NTNU

Department of Computer Science

A Data-Driven Strategy for Improving Airline Passenger Satisfaction

TDT4259 Applied Data Science – Group Project

November 27, 2022

Group 14

Last name	First name	Student E-mail
Gule	Tomas Vaagen	tomasvg@stud.ntnu.no
Jammot	Adrien	adrienj@stud.ntnu.no
Johora	Fatama Toj	fatamatj@stud.ntnu.no
Hasan	Md Anwarul	mdaha@stud.ntnu.no
Mimo	Minhaj	minhajmm@stud.ntnu.no
Rahman	Mahfujur	mahfujur@stud.ntnu.no
Vanderkelen	Maxence	maxencev@stud.ntnu.no
Wiker	Petter	pettewik@stud.ntnu.no

Contents

List of Figures

List of Tables

1 Introduction and Problem Definition

1.1 Introduction

For the past 20 years the global airline industry has been characterized by a pattern of continuous growth, only broken by the recent COVID-19 pandemic. More people than ever before choose to travel by air for both business and leisure, domestically and internationally, and ready to serve them is a growing number of airline operators looking to solidify their position in the market.

The market for airline operations is marked by high competitiveness and generally low margins on earnings, and as a consequence, many operators focus on finding a specific niche in the market where they can outperform competitors. Some airlines choose to sell passenger comfort, exclusivity and luxury at a high price point, while others opt for providing more affordable tickets at a lower standard. However, common to all airlines is the importance of having passengers satisfied with the experience provided to them, as at the end of the day, the customer is the sole revenue generator for the airline and key to ensuring future survival and growth. Satisfied customers are more likely to stay loyal to the airline and it is thus crucial that the airline enforces measures to ensure continued passenger happiness.

1.2 Problem Definition

Given the competitive nature of the airline industry, maintaining and expanding a loyal base of customers is a tough, but essential part of airline operations. Airlines are constantly on the lookout for cost-effective ways of increasing customer satisfaction. In this project, we would like to use real passenger data in combination with modern statistical and machine learning models to identify and analyze drivers for customer satisfaction. Our research questions for the project may be summarized as:

- Can we use a data-driven approach to identify decisive factors for airline customer satisfaction?
- What measures can realistically be put in place by the airline to bolster future customer satisfaction? What is the cost-effectiveness of the proposed measures?

1.3 The Team

The team that produced this report consists of eight students that shared several roles and responsibilities. All the team members showed interest in all aspects of the report and the tasks were distributed thereafter. Initially, descriptive analytic ideas were brainstormed within the team. The team was subsequently divided into three separate groups, focusing on the actual implementation of the data analysis, method and introductory parts. The larger data analysis tasks were separated into smaller, more specific tasks, with the purpose of producing plots to possibly unveil interesting aspects of the data. Further, the team decided which parts of the analysis that would be the main focus of the report.

The participating team members and their roles in the project are specified in Table 1.1. The table reflects only the concrete contributions of each team member during the data analysis and writing of this report, as naturally all group members have been involved in brainstorming the topic, data sources and scope for the project.

Table 1.1: Names of team members and their specified roles during the project work.

Name	**Role in project**
Gule, Tomas Vaagen	Tomas is in his last year on the integrated Master in computer science. He is currently working part time as a software developer. In this project he contributed with analyzing the data in Tableau,
Jammot, Adrien	Adrien is an exchange student doing a master in computer science and machine learning, he contributed by realising the data exploration and predictive models.
Johora, Fatama Toj	is a student in the Informatics department, her major is Database and Search. She contributed to the method section and some exploratory data analysis.
Hasan, Md Anwarul	Hasan is a first year masters student of Informatics. He has experience of working as IT advisor and IT auditor for different organizations. In this project he worked on method and data analysis section.
Mimo, Minhaj	Mimo is a Masters student in the informatics department and his major is Software. In this project, he worked on the method section, data analysis, and interpretation part.
Rahman, Mahfujur	Mahfujur is pursuing his masters in the department of informatics. He contributed to the method section and did some exploratory data analysis which is included in the analysis and appendix part.
Vanderkelen, Maxence	Maxence is a student on international mobility for his last year of study in computer engineering. He contributed by working on the method section and on the data set study.
Wiker, Petter	Petter is in his final year of an integrated master's degree in Nanotechnology. His contributions include shaping the project scope and writing on the project report.

2 Background

This section will cover the project objectives and how they may be resolved by applying a data-driven strategy. Both the theoretical basis and the practical implementation of our chosen data strategy, CRISP-DM, will be treated.

2.1 Purpose and Objective

As specified in Section 1.2, the purpose of this project is to provide value for decision makers within the global airline industry. The provided value will take the form of a set of recommendations for measures an airline may make in order to increase the satisfaction levels of their customers, an important metric for airlines as it is related with repeat passenger activity. By making use of a comprehensive historic data set connecting a description of a certain passenger and flight to the passenger's satisfaction level in conjunction with a chosen data strategy, one may attempt to uncover patterns and trends in the data.

