



Live07-HW01-Dataset Questions

ASK 5 Questions about this dataset

Dataset : nycflights13

ขั้นตอนที่ 1 ติดตั้ง library

```
library(nycflights13)
library(tidyverse)
```

หน้าตาของ dataset : `data("flights")`

```
> flights
# A tibble: 336,776 × 19
  year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time arr_delay carrier flight tailnum origin dest air_time distance hour minute time_hour
  <int> <int> <int> <dbl> <dbl>
1 2013     1     1    517      515      2     830     819     11  UA    1545 N14228  EWR  IAH    227    1400      5 15 2013-01-01 05:00:00
2 2013     1     1    533      529      4     850     830     20  UA    1714 N24211  LGA  IAH    227    1416      5 29 2013-01-01 05:00:00
3 2013     1     1    542      540      2     923     850     33  AA    1141 N619AA  JFK  MIA    160    1089      5 40 2013-01-01 05:00:00
4 2013     1     1    544      545     -1  1004     1022    -18  B6    725 N804JB  JFK  BQN    183    1576      5 45 2013-01-01 05:00:00
5 2013     1     1    554      600      -6     812     837    -25  DL    461 N668DN  LGA  ATL    116     762      6  0 2013-01-01 06:00:00
6 2013     1     1    554      558     -4     740     728     12  UA    1699 N39463  EWR  ORD    150     719      5 58 2013-01-01 05:00:00
7 2013     1     1    555      600     -5     913     854     19  B6    507 N516JB  EWR  FLL    158    1065      6  0 2013-01-01 06:00:00
8 2013     1     1    557      600     -3     709     723    -14  EV    5708 N829AS  LGA  IAD     53     229      6  0 2013-01-01 06:00:00
9 2013     1     1    557      600     -3     838     846     -8  B6    79 N593JB  JFK  MCO    140     944      6  0 2013-01-01 06:00:00
10 2013    1     1    558      600     -2     753     745      8  AA    301 N3ALAA  LGA  ORD    138     733      6  0 2013-01-01 06:00:00
# i 336,766 more rows
# i use `print(n = ...)` to see more rows
```

ชื่อเต็มของสายการบิน จาก `data("airlines")`

```

> airlines
# A tibble: 16 × 2
  carrier name
  <chr>   <chr>
1 9E      Endeavor Air Inc.
2 AA      American Airlines Inc.
3 AS      Alaska Airlines Inc.
4 B6      JetBlue Airways
5 DL      Delta Air Lines Inc.
6 EV      ExpressJet Airlines Inc.
7 F9      Frontier Airlines Inc.
8 FL      AirTran Airways Corporation
9 HA      Hawaiian Airlines Inc.
10 MQ     Envoy Air
11 OO     Skywest Airlines Inc.
12 UA     United Air Lines Inc.
13 US     US Airways Inc.
14 VX     Virgin America
15 WN     Southwest Airlines Co.
16 YV     Mesa Airlines Inc.

```

คำจำกัดความที่ควรรู้จากการใช้ `?flights`

dep_delay, arr_delay

Departure and arrival delays, in minutes. Negative times represent early departures/arrivals.

tailnum

Plane tail number. See `planes` for additional metadata.

hour, minute

Time of scheduled departure broken into hour and minutes.

▼ ตัวอย่างการตั้งคำถาม

Q1. most flights carrier in Sep 2013?

```

flights%>%
  filter(month == 9, year == 2013) %>%
  count(carrier) %>%
  arrange(desc(n))%>%
  head(5) %>%
  #เพื่อใช้ชื่อเติมของสายการบินจากการใช้โหลดข้อมูลมาดูจากฟังก์ชัน data("airlines")
  left_join(airlines)

```

```

> flights %>%
+   filter(month == 9, year == 2013) %>%
+   count(carrier) %>%
+   arrange(desc(n)) %>%
+   head(5) %>%
+   #เพื่อใช้ชื่อเดิมของสายการบินจากการใช้โหลดข้อมูลมาดูจากฟังก์ชัน data("airlines")
+   left_join(airlines)
joining with `by = join_by(carrier)`
# A tibble: 5 × 3
  carrier     n name
  <chr>   <int> <chr>
1 EV        4725 ExpressJet Airlines Inc.
2 UA        4694 United Air Lines Inc.
3 B6        4291 JetBlue Airways
4 DL        3883 Delta Air Lines Inc.
5 AA        2614 American Airlines Inc.

