

Final Project Proposal

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

2024-02-29

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

Data Set Introduction

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

```
# import data
airlines_data <- read.csv("train.csv")
starbucks_data <- read.csv("Starbucks satisfactory survey encode cleaned.csv")
```

Variables Description

Important Variables:

1. Gender: Gender of the passengers (Female, Male)
2. Customer Type: The customer type (Loyal customer, disloyal customer)
3. Age: The actual age of the passengers
4. Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)
5. Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)
6. Flight distance: The flight distance of this journey
7. Inflight wifi service: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
8. Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient
9. Ease of Online booking: Satisfaction level of online booking
10. Gate location: Satisfaction level of Gate location
11. Food and drink: Satisfaction level of Food and drink
12. Online boarding: Satisfaction level of online boarding
13. Seat comfort: Satisfaction level of Seat comfort
14. Inflight entertainment: Satisfaction level of inflight entertainment

15. On-board service: Satisfaction level of On-board service
16. Leg room service: Satisfaction level of Leg room service
17. Baggage handling: Satisfaction level of baggage handling
18. Check-in service: Satisfaction level of Check-in service
19. Inflight service: Satisfaction level of inflight service
20. Cleanliness: Satisfaction level of Cleanliness
21. Departure Delay in Minutes: Minutes delayed when departure
22. Arrival Delay in Minutes: Minutes delayed when Arrival
23. Satisfaction: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

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 loyalty.

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

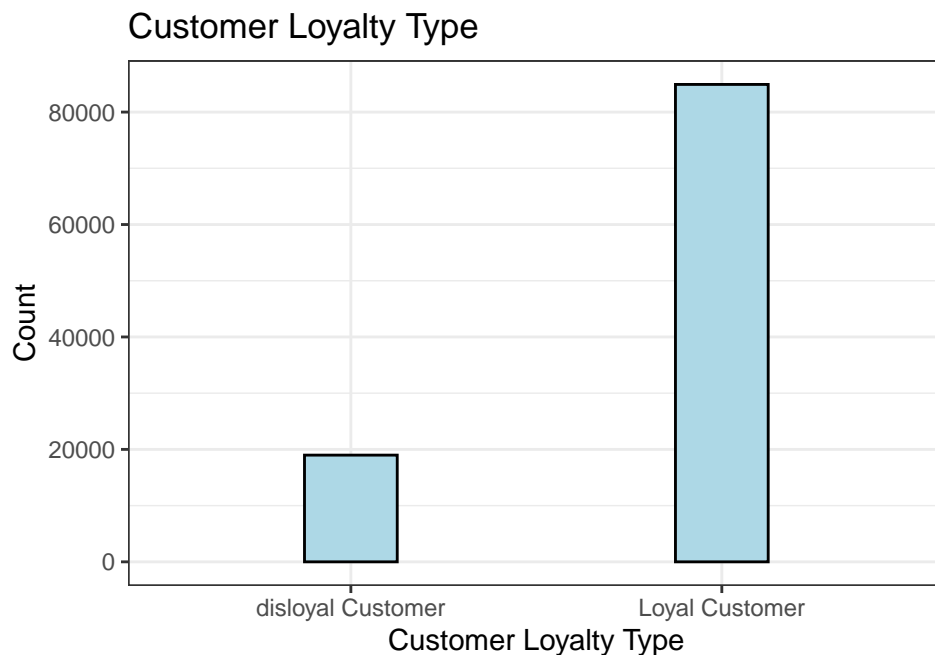
Questions to Explore

1. Which factors influence customer satisfaction the most? How to improve customer satisfaction in the future?
2. Which factors influence customer loyalty the most? How to improve customer loyalty in the future?