More specifically, we aim to use the CRISP-DM data strategy to understand and prepare the data and then implement a classification model based on modern machine learning techniques. Under the assumption that the fitted model is able to predict the passenger satisfaction level with a reasonable level of accuracy, we may then analyze its parameters to obtain an understanding of the driving factors for passenger satisfaction levels. Recommendations may then be prepared in response to the top satisfaction drivers.

2.2 The CRISP-DM Data Strategy

Since no group member has had any previous experience working with a data strategy, we chose to base our work on the CRISP-DM strategy. CRISP-DM has since its publishing in 1999 become the most common methodology[1] for data science projects, and therefore we evaluated it as a safe and simple choice for our project. An acronym for *Cross Industry Standard Process for Data Mining*, CRISP-DM consists of 6 steps, namely (1) business understanding, (2) data understanding, (3) data preparation, (4) modeling, (5) evaluation, and (6) deployment. The steps are described in detail below:

1. **Business Understanding:** To provide value to a business it is paramount to have an understanding of what the business objective to the costumer is and what they want to accomplish. So the first step revolves around understanding what the business needs and translate that to a data mining problem we can solve. A preliminary plan on how to achieve a solution to the problem is also created.

2. **Data Understanding:** With the objective and preliminary plan in place, one may start collecting the data needed for solving the problem. Taking care in exploring and describing the data is important

to garner some initial insight and make a hypothesis of what information and value is possible to extract through the subsequent data analysis.

3. **Data Preparation:** The third step revolves around preparing the data so that it is in a state ready to be fed into the data analysis tools. To provide value from the data in an effective manner, it is important to be be cautious when doing the data preparation. The data preparation step consists of:

   (a) data selection

   (b) data cleaning

   (c) construction of data features

   (d) data integration

   (e) data formatting

   A more detailed approach to each step will be covered in Section 3 where CRISP-DM is applied to the chosen passenger satisfaction data set.

4. **Modeling:** Select data features for further analysis, choose which tools and algorithms to use, define model goals and state specific model requirements. Use the selected modeling techniques to build the models. Assess and evaluate if the result of the model is viable for further use.

5. **Evaluation:** The model should be evaluated following preliminary testing. Key questions to ask are: Are the results presented in a clear way? What are the key findings? How do the findings contribute to achieving the goals of the stakeholders? It is important to critically scrutinize the model; Is there anything that can be done to improve the model results? Where does the model fail? The modeling and evaluation step may be repeated for several iterations until a satisfying, final model is found and chosen.

6. **Deployment:** With a finalized model it is important and necessary to organize findings in a way that is accessible to the stakeholders so that they may be utilized to create impact and value.

3 Method

This section will introduce the chosen Airline Passenger Satisfaction data set and describe how the CRISP-DM data strategy has been used to extract value from the data in line with the project objectives outlined in Section 2.

3.1 The Airline Passenger Satisfaction Data Set

The selected data set[2] was found and retrieved from the Kaggle platform. The data set contains around 130,000 entries, whereas each entry is associated with 23 features covering specifics of a passenger, its perceived satisfaction level faced with different aspects of a flight, and its overall satisfaction level. The features that relate to the passenger demographics and their experiences are summarized in Table 3.1 and Table 3.2, respectively.

Table 3.1: Summary of demographic features in the Airline Passenger Satisfaction data set.

Feature name	**Description**
Gender	The gender of each passenger in string format. Can have values ’Male’ or ’Female’.
Customer type	The type of customer in string format. Can have values ’Loyal customer’ or ’Disloyal customer’.
Age	The age of each passenger in integer format. Ages range from 7 to 85.
Type of travel	The travel purpose in string format e.g. personal or business travel
Customer class	The passenger’s traveling class in string format. Can have values ’Business’, ’Economy Plus’, or ’Economy’.
Flight distance	The flight distance in integer format. Values range from 31 to 4983 km.
Departure delay in minutes	The departure delay in minutes. Values range from 0 to 1592 minutes.
Arrival delay in minutes	The arrival delay time in minutes. Values range from 0 to 1592 minutes.

Table 3.2: Summary of features in the Airline Passenger Satisfaction data set pertaining to perceived customer satisfaction.

Feature	**Description**
In-flight WiFi service	Describes how the passengers rate the in-flight WiFi-service. Values range from 0 to 5.
Departure and arrival time	Describes how the passengers rate the departure and arrival time. Values range from 0 to 5.
Ease of online booking	Describes how the passengers rate the online booking procedure. Values range from 0 to 5.
Gate location	Describes how the passengers rate the location of the departure gate. Values range from 0 to 5.
Food and drink	Describes how the passengers rate onboard food and drinks services. Values range from 0 to 5.
Online boarding	Describes how the passengers rate the online boarding procedure. Values range from 0 to 5.
Seat comfort	Describes how the passengers rate the onboard seat comfort. Values range from 0 to 5.
In-flight entertainment	Describes how the passengers rate the in-flight entertainment services. Values range from 0 to 5.
On-board service	Describes how the passengers rate the onboard service. Values range from 0 to 5.
Leg room service	Describes how the passengers rate the onboard leg room. Values range from 0 to 5.
Baggage handling	Describes how the passengers rate the baggage handling. Values range from 0 to 5.
Check-in service	Describes how the passengers rate the check-in service. Values range from 0 to 5.
In-flight service	Describes how the passengers rate the in-flight service. Values range from 0 to 5.
Cleanliness	Describes how the passengers rate the cleanliness in air and on the ground. Values range from 0 to 5.
Customer satisfaction	Describes how the passengers rate the overall travel experience. Value can be either "Satisfied" or "Neutral or dissatisfied".

