Customer Satisfaction in Invistico Airline

Problem Definition

Invistico Airline Data has flight information for over 120,000 flights and their corresponding satisfaction metric. Our aim with this project is to understand the various aspects of an in-flight experience which influence the overall satisfaction level of a customer. By understanding the factors that impact satisfaction, we want to be able to help the Invistico improve and maintain the key factors. We understand that satisfaction is fairly subjective, and so this analysis will mainly be focused on factors that are in control of the airlines. We will conduct an analysis to determine the effect of various components related to flight travel that helps decide if a passenger will be satisfied or dissatisfied with the flight experience. With this analysis, we will address the following:

- Which flight components had the most influence on customer satisfaction?
- Which features an airline can consider inconsequential?
- What are the most optimum feature values to ensure a customer is satisfied?
- Make recommendations for airlines to improve features that have the most impact on customer satisfaction.

Data Source

This dataset is from Kaggle and is owned by Yakhyojon. It can be accessed here - https://www.kaggle.com/datasets/yakhyojon/customer-satisfaction-in-airline.

Dataset

Our dataset contains 21 features and 129880 entries. The 'satisfaction' column will be our target variable. It includes data points such as class, flight distance, and inflight entertainment to be used to predict whether a customer will be satisfied with their flight experience.

Field	Description	Datatype
satisfaction	Customer Satisfaction (Satisfied or Dissatisfied)	object
Customer Type	Type of customer (Loyal or Disloyal)	object
Age	Passenger Age (in Years)	int
Type of Travel	Purpose of Travel (Personal or Business)	object
Class	Flight class (Business, Eco or Eco Plus)	object
Flight Distance	Distance covered by flight (Miles)	int
Seat comfort	Level of seat comfort in flight (Rating: 0 to 5)	int
Departure/Arrival time convenient	Flight time convenience (Rating: 0 to 5)	int
Food and drink	Convenience (Rating: 0 to 5)	int
Gate location	Ease of accessibility to gates (Rating: 0 to 5)	int
Inflight wifi service	Wifi convenience (Rating: 0 to 5)	int
Inflight entertainment	Entertainment Services (Rating: 0 to 5)	int
Online support	Online customer support (Rating: 0 to 5)	int

Ease of Online booking	Ease of online booking (Rating: 0 to 5)	int
On-board service	On-board services (Rating: 0 to 5)	int
Leg room service	Leg room availability (Rating: 0 to 5)	int
Baggage handling	Baggage handling satisfaction (Rating: 0 to 5)	int
Checkin service	Check-in service (Rating: 0 to 5)	int
Cleanliness	Level of cleanliness (Rating: 0 to 5)	int
Online boarding	Ease of online boarding (Rating: 0 to 5)	int
Departure Delay in Minutes	Departure Delay (in Minutes)	int
Arrival Delay in Minutes	Arrival Delay (in Minutes)	float

Possible Analysis

Our objective in this analysis is to predict overall customer satisfaction level, categorizing passengers into 'satisfied' or 'not satisfied,' based on their interactions with the airline's end-to-end services. We'll conduct a comprehensive examination of the dataset, exploring the impact of various features on the target variable.

Our approach involves leveraging machine learning techniques, with a focus on classification models, such as Logistic Regression, to gain insights into the factors influencing passenger satisfaction. Through this analysis, we aim to uncover key drivers of customer contentment and develop a predictive model that can aid in improving the airline's services and enhancing overall passenger experience.

Potential Implications

Our hope with this analysis is to equip airlines with valuable insights regarding the impact of their services on customer satisfaction. Customer satisfaction is a combination of various amenities and services provided by the airlines. By delving into the relationships between these factors and our target variable, namely, satisfaction, we can propose enhancements and modifications to both the services and the overall customer experience.

Upon implementing our recommendations, the airline can make necessary adjustments to attain the optimal level of satisfied customers. This, in turn, will help them in achieving their goal of maximizing customer satisfaction.

Appendix A

Colab Notebook Link

Our Collab Notebook can be accessed via this link.

Data Loading and Summary Statistics

Fig.1: Data Loading:

```
# Mounting the drive and Loading the dataset
from google.colab import drive
import pandas as pd

drive.mount('/content/gdrive', force_remount=True)
path = '/content/gdrive/MyDrive/BA810-B12-Team_Project/'

airline = pd.read_csv(path + 'Airline satisfaction/Invistico_Airline.csv')
```

```
Mounted at /content/gdrive
```

Fig. 2: Summary Statistics for Numerical Variables:

```
summary_stats = pd.DataFrame(round(airline.describe(),2))
summary_stats_numeric = summary_stats.transpose()
summary_stats_numeric
```

1 to 18 of 18 entries Filter

index	count	mean	std	min	25%	50%	75%	max
Age	129880.0	39.43	15.12	7.0	27.0	40.0	51.0	85.0
Flight Distance	129880.0	1981.41	1027.12	50.0	1359.0	1925.0	2544.0	6951.0
Seat comfort	129880.0	2.84	1.39	0.0	2.0	3.0	4.0	5.0
Departure/Arrival time convenient	129880.0	2.99	1.53	0.0	2.0	3.0	4.0	5.0
Food and drink	129880.0	2.85	1.44	0.0	2.0	3.0	4.0	5.0
Gate location	129880.0	2.99	1.31	0.0	2.0	3.0	4.0	5.0
Inflight wifi service	129880.0	3.25	1.32	0.0	2.0	3.0	4.0	5.0
Inflight entertainment	129880.0	3.38	1.35	0.0	2.0	4.0	4.0	5.0
Online support	129880.0	3.52	1.31	0.0	3.0	4.0	5.0	5.0
Ease of Online booking	129880.0	3.47	1.31	0.0	2.0	4.0	5.0	5.0
On-board service	129880.0	3.47	1.27	0.0	3.0	4.0	4.0	5.0
Leg room service	129880.0	3.49	1.29	0.0	2.0	4.0	5.0	5.0
Baggage handling	129880.0	3.7	1.16	1.0	3.0	4.0	5.0	5.0
Checkin service	129880.0	3.34	1.26	0.0	3.0	3.0	4.0	5.0
Cleanliness	129880.0	3.71	1.15	0.0	3.0	4.0	5.0	5.0
Online boarding	129880.0	3.35	1.3	0.0	2.0	4.0	4.0	5.0
Departure Delay in Minutes	129880.0	14.71	38.07	0.0	0.0	0.0	12.0	1592.0
Arrival Delay in Minutes	129487.0	15.09	38.47	0.0	0.0	0.0	13.0	1584.0

Fig.3: Summary Statistics for Categorical Variables:

```
summary_stats = pd.DataFrame(airline[['satisfaction', 'Customer Type', 'Type of Travel', 'Class']].describe())
summary_stats_cat = summary_stats.transpose()
summary_stats_cat
```

1 to 4 of 4 entries Filter

index	count	unique	top	freq
satisfaction	129880	2	satisfied	71087
Customer Type	129880	2	Loyal Customer	106100
Type of Travel	129880	2	Business travel	89693
Class	129880	3	Business	62160

Fig.4: Distribution of the target variable in the dataset: