🤓

Homework

Live 07 - Data Transformation

Homework 01 - ask 5 questions about this nycflights13 dataset

▼ **Q1. Top 5 longest and shortest destination**

```
airports_origin_name <- airports %>%
  mutate(origin = faa) %>%
  select(origin, name) %>%
  mutate(origin_airport = paste(name, "(", origin, ")"))
airports_origin_name

airports_dest_name <- airports %>%
  mutate(dest = faa) %>%
  select(dest, name) %>%
  mutate(dest_airport = paste(name, "(", dest, ")"))
airports_dest_name



longest_flights <- flights %>%
  left_join(airports_origin_name, by = "origin") %>%
  left_join(airports_dest_name, by = "dest")%>%
  distinct(origin_airport, dest_airport, distance) %>%
  select(origin_airport, dest_airport, distance) %>%
  arrange(desc(distance)) %>%
  head(5)

shortest_flights <- flights %>%
  left_join(airports_origin_name, by = "origin") %>%
  left_join(airports_dest_name, by = "dest")%>%
  distinct(origin_airport, dest_airport, distance) %>%
  select(origin_airport, dest_airport, distance) %>%
  arrange(distance) %>%
  head(5)
```

Q2. Top 5 popular origin and destination

```
popular_origin <- flights %>%
  left_join(airports_origin_name, by = "origin") %>%
  group_by(origin_airport)%>%
  summarise(n = n()) %>%
  mutate(pct = n/sum(n)) %>%
  arrange(desc(n))

popular_dest <- flights %>%
  left_join(airports_dest_name, by = "dest")%>%
  group_by(dest_airport)%>%
  summarise(n = n()) %>%
  mutate(pct = n/sum(n)) %>%
  arrange(desc(n)) %>%
  head(5)
```

Q3. flight carrier in month, day

```
## no. of flights in each month
by_month <- flights %>%
  group_by(month) %>%
  summarise( n = n()) %>%
  arrange(desc(n))
```

```
## Top 10 no. of flights in each day
by_day <- flights %>%
  mutate("date" = glue("{year}/{month}/{day}")) %>%
  group_by(date) %>%
  summarise(n=n()) %>%
  arrange(desc(n)) %>%
  head(10)
```

▼ Q4. Top 5 flight that always delay

```
## arrival delay (flight)
arr_delay_flight <- flights %>%
  mutate(flight_name = glue("{carrier}{flight}")) %>%
  left_join(airlines, by="carrier") %>%
  rename(airline_name = name) %>%
  select(airline_name, flight_name, arr_delay) %>%
  arrange(-arr_delay) %>%
  head(5)

## departure delay (flight)
dep_delay_flight <- flights %>%
  mutate(flight_name = glue("{carrier}{flight}")) %>%
  left_join(airlines, by="carrier") %>%
  rename(airline_name = name) %>%
  select(airline_name, flight_name, dep_delay) %>%
  arrange(-dep_delay) %>%
  head(5)

## average arrival delay time
avg_arr_delay <- flights %>%
  filter(arr_delay > 0) %>%
  summarise(avg_arr_delay = mean(arr_delay))

## average departure delay time
avg_dep_delay <- flights %>%
  filter(dep_delay > 0) %>%
  summarise(avg_dep_delay = mean(dep_delay))
```

▼ Q5. Top 5 carrier that always delay

```
## arrival delay (carrier)
arr_delay_carr <- flights %>%
  left_join(airlines, by="carrier") %>%
  rename(airline_name = name) %>%
  select(airline_name, arr_delay) %>%
  filter(arr_delay > 0) %>%
  group_by(airline_name) %>%
  summarise(avg_arr_delay = mean(arr_delay)) %>%
  arrange(-avg_arr_delay) %>%
  head(5)

## departure delay (carrier)
dep_delay_carr <- flights %>%
  left_join(airlines, by="carrier") %>%
  rename(airline_name = name) %>%
  select(airline_name, dep_delay) %>%
  filter(dep_delay > 0) %>%
  group_by(airline_name) %>%
  summarise(avg_dep_delay = mean(dep_delay)) %>%
  arrange(-avg_dep_delay) %>%
  head(5)
```

Homework 02 - restaurant pizza SQL

create 3-5 dataframes => write table into server

▼ Connect to PostgreSQL server

```
## connect to PostgreSQL server
library(RPostgreSQL)
library(tidyverse)

## create connection
con <- dbConnect(
  PostgreSQL(),
  host = "floppy.db.elephantsql.com", ## server
  dbname = "lpcdwaks",
  user = "lpcdwaks",
  password = "LcpGHW4RQUm2m_83ZZgGMhZTib9W8nZF",
  port = 5432
)
```

