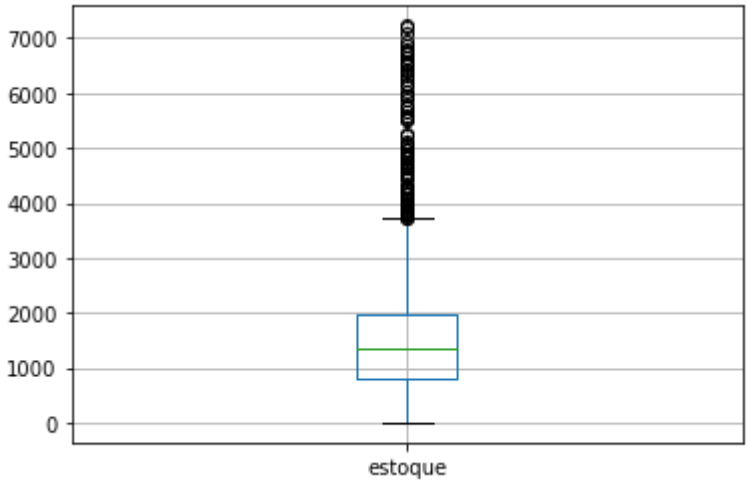


In [24]:

```
df.boxplot(column='estoque')
```

Out[24]:

<AxesSubplot:>

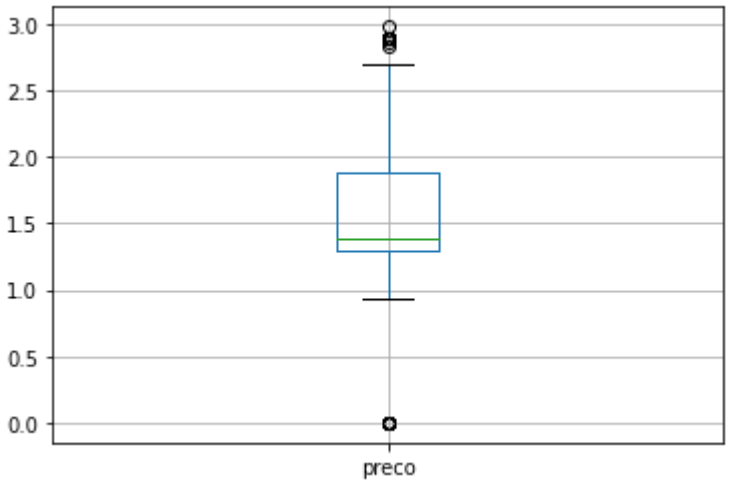


In [25]:

```
df.boxplot(column='preco')
```

Out[25]:

<AxesSubplot:>



In [27]:

```
sns.pairplot(df, x_vars=['venda','estoque','preco'], y_vars='data', height = 4)
```

Out[27]:

```
<seaborn.axisgrid.PairGrid at 0x18256974fd0>
```