Airline Passenger Satisfaction

Dani Buitrago, Grace Hyland, Ofure Udabor

Department of Data Science & Analytics, Georgetown University

April 2024

Introduction

US airlines transported 751.4 million passengers in 2023 both domestically and internationally. With such a high number of people taking flights every year, airlines have a vested interest in maintaining the satisfaction of their customers in order to ensure their profits in a competitive market. This is especially true due to the highly competitive nature of the airline industry. With numerous airlines vying for passengers' business, customer satisfaction is a key differentiator, with satisfied customers more likely to choose airline services again in the future, contributing to loyalty and repeat business.

We intend to analyze a dataset which surveys airline passengers on various different aspects of their experience flying with the given airline and then asks whether they were satisfied with the experience as a whole. Our analysis will attempt to determine whether the satisfaction of a passenger can be predicted based on the surveyed features, and if so, which aspects of a flight have the greatest level of impact on a passenger's overall satisfaction.

Our goal in this analysis would be to provide airlines with a guide on which areas would be most efficient to improve for customer retainment. This strategic approach ensures that airlines can allocate resources efficiently, focusing on areas that yield the highest returns in terms of customer satisfaction and, ultimately, loyalty. By addressing the key pain points identified through our analysis, airlines can cultivate a more positive and fulfilling travel experience for passengers, thereby bolstering their chances of retaining customers in an increasingly competitive market. The three models we intend to utilize to analyze this data are a Support Vector Machine Model, a Logistic regression model, and a Generalized Additive Model.

In conducting this analysis, we have also looked into multiple studies regarding predicting the satisfaction of airline passengers. The *J.D. Power 2023 North America Airline Satisfaction Study* discusses a paradox in the aviation industry, where, despite record revenues driven by high demand and limited supply, customer satisfaction with major airlines is declining due to soaring airfares, staffing shortages, and reduced routes. Low-cost carriers face particular challenges, where passengers seek airfare bargains. In contrast, first-class passengers' experience improved satisfaction, attributed to enhanced food and beverage services. (*2023 North America Airline Satisfaction Study*, 2023)

The study *Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model*, discusses how the civil aviation industry faced profound challenges, and how addressing passenger satisfaction became crucial for airlines in lieu of a decrease in travel demand. This study explored predictive models and factors influencing passenger satisfaction, emphasizing the importance of service quality in retaining customers. Various methodologies, including machine learning algorithms and feature selection techniques, were employed in this study to extract key factors affecting passenger satisfaction. (Jiang et al., 2024)

The study *the impact of the COVID-19 pandemic on airlines' passenger satisfaction* investigates airline passenger satisfaction trends before and during the COVID-19 pandemic by analyzing influential factors using a dataset of 9745 passenger reviews from airlinequality.com. Sentiment analysis and machine learning algorithms conducted in this study show a decline in satisfaction post-pandemic, with staff behavior emerging as a crucial factor. Addressing negative sentiments emerged, through predictive modeling, as one of the most important factors in customer satisfaction. (*The Impact of the COVID-19 Pandemic on Airlines' Passenger Satisfaction*, 2023)

Data Gathering & Cleaning

Support Vector Machines:

For the Support Vector Machine model, data processing began by taking a sample of 20,000 observations from the passenger satisfaction dataset. Because support vector machines are trained by solving a quadratic optimization problem, the original dataset, which consisted of 129,880 observations, was too large for the SVM modeling functions to process in a reasonable amount of time, and it was therefore necessary to take a sample to continue analysis with the model. After this, for the SVM modeling function to be able to process the data we were giving it, we had to change the features with character types, to factor types. Therefore, the Gender feature was encoded with 0 representing 'Male' and 1 representing 'Female', the Customer Type feature was encoded with 0 representing 'disloyal Customer' and 1 representing 'Loyal Customer', the Type of Travel feature was encoded with 0 representing 'Business Travel' and 1 representing 'Personal Travel', and the Class feature was encoded with 0 representing 'Eco', 1 representing 'Eco Plus', and 2 representing 'Business'. These features were generally encoded with higher numbers referring to the value we believed correlated to higher customer satisfaction. The rest of the features for the dataset are numeric and therefore were left unchanged. After this encoding we then removed all rows with NAN values, and then split the data into testing and training with 80 percent of the data going into training and 20 percent becoming the test set.

Logistic Regression:

For the logistic regression model, all 129,487 observations in the data set were used in the model training process. In the cleaning process we normalized all of the continuous variables to a range between 0 and 1 to match the magnitude of the binary categorical variables which had values of 0 or 1. Additionally, we converted said binary categorical variables to have the value 1 or 0 rather than the string values that were present in the data set initially. These variables were gender, customer type, type of travel, and class.

We then checked the data set for missing values and found that there were 393 rows that had a missing value for the arrival delay in minutes. To determine the meaning of the missing value, we checked the data set for cases where arrival delay was equal to 0 to determine if missing value may mean that there was no arrival delay. The data set contained 72,753 instances of a 0 for this variable. Therefore, no value for arrival delay was not equivalent to not having an arrival delay and instead likely means that we do not know the arrival delay time. Because of the size of the dataset and the uncertainty of this value, we decided to remove the 393 rows that had a missing value.

Generalized Additive Models:

We decided to focus on only monitoring the impact of the continuous variables as they pertain to the satisfaction of airline passengers. The continuous variables in the airline passenger survey dataset are: "Age", "Flight.Distance", "Departure.Delay.in.Minutes", and "Arrival.Delay.in.Minutes". The age variable is included due to the wide range of ages included; this includes a minimum age of 7 years old, a maximum age of 85 years old, and an average age of 39.4 years old.

To start, we prepare our data for analysis. The dataset is split into training and testing subsets, where the entire dataset experiences a 70 percent to 30 percent ratio split. Once split, both datasets are made into smaller subsets, containing only the continuous variables measuring: the age of each reporting passenger, the distance each reporting passenger traveled, the length of the reporting passenger's departure delay, and the length of the reporting passenger's arrival delay. Subsequently, the "satisfaction" column in the training dataset is redone. The column, which formerly had "neutral or dissatisfied" and "satisfied" as the two values exclusively contained in the column, now has a "0" in place of "neutral or dissatisfied" and a "1" in place of "satisfied".

As a part of preparing our variables for the Generalized Additive Models, we first ran a logistic regression model on the training data subset to get an idea of what variables are more impactful on the choice of satisfaction. To do this, we produced a generalized linear model in R. As seen in the visual below, the most significant predictor variables are age (significance code is approximately 0) and flight distance (significance code is approximately 0), while arrival delay in minutes is slightly less significant (significant code is approximately 0.001) and departure delay in minutes is deemed even less significant (significance code is approximately 0.05). This initial analysis assists us in gauging what variables to focus on and which to disregard. Since each variable carried some linear significance, we decided to include each variable in the future analyses to different extents.

```
Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)               -1.588e+00  5.552e-02 -28.607  < 2e-16
Age                        1.597e-02  1.201e-03  13.296  < 2e-16
Flight.Distance            6.263e-04  1.903e-05  32.903  < 2e-16
Departure.Delay.in.Minutes 3.535e-03  1.837e-03   1.924  0.05433
Arrival.Delay.in.Minutes  -5.457e-03  1.828e-03  -2.985  0.00284

(Intercept)                ***
Age                        ***
Flight.Distance            ***
Departure.Delay.in.Minutes .
Arrival.Delay.in.Minutes   **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 19209  on 13993  degrees of freedom
Residual deviance: 17711  on 13989  degrees of freedom
AIC: 17721

Number of Fisher Scoring iterations: 4
```

Summary of variables, highlighting their significance levels.

Exploratory Data Analysis

For this plot we wanted to explore the counts of the various categorical variables. We did this in order to see the distribution of the different factors that go into determining a passenger's satisfaction. As you can see from the graphs, while the gender distribution is, as expected, more or less equal, the customer type is skewed heavily, so that loyal customers are making up a far larger percentage of the dataset compared to disloyal customers. The type of travel is also skewed with business travel being twice as common among the passengers as compared to personal travel. The final graph showing the distribution of classes shows how business and economy are more or less equally distributed with economy plus making up very little of the dataset.

EDA Plots.

