

นายภากรณ์ ธนประชาพันธ์ 62010694

### Homework #3

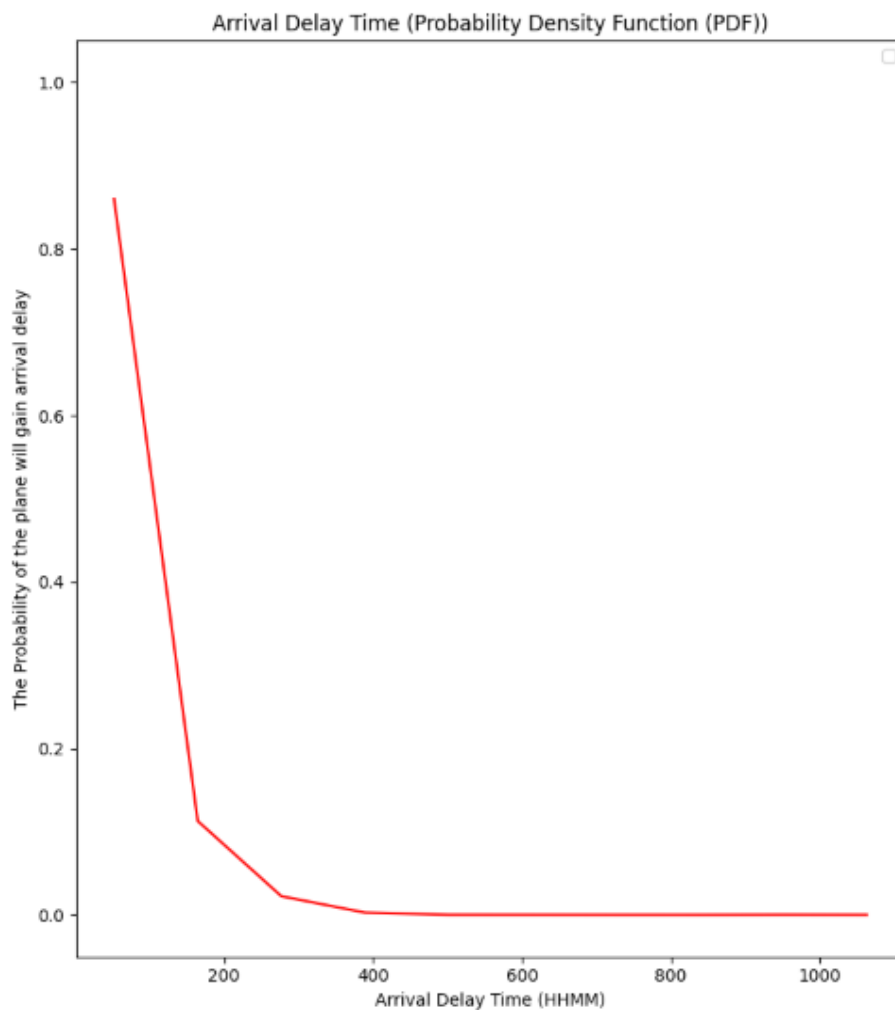
สถิติความล่าช้าของเที่ยวบินภายในประเทศสหรัฐอเมริกาเดือนสิงหาคม