3.1.1 Data Quality

Keeping in-line with CRISP-DM, data quality metrics should be evaluated at an early stage in order to garner an understanding of the strengths and potential limitations of the raw data set. Thus, we have scrutinized the data sets reliability and completeness, as well as its relevance for providing business value in the context of our defined objective:

- **Reliability:** We have chosen to trust the accuracy of the data set as it is part of a popular machine learning project on the Kaggle platform. However, the source of the data set remains undisclosed, and we cannot be fully confident that the data is real, nor account for the method used for collecting/generating the data. Still, even if the data would turn out to be synthetic, we think that the data set provides value for the project as the data analysis process closely mimics that of a real business project regardless.

  Some outliers have been identified. The treatment of these are covered more closely in Section 3.3.

- **Completeness:** The missingno Python library[3] has been used to identify outliers in the data. Conveniently, the data set is unusually complete, with only the *arrival_delay_in_minutes*-feature missing values for around 400 entries. The outliers are treated in Section 3.3.

- **Relevance:** As we are aiming to provide value to airline decision makers by identifying drivers for passenger satisfaction, we find the data set to be very relevant. With its combination of demographic data and passenger satisfaction rankings it provides a good baseline for analyzing and understanding passenger priorities.

3.2 Software Tools

To conduct the exploratory data analysis and subsequently construct the predictive model several tools were used:

- **Tableau:** Tableau is one of the marked-leading services for modern business intelligence. The platform makes it easy to explore, manage and analyze data sets and visualize them in a way that can provide business insight. On the platform you can for example aggregate data to create new feature groups and plot the relationship between features in a variety of ways, which is useful both in the initial data analysis phase of CRISP-DM as well as for visualizing insights in later stages. Tableau is also available both for Mac OS and Windows and therefore became the preferred choice for the group made up of both user types.

- **Power BI:** Power BI is a free but effective tool to represent data visually. It provides an easy to use interface for handling and aggregating large amounts of data. We used Power BI for primary analysis of the data, but as the group composition was a mix of Mac OS and Windows users, most of the work was as stated done in Tableau.
- **Python:** Python is a high level, open source, interpreted language that offers a fantastic approach to object-oriented programming. We used Python running in a Jupyter Notebook to load the data set, conduct an exploratory data analysis, and construct a predictive classification model. To handle and visualize the data we used the widely popular packages [4], seaborn[5] and matplotlib[6]. For the modeling, Scikit-learn[7], a free machine learning library, was used. It features a selection of classification, regression and clustering algorithms in a framework designed for use with pandas. We utilized seaborn, a statistical data visualization toolkit building on the plotting library matplotlib to display the model insights. It offers a sophisticated drawing tool for creating eye-catching and educational statistical visuals comparable to the excellent plotting tools found in the R programming language.

3.3 Constructing a Pipeline for Data Processing

A series of data processing operations is referred to as a pipeline. A successfully implemented pipeline offers a seamless, automated flow of data from one stage to the next by eliminating the majority of manual procedures from the process. Making a robust pipeline was important for our project work to support more efficient analysis and modeling. This section describes the steps included in our data processing pipeline.

3.3.1 Data Preparation

The data set was downloaded and initially fed into Power BI. Using its various diagrams and powerful, easy-to-use point-and-click interface we tried to correlate the data of the data-set. While doing so we realized that data cleansing, data validation and data labeling are required to further analyze the data set.

3.3.2 Data Cleansing and Validation

As noted, the native data set was unusually complete with only about 400 entries missing values, all belonging to the ’arrival_delay_in_minutes’-feature. These entries were simply removed as their small numbers were deemed insignificant for inflicting a bias in the remaining ∼130000 entries of the data set. Alternatively, and for larger amounts of missing values, one could have tried to replace the values. For example, one could have made use of the high correlation between the ’departure_delay_in_minutes’ and the ’arrival_delay_in_minutes’ to estimate missing values.

After a statistical analysis of the distribution of numeric features, some outliers were identified. All entries with a value for ’arrival_delay_in_minutes’ that exceeded 500 minutes were removed from the data set together with all entries describing a flight of over 4000 km.

3.3.3 Data Labeling

As we found the native naming of the data set features sufficient and complete, we did not relabel any of the features. However, we did see the need to segment some of the numeric variables to help with our analysis; the demographic ’age’-feature was divided into four categories, namely Children, Youths, Adults, and Seniors. Moreover, the numeric ’flight_distance’-feature was made categorical by labeling every flight shorter than 600 km as ’short’ and the remaining flights ’long’. The segment cut offs were not arbitrarily set, but rather a result of findings emerging from the exploratory data analysis.

Additionally, the ’customer_satisfaction’-feature was made numeric by encoding all ’satisfied’-valued entries as 1 and all ’neutral or dissatisfied’-valued entries as 0.

