In [1]:
```
import pandas as pd
```

In [2]:
```
data=pd.read_csv('DataQuiz.csv')
data
```

Out[2]:

	Customer	Saleorder	material	year	Pc
0	O41	161573638	material_000440	2013	350
1	K37	161575514	material_000771	2017	1
2	K37	161575514	material_000158	2017	2
3	K37	161575514	material_000422	2017	2
4	K37	161575514	material_000831	2017	4
...	...	...	...	...	...
276249	O41	164215683	material_000139	2018	20
276250	O41	164215683	material_000030	2018	25
276251	O41	164215683	material_000155	2018	25
276252	O41	164215683	material_000184	2018	40
276253	O41	164215683	material_000226	2018	2,000

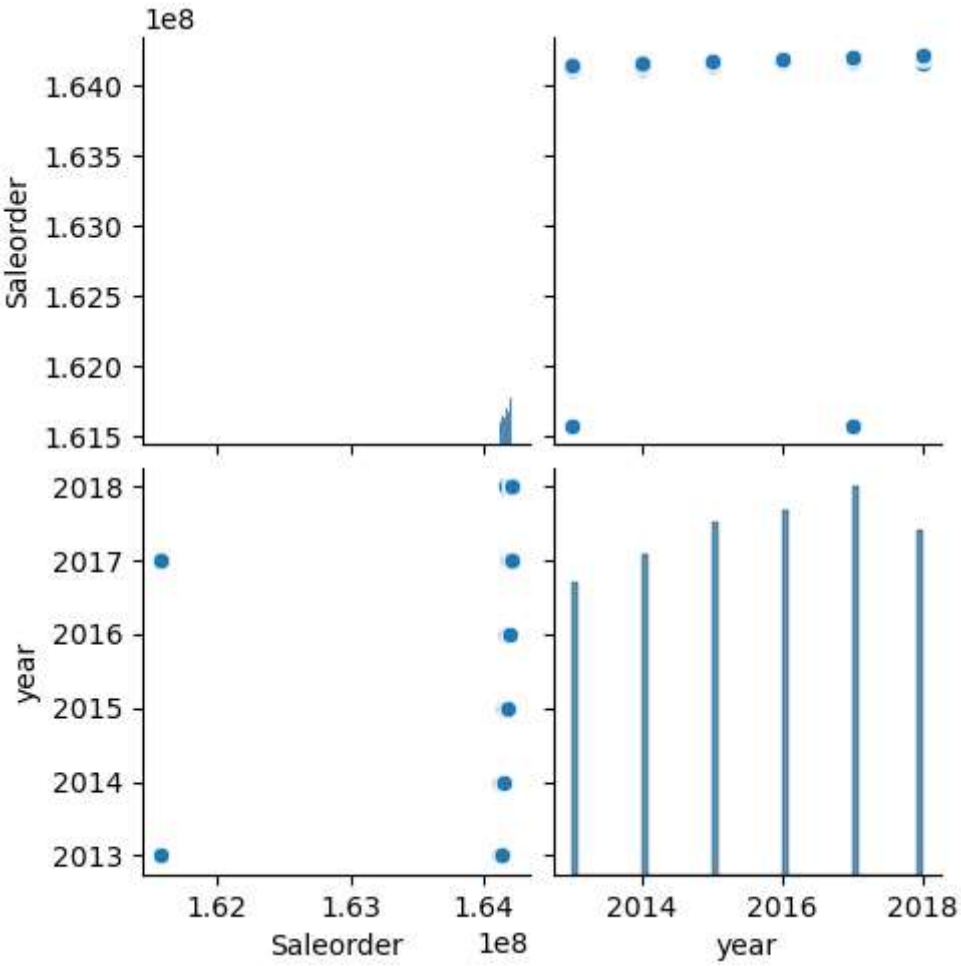
276254 rows × 5 columns

In [6]:
```
print(data.head())
```

```
  Customer  Saleorder         material  year   Pc
0      O41  161573638  material_000440  2013  350
1      K37  161575514  material_000771  2017    1
2      K37  161575514  material_000158  2017    2
3      K37  161575514  material_000422  2017    2
4      K37  161575514  material_000831  2017    4
```