```

Ans : รายชื่อสายการบินที่มีเที่ยวบินมากที่สุด 5 อันดับแรกในเดือนกันยายน 2013

Q1 : สร้างคอลัมน์ใหม่ระบุว่าเวลาเที่ยวบินที่มาถึง-ออกเดินทาง เกินหรือก่อนเวลาที่กำหนด เปรียบเทียบจำนวนครั้งที่มีการเลยกำหนดเวลา กับกันกำหนดเวลา

กำหนดให้ ไฟล์กบินที่มาถึง-ออกเดินทางเลยกำหนดระบุว่า delay และไฟล์กบินที่มาถึง-ออกเดินทางก่อนเวลาให้ระบุว่า in time

สร้างคอลัมน์ใหม่ขึ้นมา 2 คอลัมน์ เก็บ data frame ใหม่ในตัวแปร flight_new

```

flight_new <- flights %>%
  mutate(dep_delay_group = if_else(dep_delay > 0, "delay", "in time"),
        arr_delay_group = if_else(arr_delay > 0, "delay", "in time"))

```

→ คำนวณเพื่อเปรียบเทียบจำนวนครั้งการดีเลย์ ในเที่ยวบินที่ออกเดินทาง

```

flight_new %>%
  group_by(dep_delay_group) %>%
  summarise(avg_dep_delay = mean(dep_delay),
            min_dep_delay = min(dep_delay),
            max_dep_delay = max(dep_delay),
            n = n())

```

	dep_delay_group	avg_dep_delay	min_dep_delay	max_dep_delay	n
1	delay	39.4	1	1301	128432
2	in time	-4.52	-43	0	200089
3	NA	NA	NA	NA	8255

จะพบว่า ในเที่ยวบินที่ออกเดินทาง จำนวนครั้งที่เกี่ยวบินออกช้ากว่ากำหนด(delay) คือ 128,432 ครั้ง จำนวนครั้งที่เกี่ยวบินออกเร็วกว่า/กันกำหนด(in time) คือ 200,089 ครั้ง และมีเที่ยวบินที่ไม่ได้บันทึกเวลา 8,255 ครั้ง

โดยเฉลี่ย เที่ยวบินที่ออกช้ากว่ากำหนด จะออกช้ากว่ากำหนด 39.4 นาที
เวลาที่มีการ delay น้อยสุด 1 นาที และมากสุด 1,301 นาที
→ คำนวนเพื่อเปรียบเทียบจำนวนครั้งการดีเลย์ ในเที่ยวบินที่มาถึง

```
flight_new %>%
  group_by(arr_delay_group) %>%
  summarise(avg_arr_delay = mean(arr_delay),
            min_arr_delay = min(arr_delay),
            max_arr_delay = max(arr_delay),
            n = n())
```

	arr_delay_group	avg_arr_delay	min_arr_delay	max_arr_delay	n
1	delay	40.3	1	1272	133004
2	in time	-16.0	-86	0	194342
3	NA	NA	NA	NA	9430

จะพบว่า ในเที่ยวบินที่มาถึง จำนวนครั้งที่เกี่ยวบินมาถึงช้ากว่ากำหนด(delay) คือ 133,004 ครั้ง จำนวนครั้งที่เกี่ยวบินมาถึงเร็วกว่า/กันกำหนด(in time) คือ 194,342 ครั้ง และมีเที่ยวบินที่ไม่ได้บันทึกเวลา 9,430 ครั้ง

โดยเฉลี่ย เที่ยวบินที่มาถึงช้ากว่ากำหนด จะมาถึงช้ากว่ากำหนด 40.3 นาที
เวลาที่มีการ delay น้อยสุด 1 นาที และมากสุด 1,272 นาที

Q2 : สายการบินที่มีเวลาการออกเดินทางช้ากว่ากำหนด (dep_delay) โดยเฉลี่ยมากที่สุด 3 อันดับแรกคือสายการบินใด?

