Airline Passenger Satisfaction

A statistical exploration on how airline passenger satisfaction is achieved.

✈

Saim Ali
Robert Duque
Muhaimin Badar

05/04/2023

Introduction

The Airline Passenger Satisfaction dataset, available on Kaggle, encompasses comprehensive information on airline passenger satisfaction. This information is based on an airline satisfaction survey and comprises 24 attributes, which include demographic information about the passengers such as age, gender, and type of travel. Additionally, it includes various attributes associated with airline service, including inflight entertainment, seat comfort, and onboard service. The "satisfaction" attribute is the target variable, which indicates whether the passenger was satisfied or not satisfied with their overall experience. The dataset consists of approximately 129,880 rows, with each row containing 24 attributes.

We believe an in-depth analysis of the airline passenger satisfaction data can yield valuable insights that airlines can leverage to enhance their customer ecosystems to drive growth and satisfaction.

What is Airline Passenger Satisfaction?

From the time a customer enters an airport departure terminal to the time they safely arrive at your location, there are many steps of the way that determine if that customer is satisfied or not. Why care about customer satisfaction? Customers are the heart of any airline and by exceeding their satisfaction we are able to gain both a higher reputation to attract new customers and have a higher chance that that satisfied customer will return for later flights.This means that we must measure all attributes that affect this satisfaction of a given customer that include: food and drinks, leg room, seat comfort, and check-in service just to name a few. Nonetheless, if a business is to succeed to its fullest, conducting a study that helps maximize areas that we are lacking can prove to be rewarding both financially and helps boost our customer loyalty.
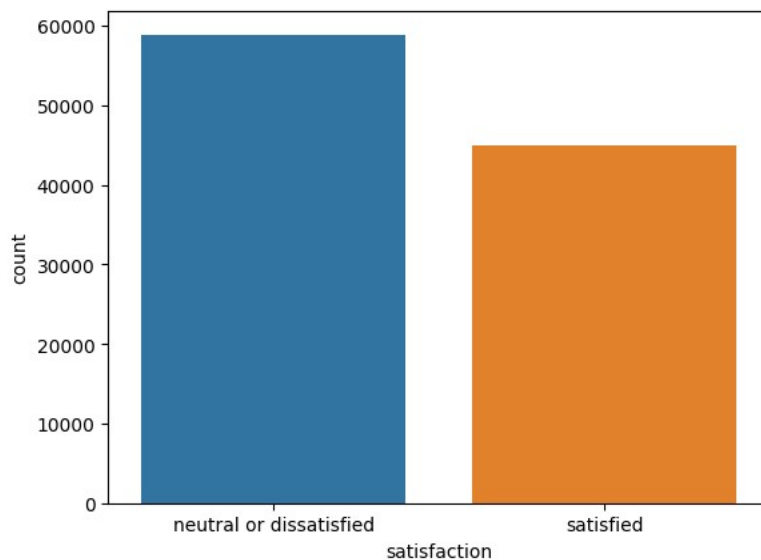
Figure 1: Satisfaction distribution

This bar plot shows the distribution of neutral or dissatisfied survey participants vs satisfied participants out of the 129,880 total participants. As seen from the neutral or dissatisfied bar, there are approximately 58,000 customers from our total participants who were not satisfied with the quality of service they received at the airline. While only about 30% of survey participants concluded that they were satisfied, this shows there is a large problem with the airline which is leaving more customers neutral or dissatisfied rather than satisfied. To find the root of this issue we must look deeper into the survey than just the concluding results to see what the customers of the airline are unhappy with in the process of their flight from departure to arrival.

Predictive Modeling Techniques

In machine learning and data science, a "model" is a mathematical representation of a dataset that can be used to make predictions or identify patterns. The effectiveness of a model depends on the type of dataset it is being used on. For example, if a dataset has labeled data, a "supervised" model can be used. On the other hand, if the data is unlabeled, an "unsupervised" model can be used. For this project on Airline Passenger Satisfaction, we commenced the data science experiment by analyzing the collected data. We chose to use a total of four distinct models: K-nearest-neighbors, Logistic Regression, Gaussian Naive-Bayes, and Decision Trees. We made a deliberate choice to employ diverse and robust predictive modeling techniques in our experiment. This approach enabled us to generate coherent results and enhanced our confidence in the statistical analysis.
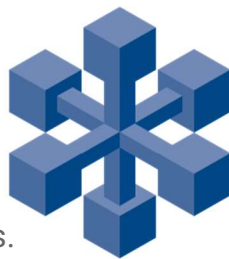
Data Analysis

The data/dataset itself derives from a survey conducted on customers who flew with this airline. To understand the study we conducted we must first explain the data we worked with to come to such results. Without a deep dive into the background analysis of this dataset our predictions for what is affecting customer satisfaction would be useless and or inaccurate. Some of the discoveries about the dataset we found relevant are listed below:

- There is almost an even number of males and females who took the survey
- The average customer age of participants is 39 years
- The flight distance varied greatly in this survey results with the average being 1189
- We also found that:
  - **Best features for prediction of satisfaction** - Online Boarding, Class, and Type of Travel
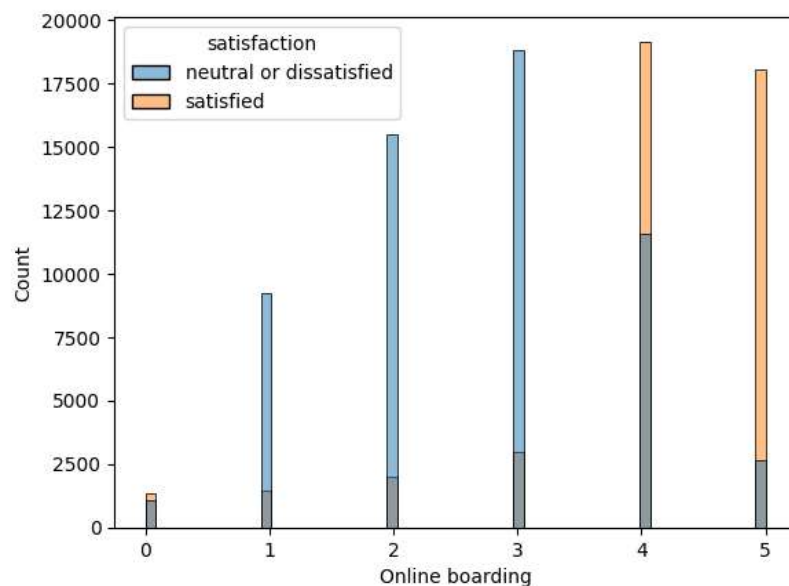

Figure 2: Online boarding

Online boarding, arguably our most important factor affecting passenger satisfaction. Notable to see how 'neutral or dissatisfied' is most noticeable in the 1-3 ratings, whereas 'satisfied' is most present in ratings of 4-5.
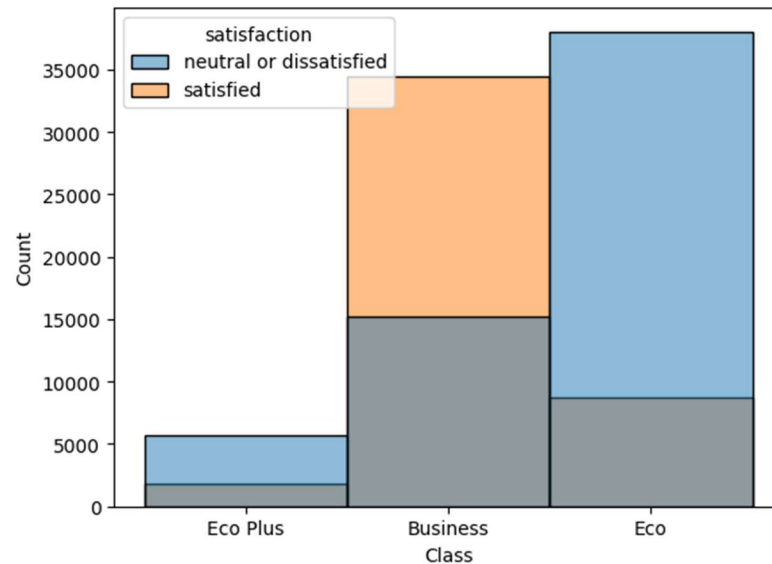
Figure 3: Class

The plot above shows that the significance in class level greatly affected the satisfaction of airline passengers. Customers who flew on lower-class flights such as Eco and Eco-plus were far more likely to be neutral or dissatisfied when compared to those who flew on Business class flights.
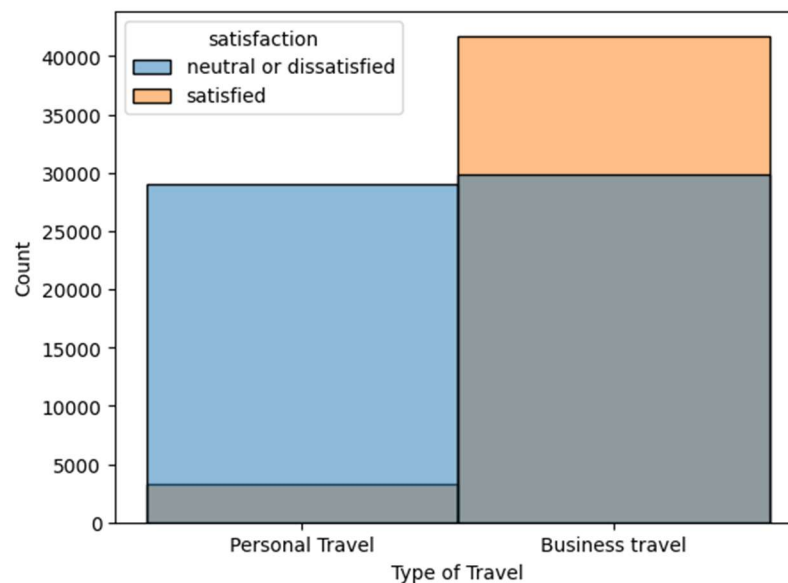

Figure 4: Type of Travel

Once again we see from this plot that our theory of Business class flyers being far more satisfied than those who flew for Personal Travel most of which were seated in the Eco or Eco-plus classes of the plane.

- ○ **Worst features for prediction of satisfaction** - Gate location, Gender, and Departure/Arrival Time Convenient
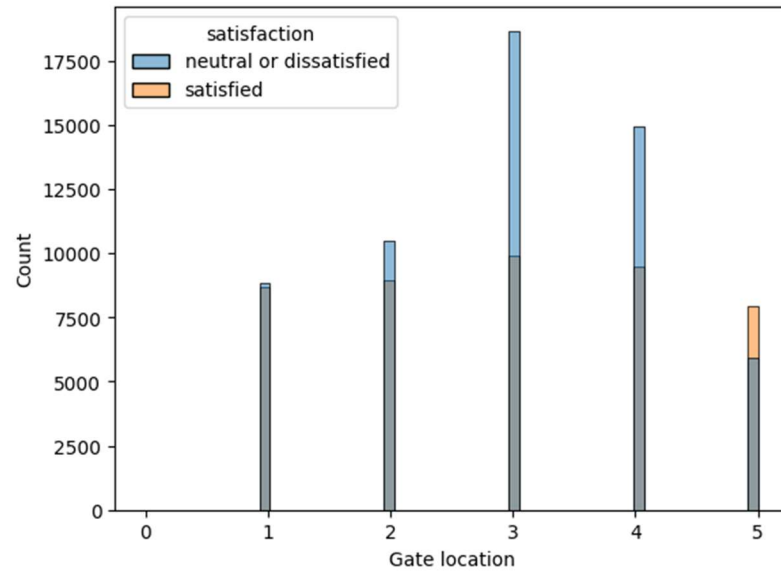
Figure 5: Gate location