For this plot we wanted to explore the distributions of satisfied vs unsatisfied passengers dependent on the categorical variables Class, Type of Travel, Customer Type, and Gender. We did this in order to get a general understanding about how each of these variables might individually affect passenger satisfaction. As you can see from the graphs, the distribution of satisfied and unsatisfied passengers is more or less equal across genders, with both genders experiencing slightly higher numbers of unsatisfied passengers compared to satisfied passengers. For customer type both loyal and disloyal customers were comparatively dissatisfied, though disloyal customers did experience comparatively higher rates of dissatisfaction. The distribution of satisfaction changes substantially however, when looking at type of travel, as business travelers were more satisfied than dissatisfied, but passengers on personal travel were overwhelmingly dissatisfied. The final graph showing the distribution of classes shows how business travelers were substantially more satisfied than economy or economy plus travelers with economy travelers showing incredibly high rates of dissatisfaction.
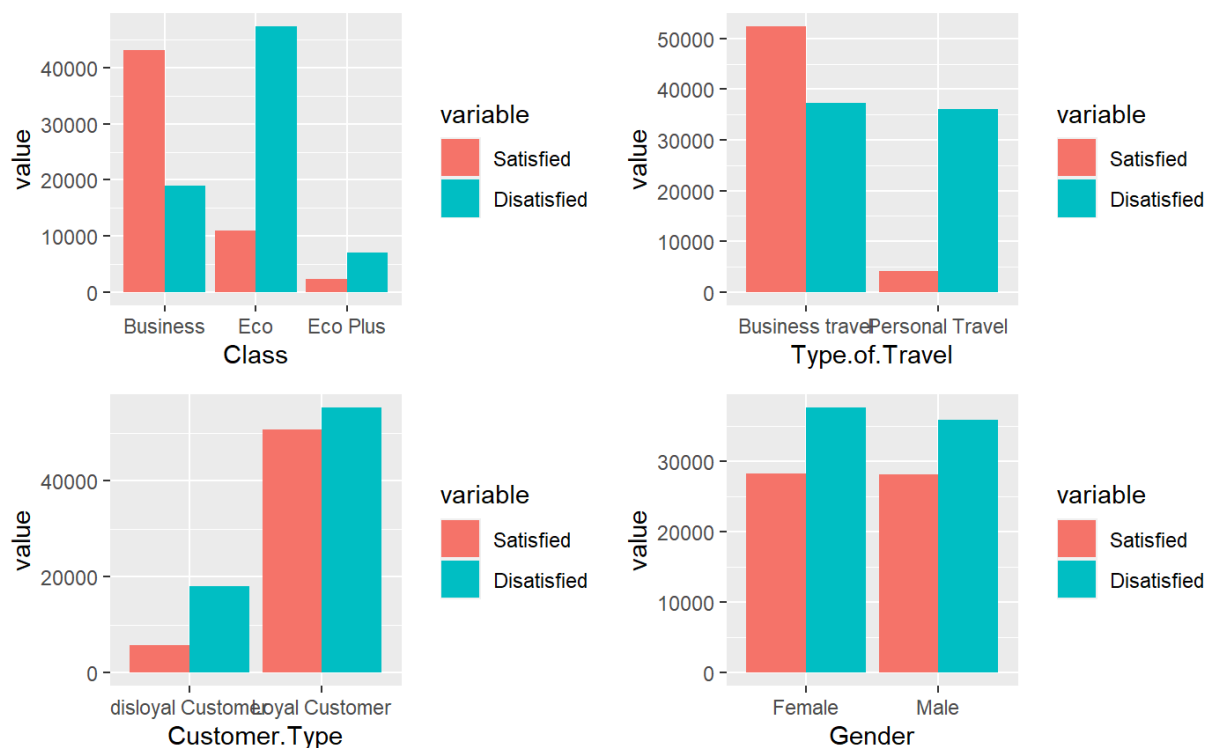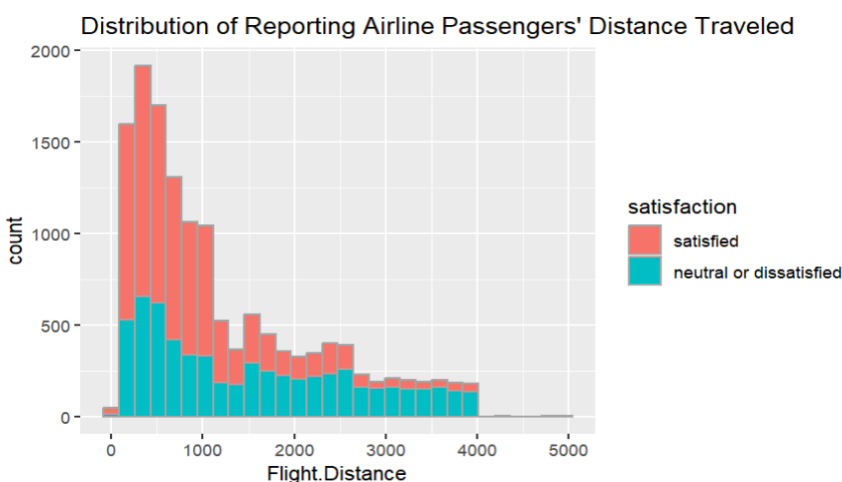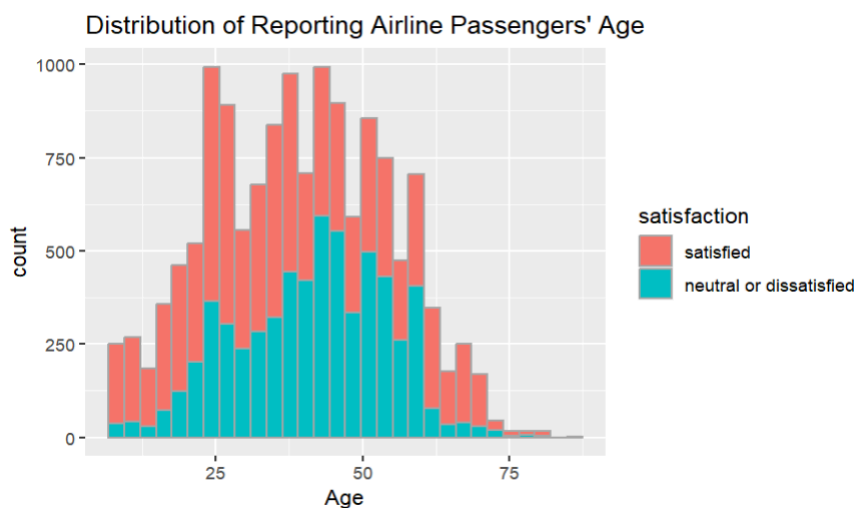
EDA Plots.

We decided to focus on the continuous variables, as we will be isolating these specific variables to serve as predictor variables later on in our analysis process. Below, we utilized histograms to display the various distributions for our dataset's continuous variables as they apply to the reporting passengers' chosen satisfaction level. This is performed to convey the age of the reporting passengers, the flight distance each reporting passenger traveled, the arrival delay each reporting passenger experienced (measured in minutes), and the departure delay experienced by each reporting passenger (measured in minutes). All distributions are drawn from the testing subset used for our future analyses, which only contains: the age of the reporting passengers, the flight distance each reporting passenger traveled, the arrival delay each reporting passenger experienced (measured in minutes), and the departure delay experienced by each reporting passenger (measured in minutes).

Our first distribution reports on the age of the reporting passengers. The minimum age of 7 years old seems to belong around 250 of the reporting passengers in the dataset. While the maximum age of the reporting passengers is 85 years old and belongs to noticeably less than 50 of the reporting passengers. The distribution has a slight right skew which displays a majority of the reporting airline passengers population being anywhere between 25 years old and 50 years old.

As shown below, the majority of the reporting passengers choose to have been satisfied with their flight.

Our second histogram distribution displays the distance traveled by the reporting passengers. The minimum distance being 0 seems to belong to distinctively less than 50 of the reporting passengers in the dataset. While the maximum distance flown by the reporting passengers is approximately 5000 miles and belongs to strikingly close to 0 of the reporting passengers. The distribution has a clear right skew which displays a majority of the reporting airline passengers population traveling less than 1500 miles. As shown below, the majority of the reporting passengers choose to have been satisfied with their flight, with a higher percentage of passengers being satisfied when their flight distance is below 1000 miles.
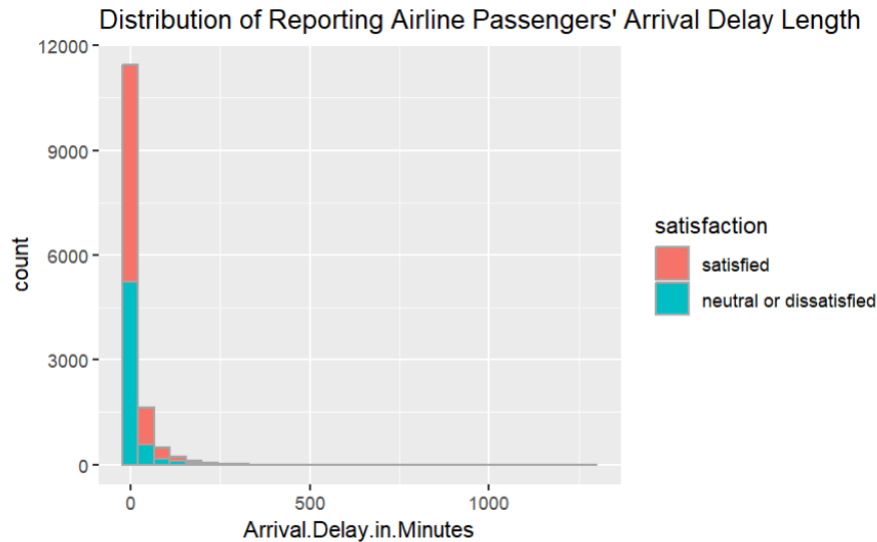

EDA Plots for Continuous Variables in Testing Subset.