หากเปรียบเทียบจากจำนวนครั้งที่มีการ delay

```
flight_new %>%
  filter(dep_delay > 0) %>%
  count(carrier) %>%
  arrange(desc(n)) %>%
  left_join(airlines)
```

จะพบว่า สายการบินที่มีการออกเดินทาง delay
มากที่สุดคือ UA (United Air Lines Inc.)

```
# A tibble: 16 x 3
  carrier   n name
  <chr>   <int> <chr>
1 UA        27261 United Air Lines Inc.
2 EV        23139 ExpressJet Airlines Inc.
3 B6        21445 JetBlue Airways
4 DL        15241 Delta Air Lines Inc.
5 AA        10162 American Airlines Inc.
6 MQ         8031 Envoy Air
7 9E         7063 Endeavor Air Inc.
8 WN         6558 Southwest Airlines Co.
9 US         4775 US Airways Inc.
10 VX        2225 Virgin America
11 FL        1654 AirTran Airways Corporation
12 F9         341 Frontier Airlines Inc.
13 YV         233 Mesa Airlines Inc.
14 AS         226 Alaska Airlines Inc.
15 HA          69 Hawaiian Airlines Inc.
16 OO             9 Skywest Airlines Inc.
```

หากเปรียบเทียบจากระยะเวลาการ delay โดยเฉพาะ เรียงลำดับจากมากไปน้อย

```
flight_new %>%
  filter(dep_delay > 0) %>%
  group_by(carrier) %>%
  summarise(avg_delay = mean(dep_delay)) %>%
  arrange(desc(avg_delay)) %>%
  left_join(airlines)
```

```
# A tibble: 16 x 3
  carrier avg_delay name
  <chr>     <dbl> <chr>
1 OO           58   Skywest Airlines Inc.
2 YV           53.0 Mesa Airlines Inc.
3 EV           50.3 ExpressJet Airlines Inc.
4 9E           48.9 Endeavor Air Inc.
5 F9           45.1 Frontier Airlines Inc.
6 MQ           44.9 Envoy Air
7 HA           44.8 Hawaiian Airlines Inc.
8 FL           40.8 AirTran Airways Corporation
9 B6           39.8 JetBlue Airways
10 DL           37.4 Delta Air Lines Inc.
11 AA           37.2 American Airlines Inc.
12 WN           34.9 Southwest Airlines Co.
13 VX           34.5 Virgin America
14 US           33.1 US Airways Inc.
15 AS           31.3 Alaska Airlines Inc.
16 UA           29.9 United Air Lines Inc.
```

จะพบว่าสายการบิน OO (SkyWest Airlines Inc.) จะมีระยะเวลาการ delay โดยเฉลี่ยมากที่สุดคือ 58 นาที รองลงมาเป็น YV(Mesa Airlines Inc.) 53.0 นาที และ EV(ExpressJet Airlines Inc.) 50.3 นาทีตามลำดับ

เพิ่มเติม : เนื่องด้วยผู้เขียนต้องการความมั่นใจในการคำนวณ จึงได้ลองเช็คค่าเฉลี่ยระยะเวลาที่มีการ delay ของสายการบินหนึ่ง ออกมานะ

```
flight_new %>%
  filter(carrier == "00" & dep_delay != is.na(dep_delay) & dep_
  summarise(avg_00_delay = mean(dep_delay))
```

จะเห็นได้ว่าผลลัพธ์ที่ได้ตรงกับ

Q3 : สายการบินที่มีเที่ยวบินน้อยที่สุด มีระยะเวลาการบินต่อ 1 ไฟล์กเฉลี่ยเท่าไร

สายการบินที่มีจำนวนเที่ยวบินน้อยที่สุด

```
flight_new %>%
  count(carrier) %>%
  arrange(n) %>%
  head(3) %>%
  left_join(airlines)
```

carrier	n	name
<chr>	<int>	<chr>
1 00	32	SkyWest Airlines Inc.
2 HA	342	Hawaiian Airlines Inc.
3 YV	601	Mesa Airlines Inc.