3.3.4 Implementing a Classification Model

To display our data in a way fitting the implementation of the scikit-learn library we used dummy variables to represent categorical variables. The dummy variables were constructed using a so-called ’one-hot’ encoding scheme. In a ’one-hot’ encoded data set, a new column is made for each unique category of a categorical column, and each entry of the column gets a value of either 0 or 1 depending on whether the categorical variable was originally represented in the entry or not. The numeric features were then normalized to avoid overestimating the importance of any variable. We then split the data set into a training set containing 80% of the data and a test set containing the remaining 20%. Finally we fit our models to the training set and evaluated them on the test set.

Several different predictive models were tried to see which one of them would yield the best classification accuracy on our data set. Out of a random forest classifier, a gradient boosting classifier, and an XGBoost classifier[8], we found the former to perform the best.

The prediction results are discussed further in Section 4. Our models certainly leave room for optimization using techniques such as model stacking and hyperparameter tuning, but we found our baseline models to perform well enough for further analysis and decided not to expend time on unnecessarily honing the prediction accuracy.

3.4 Visualizing the Data

Charts and diagrams may be a useful tool for quickly understanding large and complex data structures. To represent the data and our findings we have generated many visualizations, both to go in the project report and for internal use to help comprehend data relations and steer the project work. The visualizations were primarily made with the help of Tableau.

4 Analysis

Our analysis is divided into two major parts; first, it is important to gain an in-depth understanding of the data set and the underlying relations between features. This is primordial to the second part, which consists of successfully building the predictive classification model.

4.1 Data Summary

We want to describe the satisfaction of our customers in terms of basic characteristics such as the class they fly in, their age and the length of the flight. A curated selection of features are shown graphed in Figure 4.1. For the individual features, we find:

- **Age groups:** The adults, that is passengers in the age group 30-59, are the most satisfied with about 50% satisfaction. The seniors and children are much less satisfied, with a satisfaction level of below 25% in both segments.

- **Customer class:** The business class travelers are overall the most satisfied with a satisfaction level of around 75 %. The Eco and Eco Plus segments on the other hand have passengers with a reported satisfaction level of less than 25 %. It is hard to determine the direct cause of this, but it may hint towards customer class being an important factor for overall customer satisfaction. Another way to interpret this result could be that business class customers will say that they are more satisfied as they paid more for their ticket.

- **Gender:** There is almost no difference in satisfaction between the genders. Satisfaction levels of both men and women are found to be just below 50 %.

- **Customer type:** The loyal customers are almost twice as satisfied as the disloyal customers. One may speculate that the customers are loyal because they are satisfied. This would support the underlying industry assumption that having satisfied customers boosts customer loyalty and contributes to continued revenue generation for the airline.

- **Flight distance:** Passengers traveling longer distances tend to be more satisfied than passengers on shorter flights. Figure 4.2 and subsequent analysis delve deeper into this observation.

- **Type of travel:** Over half of passengers traveling for business purposes report being satisfied. Personal travel has a very low satisfaction rate with approximately 10 % of the customers satisfied. This result is quite surprising, and might indicate the presence of a bias in the data. For example, satisfied passengers traveling on vacation may be less inclined to answer the survey than dissatisfied