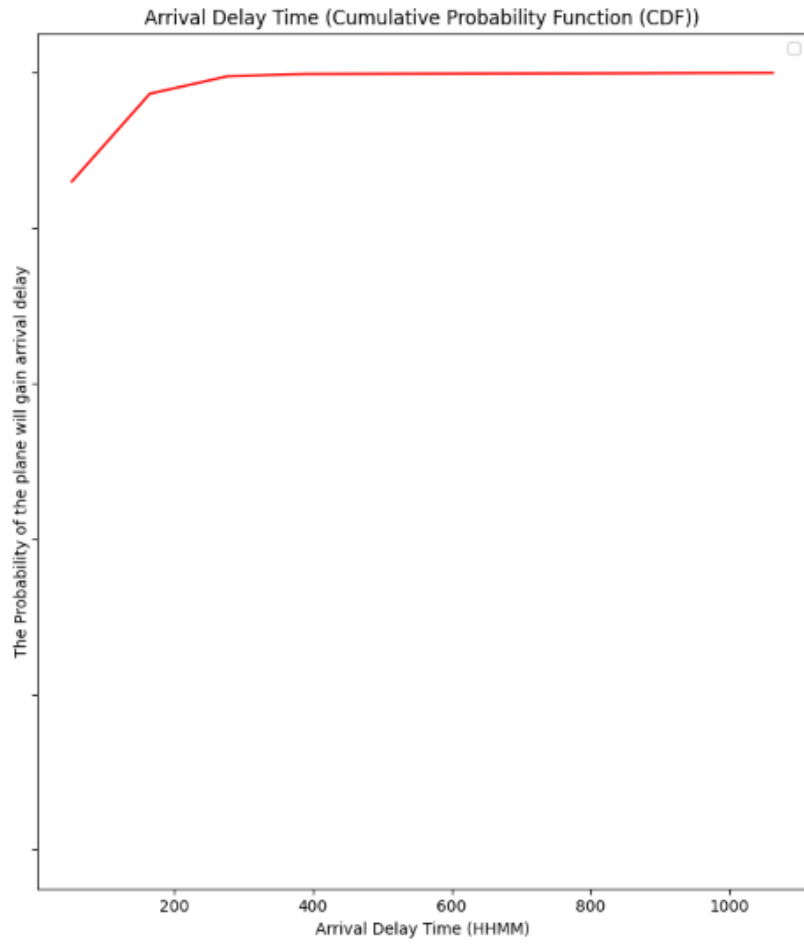
### 2018 August 2018 Nationwide Airplane Delay Statistic

Column : DEP\_DELAY ( Departure Delay (HHMM) ) และ ARR\_DELAY ( Arrival Delay (HHMM) )

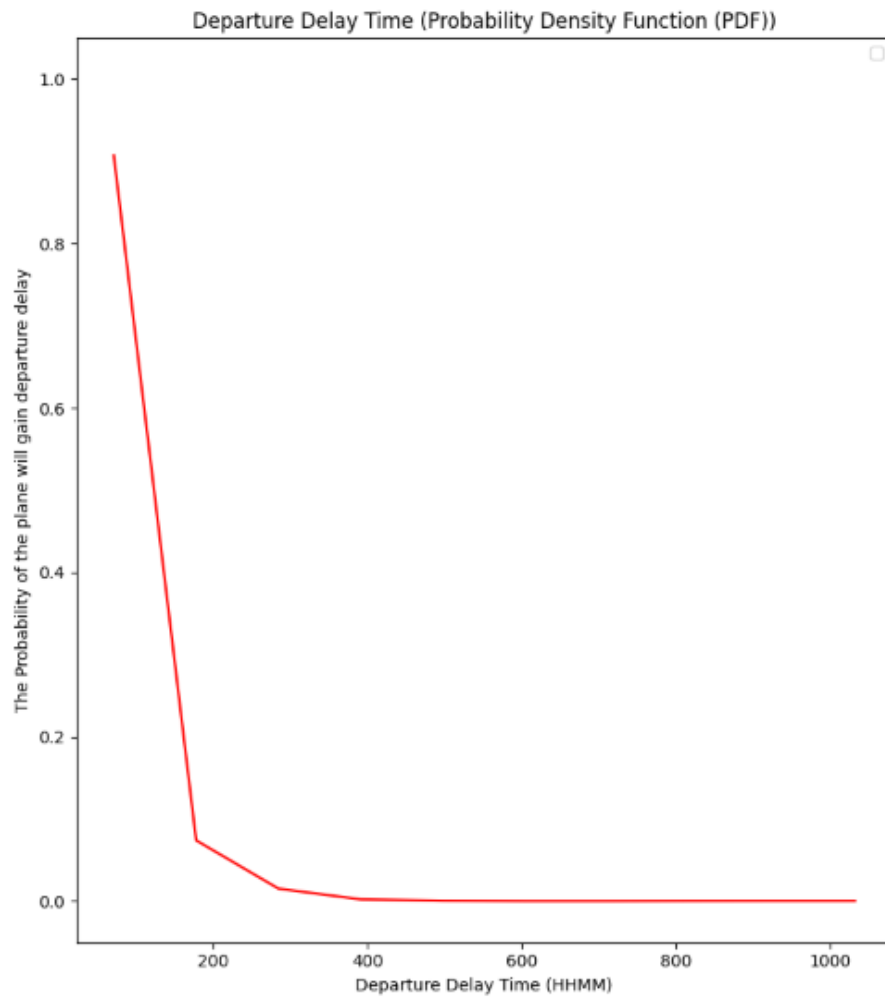
### Probability Density Function & Cumulative Probability Function

