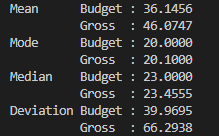
**Homework 2**

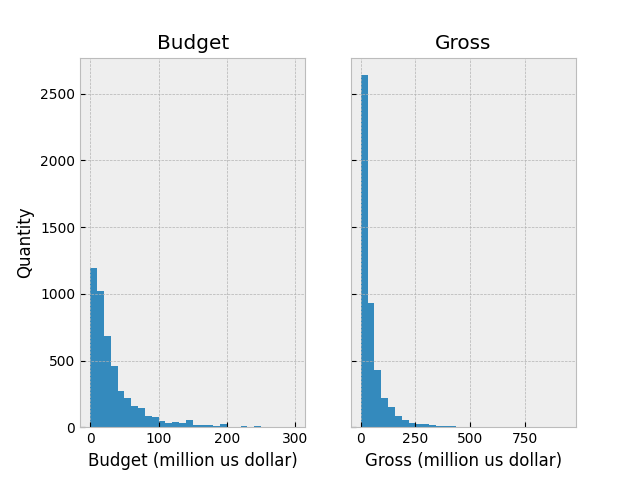
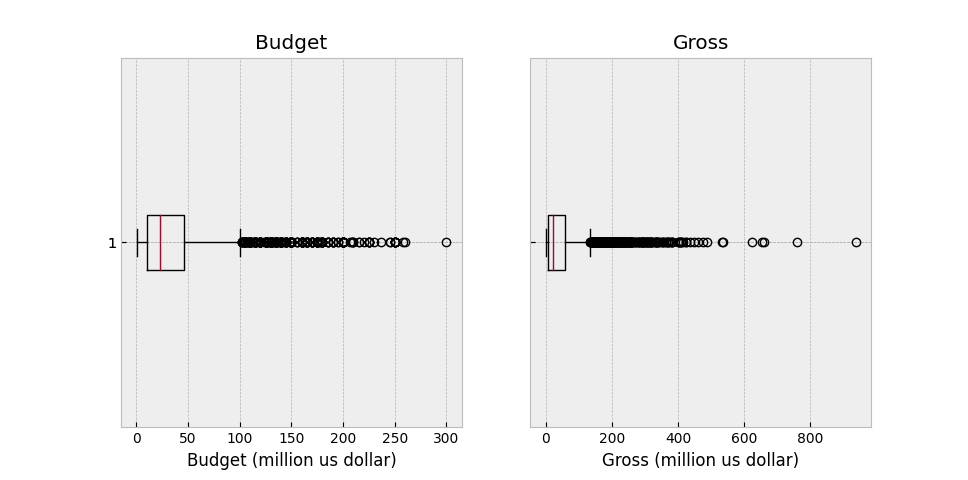
**Column ที่ใช้**

