University of Warsaw
Faculty of Economic Sciences

Adam Janczyszyn
Student book number: 419350
Jakub Żmujdzin
Student book number: 420602
Jakub Wujec
Student book number: 420463
Hubert Wojewoda
Student book number: 406420

# I appreciate a LOT traveling with Business or Premium!

## Predicting passenger's willingness to upgrade and optimal moment for sending an offer based on LOT's flight dataset.

Paper prepared for
Warsaw Econometric Challenge 2022

Warsaw, march 2022

# TABLE OF CONTENTS

## ABSTRACT

The aim of this paper is to predict the willingness of passengers of the Polish Airlines "LOT" towards upgrading their trip. An additional objective is to determine the optimal date for sending out an informational message through digital channels (sms, email, app push) addressed to passengers with active reservations who have not yet started their journey concerning the possibility of upgrading their flight class. The study used data from the Polish Airlines "LOT" database. The XGBoost model was used for prediction, to which Bayesian hyperparameter optimization and feature choice with forward selection were applied. The data was analyzed and cleaned to determine the most optimal forecast configuration. The solution used offers high prediction accuracy and was interpreted using graphs. Binary classification results confirm the significant positive effect of flight distance on LOT Polish Airlines passengers' propensity to upgrade their travel.



## KEYWORDS

Airline class upgrade prediction, machine learning, econometrics, big data, xgboost, LOT, optimal notification time

# INTRODUCTION

The high growth rate of the aviation market strongly influences the economic development of the world. In particular, the dynamic growth of the airline industry in recent decades has become a driving force for the spread of tourism in the world. The opportunities for private and business travelers to access foreign markets, distant cultures and exotic holiday destinations have increased significantly.

The aviation industry accounts for \$3.5 trillion (4.1%) of the world's gross domestic product (GDP). If aviation was a country, it would rank 17th in terms of GDP size. That's equal to the GDP of Indonesia and the Netherlands. The aviation market is expected to grow significantly in the coming decades. By 2038, global air transport is projected to provide 143 million jobs and contribute \$6.3 trillion to the global economy, according to Aviation Benefits website (source: https://aviationbenefits.org/economic-growth/).

With the rapid growth of the market, companies offering air transportation are competing strongly for the largest market share. In order to attract customers, global airline companies are improving technology, competing in prices and offering many additional services such as seat selection, baggage allowance, legroom and on-board dining. In addition, airlines offer different flight classes to allow customers to choose the right flight standard for them.

Besides the price offered, the possibility for passengers to choose personalized services is crucial when choosing a carrier. Undoubtedly, airline companies should expand the services offered to their customers and run an active advertising campaign to make passengers well informed about them. This helps to increase sales revenue and better optimize the sale of different levels of tickets. Therefore, the aim of this study is to predict the propensity of passengers of the Polish Airlines "LOT" of upgrading their trip. An additional objective is to determine the optimal date for sending out an informational message through digital channels (sms, email, app push) addressed to passengers with active reservations who have not yet started their journey, regarding the possibility of upgrading their flight class. The analysis was performed using data from the LOT Polish Airlines database. The findings of this study can be useful in assessing the customer's willingness to upgrade their flight and directing the information campaign to potentially interested individuals.

The research hypothesis of this study, states that flight length positively influences the likelihood of purchasing a flight class upgrade. It is possible the reason for this is, that since the passenger will have to spend a considerable amount of time in the aircraft it would be worthwhile to spend this time in a high-quality standard. Thus, we expect flight length to have a strongly positive effect on the prediction of the explanatory variable.

Another factor that may affect the upgrade propensity is the gender of the passenger. We believe that there is a significant difference between the number of flight class upgrades made between women and men. Due to occupations with higher prestige and higher pay, we make a side hypothesis that men will be more likely to make flight class upgrades relative to women.

In this paper we make one more side hypothesis. It is stating that when the operating carrier and the issuing carrier are two different agencies, the propensity to upgrade is likely to be lower. This is due to the existence of airline ticketing agencies. The role of such intermediaries is to mathematically model the filling of the plane by passengers, while optimizing their profits. Agencies will sell more tickets than the limit assumes, given that there is some probability that not everyone will come to the flight. With competition in the market and overselling of tickets, prices for the lowest economy classes may be reduced. We predict that consumers who are targeted by such a measure have a very high elasticity of demand with respect to price, making them unwilling to upgrade the class under any circumstances.

The entire paper is divided into five sections. The first section provides a review of the literature on the effect of variables on passenger upgrade of flight class. The next section presents the research hypotheses. This is followed by data analysis and model properties. The fourth chapter contains the verification of the hypotheses. The fifth chapter includes a sensitivity analysis of the results and an informed self-critique. The final element of the paper is the conclusion, which compiles the conclusions of the study.

# RELATED RESEARCH

## 1.1. Impact of variables on class upgrade

Introducing an additional "premium business" class according to the research work of Cui, Orhuna, Duenyas (2018) is a good way to increase profits. Diversification of classes opens new possibilities of class improvements. Consequently, according to the previously mentioned researchers, there is more diversification in prices. The procedure of adding a premium business class enabled an increase in revenue in the tourist class by an average of $850. This action ultimately leads to increased profits for air carriers.

One of the key factors considered in airline sales analysis is frequent flyer programs, also known as air miles. While initially such offers allowed the most loyal customers to buy tickets for future flights with the points they earned, over time the programs have grown significantly in terms of products and services offered. Earlier, each airline had its own individual program for frequent flyers, but these programs are now shared both among airline alliances and by third-party service providers. Such programs have been a natural motivator for consumers to increase their consumption of factors such as flights. Points can be earned not only through buying flights with a particular carrier, but also for activities such as shopping at airports or staying at hotels that support such programs. In the form of rewards, not only can you get a new ticket, but also purchase products with points or get an upgrade to your flight, which is the main focus of this research paper. However, in early 2000's, an interesting observation was made. It stated that fewer and fewer air miles are being used (Liston-Heyes, C., 2002). All in all, over time, the points have evolved from a marketing gimmick into much more serious mechanisms causing more concerns, including tax evasion, and supervisor-employee relationship problems. In addition, people were found to significantly overestimate the value of air miles, and the deviation from the true value (when air miles could be treated as a pseudo-currency), depending on how it was measured, the overestimation ranged from about 40% for the median value to over 250% for the mean(Liston-Heyes, C., 2002).

Addressing willingness to pay, it is important to mention the scientific work of Kuo and Jou (2016), about the effect of air travel distance on the willingness to pay for premium economy class. It turns out that as the distance an airplane travels increases, the proportion of people willing to pay for a premium class of tickets rises. For the purposes of our article, this could be seen as a clue towards the fact that upgrading a class at any time after purchasing a flight is more likely to happen on long-haul flights.

Returning to air miles, it is worth considering their effect on buying flight class upgrades. Referring to an article by Gossling and Nilsson (2010), who cited an IATA agency survey, nearly half of people did not use the points they collected at all in the past 12 months. Moreover, only 13% of respondents stated that they had used the points for a class upgrade. Those who traveled the most perceived lodge access and class upgrade opportunities to be the most valuable of the benefits that frequent flyer programs could offer. This could mean that there is a correlation between customers who are in the loyalty system (and potentially travel the most) and those who chose to upgrade. Furthermore, Gossling and Nilsson, citing an IATA study, mention that there is a disparity among the genders of air travelers. For the 2009 study, men made up about 75% of the sample. At the same time, among the most active flyers (more than 40 flights per year), women accounted for only 4% of the passengers. If the most active passengers were the most likely to use upgrades, it would be worthwhile to see what proportion of upgraded class flights were made by women.

It would also be interesting in the context of a study of upgrades to check the relationship between the carrier that sold the tickets and the operating carrier (MARKETING_CARRIER, OPERATING_CARRIER). Referring to the paper by Dudin, Lyasnikov (2016), air carriers resell part of their tickets to online flight sellers to optimize their pricing. In this way, both the ticket agency and the air carrier get the best price (the one at which they are ready to sell tickets most cost-effectively), at the same time, using the services of external companies whose business model is based on mathematical modeling of air travel, they reduce the ticket price for customers (expecting, among other things, that some people will not arrive for the flight). A similar phenomenon occurs through the so-called airlines alliance mentioned by Ito, Lee (2007). Carriers offer tickets for flights from the range of their fellow airline alliance. Ito and Lee in their paper say that in this way, airlines can compete with each other price-wise by convincing customers with the greatest price flexibility. Such a move can have a significant impact on the market. Travelers for whom price matters so much, in our view, will be much less likely to purchase an upgrade. Taking these two things into account, in a situation where the operating carrier and the ticketing carrier are two different entities, we hypothesize that the propensity to upgrade a ticket class will be lower.