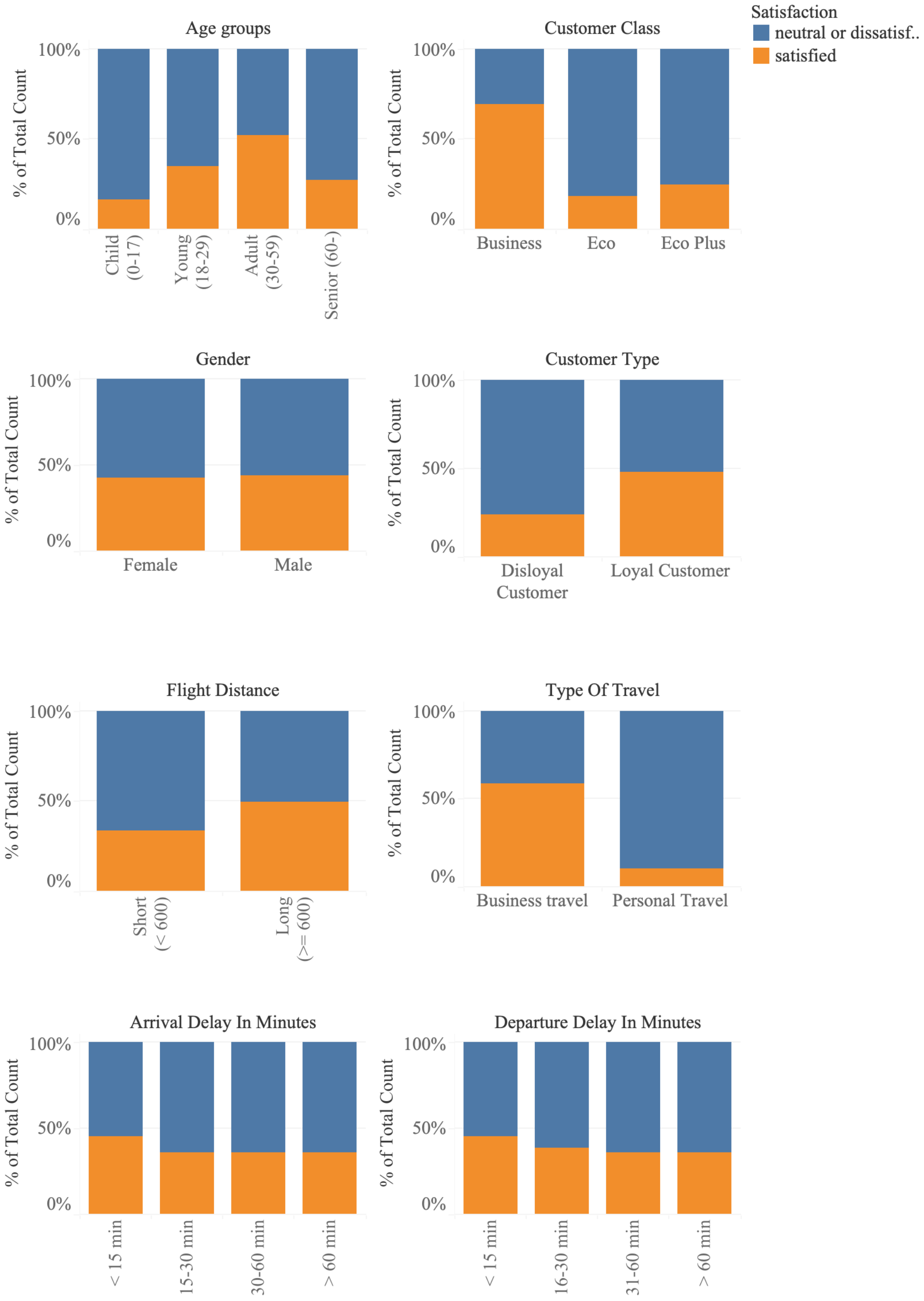


Figure 4.1: The relative amount of passengers with an overall satisfaction level of "satisfied" for a selection of curated features.

ones. Another explanation may be that passengers spending their own money on personal travel may expect more from the experience than a seasoned business traveler.

- **Arrival and departure delay:** With under 15 minutes departure and arrival delay almost half of the passengers are satisfied. The satisfaction rate decreases with increasing delays. As expected, and as can be seen in Figure A.1 of the appendix, the arrival and departure delays are highly correlated.

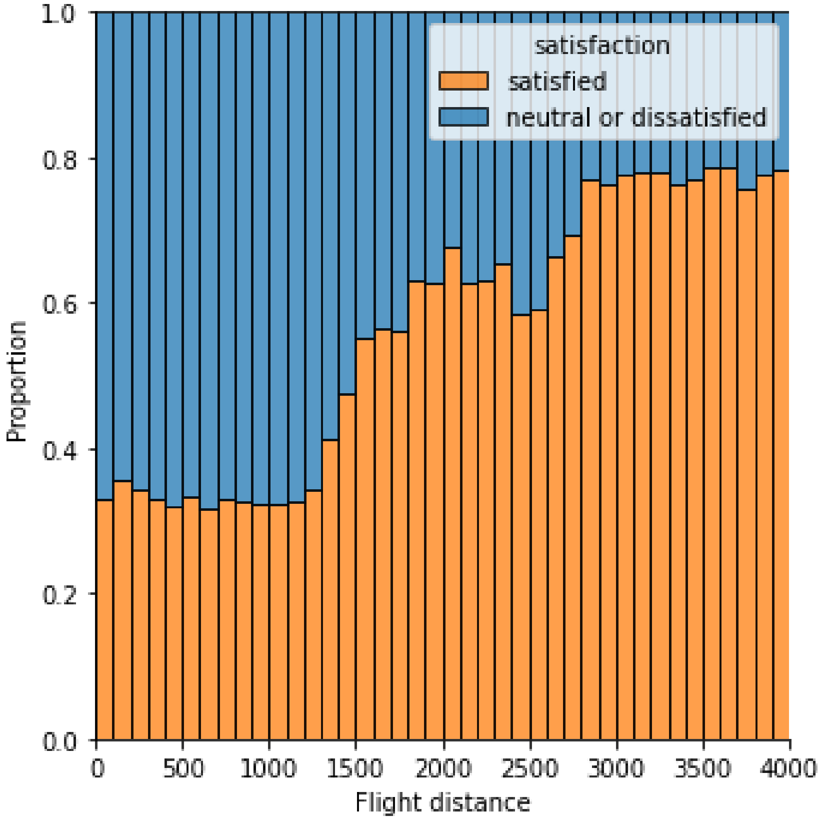


Figure 4.2: The relative amount of passengers with an overall satisfaction level of "satisfied" plotted as a function of their flight distance.

As we can see in Figure 4.2, the passengers on shorter flights are less satisfied than the passengers on longer flights. This information may indicate that airlines have room for improvement in the services provided on their short distance flights. It also tells us that we need to be careful to keep this bias in mind during our subsequent analysis. This way we are better guarded against making erroneous interpretations.

This basic data analysis results may on their own already be able to bring value and executive insight to airline decision makers, but are also paramount in constructing and interpreting the predictive model of section 4.2.

Table 4.1: Classification accuracy of the implemented machine learning models. The result of the best model, based on random forest algorithm, is emboldened.