From the above plot we can infer with confidence that the satisfaction level of Gate location did not play a significant enough role in the prediction of a satisfied vs neutral or dissatisfied customer for us to include it in the model for our study.
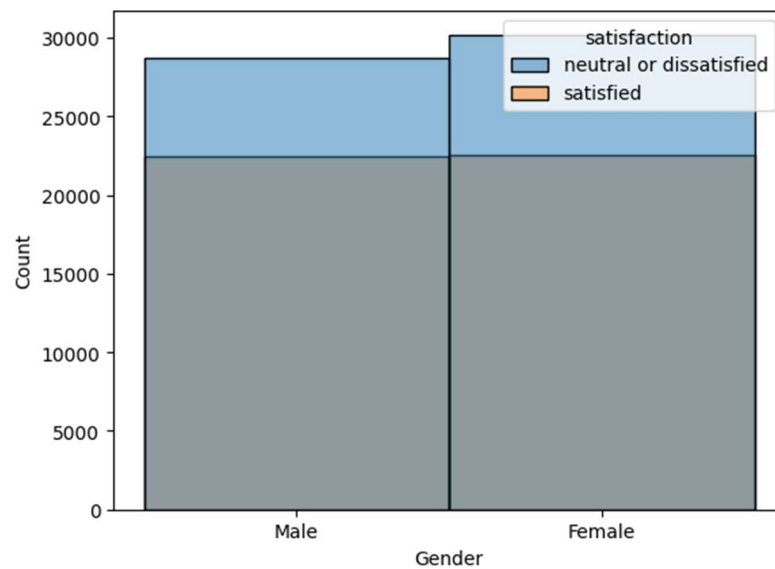
Figure 6: Gender

We previously noted that the Gender distribution of this dataset/survey is almost even between Male and Females. This is further supported by the plot above showing that there is almost no difference in the satisfaction

or dissatisfaction rates of the participants who took the survey depending on their Gender.
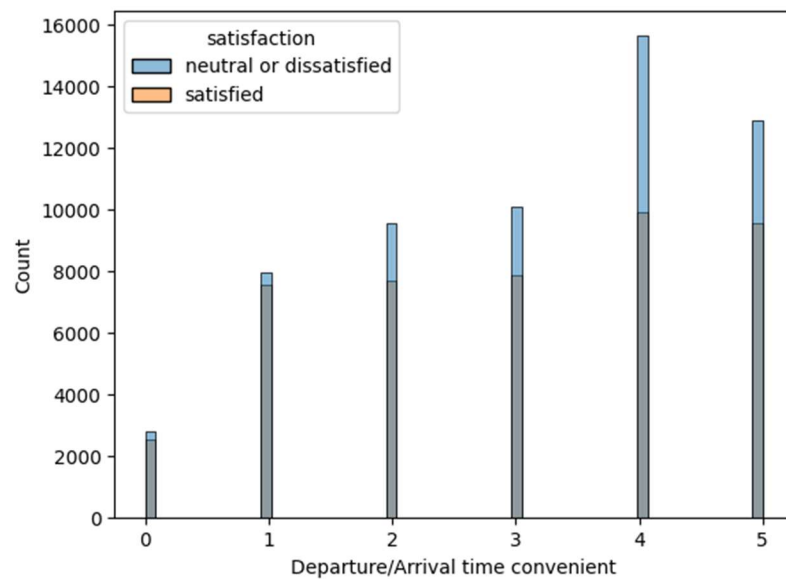

Figure 7: Departure/Arrival time convenient

Disregarding the delay on departure/arrival times of flights in the airline, when looking at the satisfaction level of the Departure/Arrival time convenient based on the survey taken, we can see that although there is some changes between the levels of satisfaction, there is not enough to determine this to be a factor for customer satisfaction in the airline.