## 1.2. Initially adopted model

In the dataset shared by LOT Polish Airlines, we find several variables that have been confirmed in the literature to influence the studied phenomenon of class upgrade choice. These are variables related to the common carrier, gender of the passenger, as well as the fact of being a member in a frequent flyer program. From the dataset we can also derive the length of the flight from the bookings number, and a variable that identifies whether the flight is perceived as long (over 5000 km). Since we want to model which passengers are most likely to purchase a flight class upgrade, how and when it is best to inform the passenger that they have this option (hoping that the passenger will be tempted by our proposal), we decided to use one of the most popular and appreciated machine learning models - XGBoost. It has very extensive applications - it can be used in both the medical industry - e.g. using the algorithm to diagnose chronic kidney disease (Ogunleye, Wang 2019) - and in intruder detection systems (Dhaliwal, Nahid, Abbas 2018). We break our model into two parts. The first helps verify which passengers are most likely to purchase a ticket upgrade, while the second helps decide when it is the best time to inform a passenger about the upgrade opportunity. For the first model, we will model the variable UPGRADED_FLAG telling whether a passenger has decided to upgrade their ticket, while model 2 will use the variable purchase_time_diff describing how much time before departure a passenger has decided to upgrade their ticket.

## HYPOTHESES

The research hypothesis of this study, states that flight length positively influences the likelihood of purchasing a flight class upgrade. It is possible that the reason for this is, that since the passenger will have to spend a considerable amount of time in the aircraft it would be worthwhile to spend this time in a high-quality standard. Thus, we expect flight length to have a strong positive effect on the prediction of the explanatory variable. We find confirmation of this thesis in the research work of Kuo and Jou (2016). The researchers show a correlation that as the distance a plane travels increases, the proportion of people willing to pay for a premium class ticket increase. For the purposes of our article, we take these results as an indication that a passenger is more likely to upgrade to a premium class ticket on long-haul flights.

Another factor that may influence the class upgrade propensity may be the gender of the passenger. We believe that there is a significant difference between the number of flight class upgrades made between women and men. This may be due to several reasons. Firstly, men generally fly more, hence a passenger is more likely to decide to upgrade their ticket class on one of these flights. Secondly, more men than women travel for business purposes due to their professional roles. Men are in the majority in professions with higher prestige and higher salaries that involve travel, as according to a Barsh and Lee report from 2012, women account for only 35% of directors in American companies. An additional influence may also be women's greater attachment to home, family, children and, at the same time, reluctance to make regular distant business trips (Collins, Tisdell 2002).

The third hypothesis related to customer's propensity to upgrade their ticker is: when the operating carrier and the issuing carrier are two different institutions, the chance of buying an upgrade is likely to be lower. This is due to the existence of airline ticketing agencies. This case is described in the scientific article by Dudin and Lyasnikov (2016). The role of such intermediaries is to mathematically model the filling of the plane by passengers, while optimizing their profits. Agencies will sell more tickets than the seat limit assumes, considering the fact that there is a certain probability that not everyone will come to the flight. With competition in the market and overselling of tickets, prices for the lowest economy classes may be reduced. We anticipate that consumers who are targeted by such a measure have a very high elasticity of demand with respect to price, making them reluctant to upgrade a class under any circumstances.

The last problem which we would like to tackle, was determining the optimal timing to send an informational message via digital channels (sms, email, app push) to passengers with an active reservation who have not yet started their trip, indicating a possibility of a flight upgrade. Assuming that ticket's prices are the highest on the day of flight's departure (based, inter alia, on Janssen's research (2014)), it would be logical for the customer not to fall for the upgrade's higher cost. Thus, for the passenger to see an opportunity in improving his seat class, we presume that the notification must be sent at least 5 days before the departure. What's more, some airline customers may treat the money spent on a basic class air ticket as a sunk cost. In that case, those passengers would be more likely to buy the class upgrade presumably close to the departure. Taking all these aspects into consideration, we make an additional thesis that the optimal time to send an information message about the possibility of upgrading is a few days before departure (but no later than required by the airline's charter).

# DATA ANALYSIS

## 3.1. DATA

*Table 1. Summary of all variables used in the model*

| Variable name | Variable description |
|---|---|
| UPGRADED_FLAG | Dependent variable used in model 1, indicates upgrade on at least one flight coupon |
| purchase_time_diff | Dependent variable used in model 2, indicates how many days before the flight an upgrade by the passenger was bought |
| TIME_DEPARTURE_LOCAL_TIME | Local departure time, measured in hours |

| | |
|---|---|
| FLIGHT_DISTANCE | Distance of the flight |
| TOTAL_PRICE_PLN | Total price of flight in PLN |
| PAX_GENDER | Gender of the passenger |
| CORPORATE_CONTRACT_FLG | Feature that indicates if it was corporate trip |
| LOYAL_CUSTOMER | Feature that indicates whether the buyer was a loyalty program member |
| BOOKING_WINDOW_D | Days between sales date and initial departure date |
| STAY_LENGTH_D | Days between first booking departure and last arrival |
| BOOKING_LONG_HOUL_FLAG | Feature that indicates Long Haul Trip |
| BOOKING_DOMESTIC_FLAG | Feature that indicates Domestic Trip |
| PAX_N | Number of passengers on reservation |
| sale_to_flight_time | Time between purchase of the ticket and departure |
| flight_len | Length of the flight in hours |

| | |
|---|---|
| if_additional_upgrade | Feature that indicates if customer has bought any additional service such as e.g. additional baggage |
| same_carrier | Feature that indicates if the ticket was sold by the same airline that operates chosen flight |
| is_sus_aircraft | Feature that indicates whether an aircraft is viable of having more than economy class. Using exploratory data analysis, we selected the top 4 aircrafts that were capable of it. |
| is_sus_payment | Feature that indicates whether a method of payment that might be affecting ticket upgrade probability. Using exploratory data analysis, we selected 9 payment methods. |
| intinerary_len | Number of flights on one ticket |
| is_sus_currency | Feature that indicates whether a currency used for transactions that might be affecting ticket's upgrade propensity. Using exploratory data analysis, we selected 6 currencies. |
| FLIGHT_RANGE | Categorical feature which indicates whether the flight was domestic, short-haul (under 5000 km) or long-haul (over 5000 km) |
| BOOKED_CABIN | Type of booked cabin |

| | |
|---|---|
| VAB | Tariff category |
| PAX_TYPE | Categorical age classification of the passenger |
| SALES_CHANNEL | Feature that indicates by whom the ticket was sold |
| TRIP_TYPE | Categorical classification of trip type. It might be One-way, Multicity or Journey. |

*Source*: *Personal work and dataset description provided by LOT*

### 3.1.2 Data preprocessing

To preprocess the data, we first selected all columns, which we wanted to include in our default model. The next step was cleaning up the data types: defining categorical variables for adequate columns, changing string inputs to numerical inputs. After that we used hot one encoding, while at the same time dropping the first categorical value to change it to a baseline value. Then, we looked at incorrect observations: we addressed missing data by replacing it with the appropriate columns mean values (except ID values and those related to loyal customers). There were also a few examples, where the data was clearly misput (i.e., negative ticker price in PLN, negative booking windows, flight distance equal to zero between two airports). Those values were dropped from the dataset. We used the train-test-split procedure to ensure that the data we trained models on wasn't somehow biased. Stratifying both samples allowed us to make precise predictions.

