import requests  
import pandas  
import numpy  
import matplotlib.pyplot as plt  
  
def get\_data():  
 data = requests.get(  
 "https://gist.githubusercontent.com/tiangechen/b68782efa49a16edaf07dc2cdaa855ea/raw/0c794a9717f18b094eabab2"  
 "cd6a6b9a226903577/movies.csv").text  
 with open('movies.csv', 'w') as f:  
 f.write(data)  
  
  
# 2. Загрузите набор данных фильмов в pandas  
data = pandas.read\_csv('movies.csv', sep=',')  
  
print('3. Перечислите все столбцы набора данных и изучите их типы. Изучите статистику по различным областям.')  
print(data.info())  
print(data.describe())  
  
count\_movies = 0  
for x in data['Film']:  
 count\_movies += 1  
print('4.Сколько всего фильмов в наборе данных?', count\_movies)  
  
  
def get\_movies\_by\_year(data\_frame):  
 result = {}.fromkeys(numpy.unique(data\_frame['Year']), 0)  
 data = numpy.array(data\_frame['Year'])  
 for x in data:  
 result.update({x: result[x] + 1})  
 return result  
  
print('5. Сколько фильмов содержится в наборе данных за каждый год?')  
print(get\_movies\_by\_year(data))  
temp\_dict=get\_movies\_by\_year(data)  
x=temp\_dict.keys()  
y=temp\_dict.values()  
plt.plot(x, y)  
  
print('6. Покажите подробную информацию о наименее и наиболее прибыльных фильмах в наборе данных.')  
f = []  
for x in numpy.array(data['Worldwide Gross']):  
 f.append(x[1:-1])  
new\_WG\_data = numpy.array(f)  
print('6. а) Подробная информация о наиболее прибыльном фильме')  
print(data.loc[new\_WG\_data.argmax()])  
print('6. б) Подробная информация о наименее прибыльном фильме')  
print(data.loc[new\_WG\_data.argmin()])  
  
# 7. Значение «Жанр» временами кажется непоследовательным; попробуйте найти  
# эти несоответствия и исправить их.  
data.loc[data['Genre'] == 'romance', 'Genre'] = 'Romance'  
data.loc[data['Genre'] == 'Romence', 'Genre'] = 'Romance'  
data.loc[(data['Genre'] == 'Comdy', 'Genre')] = 'Comedy'  
data.loc[(data['Genre'] == 'comedy', 'Genre')] = 'Comedy'  
print(f"7. Значение «Жанр» стали последовательными{numpy.unique(data['Genre'])}")

#8.Сохраните (в новом файле CSV) 10 лучших комедий по количеству просмотров;  
#укажите только название фильма, год и студию.  
  
s\_data = data.sort\_values('Profitability', ascending=False)  
s1\_data=s\_data[['Film', 'Year', 'Lead Studio','Profitability']]  
top\_movies = pandas.DataFrame(s1\_data.iloc[0:10])  
top\_movies.to\_csv('top\_10\_movies.csv')  
print()  
  
print('Прикріпляю візуальний графік коливань значень в датафреймі по рокам')  
plt.show()