

# **Business Analytics Group Project Proposal**

## **Team (12) - Space Invaders**

Ayah Assaadi

Kaya Aslanoba

Roch Jia

J.R. Noh

Lessly Rocha

### **Business Problem:**

The airline industry was devastated by COVID -19 pandemic, which forced many airlines around the globe to file for bankruptcy. Luckily, those days are past and in this time of recovery we, as Invistico Airlines, want to ensure that our customers are satisfied with the service that we provide. However, after interpreting our passenger response data, we noticed that 45% of our customers are dissatisfied with our services. From a business perspective, we want to maximize customer satisfaction in order to increase our revenue; especially in the commercial airline industry where companies face high fixed costs and low-profit margins.

To make sure we improve customer satisfaction levels, we will provide the Invistico Airlines management a two-pronged approach with preventative and immediate actions. The preventive action involves creating a predictive model to determine whether or not a future customer will be satisfied with their service. This model will allow us to measure our future customer satisfaction rates based on current practices and will further enable Investico to change those practices in response. The immediate actions are business recommendations based on the correlations between customer satisfaction and the customer experience variables in the data set. The business recommendations will provide immediate action the company can take in the interim to decrease its dissatisfied customer base.

To begin, our group looked for large trends in the dataset that would immediately point to areas of improvement for Invistico. We used Seaborn visualizations of each experience variable by satisfaction but did not find any factors that caused dissatisfaction.

Since there were no overwhelming trends, we then looked for a specific target segment to address. Using Python, we segmented customers first by satisfaction, and then segmented those unsatisfied by gender, class, and customer type. We found that dissatisfied loyal customers made up 31% of our dataset, larger than any other segment of unsatisfied customers. We were concerned by loyal customers dominating the unsatisfied segment and decided to focus on them.

### **Modeling Ideas:**

This is a classification problem to see whether a customer is satisfied or unsatisfied with the services provided [Yes=1, No=0]. Since we will be using a labeled data set to train an algorithm, with a target attribute, this is a supervised machine learning task. Our target variable is satisfaction and we believe that variables like age, gender, seat class, flight

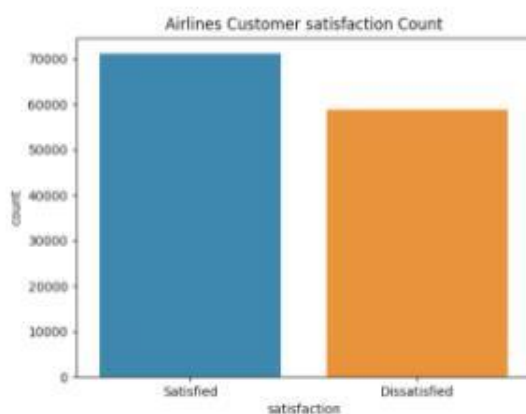
distance, seat comfort, departure/arrival time convenience, food and drink, gate location, inflight wifi service, inflight entertainment, online support, ease of online booking, on-board service, leg-room service, baggage handling, check-in service, cleanliness, online boarding, departure delay in minutes, and arrival delay in minutes are variables that will be useful in predicting the target variable, satisfaction.

## Data Details

Our group will be using an imaginary airline, named Invistico Airlines, consisting of details and feedback from customers who have already flown with them. This dataset was found on Kaggle.com.

There are 22 features/attributes, most of them being integer types (ratings of food & drinks, baggage handling, cleanliness...) , nominal (travel type and flight class), and binomial (satisfaction, gender). We have a total of 129,880 examples in the dataset.

Our group is planning on working on a classification problem (as stated above - the satisfaction of customers), and the ratio of our most prevalent class is 70%/30% for the less prevalent class. We chose this ratio but will be playing around with other ratios to see which ones would improve accuracy, recall, and f-measure in general.



```
wifi mean: 2.874438140151794
entertainment mean: 2.690344607373566
food mean: 2.778571954903839
on board service mean: 2.9234151253899245
cleanliness mean: 3.2902758332719277
legroom mean: 3.000957924987105
seat comfort mean: 2.500135091985361
convenience of departure/arrival mean: 3.3419792203964334
age mean: 40.28531427308231
online support mean: 2.94102620784516
ease of online booking mean: 2.7678382826124333
baggage handling mean: 3.275759585390416
checkin service mean: 2.931889077198929
online boarding mean: 2.788396826566453
```

```
SATISFACTION STATISTICS BY CUSTOMER TYPE
Total loyal customers: 106100
Percentage loyal customers: 81.69079149984601%
Total dissatisfied loyal customers: 40713
Percentage of loyal customers who are dissatisfied: 38.37229029217719%
Overall percent dissatisfied loyal customers: 31.34662765629812 %
```