## 3.2. Choosing features using forward selection

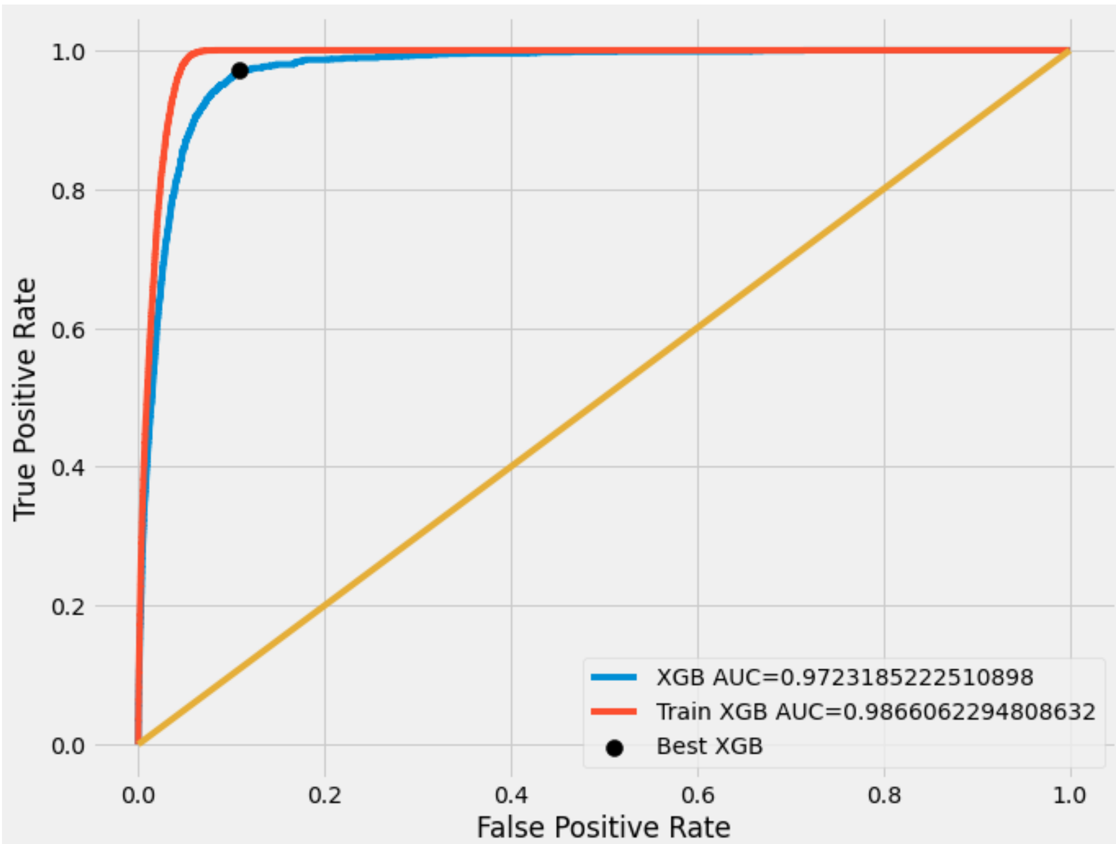
Forward selection is an iterative method of selecting variables that are the most valuable to the model. By going through all the columns, the algorithm chooses only those variables, which improve the model in terms of AUROC assessment. We start with having no features in the model. In our application of this method, we have chosen the - FLIGHT_DISTANCE - column as the first pick by manually going through feature importance analysis. The latter was assessed by using Bayesian hyperparameter optimization on our model. The choice was based on choosing the most important feature. Using repetition and only keeping the relevant variables, we make sure that our model is not overcomplicated with invaluable features, yet it's still capable of capturing interactions between regressors. We take a very similar approach while modeling the continuous variable. Though, contrary to modeling a categorical feature, the metric used for determining the loss function is MAE, not AUROC. Having both of these facts in mind, we create a general model that shouldn't be overfitted to the dataset yet should have high predictive power.

**MODEL 1. Version 1: Base XGBoost**

All the data we have was used to train model 1. About 10,000,000 million observations and 40 columns were used. The target of our prediction was UPGRADED_FLAG. Using the default XGBoost settings, we were able to achieve a cross-validated mean of ROC AUC = 0.9715 on the testing data split. The biggest difficulty in obtaining this result was the overfitting that occurred.

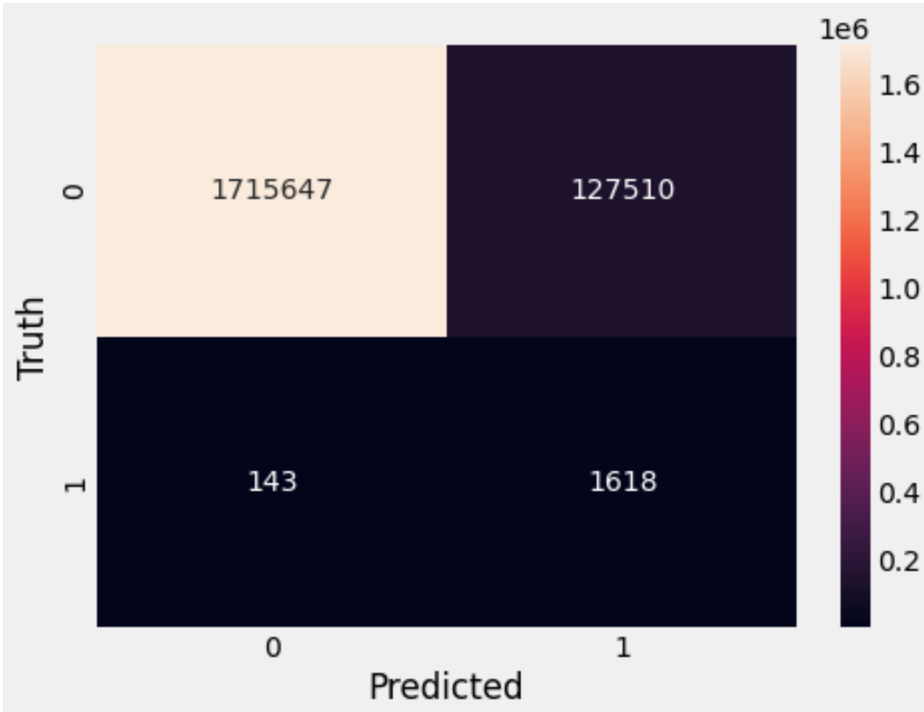
The difference between training and testing scores was equal to ~0.014 AUC, which indicates a big problem with our model. Our models often reached ~0.986 ROC_AUC on the training set and ~0.972 on the validation set. Nevertheless, due to overfitting and model complexity, our model did not perform well on the test set provided by LOT, obtaining a final ROC_AUC result of 0.932.

*Chart 1. ROC Curve of Version 1 XGBoost model.*




*Source: Personal work*

*Chart 2. Confusion matrix of Version 1 XGBoost model.*

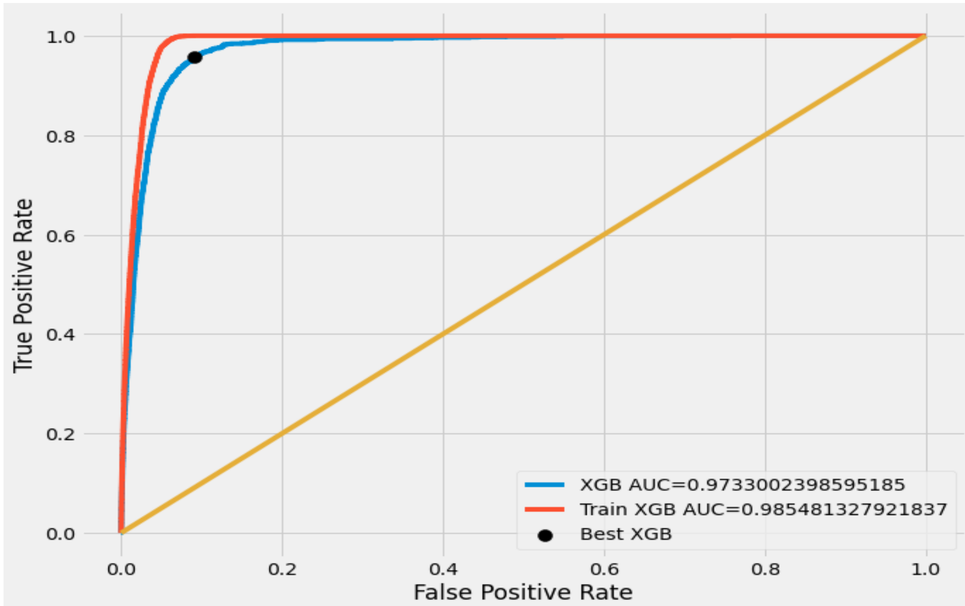



*Source: Personal work*

**MODEL 1. Version 2: Feature selection**

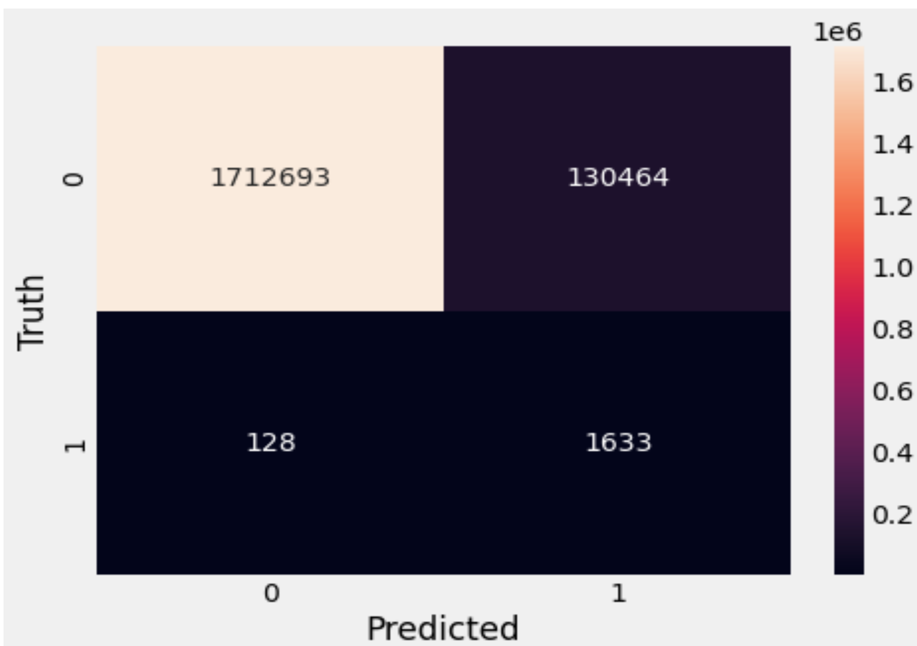
We decided that maybe the model was overfitting due to having too many features to choose from. Forward selection was the method that was chosen as our way of reducing the number of columns in the model. Using two iterations of our modified forward selection algorithm we reduced the number of features to 24, not counting the target. This improved our scoring metrics. Still using the default XGBoost settings, we were able to get a cross-validated mean of ROC AUC = 0.9712 on the testing data split. This is a loss from our first version of the model, but the difference between the splits was successfully reduced to ~0.012. This is still a big problem that needs to be addressed, but we stepped in the right direction. This version of our model performed better on the test set provided by LOT, obtaining a final ROC_AUC result of 0.943.

*Chart 3. ROC Curve of Version 2 XGBoost model*




*Source: Personal work*

*Chart 4. Confusion matrix of Version 2 XGBoost model*

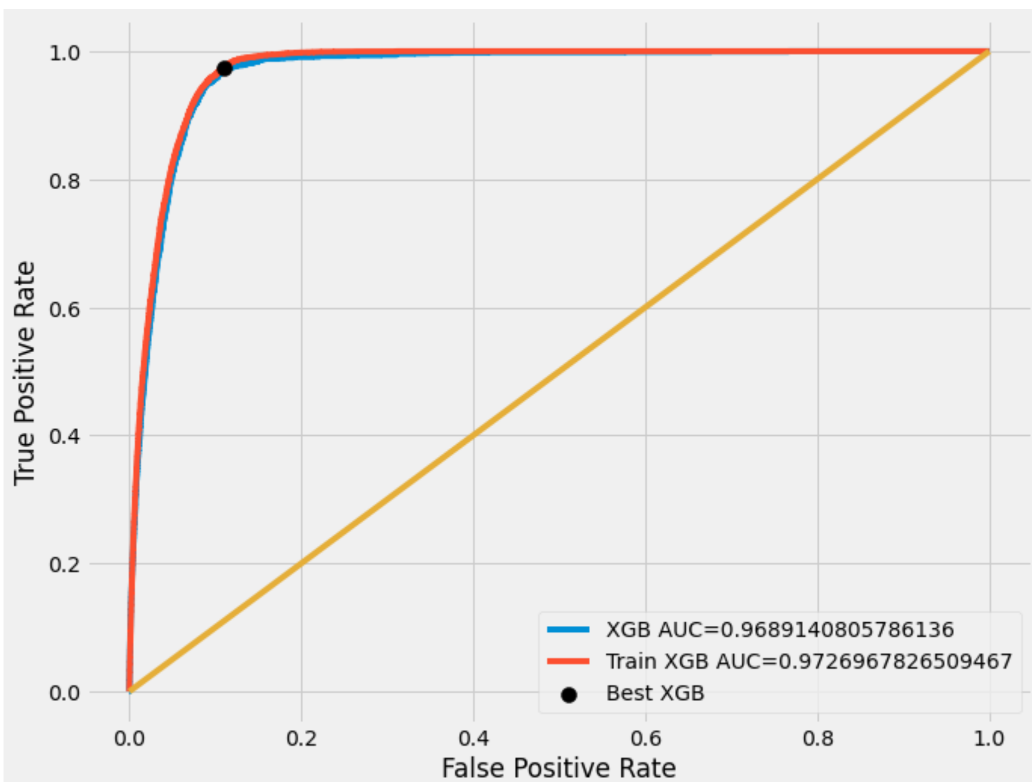



*Source: Personal work*

**MODEL 1. Version 3: Hyperparameters optimization**

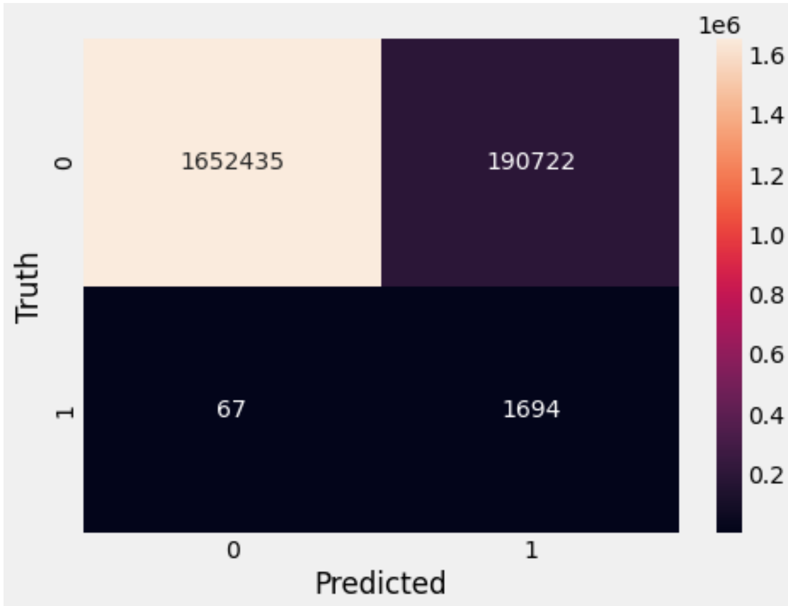
To further address the overfitting problem, we decided to perform automatic hyperparameter optimization using bayesian optimization using the python package hyperopt. We managed to find a result that satisfied our needs, and provided a satisfying solution to overfitting. After some manual tweaking, the final model achieved a cross-validated mean ROC AUC of ~0.9688, the results were also very stable, which further added value to this version. This model was generalized in a way that practically solved the overfitting problem, reducing the difference between scores in splits to a less than ~0.004, scoring 0.9689 on validation and 0.9726 on training set. As a result of these optimizations, without overfitting and with reduced model complexity, our model did finally perform best out of all versions on the test set provided by LOT, obtaining a final ROC AUC result of 0.9529.