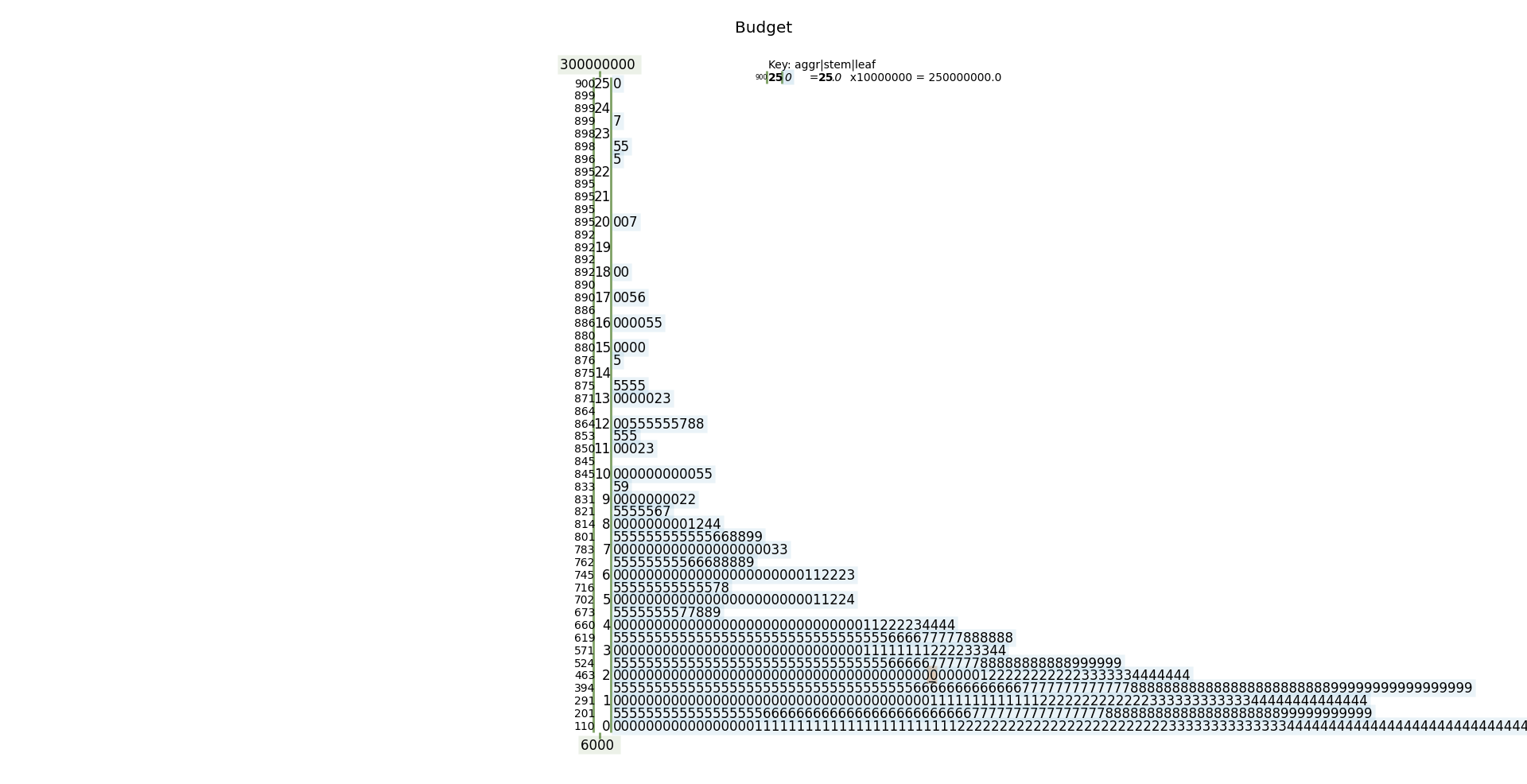
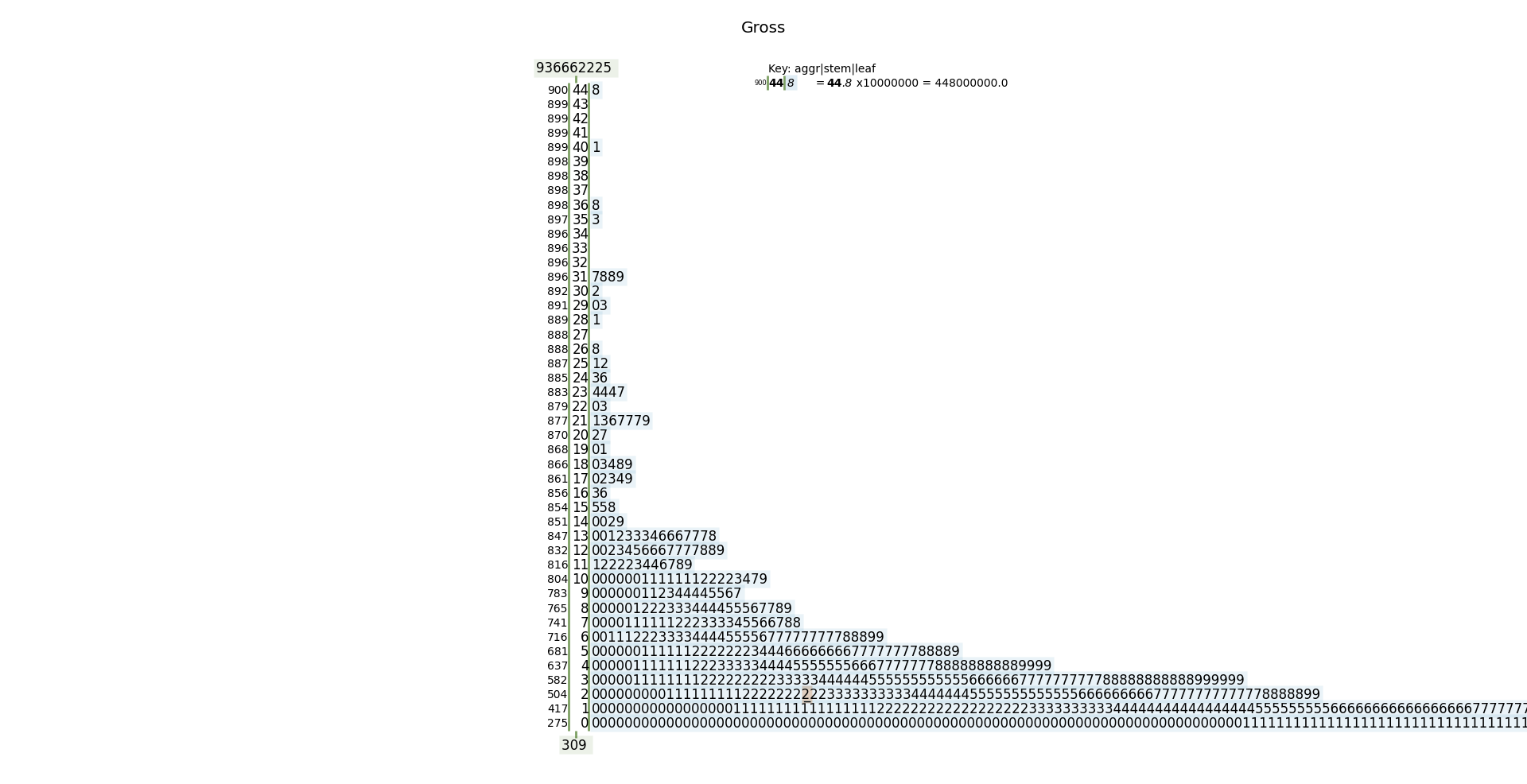
1. Budget งบประมาณในการสร้างภาพยนตร์ (million us dollar)
2. Gross รายได้ทั้งหมด (million us dollar)