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")
```



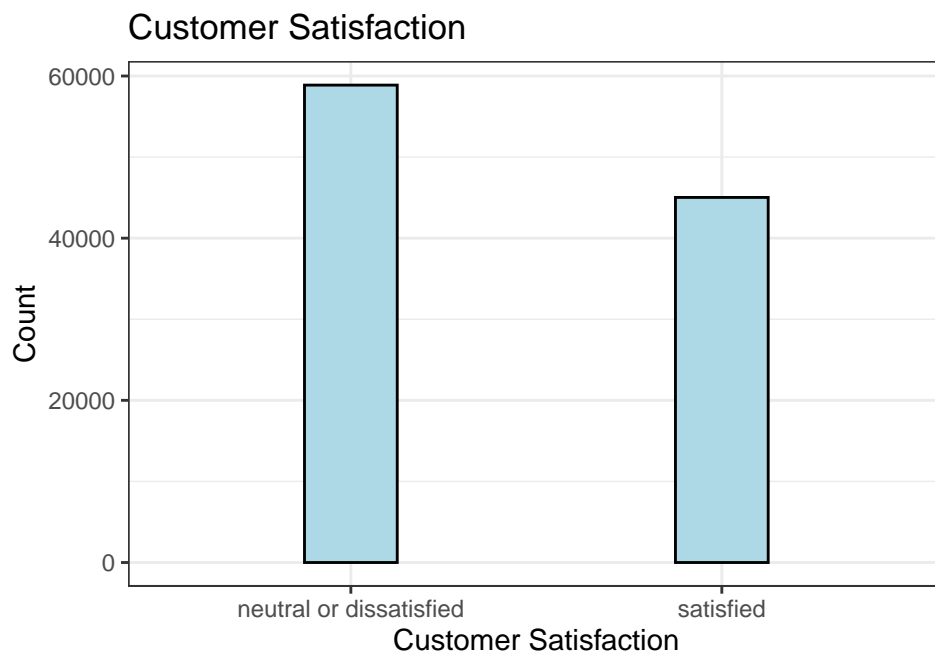
```
logit <- glm(customer_type_num ~ .-Customer.Type-id-X, family=binomial,  
  data=airlines_data)  
tidy(logit) %>% kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-1.07	0.09	-11.59	0.00
GenderMale	0.20	0.02	8.40	0.00
Age	0.08	0.00	75.15	0.00
Type.of.TravelPersonal Travel	7.11	0.09	80.79	0.00
ClassEco	-0.81	0.03	-27.91	0.00
ClassEco Plus	0.98	0.05	18.21	0.00
Flight.Distance	0.00	0.00	58.75	0.00
Inflight.wifi.service	-0.45	0.02	-29.72	0.00
Departure.Arrival.time.convenient	0.30	0.01	28.06	0.00

term	estimate	std.error	statistic	p.value
Ease.of.Online.booking	-0.06	0.01	-3.97	0.00
Gate.location	-0.02	0.01	-1.58	0.11
Food.and.drink	-0.20	0.02	-12.25	0.00
Online.boarding	0.42	0.01	28.59	0.00
Seat.comfort	0.11	0.02	7.40	0.00
Inflight.entertainment	0.60	0.02	27.83	0.00
On.board.service	-0.19	0.01	-16.66	0.00
Leg.room.service	0.00	0.01	0.04	0.97
Baggage.handling	-0.31	0.01	-22.57	0.00
Checkin.service	-0.27	0.01	-25.99	0.00
Inflight.service	-0.31	0.01	-22.25	0.00
Cleanliness	-0.43	0.02	-21.48	0.00
Departure.Delay.in.Minutes	0.00	0.00	-3.59	0.00
Arrival.Delay.in.Minutes	0.01	0.00	5.05	0.00
satisfactionsatisfied	2.16	0.03	65.07	0.00

```
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")
```



```
logit2 <- glm(satisfaction_num ~ .-Customer.Type-satisfaction-id-X,
  family=binomial,data=airlines_data)
tidy(logit2) %>% kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-7.86	0.08	-99.79	0.00
GenderMale	0.04	0.02	2.18	0.03
Age	-0.01	0.00	-11.68	0.00
Type.of.TravelPersonal Travel	-2.72	0.03	-86.49	0.00
ClassEco	-0.74	0.03	-28.79	0.00
ClassEco Plus	-0.86	0.04	-20.59	0.00
Flight.Distance	0.00	0.00	-1.58	0.11
Inflight.wifi.service	0.39	0.01	34.40	0.00
Departure.Arrival.time.convenient	-0.12	0.01	-15.13	0.00
Ease.of.Online.booking	-0.14	0.01	-12.69	0.00
Gate.location	0.03	0.01	3.18	0.00
Food.and.drink	-0.03	0.01	-2.68	0.01
Online.boarding	0.61	0.01	59.77	0.00
Seat.comfort	0.07	0.01	5.86	0.00
Inflight.entertainment	0.07	0.01	4.59	0.00
On.board.service	0.30	0.01	29.58	0.00
Leg.room.service	0.25	0.01	29.65	0.00
Baggage.handling	0.13	0.01	11.63	0.00
Checkin.service	0.32	0.01	37.76	0.00
Inflight.service	0.12	0.01	10.02	0.00
Cleanliness	0.22	0.01	18.47	0.00
Departure.Delay.in.Minutes	0.00	0.00	4.82	0.00
Arrival.Delay.in.Minutes	-0.01	0.00	-9.66	0.00
customer_type_num	2.04	0.03	67.97	0.00