*Chart 5. ROC Curve of Version 3 XGBoost model*



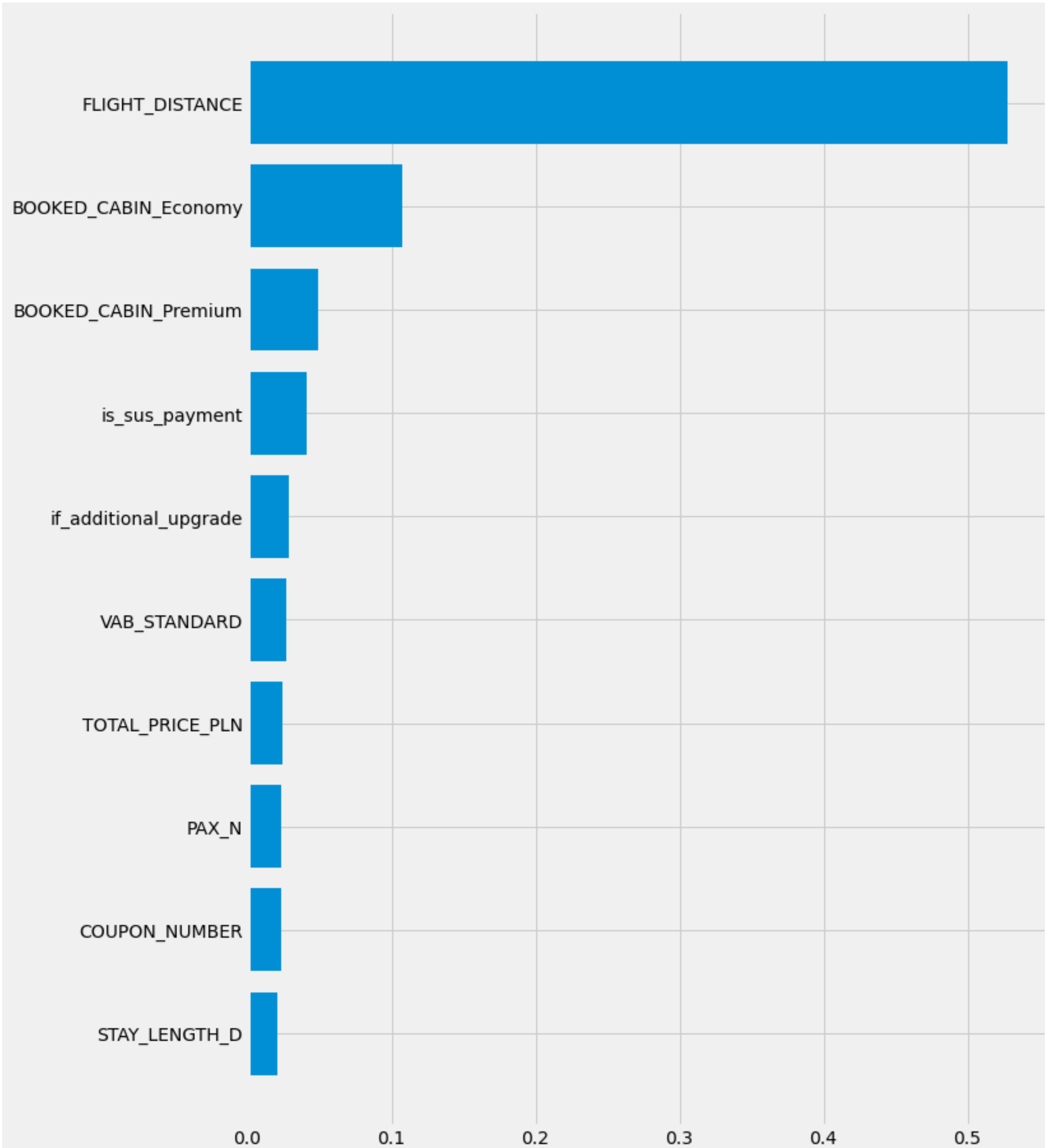

*Source: Personal work*

*Chart 6. Confusion matrix of Version 3 XGBoost model*




*Source: Personal work*

*Chart 7. Top 10 most important of Model 1. Version 3.*




*Source: Personal work*

To address the black box problem in our model, feature importance analysis was applied. Chart 7 shows that the prediction of the customer's upgrade probability is mostly influenced by flight distance. That is followed by the type of flight class booked by the customer, then by the type of payment method used.

**MODEL 2**

For training model 2, the data used in model 2 have been changed. As in this case we are only interested in people who decided to purchase an upgrade, all observations in which the passenger did not decide to upgrade were removed. Then, considering the competition requirements given by LOT airlines, only passengers who decided to upgrade at least 5 days before departure were considered.

With feed forward, we were able to change the number of explanatory variables from the original 40 to only 6. However, due to the multivariate nature of the data and the use of nonlinear estimators to capture what effect a variable has on the regressor, we modeled the predictive response to a change in the regressor ceteris paribus. We present the results in the following figures (add legitimate captions and illustrations and sources). To evaluate model 2, we used the following characteristics:

Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE).

The best model that we were able to create has the following characteristics.

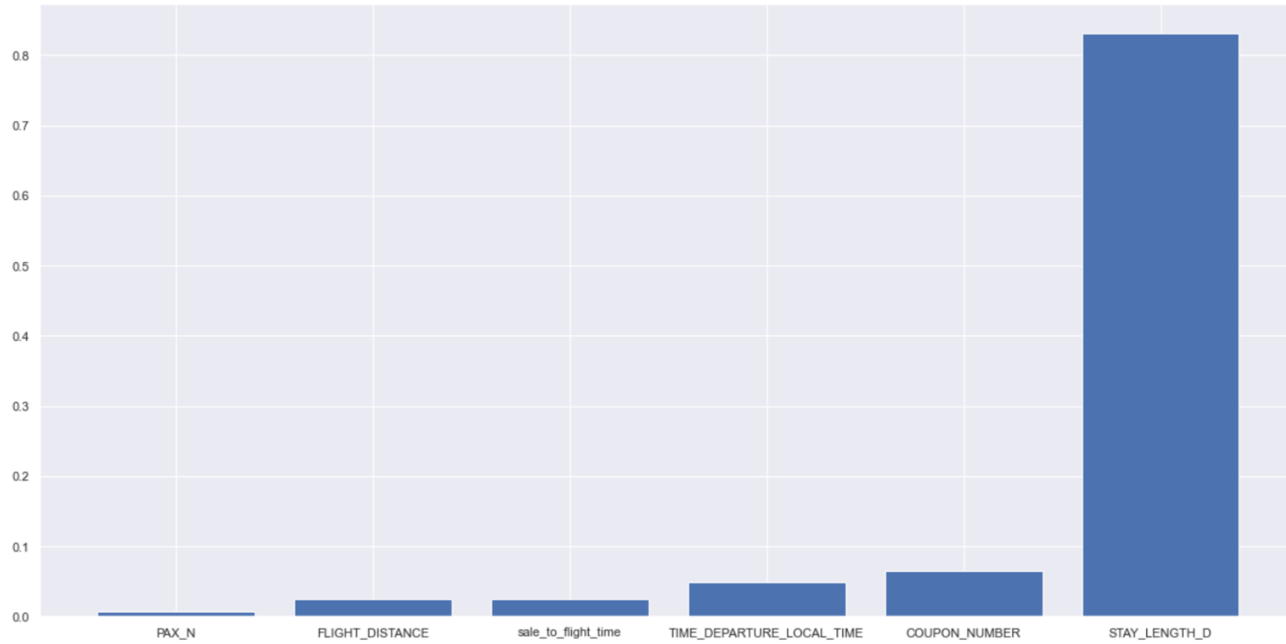
MAE: 1.87

RMSE: 5.27

MSE: 33.9

This means that our proposed model was wrong by 2.28 days on average, which we consider a very good result.

*Chart 8. Feature importance of Model 2.*




*Source: Personal work*

To consider the black box problem, the feature importance implemented in XGBoost library was used. As you can see, the prediction of the correct upgrade date is by far the most influenced by the length of the departure. It is followed by the number of tickets on the order and the local departure time.