As previously noted, we opted to utilize a range of modeling techniques that respond very differently to the bias and variance curve. In the realm of data science, this constitutes a key concept that facilitates optimal model fine-tuning to mitigate the impact of "noisy data." This "noisy data" in a real word scenario can cloud our judgment when we try to depict the reasons for negative customer satisfaction ratings. After removing those data points, we got to analyzing which aspects of the customer experience will help us actually predict whether a customer is satisfied or not. Not every attribute from our datasets customer feedback would be useful when we attempt to predict customer satisfaction. Leaving these attributes in there can also result in an output that is similar to the previously described "noisy data" which in turn, leads to inaccurate predictions on what is driving customers to answer our survey with a negative satisfaction rating.

Lastly, another problem with the data collected from the survey was that some of the answers provided from the survey participants were empty or null. This was due to two reasons: either the survey was filled out inadequately with missing information or
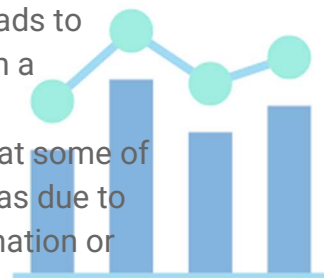
the attribute/question did not apply to the participant such as the "Delayed Flight in Minutes" question being asked to customers whose flight was not delayed at all.

Key Visualizations
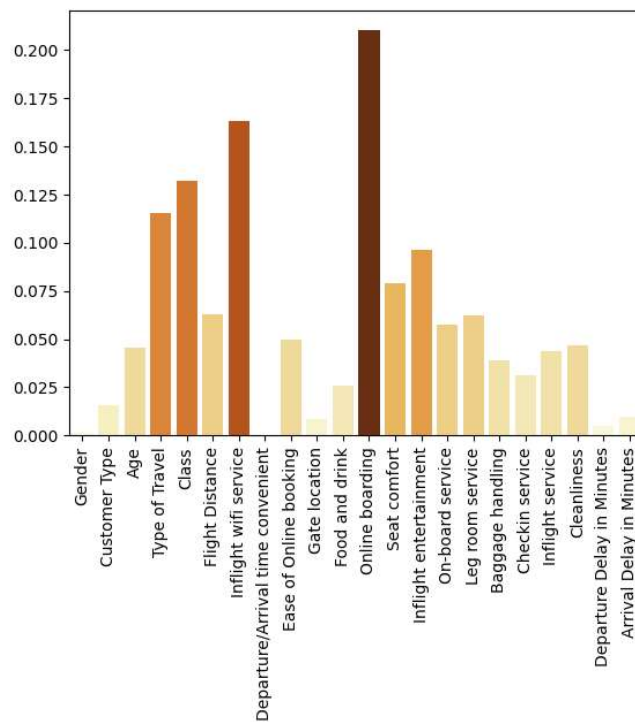
Figure 8: Attribute makeup

This is an excellent visualization of what factors affect customer satisfaction. Based on our mutual information regression model, the visualization above shows which features are relevant and/or influence the value of customer satisfaction. Using this model/visualization technique, a company can prioritize improving the features with a higher relevance and put less emphasis on features with a lower relevance. Here, we can easily obtain from the bar plot that online boarding is by far the most important factor that determines customer satisfaction along with inflight wifi service, class of airline ticket, type of travel and inflight entertainment.
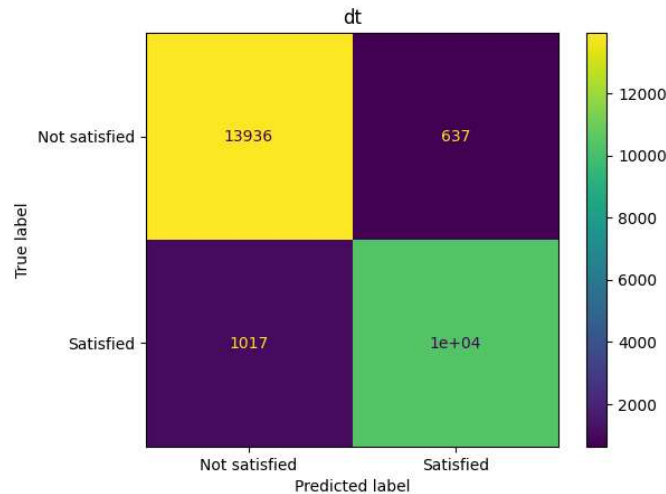
Figure 9: Accuracy visual