Our third distribution reports on the length of the departure delays in minutes of the reporting passengers. The minimum duration being 0 minutes seems to belong to almost 12,000 of the reporting passengers in the dataset. While the maximum duration of the departure delay of the reporting passengers is barely visible past approximately 250 minutes. The distribution has a clear right skew which displays a majority of the reporting airline passengers population not experiencing much of a flight departure delay at all. As shown below, the majority of the reporting passengers choose to have been satisfied with their flight.

Our fourth distribution reports on the length of the arrival delays in minutes of the reporting passengers. Just like the previous histogram, the minimum duration of arrival delay is 0 minutes and apparently is experienced by almost 12,000 of the reporting passengers in the dataset. Whereas, just like in the previous plot, the maximum duration of the departure delay of the reporting passengers is barely visible past approximately 250 minutes. The distribution also has a clear right skew which displays a majority of the reporting airline passengers population not experiencing much of a flight departure delay at all. As in the last histogram for Exploratory Data Analysis below, the majority of the reporting passengers report to have been satisfied with their flight.
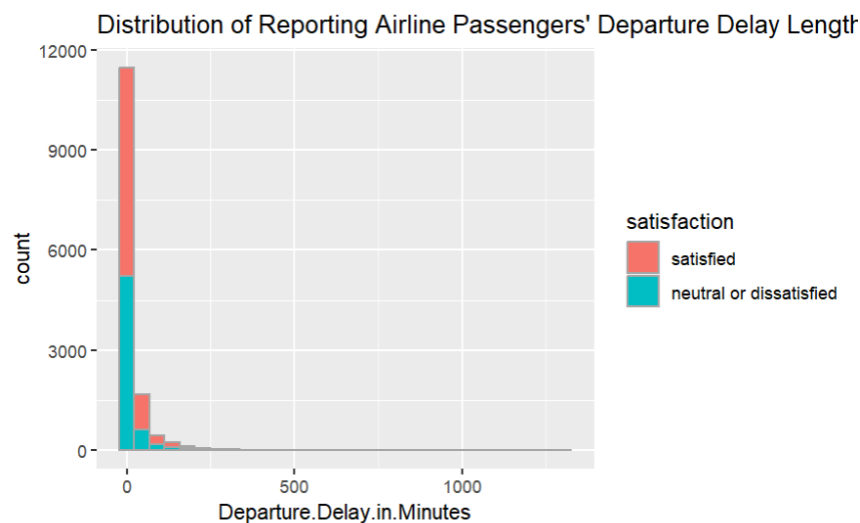
EDA Plots for Continuous Variables in Testing Subset.

Statistical Methods & Analysis

Support Vector Machines:

Support Vector Machines (SVMs) are powerful supervised learning models used for classification and regression tasks. They are particularly effective in high-dimensional spaces and when the number of dimensions exceeds the number of samples. SVMs work by finding the hyperplane that best separates different classes in the input space. Support Vectors are the data points that are closest to the hyperplane, while the hyperplane is a decision boundary that divides the feature space into two parts, one for each class. SVM aims to maximize the distance between the hyperplane and the support vectors to improve the model's generalization ability and its resistance to overfitting.

SVMs can efficiently perform non-linear classification using a kernel. Instead of directly finding the hyperplane in the original input space, SVMs can implicitly map the input vectors into high-dimensional feature spaces. Common kernels include linear, polynomial, radial basis function (RBF), and sigmoid kernels. SVMs also incorporate regularization parameters that control the trade-off between maximizing the margin and minimizing classification errors. This helps prevent overfitting and improves the model's performance. SVM utilizes a loss function which penalizes misclassification and aims to minimize classification errors while maximizing

the margin. Common loss functions include hinge loss for classification tasks and epsilon-insensitive loss for regression tasks.

In this case we will create a model that will divide the observations into two groups in order to predict which observations belong in the satisfied class and which belong in the non-satisfied class based on the data features.

Logistic Regression:

A logistic regression model is a machine learning model that is used in classification problems to assign one of two or more labels to each data point. Logistic regression uses a discriminative modeling approach where the probabilities of classes are evaluated using X directly which differentiates it from the similar naive Bayes classifier. A logistic regression model takes training data with both inputs and outputs and is trained using stochastic gradient descent and cross-entropy loss. The trained model then applies a label to each input of the test data set by determining the probability of each label being correct and assigning the label with the highest probability. In binary classification, which we will do in this analysis, the decision boundary is such that if the probability of one label is higher than 0.5 then that label will be assigned (Jurafsky & Martin, 2008).

Logistic regression is a good classifier to use for large data sets with potentially correlated features. The model can detect correlated features and will split part of the weight between each of the correlated features rather than treating them as independent features like a naive Bayes classifier (Jurafsky & Martin, 2008). Our dataset has many features that have potential to be highly correlated such as satisfaction with the check-in service and satisfaction with the boarding service so a logistic regression model is best for the data set we are working with.

In this analysis, we will train a logistic regression model that can identify whether or not a passenger was satisfied with their flight experience based on responses to survey questions. Through this model we will be able to identify the most important features for airline passenger satisfaction.

Generalized Additive Models:

Generalized Additive Models are models that allow for non- linearity among quantitative and

qualitative variables. Specifically, Generalized Additive Models are composed of," non- linear functions of each of the [predictor] variables" (James, et al., 2021). A step above a simple generalized linear model, Generalized Additive Models display a response variable against its predictors in a way that allows one to gain a better idea of the non- linearity within the model's relationships.

The additive nature of Generalized Additive Models is implemented through the function of each explanatory variable being added together to form a relatively smoothed out non- linear line. This allows for a more comprehensible perspective of the feature's influences on the response variable. In the case of airline passenger satisfaction, the only continuous variables in the dataset are: Age, Arrival Delays, Departure Delays, and Flight Distance. Using these features as predictor variables against our response variable (satisfaction) will now provide insight into how these specific variables influence passenger satisfaction.

Using a Generalized Additive Model for our analysis provides a few advantages for viewing the relationship in our data set. One benefit is the efficiency we are afforded when we implement the model, instead of being forced to resort to manually trying out different transformations on each predictor variable to make room for non- linearity (James, et al., 2021). Another advantage is, due to the model being additive, we can now see what effect each individual predictor variable (age, arrival delays in minutes, departure delays in minutes, and the distance flown by the passenger) will have on satisfaction. With that said, these benefits heavily influenced our choice to include the model in our analyses.

To implement our Generalized Additive Models, the package, "mgcv", from our R library is imported. For reproducibility, a seed of 100 is set for each of our Generalized Additive Model functions. From the "mgcv" package in R, the "gam" function is utilized on our training subset. Where the satisfaction column is our response variable and the columns for age, flight distance traveled, departure delay in minutes, and arrival delay in minutes are the predictor variables. Each predictor variable is affected by R's smoothing function, which causes each function to be affected by smoothing splines to give a more summarized view of each observed relationship.

```
Family: binomial
Link function: logit

Formula:
satisfaction ~ s(Age) + s(Flight.Distance) + s(Arrival.Delay.in.Minutes)
+
    s(Departure.Delay.in.Minutes)
```

Initial Generalized Additive Model.

From the "mgcv" package in R, the "predict.gam" function is utilized to predict the choice of satisfaction for our testing subset using one of our previous Generalized Additive Models on our training data subset. This includes inputting the chosen previous model, our testing subset, and our preference to have the predictions be returned as probabilities. This creates a new column in our testing subset, called, "prob," which will house the prediction probabilities.

```{r}
subTest$prob <- predict.gam(arrivalExperience, newdata = subTest, type =
"response")

summary(subTest$prob)
```

Prediction function.

Results + Discussion

Support Vector Machines:

For the Support Vector Machines Model we first implemented Support Vector Machine (SVM) classifiers with various kernel functions (linear, polynomial, radial basis function (RBF), and sigmoid) to perform a classification task.

The first step involved importing necessary modules from scikit-learn (sklearn). This includes SVC for implementing the SVM classifier, classification_report and confusion_matrix for evaluating model performance, make_pipeline for constructing a pipeline, and StandardScaler for preprocessing the data.

After importing the required modules, we constructed a pipeline for each SVM classifier. Within each pipeline, we standardized the features using StandardScaler() to ensure all features have the same scale, which can improve the performance of the SVM algorithm. The SVM classifiers are configured with different kernel functions (kernel="linear", kernel="poly", kernel="rbf", and kernel="sigmoid") to explore different decision boundaries.