สายการบินที่มีจำนวนเที่ยวบินน้อยที่สุด คือ OO (SkyWest Airlines Inc.) จำนวน 32 เที่ยวบิน มีระยะเวลาการบินต่อ 1 รอบเฉลี่ย

```
flight_new %>%
  select(year, month, day, carrier, distance, air_time) %>%
  filter(carrier == "00" & air_time != is.na(air_time)) %>%
  summarise(avg_air_time = mean(air_time))
```

avg_air_time	>	83%60
<dbl>	[1]	23
83.5		

มีระยะเวลาการบินต่อ 1 รอบเฉลี่ย 83.5 นาที คิดเป็น 83 นาที 30 วินาที คิดเป็น 1 ชั่วโมง 23 นาที 30 วินาที

Q4 : ช่วงเดือนไหนมีเที่ยวบินอุกมากที่สุด จุดหมายปลายทาง(dest)ที่นิยมไปในเดือนนั้น คือที่ไหน

```
flight_new %>%
  select(year, month, day, flight) %>%
  group_by(month) %>%
  count(month) %>%
  arrange(desc(n)) %>%
  head(3)
```

```
# A tibble: 3 × 2
# Groups:   month [3]
  month     n
  <int> <int>
1      7 29425
2      8 29327
3     10 28889
```

จะพบว่าเดือนที่มีเที่ยวบินอุกมากที่สุดคือ เดือน 7 หรือเดือนกรกฎาคม จำนวน 29,425 เที่ยวบิน หาจุดหมายปลายทางที่นิยมไป

```
flight_new %>%
  filter(month == 7) %>%
  count(dest) %>%
  arrange(desc(n)) %>%
  left_join(airports,
             by = c("dest" = "faa")) %>%
  select(dest, name, n)
```

```
# A tibble: 94 × 3
  dest    name          n
  <chr> <chr>        <int>
1 ORD    Chicago O'Hare Intl 1573
2 ATL    Hartsfield Jackson Atlanta Intl 1511
3 LAX    Los Angeles Intl 1500
4 BOS    General Edward Lawrence Logan Intl 1378
5 CLT    Charlotte Douglas Intl 1260
6 MCO    Orlando Intl 1257
7 SFO    San Francisco Intl 1231
8 MIA    Miami Intl 978
9 FLL    Fort Lauderdale Hollywood Intl 938
10 DTW   Detroit Metro Wayne Co 837
# i 84 more rows
# i Use `print(n = ...)` to see more rows
```

จุดหมายปลายทางที่นิยมไปมากที่สุดในเดือนกรกฎาคม คือ ORD (Chicago Ohare Intl) 1,573 เที่ยวบิน

Q5 : ระยะทางระหว่างสนามบินใดไกลที่สุด และมีสายการบินใดให้บริการ

จัดอันดับระยะทางที่ไกลที่สุด

```
flight_new %>%
  select(carrier, origin, dest, distance) %>%
  arrange(desc(distance))
```

```
# A tibble: 336,776 x 7
  carrier name      origin name.x      dest name.y      distance
  <chr>   <chr>     <chr>  <chr>     <chr>  <chr>     <dbl>
1 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
2 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
3 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
4 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
5 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
6 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
7 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
8 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
9 HA     Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
10 HA    Hawaiian Airlines Inc. JFK  John F Kennedy Intl HNL  Honolulu Intl  4983
# i 336,766 more rows
# i Use `print(n = ...)` to see more rows
```

ระยะทางที่ไกลที่สุด 4,983 ไมล์ จากสนามบิน JFK (John F Kennedy Intl) ไปสนามบิน HNL (Honolulu Intl)

```
flight_new %>%
  filter(distance == 4983) %>%
  select(carrier, origin, dest, distance) %>%
  count(carrier) %>%
  left_join(airlines)
```

```
# A tibble: 1 x 3
  carrier  n name
  <chr>   <int> <chr>
1 HA       342 Hawaiian Airlines Inc.
```

จะพบว่ามีเพียงสายการบินเดียวคือ HA (Hawaiian Airlines Inc.) ก็งหมด 342 เที่ยวบิน