▼ **Create Dataframe**

```
## create dataframe
customers <- tribble(
  ~id, ~name, ~gender, ~age, ~birthdate, ~phone,
  '001', 'John', 'M', 23, '2000-01-01', '0123456789',
  '002', 'Thomas', 'M', 21, '2002-10-27', '0845569987',
  '003', 'Alex', 'M', 26, '1997-04-06', '0658745123',
  '004', 'Lily', 'F', 30, '1993-08-19', '0875569412',
  '005', 'Sonia', 'F', 53, '1970-06-08', '0569987451',
  '006', 'Zack', 'M', 46, '1977-02-28', '0547891236'
)
customers

menus <- tribble(
  ~menu_id, ~menu, ~price, ~menu_group,
  'M001', 'Smoked Salmon', 350, 'main dish',
  'M002','Margherita', 150, 'main dish',
  'M003','Pepperoni', 180, 'main dish',
  'M004','Hawaiian', 180, 'main dish',
  'M005','Vegetarian', 150, 'main dish',
  'M006','Quattro Formaggi', 200, 'main dish',
  'M007','BBQ Chicken', 230, 'main dish',
  'M008','Supreme', 300, 'main dish',
  'M009','White Pizza', 380, 'main dish',
  'M010','Spinach and Feta', 280, 'main dish',
  'M011','Mushroom and Onion', 200, 'main dish',
  'M012', 'Burger', 360, 'main dish',
  'M013', 'French Fried', 120, 'appitizer',
  'M014', 'Spaghetti', 250, 'main dish',
  'M015', 'Salad', 190, 'appitizer',
  'M016', 'Ice cream', 80, 'dessert',
  'M017', 'Cheesecake', 120, 'dessert',
  'M018', 'Taco', 150, 'appitizer',
  'M019', 'Spinach-Cheesse Bake', 180, 'appitizer',
  'M020', 'Fried Chicken', 160, 'appitizer',
  'M021', 'Coke', 20, 'beverage',
  'M022', 'Mineral water', 45, 'beverage',
  'M023', 'Sparkling water', 25, 'beverage'
)
menus

orders <- tribble(
  ~order_id, ~order_date, ~customer_id, ~menu_id, ~quantity
  'O001', '2022-01-01', '001', 'M001', 1, F,
  'O001', '2022-01-01','001', 'M002', 1, F,
  'O001', '2022-01-01','001', 'M015', 2, F,
  'O001', '2022-01-01','001', 'M016', 5, F,
  'O001', '2022-01-01','001', 'M023', 10, F,
  'O002', '2022-01-01','002', 'M001', 2, T,
  'O003', '2022-02-14','003', 'M018', 3, T,
  'O003', '2022-02-14','003', 'M022', 1, T,
  'O004', '2022-03-01','004', 'M006', 2, F,
  'O004', '2022-03-01','004', 'M002', 2, F,
  'O004', '2022-03-01','004', 'M003', 2, F,
  'O004', '2022-03-01','004', 'M001', 1, F,
  'O004', '2022-03-01','004', 'M021', 4, F,
  'O004', '2022-03-01','004', 'M023', 3, F,
  'O005', '2022-03-01','004', 'M001', 2, T,
  'O006', '2022-04-12','005', 'M001', 1, T,
  'O006', '2022-04-12','005', 'M002', 1, T,
  'O006', '2022-04-12','005', 'M003', 1, T,
  'O006', '2022-04-12','005', 'M004', 1, T,
  'O006', '2022-04-12','005', 'M005', 1, T,
  'O006', '2022-04-12','005', 'M006', 1, T,
  'O006', '2022-04-12','005', 'M007', 1, T,
  'O006', '2022-04-12','005', 'M008', 1, T,
  'O006', '2022-04-12','005', 'M009', 1, T,
  'O006', '2022-04-12','005', 'M010', 1, T,
  'O006', '2022-04-12','005', 'M011', 1, T,
  'O006', '2022-04-12','005', 'M014', 1, T,
  'O006', '2022-04-12','005', 'M020', 1, T,
  'O006', '2022-04-12','005', 'M017', 3, T,
  'O006', '2022-04-12','005', 'M019', 2, T,
  'O007', '2022-04-30','006', 'M017', 1, F,
  'O007', '2022-04-30','006', 'M001', 1, F,
  'O007', '2022-04-30','006', 'M023', 1, F,
  'O007', '2022-04-30','006', 'M015', 1, F,
  'O007', '2022-04-30','006', 'M023', 1, F
)
orders
```

Write tables in DB

```
## db write tables
dbWriteTable(con, "customers", customers)
dbWriteTable(con, "menus", menus)
dbWriteTable(con, "orders", orders)

## see the tables in db
dbListTables(con)

## remove tables
dbRemoveTable(con, "products")
```

Write SQL by using dbGetQuery

```
dbGetQuery(con, "select * from customers")

## find overall earning
quant_menus <- dbGetQuery(con, "select
                  a.menu_id,
                  b.menu,
                  sum(a.quantity) as total,
                  sum(a.quantity)* b.price as grand_total
                  from orders a join menus b
                  on a.menu_id = b.menu_id
                  group by a.menu_id, b.menu, b.price
                  order by total desc")
```

```
dbWriteTable(con, "quant_menus", quant_menus)

dbGetQuery(con, "select sum(grand_total) from quant_menus"
```