Next, we fit each SVM model to the training data (x_train, y_train). Once the models were trained, we made predictions on the test data (x_test) using the predict() method. Subsequently, we computed the confusion matrix and classification report to evaluate the performance of each SVM model.

The confusion matrix provides a tabular representation of the true positive, true negative, false positive, and false negative predictions made by the classifier. The classification report

summarizes various performance metrics such as precision, recall, and F1-score for each class, as well as overall accuracy, macro average, and weighted average metrics.

Finally, we visualized the confusion matrix using ConfusionMatrixDisplay from scikit-learn and matplotlib.pyplot. This graphical representation aids in interpreting the performance of each classifier and identifying any misclassifications.

Linear Kernel:

	precision	recall	f1-score	support
neutral or dissatisfied	0.868901	0.911437	0.889661	2247.000000
satisfied	0.878138	0.822719	0.849526	1743.000000
accuracy	0.872682	0.872682	0.872682	0.872682
macro avg	0.873520	0.867078	0.869594	3990.000000
weighted avg	0.872936	0.872682	0.872128	3990.000000
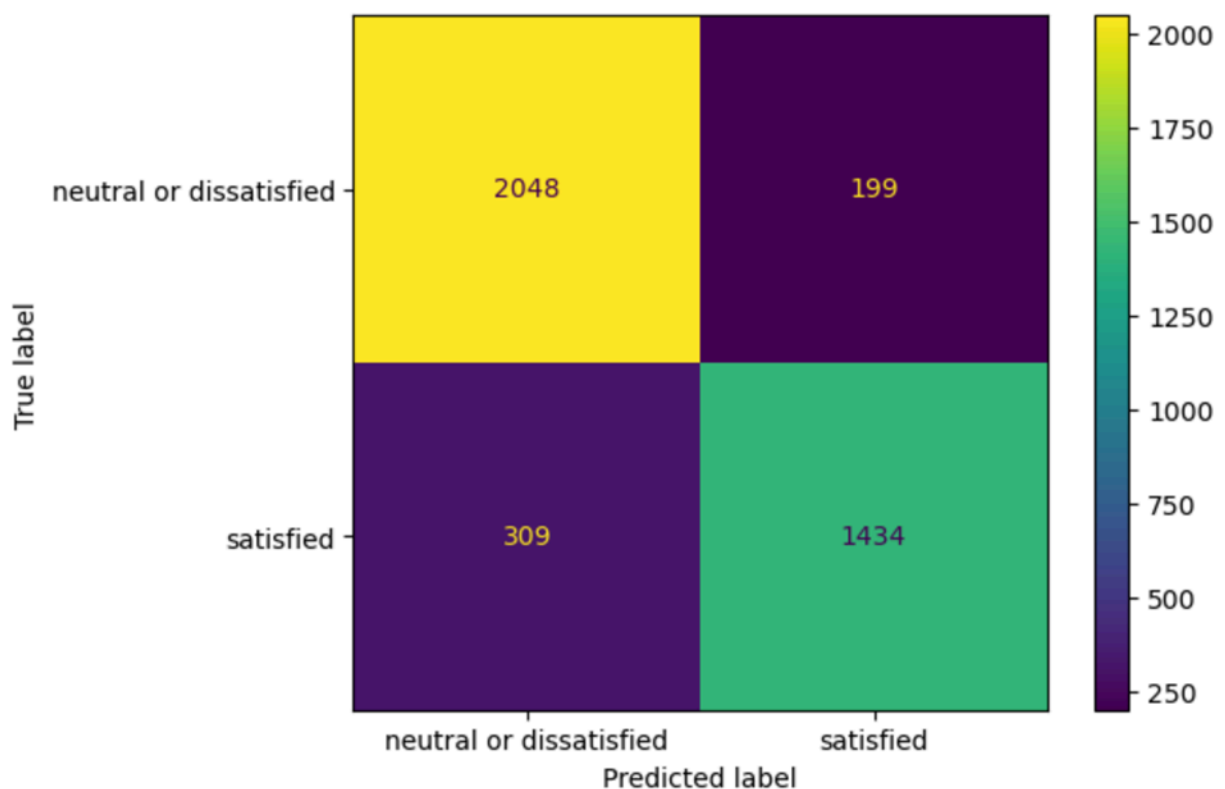
The Linear Kernel with an accuracy of 87% was by far the best performing kernel out of the 4.
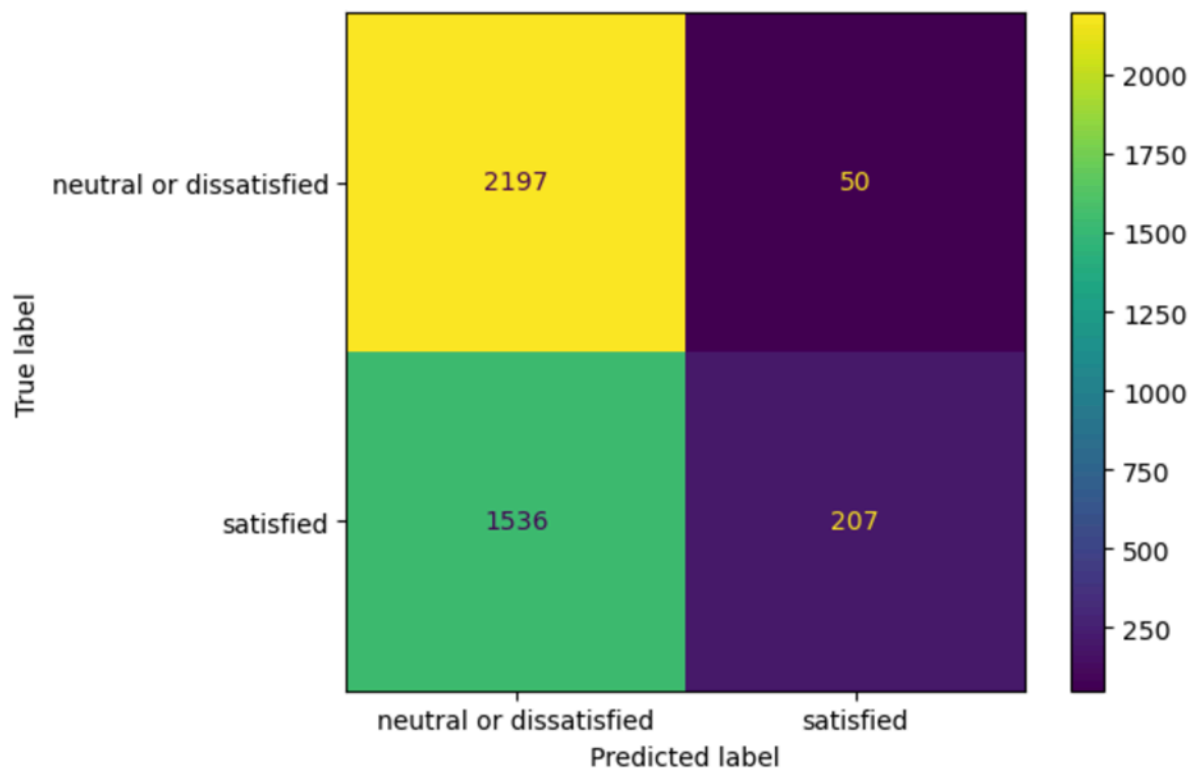
Polynomial Kernel:

	precision	recall	f1-score	support
neutral or dissatisfied	0.595633	0.959057	0.734868	2247.000000
satisfied	0.752688	0.160643	0.264775	1743.000000
accuracy	0.610276	0.610276	0.610276	0.610276
macro avg	0.674161	0.559850	0.499822	3990.000000
weighted avg	0.664241	0.610276	0.529512	3990.000000
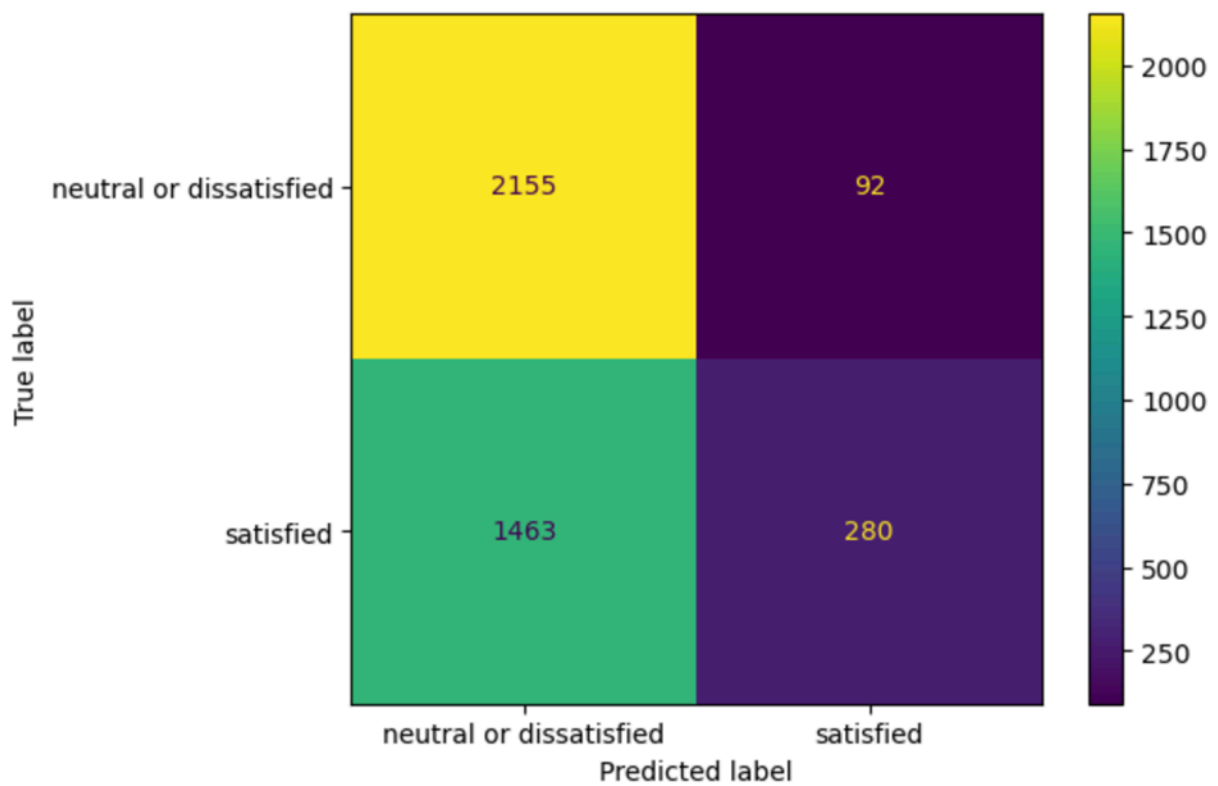
The polynomial kernel was second best with 61% accuracy, but suffered from mislabelling most observations as unsatisfied.

RBF Kernel:

	precision	recall	f1-score	support
neutral or dissatisfied	0.588535	0.977748	0.734783	2247.000000
satisfied	0.805447	0.118761	0.207000	1743.000000
accuracy	0.602506	0.602506	0.602506	0.602506
macro avg	0.696991	0.548254	0.470891	3990.000000
weighted avg	0.683291	0.602506	0.504225	3990.000000

The RBF kernel had the same issue as the polynomial kernel where most observations were labeled as unsatisfied regardless of their true label.

Sigmoid Kernel:

	precision	recall	f1-score	support
neutral or dissatisfied	0.562218	0.554962	0.558567	2247.000000
satisfied	0.435666	0.442915	0.439260	1743.000000
accuracy	0.506015	0.506015	0.506015	0.506015
macro avg	0.498942	0.498938	0.498913	3990.000000
weighted avg	0.506935	0.506015	0.506449	3990.000000
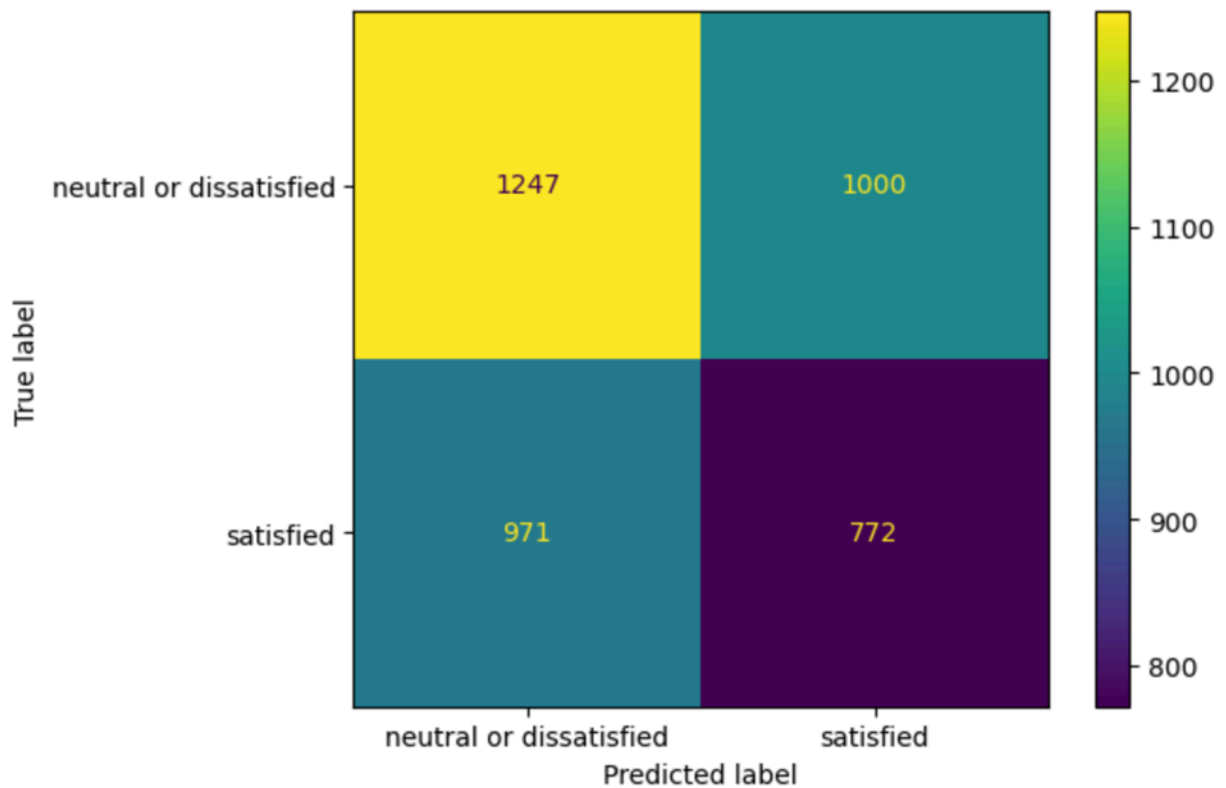
The sigmoid kernel had the worst accuracy at 50%.

Logistic Regression:

For the Logistic Regression model we used all 22 features in the data set to make predictions on whether the customer would be satisfied or dissatisfied. We split the data into a training and test set with an 80-20 split.

To initialize and train the logistic regression model, we used SkLearn's linear_model package. The initial model we created was a model with SkLearn's default hyperparameters (L2 regularization and lbfgs solver). After training the model on the training data set, we predicted labels for the test data and constructed a confusion matrix for the predictions.
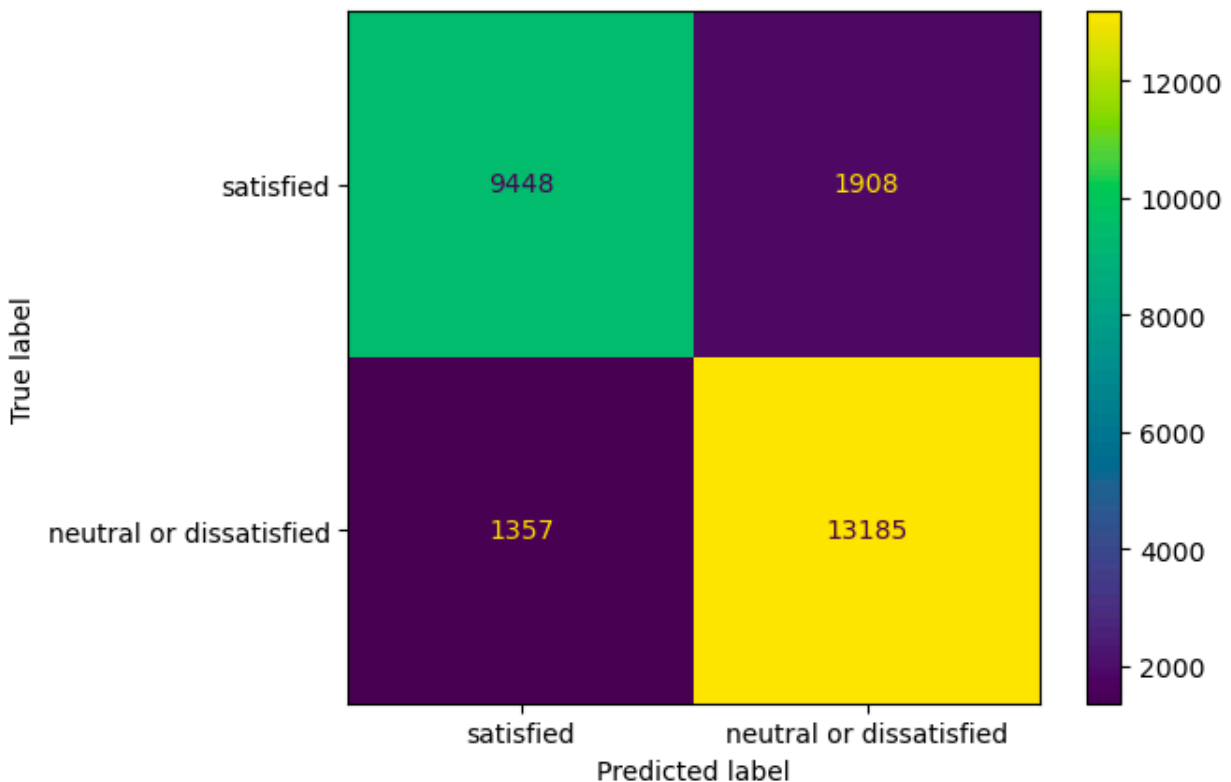

As seen above, the true labels match the predicted labels in most cases. The model is more prone to assigning the "neutral or dissatisfied" label than the "satisfied" label with a recall score for "neutral or dissatisfied" of 0.91 while the recall score for "satisfied" is 0.83. The model predicted the proportion of satisfied customers to be 41.7% while the true proportion of satisfied customers in the data set is 43.4%. This model achieved an accuracy score of 87.39%.