## HYPOTHESES VERIFICATION

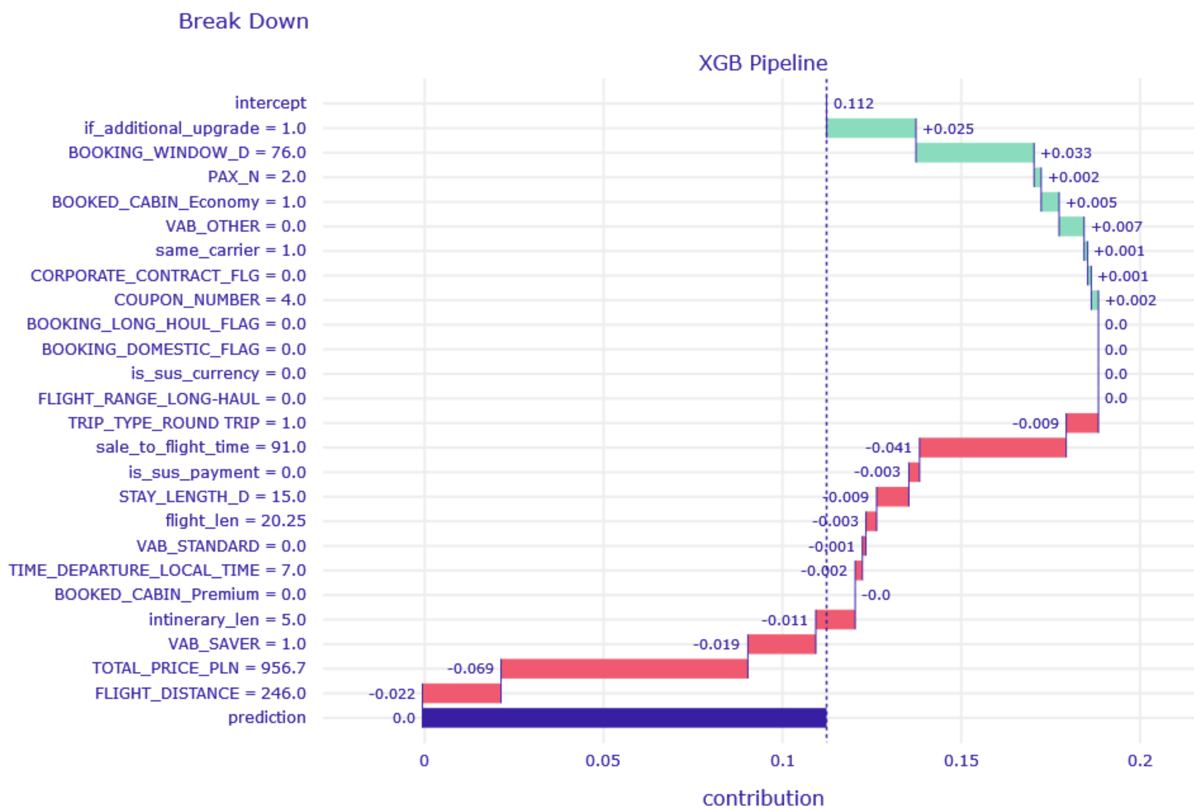
To verify the hypotheses raised, we used Model 1 Version 3 for the three first postulates and Model 2 for the last hypothesis.

Hypothesis 1: flight length positively influences the likelihood of purchasing a flight class upgrade.

To address this problem, we plotted a breakdown graph of key features that our model used to categorize the data. First, we look at random observation, where the passenger did not buy the class upgrade. We calculated this customer's propensity of buying the upgrade at 0.11. Addressing graph 9, we can see that flight distance played a key role in predicting the probability. Since the distance between arrival and departure city is low, the customer had no need to buy an upgrade.

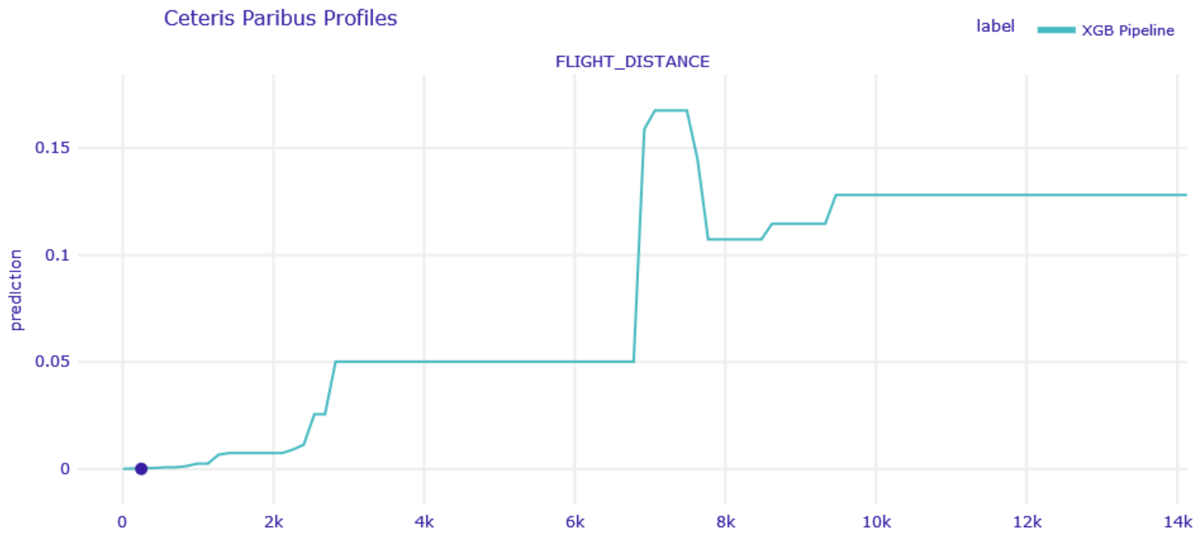
The next chart (10) plots flight distance against prediction for buying the upgrade. Looking at the light blue line, we can see that the trend is: the higher the distance, the higher the prediction for upgrading the ticker.

*Chart 9. Model 1 sample prediction breakdown for customer who didn't buy an upgrade*




*Source: Personal work*

*Chart 10. Flight distance versus propensity to buy ticket upgrade for an upgrade non-buyer*

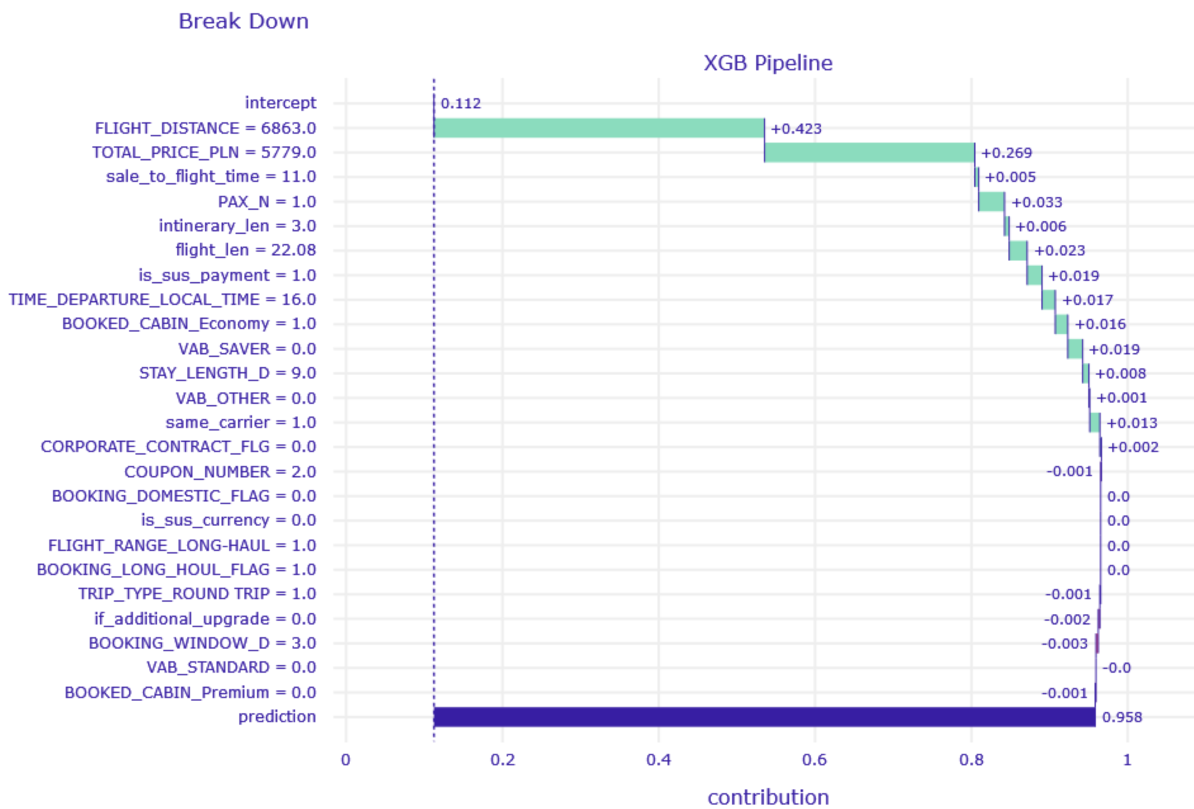



*Source: Personal work*

Though we should look at a passenger, whose decision was to update the flight, to make sure that the predictions are right for both sides.