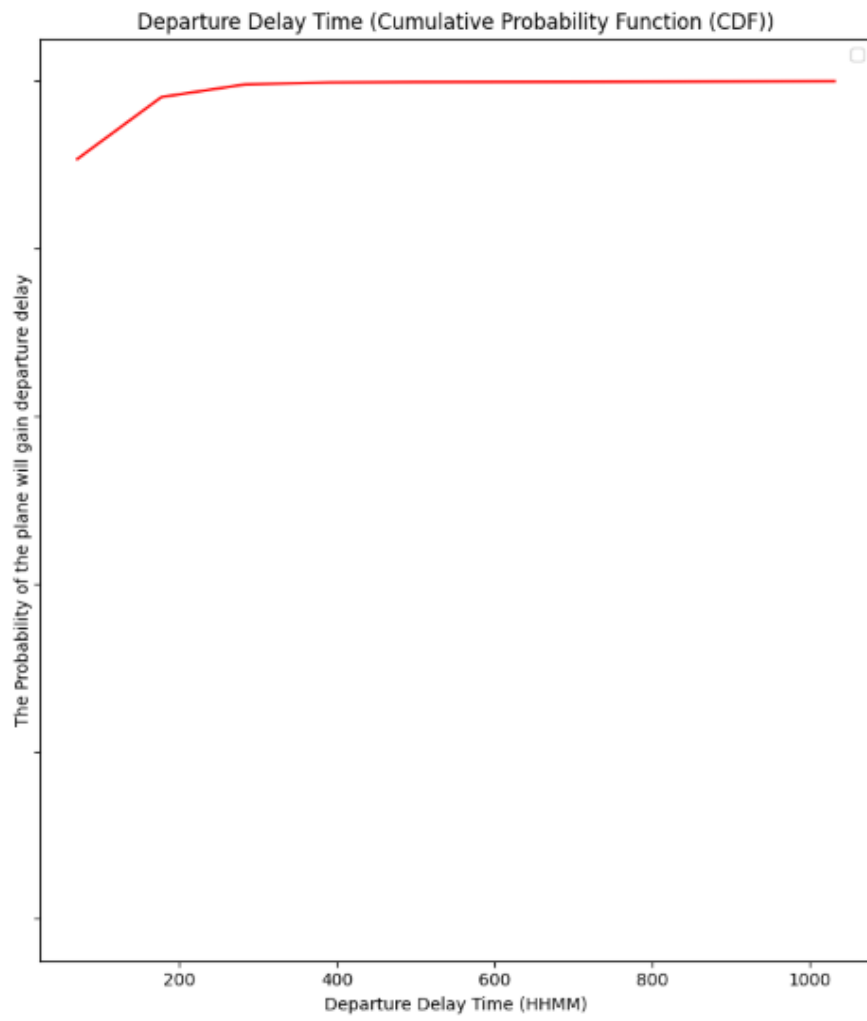
ARR\_DELAY ( Arrival Delay (HHMM) )





## DEP\_DELAY ( Departure Delay (HHMM))





**บทวิเคราะห์ข้อมูลจากกราฟ :** จากชุดข้อมูลที่ได้นำมาจะสามารถเห็นได้ว่า ทั้งกราฟ Probability Density Function และ Cumulative Probability Function ของทั้งสองค่าที่ได้เลือกมา ( เวลาล่าช้าขาออก ( Departure Delay (HHMM) ) และเวลาล่าช้าขาเข้า ( Arrival Delay (HHMM) ตามลำดับ ) มีรูปร่างคล้ายคลึงกันเป็นอย่างมาก เนื่องจากกราฟที่ได้ทำการวาดไปใน HW ก่อนหน้า ได้ให้ข้อสรุปว่า ค่าเวลาล่าช้าขาออก ( Departure Delay (HHMM) ) ส่งผลโดยตรงกับ ค่าเวลาล่าช้าขาเข้า ( Arrival Delay (HHMM) ) อย่างชัดเจนในรูปแบบแปรผันตรง หากเที่ยวบินออกช้า เวลาถึงจะช้าตามไปด้วย จากกราฟยังสามารถเห็นได้อีกว่า ในเดือนสิงหาคม 2018 เวลาที่เครื่องบินส่วนใหญ่จะเกิดการล่าช้า ไม่ว่าจะเป็นขาเข้าหรือขาออกก็ตาม จะมีค่าอยู่ที่ไม่เกิน 200 หรือ สองชั่วโมง ซึ่งส่วนนี้มีความเป็นไปได้ที่สูงมากซึ่งมีค่าตั้งแต่ประมาณ 0.9-0.1 เลยทีเดียว

