Report

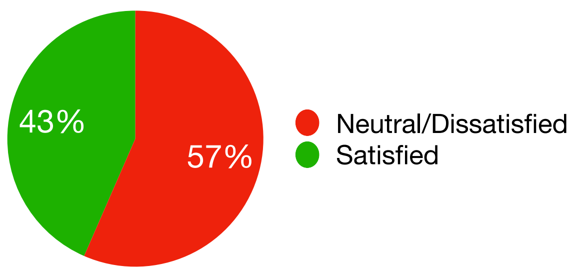
INITIAL DATA EXPLORATION:

My initial data exploration revealed that the testing data contains 20% of the total records in the dataset. This is a good split for the training and testing data.

Next, I noticed there were 393 missing values in the "Arrival Delay in Minutes" feature. After evaluating the satisfaction of passengers with these missing values, I found that about 43.3% of those passengers were satisfied. The meaning of the missing values were hard to interpret due to the various possibilities and the inconclusive statistic.

STATISTICAL ANALYSIS:

The following analyses were done after merging the test and training dataset into one dataframe. An initial analysis on the number of satisfied passengers revealed that only about 43.446% of passengers were satisfied, while 56.554% were neutral or dissatisfied.



This simple statistic suggests that the majority of passengers are not satisfied.

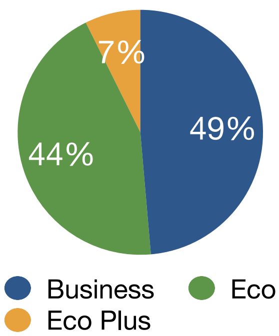
Next, the features rated on a scale of 1 to 5 were evaluated. First, I removed every row from the dataset whose features had a value of 0 since they signified missing responses. Not removing or imputing these 0 values could skew the mean of the features. After calculating the mean of the values in every such feature, I noticed that none of these features had a mean rating greater than 4. Inflight wifi service had the lowest mean of 2.82, with Ease of Online booking (mean: 2.88) and Gate location (mean: 2.99) following closely behind. This data is presented in the table to the right. Improving these services, especially ones with a lower mean score, could result in higher overall customer satisfaction.

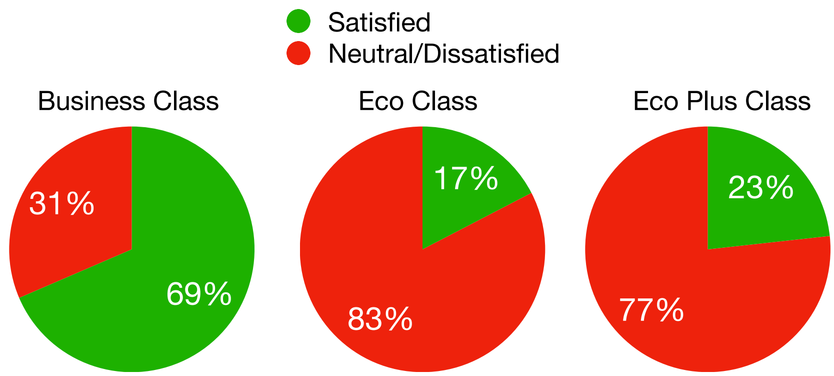
		Mean	Std. Dev.
0	Inflight wifi service	2.818924	1.253488
1	Departure/Arrival time convenient	3.206529	1.386967
2	Ease of Online booking	2.879038	1.298530
3	Gate location	2.987187	1.282230
4	Food and drink	3.212538	1.324593
5	Online boarding	3.329783	1.263577
6	Seat comfort	3.456304	1.312268
7	Inflight entertainment	3.379790	1.327307
8	On-board service	3.386559	1.285585
9	Leg room service	3.381562	1.295240
10	Baggage handling	3.638669	1.167754
11	Checkin service	3.293592	1.265982
12	Inflight service	3.647271	1.165077
13	Cleanliness	3.293557	1.310130

After this, I calculated the satisfaction rate of two different features. The first is the age of passengers. The average age of a passenger who participated in the survey is about 39. Additionally, the average age of a satisfied passenger is around 42, while the average age of a dissatisfied passenger is around 37 to 38. Catering services and promotions towards the younger people could help improve the satisfaction amongst these age groups. Improving upon poorly rated services such as inflight wifi service and ease of online booking, as discussed previously, could be effective approaches to raising the satisfaction levels.

Finally, I calculated the satisfaction levels of passengers in the three classes — business class, eco class, and eco plus class. These are my findings:

Passenger distribution:





As evident from the charts above, business class has the highest satisfaction rate (69%), while Eco and Eco Plus have a very low satisfaction rate (below 25%). Hence, improving the passenger experience and services in the Eco class, which has a significant passenger distribution (44%), can yield higher satisfaction rates. Although the Eco Plus class also requires improvement, it only has a passenger distribution of 7% — much lower than the other classes. Thus, focusing more on the other classes would be more beneficial.

DATA PREPROCESSING FOR MODEL BUILDING:

After creating the test and train dataframes, I set the first row as the column headers. Next, I removed the first two columns (unnamed column and id column), which seems insignificant for determining the target attribute.

I then created the x_train, y_train, x_test, and y_test dataframes, to be used for the model building process. The x dataframes for both train and test contain all the features except for the target feature (satisfaction), while the y (target) dataframes contain only the target feature. After this, I encoded the categorical features using an ordinal encoder.

Lastly, I imputed the missing values in the "Arrival Delay in Minutes" feature in both the train and test dataframes with the mean for its column.

CLASSIFICATION MODELS:

I chose to apply three different classification models to this dataset — Decision Tree, Random Forest, and Multilayer Perceptron classifiers.

First, I tested these classifiers without tuning any parameters. The Decision Tree classifier yielded an accuracy score of 0.9461, Random Forest classifier yielded an accuracy score of 0.9565, and Multilayer Perceptron classifier yielded an accuracy score of 0.9279. Thus, the Random Forest classifier produced the best model, and Multilayer Perceptron the worst.

After the untuned tests, I ran grid search on a variety of parameters to determine the best parameters for each classifier. The parameters for the Decision Tree grid search built 1280 models, the Random Forest grid search built 30 models, and Multilayer Perceptron built 216 models.

A model was built for each of the classifiers again with tuned parameters. The tuned Decision Tree classifier yielded an accuracy score of 0.9539, the tuned Random Forest classifier yielded an accuracy score of 0.9585, and the tuned Multilayer Perceptron classifier yielded an accuracy score of 0.9306. Therefore, even with tuned parameters, the Random Forest classifier prevails with the most effective model. These scores are presented in the table below:

	Untuned Parameters accuracy score	Tuned Parameters accuracy score	Difference (Tuned - Untuned)
Decision Tree Classifier	0.9461	0.9539	0.0078
Random Forest Classifier	0.9565	0.9585	0.0020
Multilayer Perceptron Classifier	0.9279	0.9306	0.27

The tuned parameters resulted in a better model for each classifier. Decision Tree had the highest accuracy gain (0.7738%).

Finally, I checked the tuned models for over-fitting and under-fitting by evaluating the difference between the accuracy scores for their training and test datasets. These are the results:

	Training-set accuracy score	Testing-set accuracy score	Difference
Decision Tree Classifier	0.9738	0.9539	0.0199
Random Forest Classifier	0.9973	0.9577	0.0396
Multilayer Perceptron Classifier	0.9323	0.9306	0.0017

Since the testing and training set accuracy scores do not present a significant difference, there is no over-fitting or under-fitting present in any of the tuned models.