# Executive Summary

The main objective of this study was to create a consultancy report for Skywards International Airlines Corporation using advanced data analytics strategies and machine learning algorithms to objectively identify and predict factors influencing customer satisfaction. A dataset comprising 103903 rows and columns with features such as customer ID, gender, age, type of travel, flight distance, inflight Wi-Fi service, type of onboarding, and delays, among other factors, was analyzed. Key variables that impact customer satisfaction were first analyzed, showing various factors influencing customer satisfaction, such as the booking processes, travel experience, delays, and the quality of inflight services.

Table of Contents

[Executive Summary 1](#_Toc178155700)

[Introduction 3](#_Toc178155701)

[Key factors that impact on passage satisfaction 3](#_Toc178155702)

[Tasks 3](#_Toc178155703)

[Extract, Transform, and Load (ETL) Phase 3](#_Toc178155704)

[Exploratory Visualizations 4](#_Toc178155705)

[Analytical Models Selection 7](#_Toc178155706)

[Models Evaluation 8](#_Toc178155707)

[Recommendations 10](#_Toc178155708)

[Next Steps 11](#_Toc178155709)

# Introduction

Skywards International Airlines Corporation embarked on a journey to elevate its levels of customer satisfaction by leveraging the power of data analytics and machine learning models. This decision was in line with previous studies that have shown that airlines prioritizing customer satisfaction end up having more customers and becoming profitable (Tahanisaz, 2020). Through this comprehensive data analysis report, the management of Skywards International Airlines Corporation was able to unveil the key factors contributing to passenger satisfaction and formulate actionable recommendations and next steps for the organization going forward.

# Key factors that impact on passage satisfaction

Passenger satisfaction in airlines is impacted by a wide array of factors that, when combined, collectively determine if they will be repeat customers or otherwise. Because of this, the majority of airlines across the globe, including Skywards International Airlines Corporation, strive to ensure that the experience for their clients remains positive. One critical component influencing satisfaction among the client is timeliness, punctuality, and reliability.

# Tasks

## Extract, Transform, and Load (ETL) Phase

The ETL phase is critical in the data analysis and machine learning process and helps the analysts clean and decipher critical elements of the dataset. The primary objective for this phase is to ensure that there is an exploration of the ambiguities, assumptions, and anomalies surrounding the data. The project was done using Python, and the first step of the analysis is always supposed to be data extraction.

## Exploratory Visualizations

There were female flight attendants 52727 compared to male flight attendants 51,177.

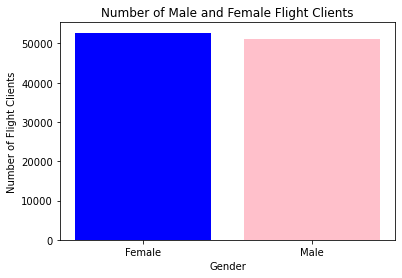


Figure 1 A comparison between male and female attendants

Data analysis based on the type of travel revealed that 69.0% of all the clients were on business travel while only 31.0% of the passengers were on the airlines for personal reasons.

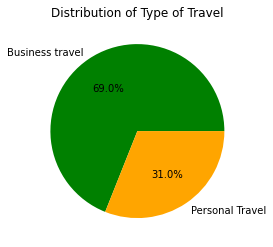


Figure 2 Comparison between airline passengers for business and personal travel reasons.

Comparing the reasons for travel for an individual based on class revealed several interesting insights. First, it was revealed that many passengers preferred the business class followed by eco, and the eco-plus class had the least number of passengers. It was discovered that most business travelers preferred the business class.

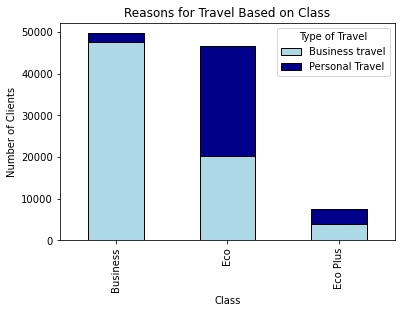


Figure 3 Analyzing the clients' preferred class based on the travel type.

Based on the frequency distribution of the table, the respondents showed that the Wi-Fi services provided by the Airline were mostly shaky and volatile. A few clients felt that the services provided were excellent, but the majority said that the services were either poor or just fair.

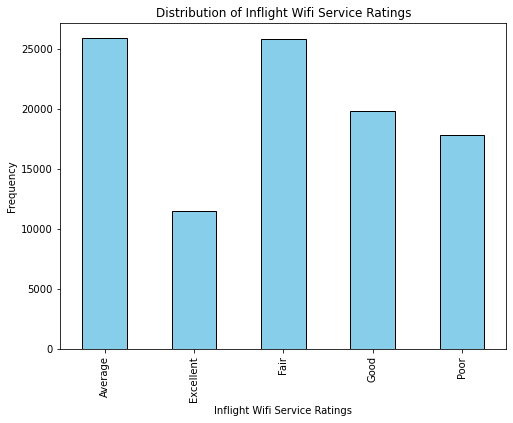


Figure 4 Comparing the quality of the Wi-Fi services based on the client's ratings

Additionally, analysis of the level of satisfaction the clients had based on the arrival or departure time showed that the Airline was doing great work in this regard. Data revealed that most respondents reviewed that the departure/arrival time convenience was excellent.