Looking at feature breakdown for a random user who did buy and upgrade, we can see that that flight distance was most influential in determining the propensity of customers buying an upgrade. We assessed with 96% probability that had taken this action, and we proved right.

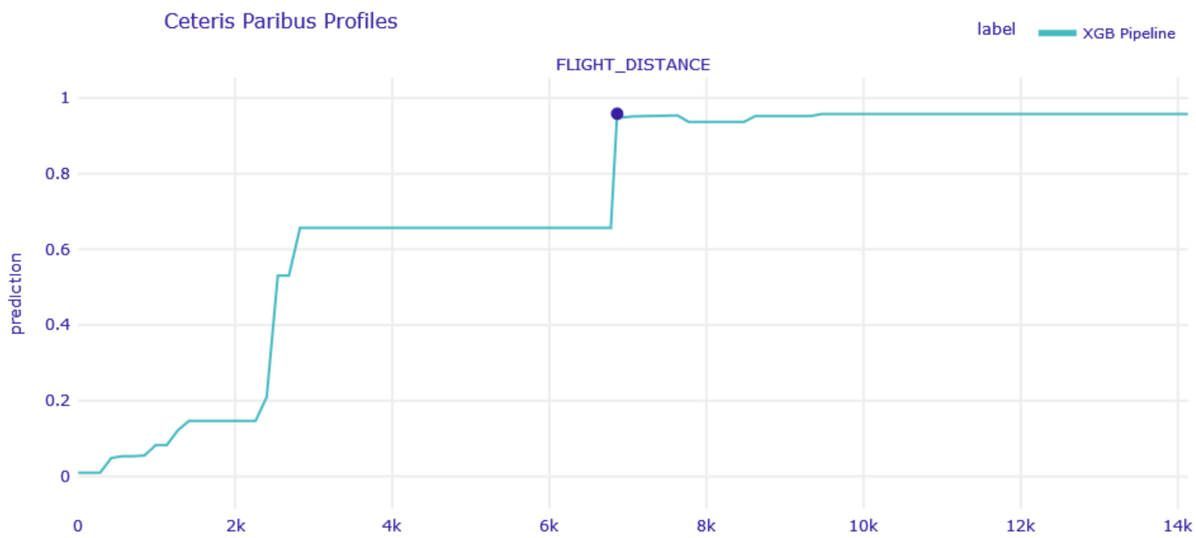
*Chart 11: Model 1 sample prediction breakdown for customer who bought an upgrade*




*Source: Personal work*

Let us see the ceteris paribus flights' distance influence on propensity to buy a class upgrade. Plot on Chart 12 is similar to the one on Chart 10, though a bit different. In this case, the probability of buying a ticket upgrade follows a clear trend and allows only one propensity value drop while at the time rising in value. It is important to state that the ceteris paribus profiles are very similar, they don’t depend on the observations. We believe that hypothesis 1 which states flight length positively influences the likelihood of purchasing a flight class upgrade shall not be rejected.

*Chart 12: Flight distance versus propensity to buy ticker upgrade for an upgrade buyer*

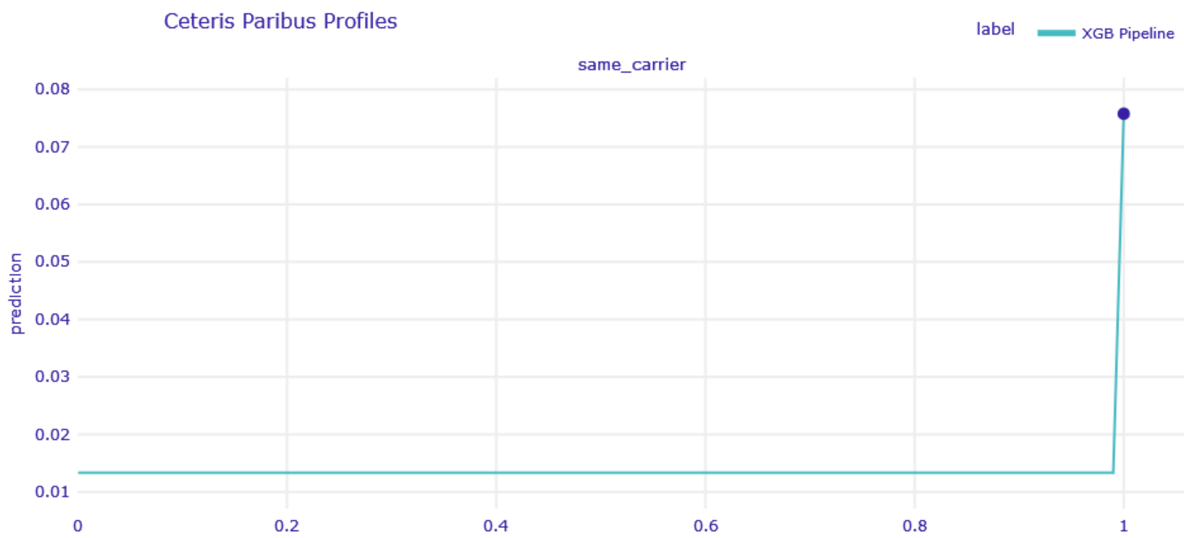



*Source: Personal work*

Addressing hypothesis 2, we can use charts 9 and 11, similarly to hypothesis 1. While flight distance did play an important role in predicting propensity of buying a class upgrade, we can see that there is no variable indicating passenger's gender. It is because that feature did not play and important role during the prediction modeling. Based on that, we can assume that gender does not play an important role in assessing passenger's propensity to buy an upgrade, thus rejecting hypothesis 2.

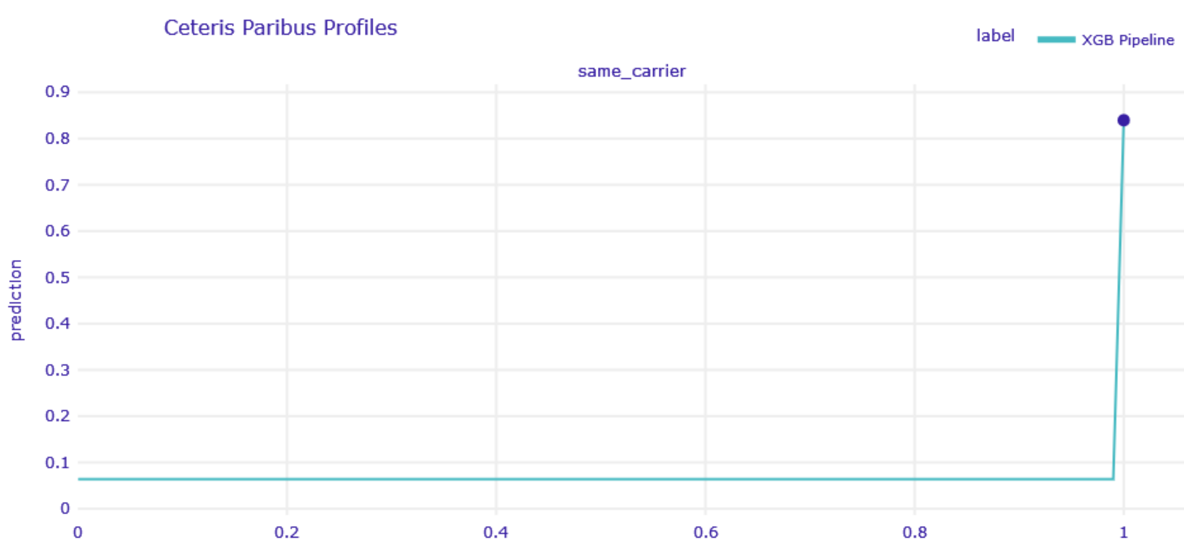
While discussing hypothesis 3, it is important to bring up another two charts - 13 and 14. Firstly, looking at charts 9 and 11, we can see those tickets having the same carrier do influence the propensity, they account for respectively: 1.3% and 0.1%. Though the values are marginally low. The values somehow represent what we stated in this hypothesis. Looking at the ceteris paribus profiles for non-buyer and buyer (sequentially, chart 13 and 14), we can see that when the carrier is the same, it increases the propensity for non-buyers of upgrading their class by about 7.5%. Although, when we change the same_carrier variable, ceteris paribus, for a buyer that is likely to buy the class upgrade, we can see that it increases by about 80%. We believe that such results should not cause the rejection of hypothesis 3, thus we believe that while the operational and marketing carrier is the same, it increases the consumer propensity to buy the class upgrade.

*Chart 13: Flight distance versus propensity to buy ticker upgrade for a non-upgrade buyer*




*Source: Personal work*

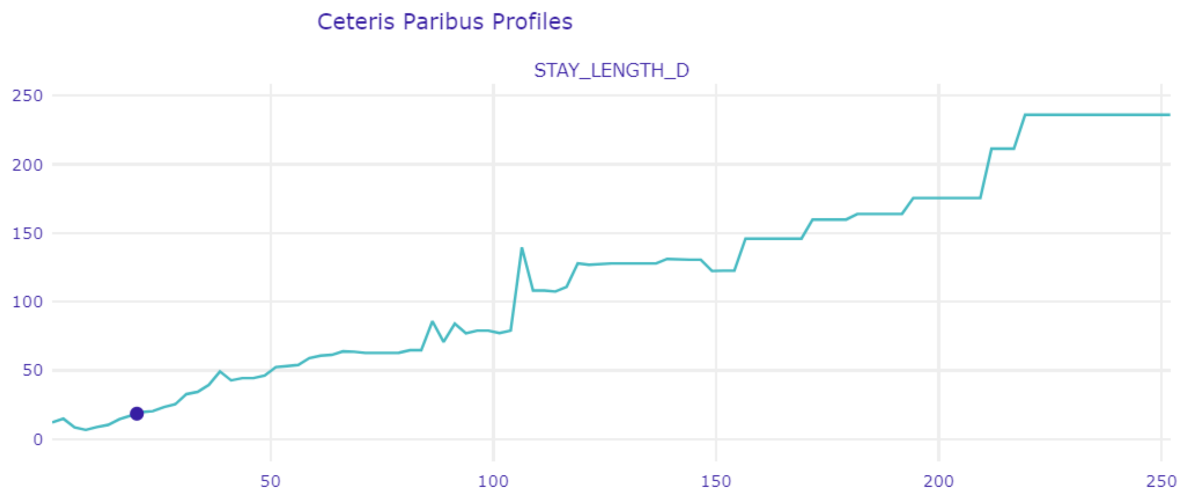
*Chart 14: Flight distance versus propensity to buy ticker upgrade for an upgrade buyer*




*Source: Personal work*

Addressing the second problem: verifying hypothesis 4. The effect of the passenger's length of stay at the destination on the day to be notified is affected as follows: As the length of planned stay increases it steadily notes that the potential customer should be notified earlier and earlier, based on chart 15.

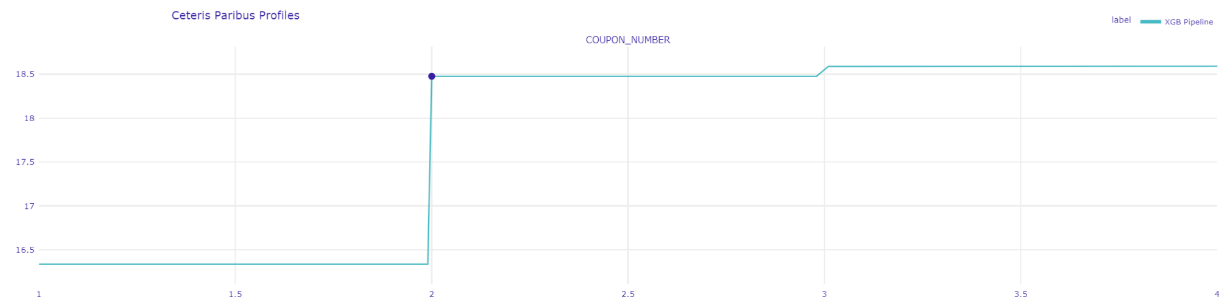
*Chart 15: Ceteris paribus profiles for length of stay against how early the consumer should be reminded of class upgrade option*




*Source: Personal work*

The trend is upward, and we expect that the longer the stay a consumer plan, the earlier they should be notified. Business Interpretation: preferred date of customer notification is as early as possible after the flight reservation. Looking at the data, we believe that we cannot reject the null hypothesis - the customers prefer to buy tickets early.

*Chart 16: Ceteris paribus profiles for coupon number against how early the consumer should be reminded of class upgrade option*




*Source: Personal work*

What's more, looking at the data, we noticed that holders of one-way tickets, on the other hand, prefer to be notified much earlier about the possibility of class upgrade before departure than those of two-way or connecting flights.

## INTENTIONAL SELF-CRITICISM

The model we managed to construct is by no means perfect. It lacks a deeper look at the data itself, and statistical tests should be performed to get a better idea of the relationships that exist in this database. But data is not the only shortcoming. It would also be worthwhile to test more than one machine learning model to select the most optimal one. Certainly, neural networks would be a promising choice, but they would in turn require normalization and more careful transformation of the data. The XGBoost model has been used in this paper because it is a model that solves many similar problems in a promising way, while not requiring much data processing. We were not given the opportunity to test whether this model does the best job on this problem, yet we are pleased with the results that were achieved in these 24 hours of work.

## SUMMARY

Due to the high growth dynamics of the airline market and the consequent development of tourism, the competition between airline companies for their share of the public transport market is intensifying. Global airlines are now facing the challenge of how-to best tailor their service offerings to their passengers. To succeed, an active information campaign is needed to spread awareness among customers of the possibility of customizing services and upgrading the quality of the flight at additional cost. It is important that in the current process passengers receive a lot of pre-departure messages regarding the itinerary and the airline's offerings. Caring for a positive customer experience, the carrier cannot allow too many messages to be sent to passengers. The outreach campaign should be targeted to those who have a high propensity to upgrade. Hence, the purpose of this study was to determine the propensity of LOT Polish Airlines passengers to upgrade their travel. While an additional objective was to determine the optimal timing for sending an informational message. Based on data analysis and the hypotheses formation, we reject hypothesis 2, which stated that gender has and influence of customer’s propensity to buy a class upgrade. We do not reject hypotheses 1, 3, 4, thus concluding that flight’s distance and whether the same carrier operates and sells the tickets for it has affects customer’s likelihood to buy a class upgrade. Addressing hypothesis 4, we believe that a customer should be notified about an upgrade possibility rather earlier than sooner, though we provide a model to predict strict values for certain customers. What’s more, we see a connection between coupon number and when to alert customer of class upgrade possibility, remarking that one-way passengers prefer to be notified later than those who had coupon number more than 2.

Further work on the study should extend the analysis by preparing the data more thoroughly, creating more additional variables, and using a more complex model architecture. Due to the large number of both observations and variables, using a neural network architecture seems to be a good step forward.

## REFERENCES

Cui Y., Orhun A., Duenyas I., 2018 How Price Dispersion Changes When Upgrades Are Introduced: Theory and Empirical Evidence from the Airline Industry

Dhaliwal S., Nahid A., Abbas R., 2018 Effective Intrusion Detection System Using XGBoost

Dudin M., Lyasnikov N., 2016 Ticket sales optimization in the conditions of the independent and crossover demand on the basis of economic and mathematical modeling

Gossling S., Nilsson J., 2010 Frequent flyer programmes and the reproduction of aeromobility

Ito H., Lee D., 2007 Domestic Code Sharing, Alliances, and Airfares in the U.S. Airline Industry

Janssen, T. 2014 A Linear Quantile Mixed Regression Model for Prediction of Airline Ticket Prices

Kuo C., Jou R., 2016, Willingness to pay for airlines' premium economy class: The perspective of passengers

Liston-Heyes C., 2002, Pie in the Sky? Real Versus Perceived Values of Air Miles

Ogunleye A., Wang Q., 2019 XGBoost Model for Chronic Kidney Disease Diagnosis

Tisdell C., Collins D., 2002 Gender and Differences in Travel Life Cycles

Yee L., Barsh J., 2012 Unlocking the full potential of women at work, McKinsey