Model type	**Train set accuracy (%)**	**Test set accuracy (%)**
Random Forest	99.8	**91.0**
Gradient Boosting	89.3	88.7
XGBoost	99.7	89.4

4.2 Results of the Predictive Modeling

The accuracy of the various implemented machine learning models are summarized in Table 4.1. From these results we chose to use the random forest classifier to make our predictions. The accuracy of 91.0 % means that, given the survey filled by a new customer, 9 times out of 10 the model will successfully predict the overall satisfaction level of the new customer.

Figure 4.3 shows the results of conducting a feature importance analysis on the chosen machine learning model. The feature importance is a measure of the relative importance of each feature for the classifier when making decisions. We find that the "inflight_wifi_service§-feature is the clearly most influential discriminator between satisfied and unsatisfied customers. The "ease_of_online_booking" and "online_boarding" features are also found to be significantly distinguished from the rest in their importance.

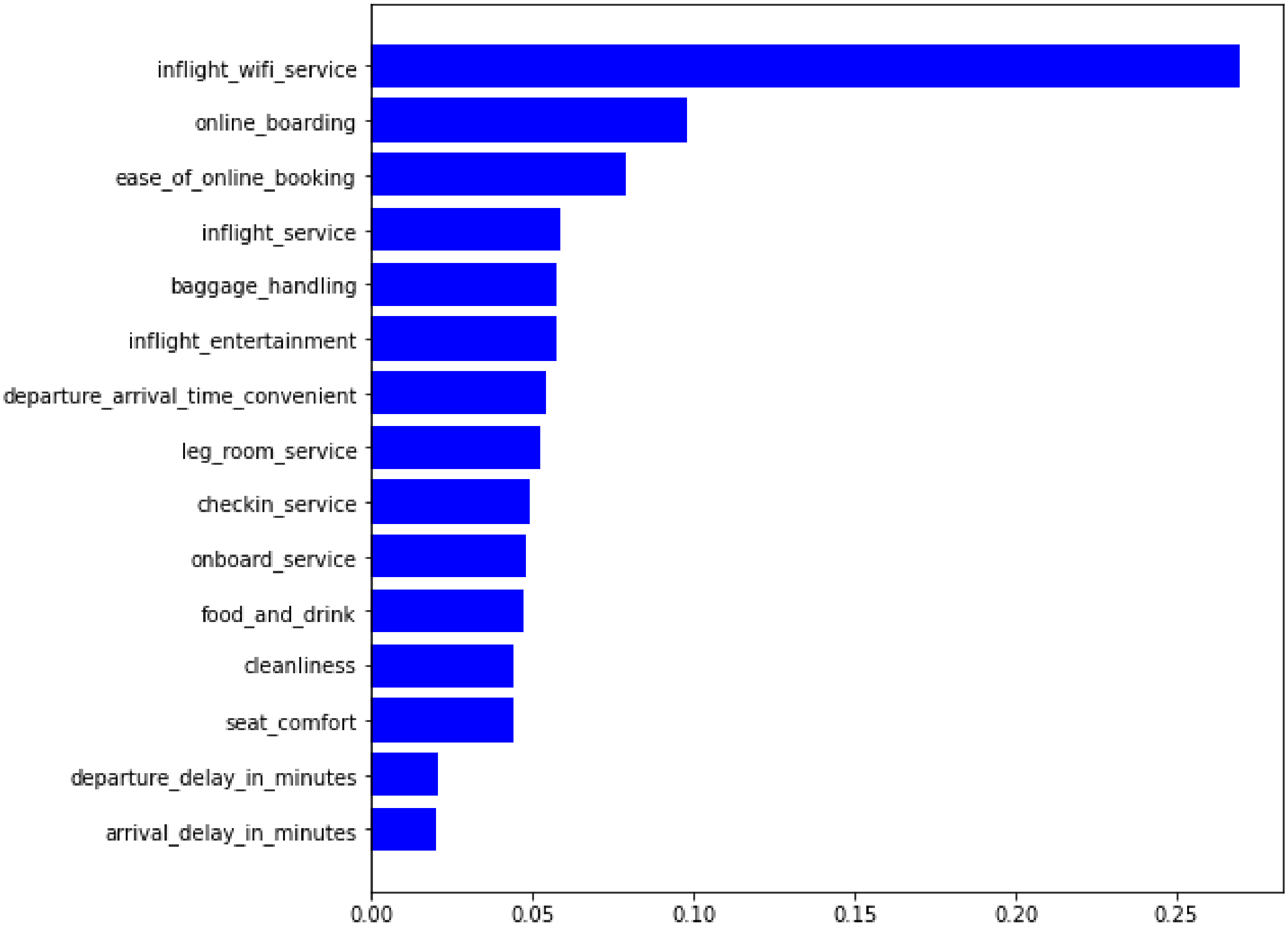


Figure 4.3: Feature importances found for the chosen Random Forest machine learning model.

5 Interpretation and Recommendation

In this section we will come up with recommendations for increasing the airline passenger satisfaction. We will also present an implementation plan for the recommendations. To conclude we discuss the limitations of the analysis and what further work could be done to bolster the future analysis.