**สถิติต่างๆ**

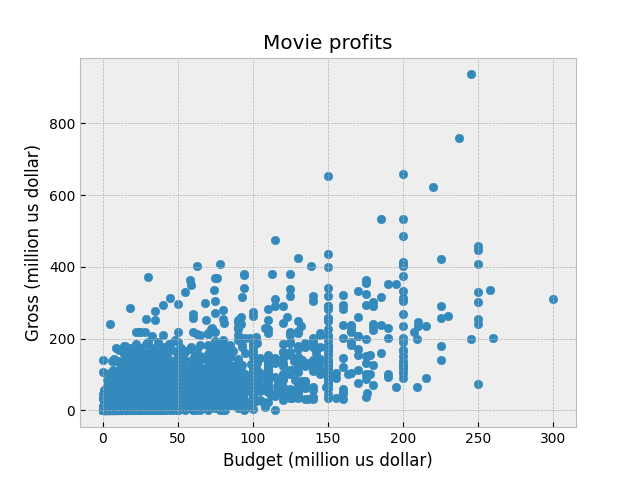


**กราฟต่างๆ**

1. **Histogram**
2. **Box plots**
3. **Stem and Leaf**



1. **Scatter**



**ตัวแปรต้น :** ทุนสร้าง (Budget)

**ตัวแปรตาม :** รายได้ (Gross)

**Outlier :**

1. Budget : ทุกข้อมูลที่มากกว่า 100 ล้านดอลลาร์
2. Gross : ทุกข้อมูลที่มากกว่า 138 ล้านดอลลาร์

**เหตุผล :** เพราะในการสร้างภาพยนตร์จำเป็นต้องอาศัยทุนในการสร้าง ทางผู้ค้นคว้าจึงต้องการทราบว่าทุนในการสร้างภาพยนตร์ส่งผลต่อรายได้ของภาพยนตร์หรือไม่

**บทวิเคราะห์**

จากการวิเคราะห์ข้อมูลจากกราฟ จะเห็นได้ว่าเมื่อภาพยนตร์มีงบประมาณในการสร้างน้อย ก็จะมีรายได้ใกล้เคียง หรือได้กำไรใกล้เคียงกับทุนเป็นส่วนใหญ่ แล้วก็มีบางส่วนที่ได้กำไรจำนวนมาก ก็มักจะมีทุนสร้างที่มากเช่นกัน

สรุปได้ว่า ภาพยนตร์ที่มีรายได้สูง มักจะเป็นภาพยนตร์ที่มีทุนสูงเช่นกัน แต่มีภาพยนตร์จำนวนน้อยที่สามารถทำกำไรจากทุนสร้างได้หลายเท่า ดังนั้นทุนในการสร้างภาพยนตร์จะแปรผันตรงกับรายได้ของภาพยนตร์

**Python code**

import statistics as stc

import matplotlib.pyplot as plt

import pandas as pd

import stemgraphic

plt.style.use('bmh')

df = pd.read\_csv('moviesfilter.csv')

# budget gross company name

x = df['budget']

y = df['gross']

z = df['company']

budget = x.to\_list()

gross = y.to\_list()

company = z.to\_list()

#format data to million dollar

for i in range(0, len(budget)):

    budget[i] = budget[i]/1000000

for i in range(0, len(gross)):

    gross[i] = gross[i]/1000000

#Print all detail

def detail():

    print("Mean      Budget :",str(stc.mean(budget)))

    print("          Gross  :",str(stc.mean(gross)))

    print("Mode      Budget :",str(stc.mode(budget)))

    print("          Gross  :",str(stc.mode(gross)))

    print("Median    Budget :",str(stc.median(budget)))

    print("          Gross  :",str(stc.median(gross)))

    print("Deviation Budget :",str(stc.stdev(budget)))

    print("          Gross  :",str(stc.stdev(gross)))

def histogram():

    fig, ax = plt.subplots(1, 2, sharey=True)

    ax[0].set\_xlabel('Budget (million us dollar)')

    ax[0].set\_ylabel('Quality')

    ax[0].set\_title('Budget')

    ax[0].hist(budget, bins=30)

    ax[1].set\_title('Gross')

    ax[1].hist(gross, bins=30)

    ax[1].set\_xlabel('Gross (million us dollar)')

    plt.show()

def boxplot():

    fig, ax = plt.subplots(1, 2, sharey=True)

    ax[0].set\_title('Budget')

    ax[0].boxplot(budget, vert=False)

    ax[0].set\_xlabel('Budget (million us dollar)')

    ax[1].set\_title('Gross')

    ax[1].boxplot(gross, vert=False)

    ax[1].set\_xlabel('Gross (million us dollar)')

    plt.show()

def stem():

    stemgraphic.stem\_graphic(df['budget'])

    plt.title('Budget')

    plt.show()

    stemgraphic.stem\_graphic(df['gross'])

    plt.title('Gross')

    plt.show()

def scatter():

    plt.xlabel('Budget (million us dollar)')

    plt.ylabel('Gross (million us dollar)')

    plt.title('Profit')

    plt.scatter(budget,gross)

    plt.show()

if \_\_name\_\_ == "\_\_main\_\_":

    detail()

    histogram()

    boxplot()

    stem()

    scatter()