**Team Alpha**

Team members**:** *Yongshi Qiu, Hoihin Lau(Edwin), Wanying Li, Eric Sedaghat, Wanchang Wu*

**Business Problem and Overview**

As of late airlines have been struggling more than ever due to the recent pandemic. As a result, the margin for error in the airline industry has reduced immensely. With the relative value for each customer has grown due to the drastic decrease in demand for travel. Now airlines must be sure to optimize each customer's experience in order to ensure that those who fly with them are satisfied and continue to fly with them through these trying times. Being that airlines only have so many resources and cannot optimize every aspect of a flight experience, which is important to identify what factors carry the most weight in regard to satisfying the customers' needs and wants so that they may allocate their resources more efficiently.

Our team intends to utilize effective data mining techniques on the data available in order to ensure that a customer is satisfied and identify which factors contribute the most to a satisfactory experience. If it is done well, we will be able to select a random customer and identify their satisfaction level based on a myriad of attributes provided such as age, class, distance, and a myriad of other factors. This can then help operations to maintain customers’ loyalty and attract potential newcomers. We believe that this analysis can significantly help airlines efficiently allocate their resources to best serve their customers.

**Data**

The data was given by an airline organization, it consists of the details of customers who have already flown with them and whether they were satisfied with the services. This dataset contains 129881 rows and a total of 23 attributes, with 393 missing values in the column of “Arrival Delay in Minutes”. There are 18 numerical variables such as flight distance and ratings of food and drink and 5 categorical variables such as gender, customer type. Our target variable is satisfaction, which is a binary variable that only has two values: satisfied and dissatisfied. In this project, we will leverage 22 attributes consisting of customer characteristics such as age, gender, customer type and flight information, ratings of services provided to determine the key factors that affect customer satisfaction and help the airline organization to emphasize on those aspects of services to make their customers satisfied.

To sum up, we will be using data mining tools to identify customer satisfaction levels for airline customers: a classification analysis using supervised learning models to find out the most accurate prediction of our target outcome, satisfaction(1). The independent variables(2-23) are listed in the appendix, and the dependent outcome is (1) satisfaction. For the full list, please see Appendix A.

**Appendix A: Attributes**

1.satisfaction

2.Gender

3.Customer Type

4.Age

5.Type of Travel

6.Class

7.Flight Distance

8.Seat comfort

9.Departure/Arrival time convenient

10.Food and drink

11.Gate location

12.Inflight wifi service

13.Inflight entertainment

14.Online support

15.Ease of Online booking

16.On-board service

17.Leg room service

18.Baggage handling

19.Checkin service

20.Cleanliness

21.Online boarding

22.Departure Delay in Minutes

23.Arrival Delay in Minutes