5.1 Recommendations

Based on the data analysis in Section 4 we are able to identify some key measures than can be put in place by airline decision makers to improve customer satisfaction:

- **Improve in-flight WiFi-coverage.** The feature importance analysis in Section 4.2 clearly shows the importance of having a satisfactory in-flight WiFi-coverage. We would recommend airline decision makers to focus on bettering the WiFi-coverage on flights, as this would have a relatively large impact on overall customer satisfaction. The improvements may for example involve upgrading WiFi-availability, stability, speed, and bandwidth to match modern technical and customer standards. Implementing this measure would most likely be very cost-effective, with a large expected satisfaction increase, and thus an increase in the number of repeat customers, per invested dollar.
- **Better the online boarding procedures.** Another important feature found to drive overall customer satisfaction was the passengers satisfaction with the online boarding process. Online boarding has emerged as an alternative to the manual check-in process, and is generally a convenient way for passengers to organize travel documents, seating arrangements, and information on their personal devices. However, to meet customer expectations, care must be put into making the online boarding procedure as accessible and streamlined as possible. We recommend that airline decision makers review their online boarding services and ensure that they are up-to-date with the latest public expectations.
- **Ease online booking.** In Section 4.2 we can see that ease of online booking is the third most important feature for predicting customer satisfaction. The airlines should therefore put a lot of effort into maintaining an easy-to-use online booking tool. For example, the airlines can improve the existing website to meet modern expectations in responsiveness and ease of use. As seen in Figure 4.1, senior people are less satisfied than adults and youth. A focus area could therefore be to make online services more accessible for older people that may struggle with modern technology tools.

5.2 Implementation

Table 5.1 provides a possible timeline for the implementation of the recommended measures presented in section 5.1. Following their implementation, each measure may be evaluated by the proposed success criteria.

Table 5.1: An overview of the recommendations and their time frame and success criteria.

Recommendation	Time frame	Success criteria
Improve in-flight WiFi-coverage	1-2 years	Increase average satisfaction with WiFi-coverage by x%.
Better the online boarding procedures	6-12 months	Increase average satisfaction with the online boarding service by y%.
Ease online booking	6-12 months	Improve online booking satisfaction in the senior segment with z%.

5.3 Limitations

As discussed in section 3.1.1, the data set was found on Kaggle with an unknown source and origin. therefore, it is unclear what airlines have been a part of the survey or in which country/countries the flights have taken place. This may make the data biased. There is no certainty that the survey would give the same results in different parts of the world, and subsequently, that or proposed recommendations will be valid across regions and airlines. The data set was also uploaded approximately two years ago, and given the impacts of the COVID-19 pandemic, our analysis of the air travel domain may be outdated. To make more precise and pertinent recommendations, it would also be helpful to have a deeper business understanding of the industry. This could for example have been achieved by interviewing a subject matter expert.

5.4 Suggestions for Further Analysis

As discussed, a major limitation for the project was the limited background information available for the data set and therefore the risk of an underlying bias. To solve this, a new survey encompassing multiple airlines in multiple countries could be done. This would also solve the question of the unclear data origin. In such a survey one could include additional features like ticket price and the passengers satisfaction level with the price point. With this additional information it would be possible to do a new analysis were you could differentiate the recommendations for different parts of the world. The ticket price feature could be used to see if the there is an correlation between the price and to what standard passengers need to have in order to be satisfied.

References

[1] Jeff Saltz. CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects. https://www.datascience-pm.com/crisp-dm-still-most-popular/.

[2] Airline Passenger Satisfaction. https://www.kaggle.com/datasets/binaryjoker/airline-passenger-satisfaction.

[3] missingno — missing data visualization module for python. https://github.com/ResidentMario/missingno, August 2022.

[4] pandas.DataFrame — pandas 1.5.1 documentation. https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html.

[5] Michael Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, April 2021.

[6] Matplotlib — Visualization with Python. https://matplotlib.org/.

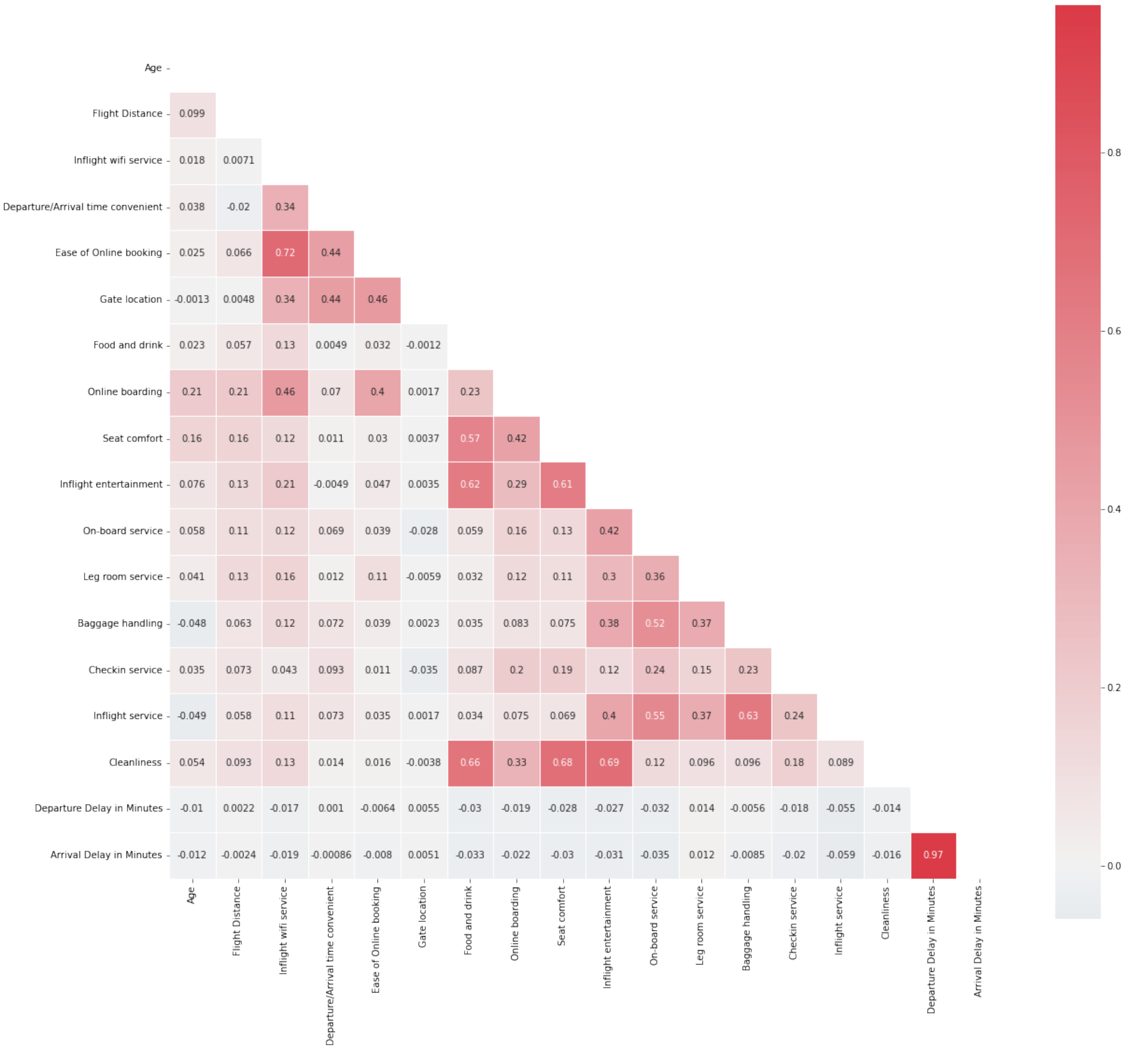
[7] scikit-learn. https://scikit-learn/stable/modules/classes.html.

[8] XGBoost Python Package — xgboost 2.0.0-dev documentation. https://xgboost.readthedocs.io/en/latest/python/index.html.

Appendices

Appendix A: Feature Correlations

A.1 Correlation Matrix for the Features of the Airline Passenger Satisfaction Data Set



	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes
Age																		
Flight Distance	0.099																	
Inflight wifi service	0.018	0.0071																
Departure/Arrival time convenient	0.038	-0.02	0.34															
Ease of Online booking	0.025	0.066	0.72	0.44														
Gate location	-0.0013	0.0048	0.34	0.44	0.46													
Food and drink	0.023	0.057	0.13	0.0049	0.032	-0.0012												
Online boarding	0.21	0.21	0.46	0.07	0.4	0.0017	0.23											
Seat comfort	0.16	0.16	0.12	0.011	0.03	0.0037	0.57	0.42										
Inflight entertainment	0.076	0.13	0.21	-0.0049	0.047	0.0035	0.62	0.29	0.61									
On-board service	0.058	0.11	0.12	0.069	0.039	-0.028	0.059	0.16	0.13	0.42								
Leg room service	0.041	0.13	0.16	0.012	0.11	-0.0059	0.032	0.12	0.11	0.3	0.36							
Baggage handling	-0.048	0.063	0.12	0.072	0.039	0.0023	0.035	0.083	0.075	0.38	0.52	0.37						
Checkin service	0.035	0.073	0.043	0.093	0.011	-0.035	0.087	0.2	0.19	0.12	0.24	0.15	0.23					
Inflight service	-0.049	0.058	0.11	0.073	0.035	0.0017	0.034	0.075	0.069	0.4	0.55	0.37	0.63	0.24				
Cleanliness	0.054	0.093	0.13	0.014	0.016	-0.0038	0.66	0.33	0.68	0.69	0.12	0.096	0.096	0.18	0.089			
Departure Delay in Minutes	-0.01	0.0022	-0.017	0.001	-0.0064	0.0055	-0.03	-0.019	-0.028	-0.027	-0.032	0.014	-0.0056	-0.018	-0.055	-0.014		
Arrival Delay in Minutes	-0.012	-0.0024	-0.019	-0.00086	-0.008	0.0051	-0.033	-0.022	-0.03	-0.031	-0.035	0.012	-0.0085	-0.02	-0.059	-0.016	0.97	

Figure A.1: A correlation plot showing the correlation between the various features of the airline passenger satisfaction data set.