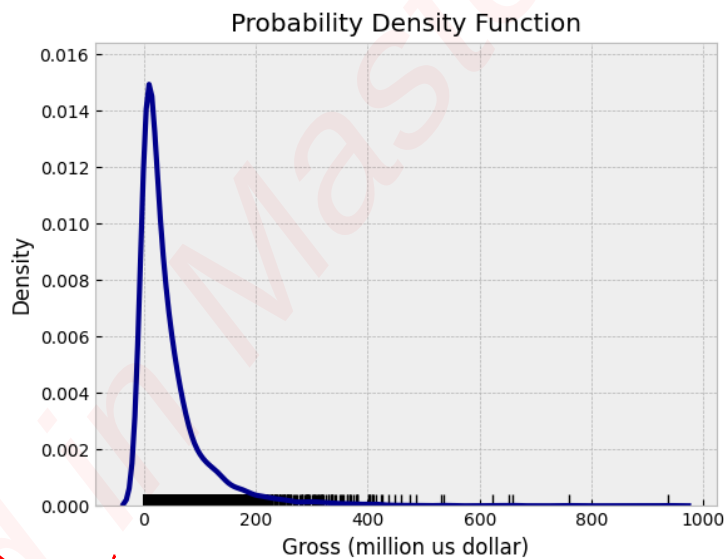
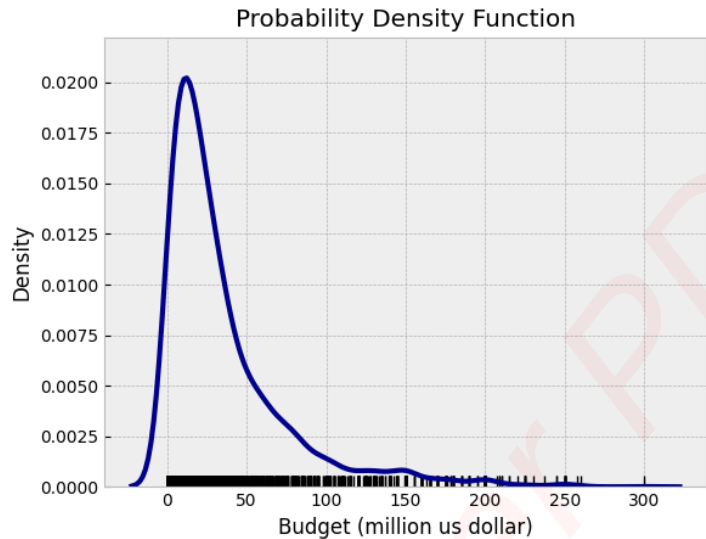


## Homework 3

## 1. Probability Density Function



PDF

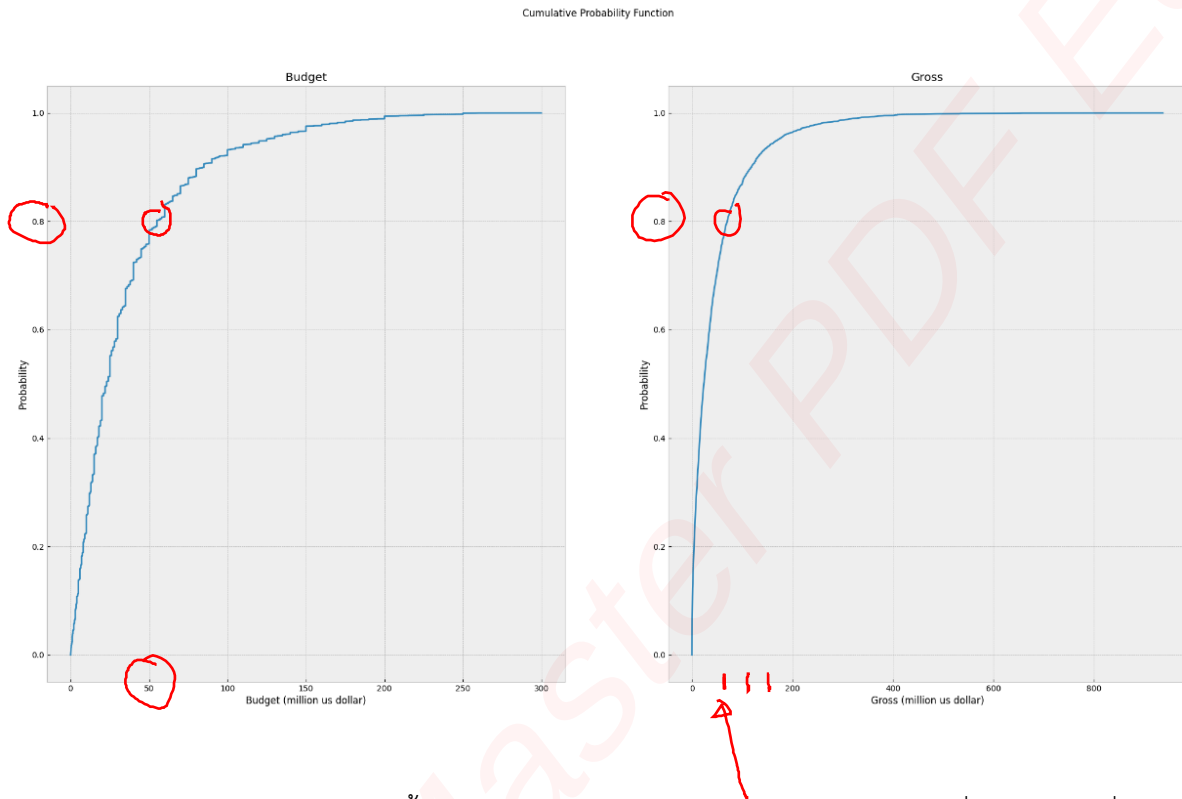
จุดยอดของกราฟต้นทุนประมาณ 12 ล้านดอลลาร์ และรายได้ทั้งหมดอยู่ที่ประมาณ 9 ล้านดอลลาร์ จากกราฟแสดงว่า ภาพยนตร์ในช่วงประมาณ 20% ใช้งบประมาณที่ 12 ล้านดอลลาร์ และมีภาพยนตร์ประมาณ 15% มีรายได้อยู่ที่ประมาณ 9 ดอลลาร์ ดังนั้นจะสรุปได้ว่าอุตสาหกรรมภาพยนตร์ มีการใช้งบประมาณมากกว่ารายได้ของภาพยนตร์