In [9]:
```
import matplotlib.pyplot as plt
import seaborn as sns

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



In [21]:
```
%pylab inline
import pandas as pd
import seaborn as sns
from sklearn.model_selection import train_test_split
df = pd.read_csv('DataQuiz.csv', usecols=[1,2,3,4,5])
X = df[['Customer','Saleorder','material','Pc']]
y = df[['year']]
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.3,random_stat
X_train.head()
X_test.head()
y_train.head()
y_train.head()
```

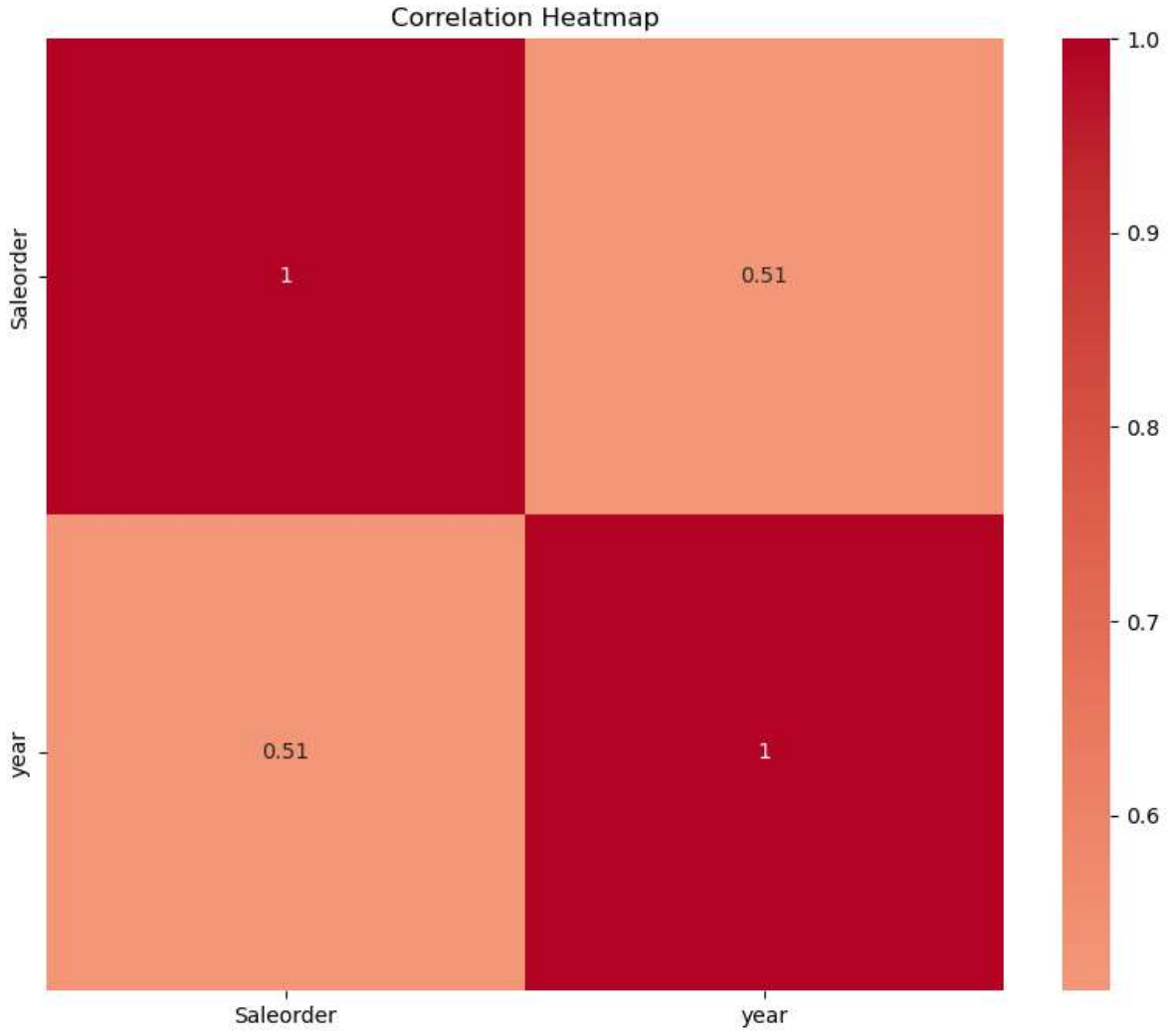
```
%pylab is deprecated, use %matplotlib inline and import the required librarie
s.
Populating the interactive namespace from numpy and matplotlib

C:\Users\User\AppData\Local\Temp\ipykernel_18524\2579233424.py:5: FutureWarni
ng: Defining usecols with out of bounds indices is deprecated and will raise
a ParserError in a future version.
  df = pd.read_csv('DataQuiz.csv', usecols=[1,2,3,4,5])
```

Out[21]:

	year
129847	50
176542	50
40324	20
92659	8
257471	100

In [23]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
file_path = 'path/to/your/file.csv'
data = pd.read_csv('DataQuiz.csv')
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', center=0)
plt.title('Correlation Heatmap')
plt.show()
```



In [28]:
```
correlation_matrix = data.corr()
print(correlation_matrix)
```

```
           Saleorder      year
Saleorder   1.000000  0.509506
year        0.509506  1.000000

C:\Users\User\AppData\Local\Temp\ipykernel_18524\4203189485.py:1: FutureWarni
ng: The default value of numeric_only in DataFrame.corr is deprecated. In a f
uture version, it will default to False. Select only valid columns or specify
the value of numeric_only to silence this warning.
  correlation_matrix = data.corr()
```

In [ ]: