# Progress Report

Team 12: Joanna Fang, Wenjin Jiang, Eri Jibiki, Roy Song, Michael Wang, Kate Zhang

2024-04-10

```
# load packages
library(tidyverse)
library(dplyr)
library(knitr)
library(broom)
library(ggplot2)
```

## Background

This dataset, titled 'Airline Passenger Satisfaction', is a comprehensive collection of data that aims to capture and analyze the satisfaction levels of airline passengers. It includes 23 variables such as age, gender, type of travel (personal or business), class of service (Economy, Business, etc.), flight distance, in-flight services, cleanliness, and overall satisfaction. It consists of survey responses from over 130,000 participants, which is divided into training (80%) and testing (20%) sets.

This dataset is likely from an airline or research group focused on enhancing passenger satisfaction. In the competitive airline industry, understanding and improving passenger satisfaction is crucial for retaining loyalty, enhancing brand reputation, and securing a competitive edge. The dataset serves as a critical tool for improving strategic plan, identifying key satisfaction drivers, and facilitating improvements to better meet and exceed passenger expectations.

This dataset is hosted on Kaggle. (https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction?select=train.csv)

```
# import data
airlines_data <- read.csv("train.csv")
```

## Hypothesis List

1. "Inflight Entertainment" is more important for the satisfaction rate of vacation travelers than it is for business travelers.
2. Online booking convenience is more important for younger customers than older customers.
3. "Seat Comfort" is more important for travelers in higher classes (e.g., Business, Eco Plus) than for those in lower classes.
4. Customers in loyal program are more satisfied with the flight experience.

## Statistical Models

We will conduct a logistic regression using Customer Type as the response variable, and all other variables as the independent variables to explore factors influencing customer satisfaction.

To test our hypotheses, we will add interaction terms of "inflight entertainment x type of traveler", "ease of online booking x age", and "seat comfort x class" to the logistic regression model. (See Appendix)

**Statistical Model for Customer Satisfaction**:

$$log(\frac{P}{1-P}) = \beta_0 + \beta_1 Gender + \beta_2 Age + \beta_3 Class + \beta_4 Type\_of\_Travel + \beta_5 Gate\_location$$

$$+\beta_6 Inflight\_wifi\_service + \beta_7 Departure\_Arrival\_time\_convenient + \beta_8 Ease\_of\_Online\_booking$$

$$+\beta_9 Food\_and\_drink + \beta_{10} Online\_boarding + \beta_{11} Seat\_comfort + \beta_{12} Inflight\_entertainment$$

$$+\beta_{13} Onboard\_service + \beta_{14} Baggage\_handling + \beta_{15} Checkin\_service + \beta_{16} Inflight\_service$$

$$+\beta_{17} Cleanliness + \beta_{18} Departure\_Delay\_in\_Minutes + \beta_{19} Arrival\_Delay\_in\_Minutes + \beta_{20} satisfaction$$

$$+\beta_{21} Inflight\_entertainment \times Type\_of\_Travel + \beta_{22} Ease\_of\_Online\_booking \times Age + \beta_{23} Seat\_comfort \times Class$$

## Preliminary Results

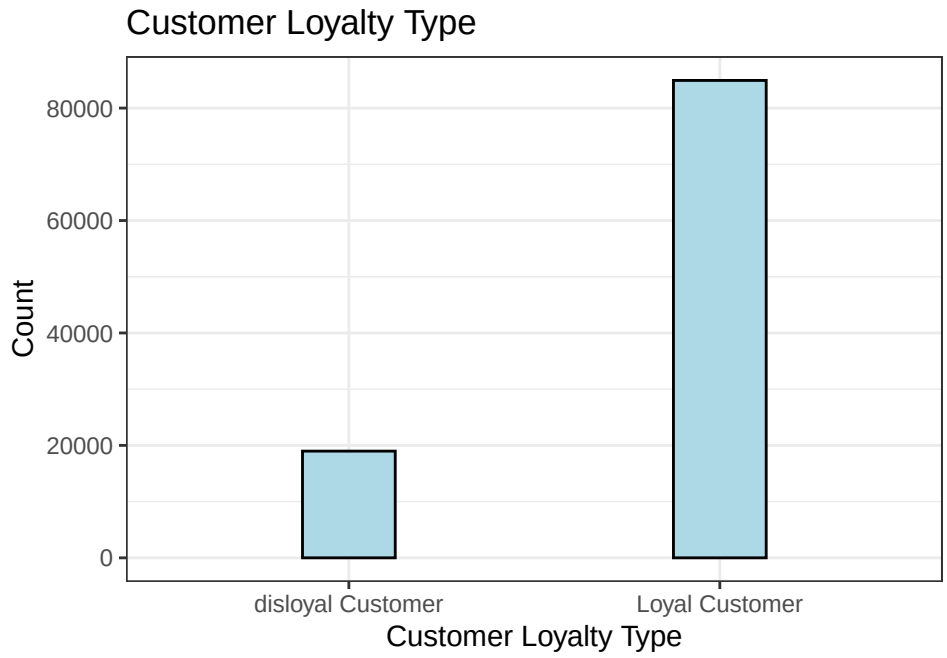
1. The interaction between "Type.of.Travel(Personal Travel)" and "Inflight.entertainment" ($\beta_{21}$) is statistically significant and negative, meaning inflight entertainment is actually more important for business traveler than for personal traveler. This contradicts our hypothesis.
2. The interaction between "Age" and "Ease of Online Booking" ($\beta_{22}$) is statistically significant and positive, but it is very small (0.0031). The ease of online booking is more important for elderly people than young people, but the influence is small.
3. The interaction between "Class" and "Seat Comfort" ($\beta_{23}$) is statistically significant and negative, meaning seat comfort is more important for business traveler than for Eco Plus traveler than for Eco traveler. This is consistent with our hypothesis.
4. The coefficient for customer type ($\beta_{20}$) is positive, meaning loyal customers do have a higher satisfaction rate than disloyal customers. This is consistent with our hypothesis

## Appendix: Exhibition

### Preliminary Data Exploration and Regression

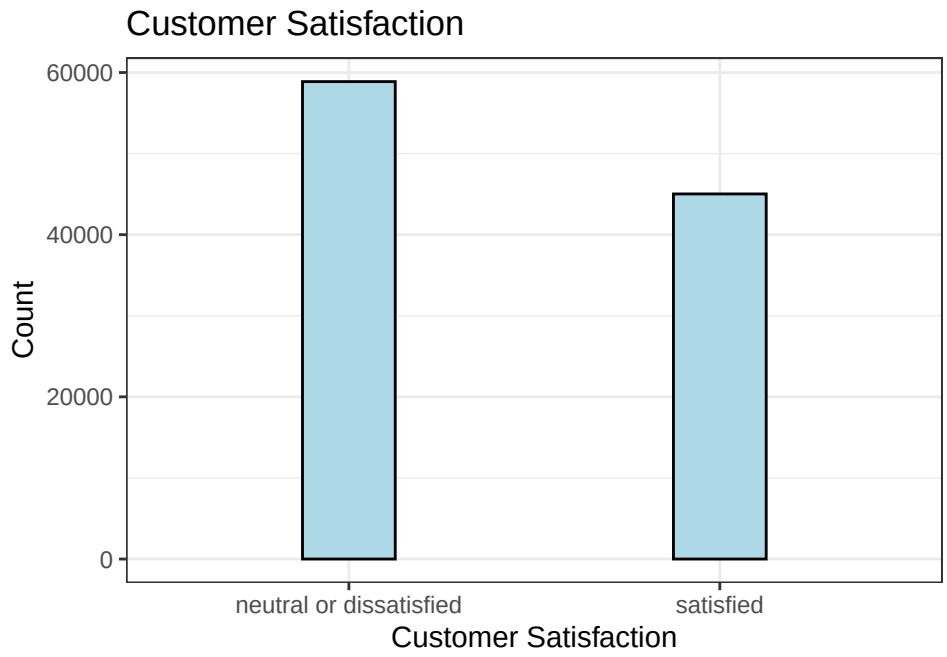
```
airlines_data <- airlines_data %>%
  mutate(customer_type_num = if_else(Customer.Type == "Loyal Customer", 1, 0))
```

```
ggplot(airlines_data, aes(x = Customer.Type)) + geom_bar(width = 0.25,
                                          fill = "lightblue", color = "black")+
  theme_bw()+
  labs(title = "Customer Loyalty Type",
       x = "Customer Loyalty Type",
       y = "Count")
```

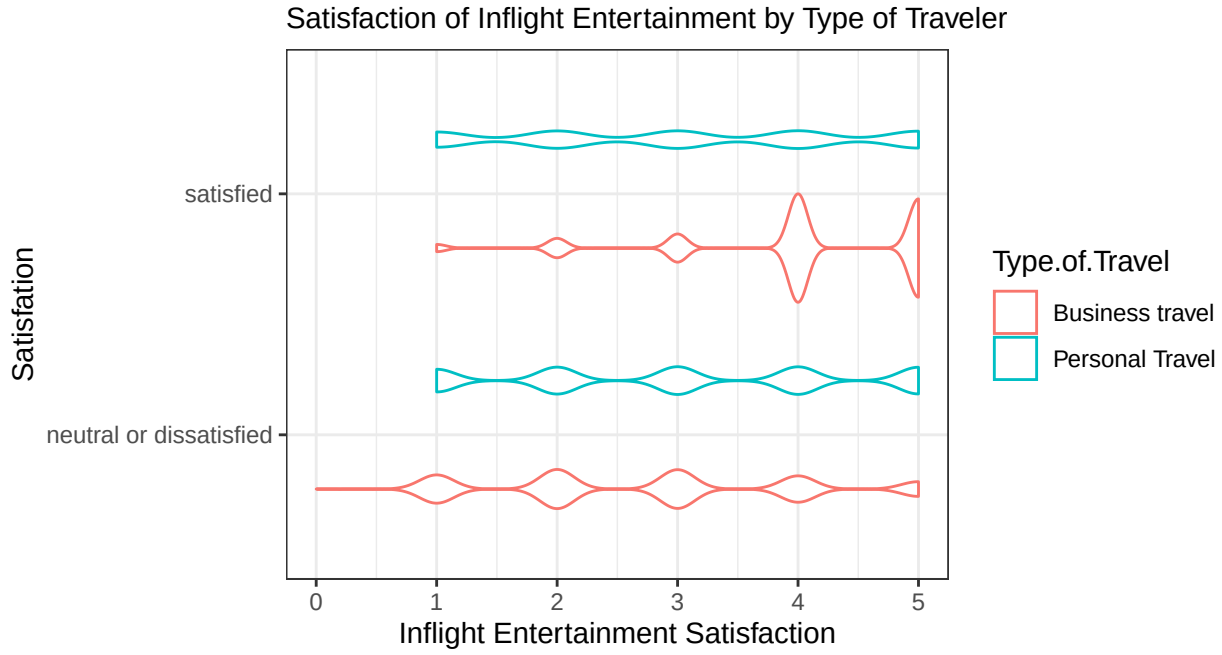



```
airlines_data <- airlines_data %>%
  mutate(satisfaction_num = if_else(satisfaction == "satisfied", 1, 0))
```

```
ggplot(airlines_data, aes(x = satisfaction)) + geom_bar(width = 0.25,
                                          fill = "lightblue", color = "black")+
  theme_bw()+
  labs(title = "Customer Satisfaction",
       x = "Customer Satisfaction",
       y = "Count")
```

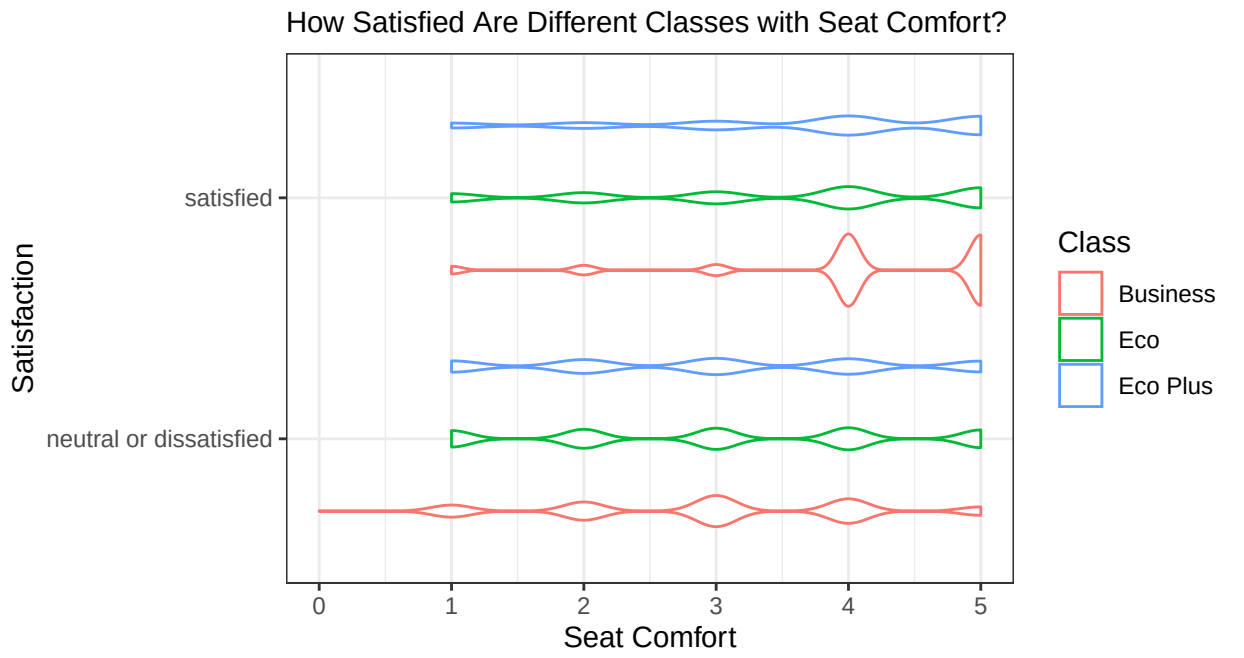



```
ggplot(airlines_data, aes(x = Inflight.entertainment, y = satisfaction, color = Type.of.Travel)) +
  geom_violin()+
  theme_bw()+
  labs(subtitle = "Satisfaction of Inflight Entertainment by Type of Traveler",
       x = "Inflight Entertainment Satisfaction",
       y = "Satisfation")
```




Business travelers who are satisfied with their trip tend to have a higher satisfaction rate with inflight entertainment compared to personal travelers. This indicates interactions between "inflight entertainment" and "type of traveler"

```
ggplot(airlines_data, aes(x = Seat.comfort, y = satisfaction, color = Class)) +
  geom_violin()+
  theme_bw()+
  labs(subtitle = "How Satisfied Are Different Classes with Seat Comfort?",
       x = "Seat Comfort",
       y = "Satisfaction")
```




Business class travelers tend to have a different pattern of seat comfort satisfaction in their overall satisfaction rate compared to Eco and Eco Plus travelers. This indicates potential interactions between "seat comfort" and "class"

### Model

```
logit2 <- glm(satisfaction_num ~ Gender + Age + Type.of.Travel + Class + Flight.Distance
              + Inflight.wifi.service + Departure.Arrival.time.convenient + Ease.of.Online.booking
              + Food.and.drink + Online.boarding + Seat.comfort + Inflight.entertainment
              + On.board.service + Baggage.handling + Checkin.service + Inflight.service
              + Cleanliness + Departure.Delay.in.Minutes + Arrival.Delay.in.Minutes
              + customer_type_num + Inflight.entertainment*Type.of.Travel +
                Ease.of.Online.booking*Age + Seat.comfort*Class,
              family=binomial,data=airlines_data)
tidy(logit2) %>% kable(digits = 4)
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -7.8597 | 0.1007 | -78.0592 | 0.0000 |
| GenderMale | 0.0197 | 0.0197 | 1.0002 | 0.3172 |
| Age | -0.0168 | 0.0017 | -9.7264 | 0.0000 |

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| Type.of.TravelPersonal Travel | -0.7506 | 0.0653 | -11.4957 | 0.0000 |
| ClassEco | 0.0110 | 0.0626 | 0.1759 | 0.8604 |
| ClassEco Plus | -0.4857 | 0.1175 | -4.1348 | 0.0000 |
| Flight.Distance | 0.0000 | 0.0000 | -1.0644 | 0.2871 |
| Inflight.wifi.service | 0.3804 | 0.0118 | 32.3134 | 0.0000 |
| Departure.Arrival.time.convenient | -0.1282 | 0.0078 | -16.4282 | 0.0000 |
| Ease.of.Online.booking | -0.2172 | 0.0240 | -9.0541 | 0.0000 |
| Food.and.drink | -0.0334 | 0.0108 | -3.0962 | 0.0020 |
| Online.boarding | 0.5763 | 0.0106 | 54.4926 | 0.0000 |
| Seat.comfort | 0.1700 | 0.0138 | 12.3254 | 0.0000 |
| Inflight.entertainment | 0.3063 | 0.0151 | 20.2447 | 0.0000 |
| On.board.service | 0.3117 | 0.0102 | 30.4184 | 0.0000 |
| Baggage.handling | 0.1507 | 0.0115 | 13.1312 | 0.0000 |
| Checkin.service | 0.3284 | 0.0086 | 37.9954 | 0.0000 |
| Inflight.service | 0.1456 | 0.0121 | 12.0392 | 0.0000 |
| Cleanliness | 0.2206 | 0.0122 | 18.1390 | 0.0000 |
| Departure.Delay.in.Minutes | 0.0050 | 0.0010 | 4.9586 | 0.0000 |
| Arrival.Delay.in.Minutes | -0.0091 | 0.0010 | -9.2024 | 0.0000 |
| customer_type_num | 1.9850 | 0.0303 | 65.5401 | 0.0000 |
| Type.of.TravelPersonal Travel:Inflight.entertainment | -0.5892 | 0.0181 | -32.4965 | 0.0000 |
| Age:Ease.of.Online.booking | 0.0031 | 0.0005 | 6.1089 | 0.0000 |
| ClassEco:Seat.comfort | -0.2388 | 0.0174 | -13.7457 | 0.0000 |
| ClassEco Plus:Seat.comfort | -0.1223 | 0.0329 | -3.7175 | 0.0002 |