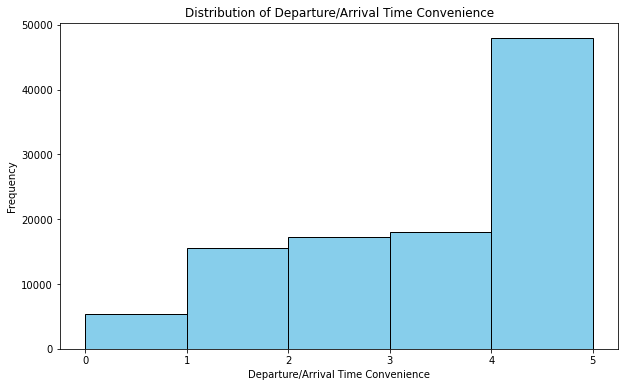


Figure 5 Convenience ratings of the clients based on the departure/arrival time

## Analytical Models Selection

To ensure that customer satisfaction at Skywards International Airlines Corporation is enhanced, two analytical models were selected to help reveal patterns within the vast dataset being analyzed. The two selected models were decision trees and random forest classifiers, and the merit was based on their distinct characteristics and proven capabilities to handle the dataset. First, studies have shown that decision trees can handle both classification and regression tasks, which is needed for modeling the dataset provided. The trees also possess the advantage of being capable of mimicking the way humans make their decisions

The selection of the decision tree and random forest classifier is justified by their suitability to perform the task at hand. The model is complicated and has many non-linear features, making the two complementary predictive models a perfect fit. The two models are known to be relatively scalable and rely on computationally efficient algorithms, making them perfect for analyzing large datasets like the one for Skywards International Airlines. Other advantages of the two selected machine learning models are their abilities to handle imbalanced classes, their robustness when dealing with outliers, and their efficient handling of high-dimensional datasets like the one provided.

## Models Evaluation

Decision trees, especially when conducting classification tasks, as was the case with the data provided, do not use a loss function. Instead, the trees rely on impurity measures at each node of a tree to ensure that the loss function is minimized. The most common impurity measures associated with decision trees are gini impurity, entropy, and the miscalculation rate.

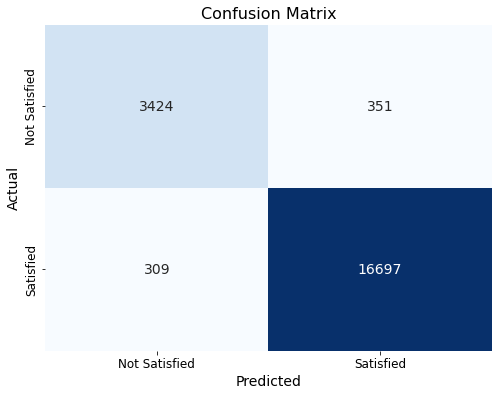


Figure 6 Confusion matrix for the decision tree predictive model test results

The random forest classifier had an accuracy of about 98.48%, which was even more impressive than the levels of accuracy that the results for the decision tree showed. 99.25 of the test data had correct optimistic predictions, also called a model's precision.

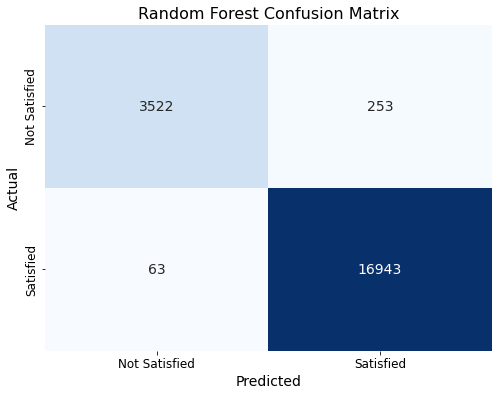


Figure 7 Confusion matrix for the decision tree predictive model test results

# Recommendations

Based on the analysis, it is recommendable for Skywards International Airlines Corporation to focus on areas where low ratings were observed, such as the quality of in-flight entertainment, the consistency of the Wi-Fi provided, onboard services, and the levels of cleanliness in the airlines. The management of the Airline must also pay close attention to the delays in departure and arrivals as they significantly determine if the client is satisfied or otherwise.

# Next Steps

Going forward, Skywards International Airlines Corporation should regularly update the two trained predictive models to ensure they remain effective based on current data. More supervised learning should also be done as they are critical in training the predictive models.