We took the absolute value of the feature coefficients for each of the 22 features to determine the relative importance of each feature in predicting the label. For the initial model, we found that the most important features were type of travel, online boarding satisfaction, and customer type. The least important features were flight distance, gender of passenger, and satisfaction with food and drink service.

**Grid Search for Best Parameters:**

After creating an initial model with the default hyperparameters we attempted to create a better model by tuning the hyperparameters to find the ideal hyperparameters to use during the training process. To accomplish this we did a grid search of every combination of several of the hyperparameters to test which combination obtains the best result. The grid space we searched over included the solver ('lbfgs', 'liblinear', 'newton-cholesky', 'sag', and 'newton-cg'), penalty (None, 'l1', and 'l2') and C (100, 10, 1, 0.1, and 0.01). The grid search tests each combination and returns the parameters that have resulted in the best accuracy score as well as the accuracy score that is achieved. The best parameters for this data set were sag solver, l2 penalty and a C value of 0.01 which achieved an accuracy of 87.45%.

Given the best hyperparameters from the grid search, we then trained a model with these hyperparameters. We used the same training and test split from the above model and trained the Logistic Regression model on the training data. After training the model, we generated label predictions for all of the test data and constructed the following confusion matrix.
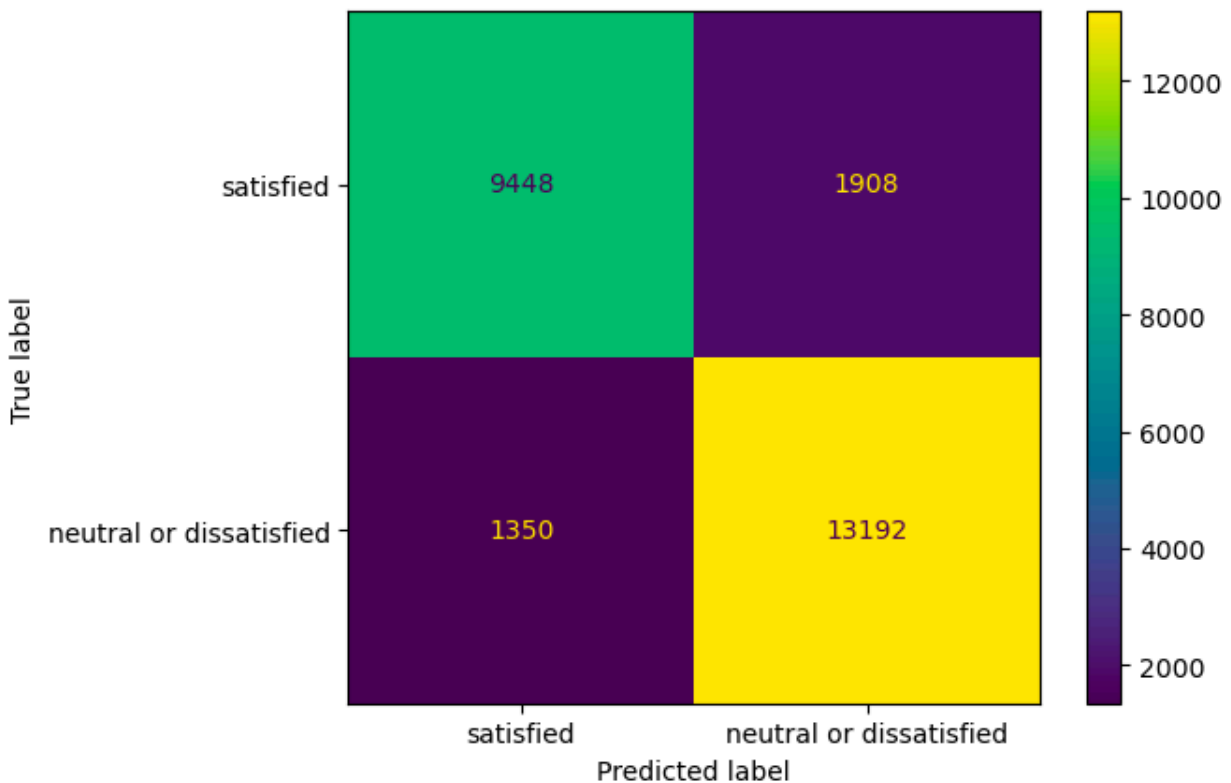
This tuned model performed very similarly to the model with default hyperparameters with the majority of predicted labels being correct. Again the model is more prone to labeling passengers as neutral or dissatisfied (Recall=0.91) rather than satisfied (Recall=0.83). This model achieved an accuracy score of 87.42%.

When looking at the absolute values of coefficients in the model, we see similar importance rankings for the factors as we did in the initial model.

Importance of Each Factor in Predicting Passenger Satisfaction
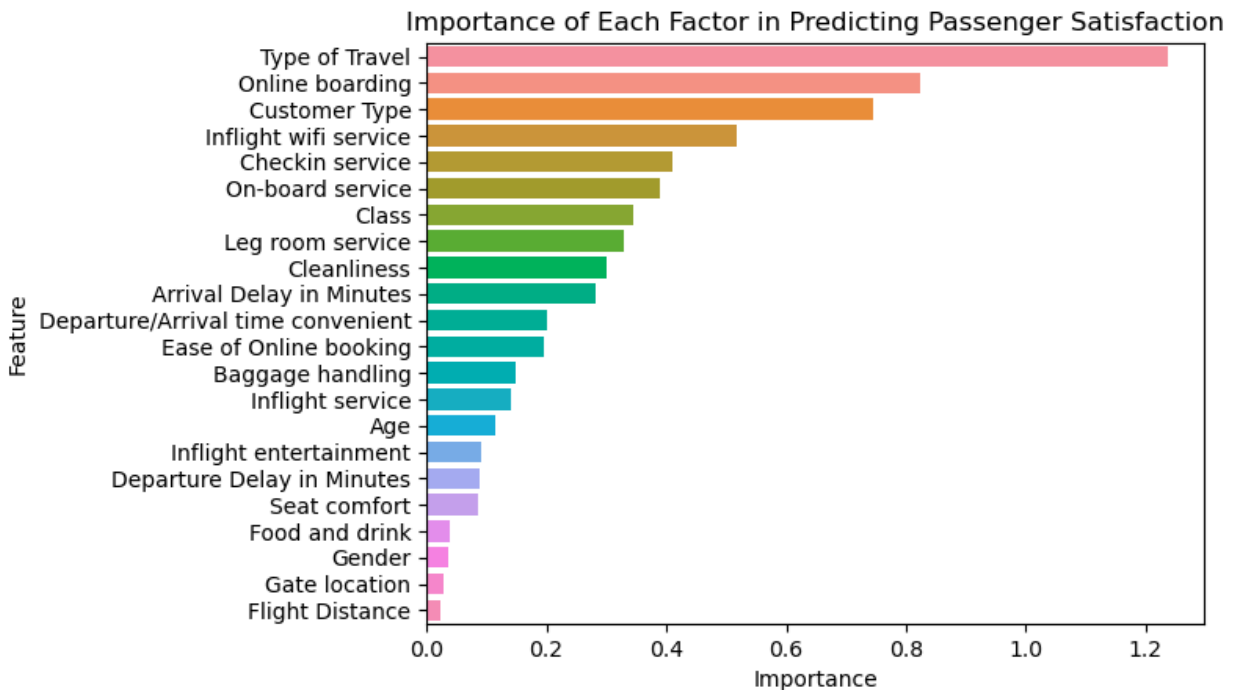
The most important factors for predicting passenger satisfaction were type of travel, online boarding satisfaction, and customer type. The least important factors for predicting passenger satisfaction were flight distance, gate location, and passenger gender.
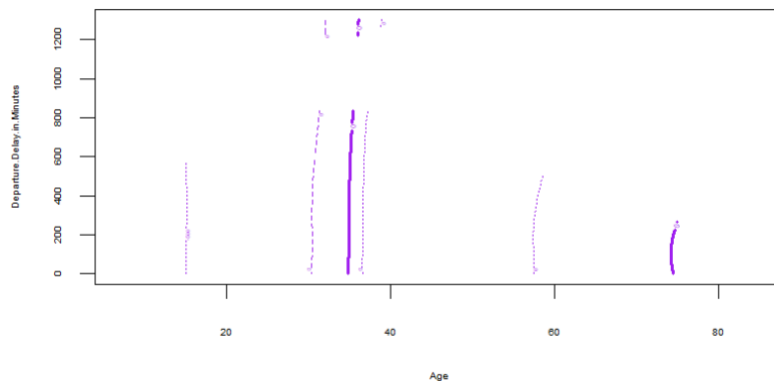
Generalized Additive Models:

For the Generalized Additive Model, we conducted three different analyses using Generalized Additive Models. Each model has all continuous variables (the age of each reporting passenger, the distance each reporting passenger traveled, the arrival delay in minutes that each reporting passenger experienced, and the departure delays in minutes that each reporting passenger experienced) included as predictor variables against the response variable, satisfaction. Due to the departure delay in minutes not being as significant, we decided to only include that as part of the interaction term with every other predictor variable considered to see how the departure delay influences the observed relationship of the other variables.

Our first model includes the interaction between age and departure delays. Each variable is affected by a smoothing variable and is considered against satisfaction. Age yielded a p- value of less than 2e-16, flight distance yielded a p- value of less than 2e-16, arrival delay in minutes yielded a p-value of less than 2e-16, and our interaction term between age and departure delay yielded a p-value of 3.95e-06. Interestingly enough, each predictor variable considered in the model happened to significantly influence the chosen satisfaction level of reporting passengers.
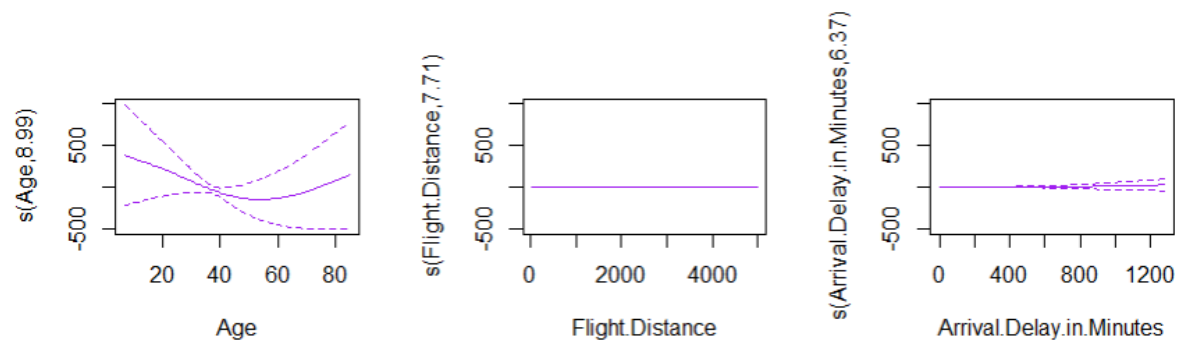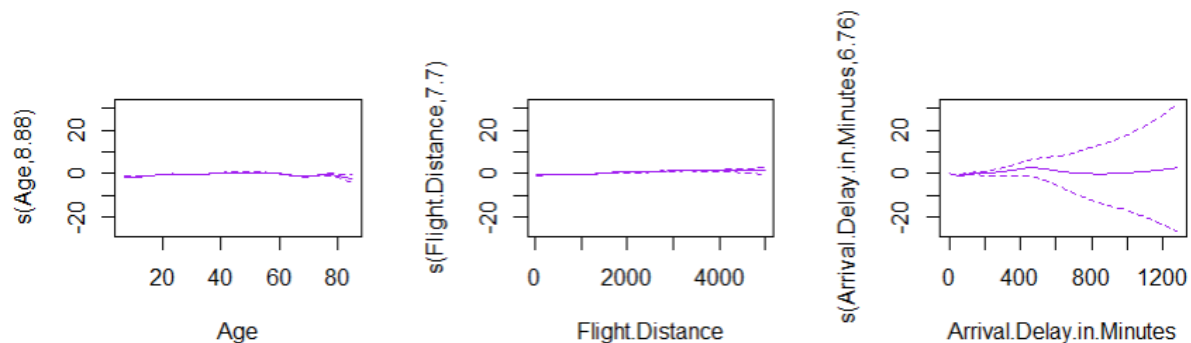

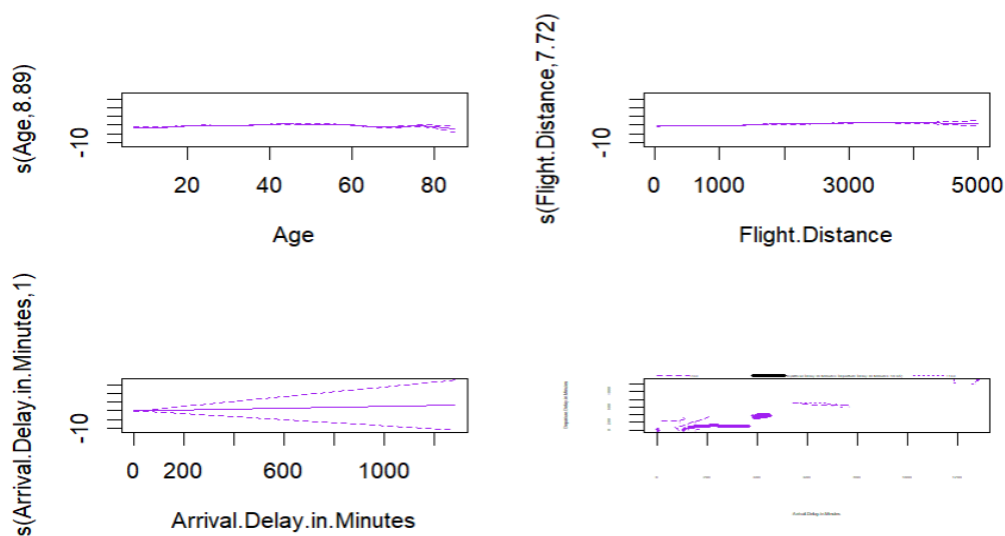*First Generalized Additive Model plot.*

Our second model includes the interaction between flight distance and departure delays. Each variable is affected by a smoothing variable and is considered against satisfaction. Age yielded a p- value of less than 2e-16, flight distance yielded a p- value of less than 2e-16, arrival delay in minutes yielded a p-value of less than 2e-16, and our interaction term between flight distance and

departure delay yielded a p-value of 0.404. Evidently, while most predictor variables considered in the model happened to significantly influence the chosen satisfaction level of reporting passengers, the interaction term does not significantly affect satisfaction.
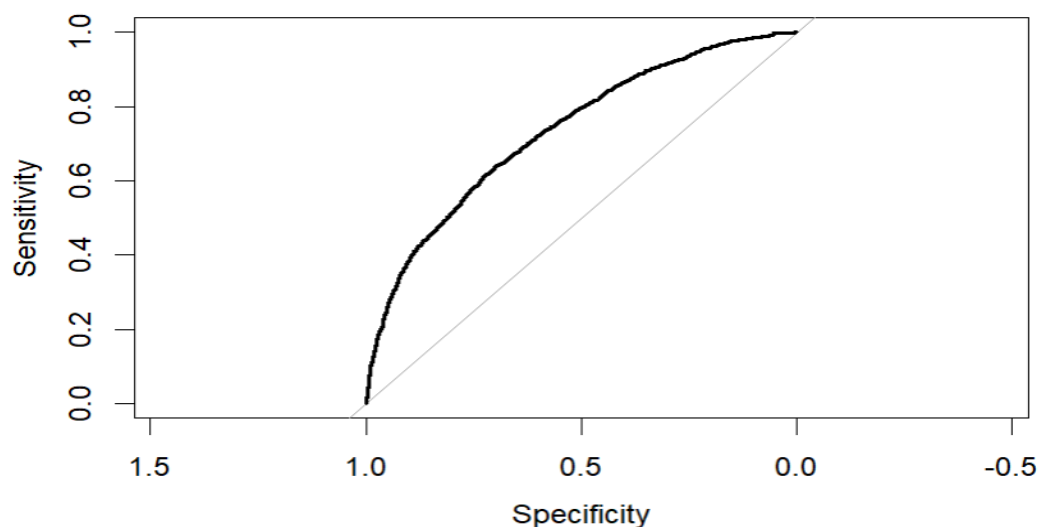

Second Generalized Additive Model plot.

