```
#!/usr/bin/env python
# coding: utf-8

# In[211]:


import pandas as pd
import matplotlib.pyplot as plt
import re
import numpy as np
import seaborn as sns
import calendar

# HELPER FUNCTIONS
def FindNums(list):
    processed_list = []
    for value in list:
        value = value.replace("$", "")
        value = value.replace(",", "")
        processed_list.append(value)
    return processed_list

def new(list):
    processed_list = []
    for value in list:
        new_val = ''.join(filter(str.isdigit, value))
        if new_val == '':
            new_val = 0
        processed_list.append(new_val)
    return processed_list

def getMonth(list):
    month_list = []
    for val in list:
        month,day,year = val.split('/')
        month = calendar.month_name[int(month)]
        month_list.append(month)
    return month_list


# In[239]:


# UPLOAD DATA

data = pd.read_csv("ComcastDataSet")
data = data.rename(columns={"SALE MONTH": "month", "NUMBER OF UNITS OFFERED":
"unitsOffered",
                    "NUMBER OF UNITS SOLD": "unitsSold","PRODUCT": "product",
                    "DIVISION": "division", "SALES CHANNEL": "salesChannel", "TOTAL REVENUE ":
"totalRevenue"})

data.info()
data['totalRevenue']
```

```
#CONVERT OBJECTS INTO INTS
# convert sale month into month number

data['month'] = getMonth(data['month'])

#convert number of units offered into int
data['unitsOffered'] = FindNums(data['unitsOffered'])
data['unitsOffered'] = data['unitsOffered'].astype(int)

#convert number of units sold into int
data['unitsSold'] = FindNums(data['unitsSold'])
data['unitsSold'] = data['unitsSold'].astype(int)

#convert total revenue into int
data['totalRevenue'] = FindNums(data['totalRevenue'])
data['totalRevenue'] = new(data['totalRevenue'])
data['totalRevenue'] = data['totalRevenue'].astype(int)


#REMOVE DUPLICATES
data.drop_duplicates(inplace=True)

#REMOVE MISSING DATA
data = data.drop(data[data['totalRevenue'] == 0].index)

data.info()
data.head()

data.to_csv('clean_data.csv', sep =',', index = None)

#NEW FEATURES

# Conversion Rate = ratio of NUMBER OF UNITS SOLD to NUMBER OF UNITS OFFERED

# Revenue per Unit = dividing TOTAL REVENUE by the NUMBER OF UNITS SOLD


# In[219]:


print(data.groupby(data['month']).mean())

#DATA VISUALIZATION

#plt.hist(data['totalRevenue'])
#plt.show()

rep_plot = data['unitsSold'].groupby(data['month']).sum().plot(kind='bar')
rep_plot.set_xlabel("month")
rep_plot.set_ylabel("unitsSold")

#sns.boxplot(data)
#sns.pairplot(data)
```

```
# In[188]:


rep_plot = data['unitsSold'].groupby(data['salesChannel']).sum().plot(kind='bar')
rep_plot.set_xlabel("salesChannel")
rep_plot.set_ylabel("unitsSold")


# In[193]:


sns.pairplot(data)


# In[ ]:
```