# coding=utf-8  
#1. Create a new script called Script14Lab.py.  
#2. For this lab, we will be using a dataset from data.gov. Go to the following page and download the csv file in the screenshot below.  
#https://catalog.data.gov/dataset/consumer-complaint-database#topic=consumer\_navigation  
#3. Load the file into a Pandas dataframe and answer the following questions  
import pandas as pd  
cc = pd.read\_csv('http://bit.ly/consumercomplaints')  
#4. Retrieve the 1st 5 rows of the dataset  
print(cc.head())  
#5. Delete the following columns and then display the 1st 5 rows in the new dataset  
#a. Consumer complaint narrative  
#b. Company public response  
#c. Consumer consent provided  
#d. Consumer disputed?  
#e. Complaint ID  
#f. Tags  
#g. Issue  
cc.drop('Date received', axis =1, inplace=True)  
cc.drop('Product', axis=2, inplace=True)  
cc.drop('Sub Product', axis=3, inplace=True)  
cc.drop('Issue', axis=4, inplace=True)  
cc.drop('Sub Issue', axis=5, inplace=True)  
print(cc.head())  
#6. How many rows and columns are in the dataset.  
print(cc.shape)  
#7. Delete rows 0 – 1000. (We did not cover this in the class. You will need to use the Range function. Go look at Script3.py for help)  
cc.drop([0,1000], axis=0, inplace=True)  
print(cc.head())  
#8. How many rows and columns are in the dataset.  
print(cc.shape)  
#9. Sort the dataframe by Date Received descending – Print Dataframe  
print(cc['Date Received'].head().sort\_values(ascending=False))  
#10. Find the number of disputes by Company  
print(cc.head().sort\_values('Company'))  
#11. Find the number of disputes by State  
print(cc.head().sort\_values('State'))  
#12. Find the number of disputes by DateReceived Year, Month descending  
print(cc.head().sort\_values('DateReceived Year'))  
print(cc['Month'].head().sort\_values(ascending=False))