ตามหลักเป็นอันดับ

ประมาณ 20 จาก 100 เรื่อง

15 จาก 100 เรื่อง

## 2. Cumulative Prob Function



ความชันของกราฟต้นทุนในช่วงตั้งแต่ไม่ถึง 1 ล้าน จนถึงช่วง 50 ล้านดอลลาร์จะเริ่มมีความชันที่น้อยลงอย่างเห็นได้ชัด ทำให้เห็นว่าข้อมูลต้นทุนของภาพยนตร์จะอยู่ที่ประมาณ 1-50 ล้านดอลลาร์ เช่นเดียวกับกราฟรายได้ของภาพยนตร์ แต่จะเห็นได้ชัดว่าความชันของต้นทุนจะเริ่มน้อยลงที่ประมาณ 150 ล้านดอลลาร์ และรายได้จะเริ่มน้อยลงที่ 250 ล้านดอลลาร์ แสดงให้เห็นภาพยนตร์จำนวนน้อยมากที่ใช้ทุนมากกว่า 150 ล้านดอลลาร์ และมีรายได้มากกว่า 250 ล้านดอลลาร์

ภาพยนต์ ส่วนใหญ่ ถึง 50% (0.8) ใช้ทุน \_\_\_\_\_ เลื่อนขั้ว \_\_\_\_\_

## Source code

```
import statistics as stc
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import numpy as np

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 densityplot():
    sns.distplot(budget, hist = False, kde = True, rug = True,color = 'darkblue',
kde_kws={'linewidth': 3},rug_kws={'color': 'black'})

    # Plot formatting
    plt.title('Probability Density Function')
    plt.xlabel('Budget (million us dollar)')
    plt.ylabel('Density')
    plt.show()

    sns.distplot(gross, hist = False, kde = True, rug = True,color = 'darkblue',
kde_kws={'linewidth': 3},rug_kws={'color': 'black'})

    # Plot formatting
    plt.title('Probability Density Function')
    plt.xlabel('Gross (million us dollar)')
    plt.ylabel('Density')
    plt.show()

def cumulative():
    budgetData = sorted(np.array(budget))
    grossData = sorted(np.array(gross))
    budgetProb = 1. * np.arange(len(budgetData)) / (len(budgetData)-1)
    grossProb = 1. * np.arange(len(grossData)) / (len(grossData)-1)
    fig, ax = plt.subplots(1, 2)
    fig.suptitle('Cumulative Probability Function')
    ax[0].set_title('Budget')
    ax[0].plot(budgetData, budgetProb)
    ax[0].set_xlabel('Budget (million us dollar)')
    ax[0].set_ylabel('Probability')
    ax[1].set_title('Gross')
    ax[1].plot(grossData, grossProb)
    ax[1].set_xlabel('Gross (million us dollar)')
    ax[1].set_ylabel('Probability')
    plt.show()

if __name__ == "__main__":
    detail()
    densityplot()
    cumulative()

```