ทำให้ชุดข้อมูลส่วนมากจะกระจุกอยู่บริเวณฝั่งซ้ายของทั้งสองกราฟ ส่วนเวลาล่าช้าที่มากกว่า 200 นั้นส่วนใหญ่จะไม่ค่อยเกิดขึ้นจนถึงขั้นแทบไม่มีโอกาสเกิดขึ้นเลย เช่นการล่าช้าไป 1000 หรือ 10 ชั่วโมง ถือว่าโอกาสเกิดเหตุการณ์นี้มีเป็นจำนวนน้อยมากหากเทียบกับจำนวนข้อมูลที่อยู่ในกลุ่มอยู่ทางด้านซ้าย

**เราจึงวิเคราะห์ได้ว่า เวลาล่าช้าทั้งขาออกกับเวลาล่าช้าขาเข้าในเกือบจะทุกกรณีจะมีค่าอยู่ไม่มากกว่า 200 หรือ 2 ชั่วโมง ส่วนโอกาสที่เที่ยวบินจะมีเวลาล่าช้าทั้งขาเข้ากับเวลาล่าช้าขาเข้า มากกว่า 200 หรือ 2 ชั่วโมง จะมีโอกาสได้น้อยแล้วค่อยๆลดลงไปเรื่อยๆจนถึงขั้นไม่แทบมีโอกาเกิดขึ้นเลย**

## Source Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

sample_data = pd.read_csv("August_2018_Nationwide_Airplane_Delay_Statistic.csv")

x = sample_data["DEP_DELAY (HHMM)"]
y = sample_data["ARR_DELAY (HHMM)"]

DEP_DELAY_COUNT, DEP_DELAY_BIN = np.histogram(x, bins=10)
ARR_DELAY_COUNT, ARR_DELAY_BIN = np.histogram(y, bins=10)

DEP_DELAY_PDF = DEP_DELAY_COUNT / sum(DEP_DELAY_COUNT)
DEP_DELAY_CDF = np.cumsum(DEP_DELAY_PDF)
ARR_DELAY_PDF = ARR_DELAY_COUNT / sum(ARR_DELAY_COUNT)
ARR_DELAY_CDF = np.cumsum(ARR_DELAY_PDF)

figure, DEP_DELAY = plt.subplots(1, 2, figsize=(18, 9), sharey=True)
DEP_DELAY[0].set_title("Departure Delay Time (Probability Density Function (PDF))")
DEP_DELAY[0].set_xlabel("Departure Delay Time (HHMM)")
DEP_DELAY[0].set_ylabel("The Probability of the plane will gain departure delay")
DEP_DELAY[0].plot(DEP_DELAY_BIN[1:], DEP_DELAY_PDF, color="red")
DEP_DELAY[0].legend()

DEP_DELAY[1].set_title("Departure Delay Time (Cumulative Probability Function (CDF))")
DEP_DELAY[1].set_xlabel("Departure Delay Time (HHMM)")
DEP_DELAY[1].set_ylabel("The Probability of the plane will gain departure delay")
DEP_DELAY[1].plot(DEP_DELAY_BIN[1:], DEP_DELAY_CDF, color="red")
DEP_DELAY[1].legend()

figure, ARR_DELAY = plt.subplots(1, 2, figsize=(18, 9), sharey=True)
ARR_DELAY[0].set_title("Arrival Delay Time (Probability Density Function (PDF))")
ARR_DELAY[0].set_xlabel("Arrival Delay Time (HHMM)")
ARR_DELAY[0].set_ylabel("The Probability of the plane will gain arrival delay")
ARR_DELAY[0].plot(ARR_DELAY_BIN[1:], ARR_DELAY_PDF, color="red")
ARR_DELAY[0].legend()

ARR_DELAY[1].set_title("Arrival Delay Time (Cumulative Probability Function (CDF))")
ARR_DELAY[1].set_xlabel("Arrival Delay Time (HHMM)")
ARR_DELAY[1].set_ylabel("The Probability of the plane will gain arrival delay")
ARR_DELAY[1].plot(ARR_DELAY_BIN[1:], ARR_DELAY_CDF, color="red")
ARR_DELAY[1].legend()

plt.show()
```