NYC FLIGHT 2013 BUSINESS ANALYSIS

```
library(dplyr)
library(tidyverse)
```

```
Attaching package: ‘dplyr’


The following objects are masked from ‘package:stats’:

    filter, lag


The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union


Warning message in system("timedatectl", intern = TRUE):
“running command 'timedatectl' had status 1”
Warning message:
“Failed to locate timezone database”
── Attaching packages ─────────────────────────────────────── tidyverse 1.3
```

```
flights <- read.csv("flights.csv")
```

```
complete.cases(flights)
```

```
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · ⋯ · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE
```

```
sum(complete.cases(flights)/nrow(flights))
```

```
0.971999192341495
```

Complete case `97.19%`

DROP MISSING VALUES

```
flights <- drop_na(flights)
```

```
glimpse(flights)
```

```
Rows: 327,346
Columns: 19
$ year           <int> 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013
$ month          <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
$ day            <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
$ dep_time       <int> 517, 533, 542, 544, 554, 554, 555, 557, 557, 558, 55
$ sched_dep_time <int> 515, 529, 540, 545, 600, 558, 600, 600, 600, 600, 60
$ dep_delay      <int> 2, 4, 2, -1, -6, -4, -5, -3, -3, -2, -2, -2, -2, -2,
$ arr_time       <int> 830, 850, 923, 1004, 812, 740, 913, 709, 838, 753, 8
$ sched_arr_time <int> 819, 830, 850, 1022, 837, 728, 854, 723, 846, 745, 8
$ arr_delay      <int> 11, 20, 33, -18, -25, 12, 19, -14, -8, 8, -2, -3, 7,
$ carrier        <chr> "UA", "UA", "AA", "B6", "DL", "UA", "B6", "EV", "B6"
$ flight         <int> 1545, 1714, 1141, 725, 461, 1696, 507, 5708, 79, 301
$ tailnum        <chr> "N14228", "N24211", "N619AA", "N804JB", "N668DN", "N
$ origin         <chr> "EWR", "LGA", "JFK", "JFK", "LGA", "EWR", "EWR", "LG
$ dest           <chr> "IAH", "IAH", "MIA", "BQN", "ATL", "ORD", "FLL", "IA
$ air_time       <int> 227, 227, 160, 183, 116, 150, 158, 53, 140, 138, 149
$ distance       <int> 1400, 1416, 1089, 1576, 762, 719, 1065, 229, 944, 73
$ hour           <int> 5, 5, 5, 5, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6
$ minute         <int> 15, 29, 40, 45, 0, 58, 0, 0, 0, 0, 0, 0, 0, 0, 0, 59
```

5 BUSINESS QUESTIONS

Q1 : Top 3 months had most flights

```
flights %>%
    count(month) %>%
    arrange(desc(n)) %>%
    head(3)
```

A data.frame: 3 × 2

	month	n
	<int>	<int>
1	8	28756
2	10	28618
3	7	28293

ANS : Aug , Oct , Jul had most flights in 2013.

 # Q2 which carrier had most flights in august 2013

```
flights %>%
  filter(month==8, year==2013) %>%
  count(month,carrier) %>%
  arrange(desc(n)) %>%
  head(3)
```

A data.frame: 3 × 3

	month	carrier	n
	<int>	<chr>	<int>
1	8	UA	5085
2	8	B6	4941
3	8	EV	4337

```
ANS : UA , B6 , EV is Top3 carrier in august 2013.
```

Q3 WHERE IS TOP DESINETION IN AUGUST 2013

```
flights %>%
  filter(month==8, year==2013) %>%
  count(dest) %>%
  arrange(desc(n)) %>%
  head(5)
```

A data.frame: 5 × 2

	dest	n
	<chr>	<int>
1	ORD	1569
2	LAX	1500
3	ATL	1488
4	BOS	1351
5	MCO	1256

ANS : O'Hare International Airport

Q 4 : Top 3 destinatination had longgest air time

```
flights %>%
    group_by(dest) %>%
    summarise(mean_arrtime = mean(arr_time)) %>%
    arrange(desc(mean_arrtime )) %>%
    head(3)
```

A tibble: 3 × 2

dest	mean_arrtime
<chr>	<dbl>
LEX	2227.000
SJC	2097.015
CRW	2057.806

ANS : LEX , SJC , CRW

Q5: What was top 5 popular destination of UA ?

```
flights %>%
  filter(carrier == "UA") %>%
  group_by(dest) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  head(5)
```

A tibble: 5 × 2

dest	n
<chr>	<int>
IAH	6814
ORD	6744
SFO	6728
LAX	5770
DEN	3737