Our third model includes the interaction between arrival and departure delays. Each variable is affected by a smoothing variable and is considered against satisfaction. Age yielded a p- value of less than 2e-16, flight distance yielded a p- value of less than 2e-16, arrival delay in minutes yielded a p-value of 0.651, and our interaction term between arrival and departure delays yielded a p-value of less than 2e-16. Unlike the previous models, this model shows two individual predictor variables and the interaction term are significantly affecting satisfaction choice. This change in significance for arrival delay suggests that its influence on the choice in satisfaction significantly depends on the presence and length of the departure delay.

Interaction: Arrival and Departure Delay
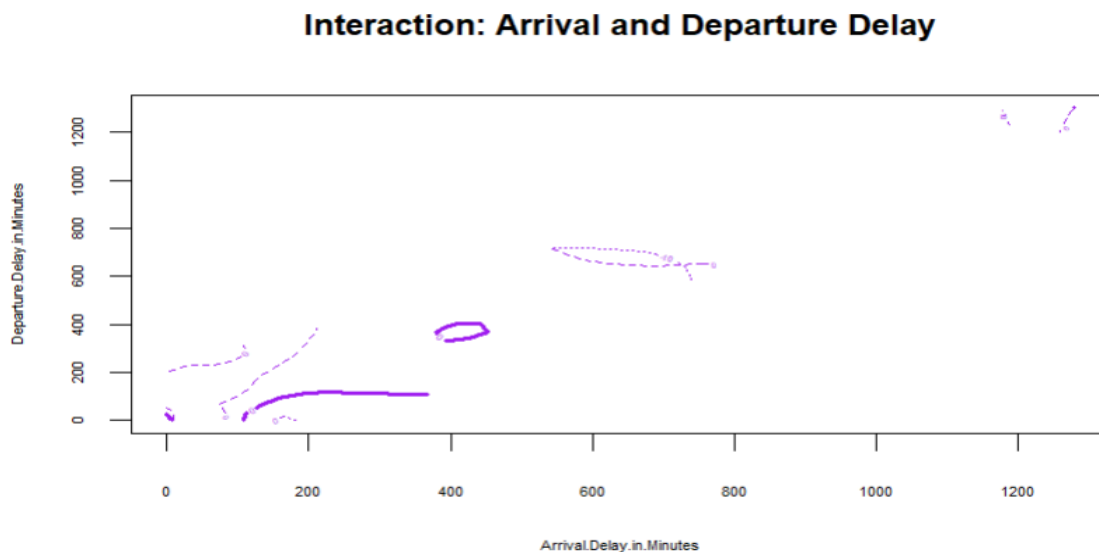
Third Generalized Additive Model plot.

Out of curiosity, we decided to look at the predictions of the reporting passengers in our test subset being satisfied. After running "predict.gam", we compared the actual reports of satisfaction to our testing subset- based predictions. Our prediction produced a Receiver Operating Characteristic (ROC) curve with an associated Area Under the Curve (AUC) value of 0.7328. Also, we looked deeper into the relationship between the arrival and departure times and plotted the probability of a reporting passenger being satisfied below.
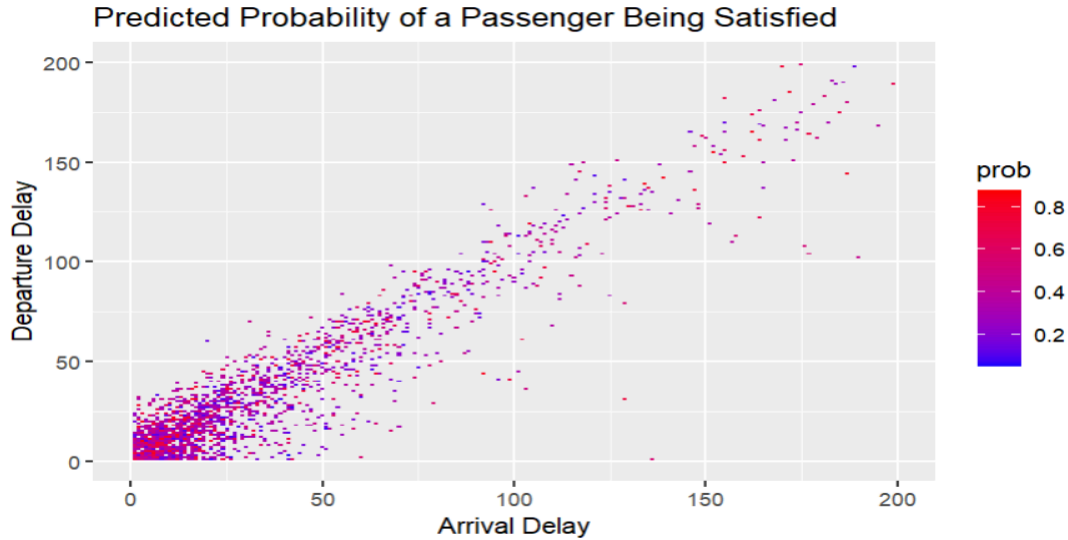
Predicted Generalized Additive Model plot.

Conclusion

As seen in the results above, there are several features of a flight that can predict whether or not a passenger will report satisfaction after their flight experience. Enhancing customer satisfaction should be a priority for airlines so that they can have customer retention as well as gain new customers. Looking at what determines the satisfaction levels of customers can be critical information for airlines as they look to improve their businesses. As reported above, the most important factors for customer satisfaction are type of travel, online boarding satisfaction, and customer type. There seems to be a bit of a trend in the type of factors that are important to passengers and their satisfaction. A lot of the lower ranked factors have to do with the luxuriousness of the experience. This category includes things like food and drink service, in-flight entertainment, and comfortability of the seat. On the other hand, factors having to do with convenience and efficiency rank higher in the importance for customer satisfaction. These include factors such as check-in service, boarding service, inflight internet, and convenience of the departure and arrival time. This pattern indicates that there has been a possible shift from the majority of passengers looking for a luxury flight experience to the majority of passengers looking for ease and convenience in their travels.

Type of travel is a factor that refers to whether the passenger is traveling for business or for leisure. It is interesting to note that the most important factor for predicting satisfaction is one

that is not in the airline's control– people traveling for business will travel for business and people traveling for leisure will travel for leisure. While this alone does not give an airline a lot to go off of for improving their customer satisfaction, it may be interesting to use this information to segment the customer base into different groups. Repeating this analysis with the data separated into people traveling for business and people traveling for leisure may show which segment of the consumer base is generally more satisfied. This breakdown can also show how the importance of factors varies amongst these two different groups. Some factors may be much more important for satisfaction for those traveling for leisure than they are for those traveling for business. If this is the case, then it would be the most impactful for the airlines to target the factors that are most important to the segment of the consumer base that tends to be less satisfied with their flight experiences.

Another important factor of note is the type of customer that is traveling. The two types of customer that this analysis examines are loyal customers and non-loyal customers. Looking at which of the two groups of customers is generally more satisfied with their experience could provide valuable insight to airlines. If loyal customers are the less satisfied group, then the airline should look into improving their frequent flier loyalty programs. If non loyal customers are the less satisfied group, then the airline could look into expanding their frequent flier loyalty programs to include a larger base that will be more likely to be satisfied with their experience.

Delays (both in arrival time and departure time) also have a pretty great impact on a passenger's satisfaction as we can see from the Generalized Additive Models. This is a factor that could be in an airline's control depending on the context but oftentimes is also due to external factors such as weather and air traffic. Despite not necessarily being in the control of the airline, the probability of dissatisfaction among passengers who experienced delays means that airlines should probably look further into what possible options there are for improving the experience of customers during some of the inevitable delays.

References

Klein, T. (2020, February 20). Airline passenger satisfaction. Kaggle.

https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction/data?select

=train.csv

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning:

With applications in R. Springer.

Jiang, X., Zhang, Y., Li, Y., & Zhang, B. (2024, March 5). *Forecast and analysis of aircraft*

passenger satisfaction based on RF-RFE-LR model. Scientific Reports. Retrieved April

30, 2024, from https://www.nature.com/articles/s41598-022-14566-3

Jurafsky, D., & Martin, J. (2008). Logistic Regression. In *Speech and Language Processing, 2nd*

Edition (pp. 1-8). Pearson Prentice Hall. https://web.stanford.edu/~jurafsky/slp3/5.pdf

Pereira, F., Costa, J. M., Ramos, R., & Raimundo, A. (2023). The impact of the COVID-19

pandemic on airlines' passenger satisfaction. Journal of air transport management, 112,

102441. https://doi.org/10.1016/j.jairtraman.2023.102441

Troy, M., (2023, May 10). *2023 North America Airline Satisfaction Study*. J.D. Power. Retrieved

April 30, 2024, from

https://www.jdpower.com/business/press-releases/2023-north-america-airline-satisfaction

-study