Given Decision Trees were the best performing model, we can explore the final results on the data. There were 1654 total instances where our model predicted a customer the opposite of their satisfaction compared to their actual recorded satisfaction. We can also see that our model is more conservative when giving out *true* satisfied labels which means we can be much more confident when a customer prediction is labeled as 'satisfied.'

Importance of Airline Passenger Satisfaction

For the airline to continue serving customers at a high standard we must understand how each step of a customer's experience is going. Through the use of a survey and the study of those recorded results we did just that, we created a mathematical model which can accurately predict relative customer satisfaction given the results of a survey. Using such a machine learning model can be beneficial in various different ways. One of the most important ways using such a model can be by allocating additional resources to parts of the customer experience that have a high significance on the outcome of satisfaction in the airline if that attribute is performing poorly as we can see through the survey. These models can give us results that we can rationalize into business models that drastically improve the airline experience, with a goal of improving satisfaction rate in mind we should aim for at least 60% of all customers to be satisfied. Without the use of Data Science for this scenario we would have had to guess what is causing customer dissatisfaction but through our study we can see exactly which factors we will need to look over to achieve the goals we set for this airline.

Business Strategy

There are a couple of ways airlines or aviation industry businesses can leverage our research on Airline Passenger Satisfaction:

- Identify areas where the business can improve satisfaction: Begin by analyzing factors that affect the customer's satisfaction ratings and actively work to improve in those areas. For example, if the airline determines that seat comfort plays a major role in the satisfaction ratings, they can proactively offer other seats based on availability or invest in a more comfortable experience overall.
- By understanding what factors are most important to customers, the airline company can tailor their marketing/promotions to highlight said features. For example, from the feature selection visualization (**Figure 8: Attribute makeup**), we can see that online boarding has the highest relevance with respect to customer satisfaction. The company should promote online boarding with special deals and/or promotions to customers who use online boarding.
- Work on improving customer retention: by improving customer satisfaction ratings, they are much more likely to improve their customer retention rates as well. Satisfied customers are highly likely to work with the business in the future and likely recommend it to their friends and family. This can drive profits and create a stronger image of the business' brand.
- Optimization of pricing model: By understanding what aspects a customer values most, businesses can adjust business capital and pricing model to reflect the value of the identified factors.
- Evaluate their performance: Customer satisfaction is an excellent way to determine how well the business is doing. Unhappy customers generally indicate poor business performance whereas happy customers indicate a well performing business.

With the given business strategies, the airline company can cater to customer satisfaction by leveraging relevant features to better the experience and overall profits. Since online boarding is a relevant factor to the company (**Figure 8: Attribute makeup**), there should be more marketing on the online services of the business as well as promote deals on online memberships that come with a multitude of benefits that would be hard for frequent fliers to turn down. Along with marketing online services of the airline, some accommodations can be made for inflight services to help ensure customer satisfaction. Inflight wifi service is a relevant aspect to customer satisfaction so the company should take surveys from customers to ask if the inflight wifi service is reasonably priced/fast. Based on the overall response, the company can adjust the prices of inflight wifi services and/or upgrade wifi infrastructure to accommodate customer needs. Given **Figure 4: Type of Travel,** we can also see that type of travel is

partitioned to where economy fliers tend to be more dissatisfied than business fliers. The company should consider special deals to entice economy fliers to use our services more and increase profits and customer retention. Loyalty to the company should be the number one priority for a business to focus on because customer retention can be beneficial to the spreadability and accountability of the company's services.

Conclusion

In conclusion, the analysis of the Airline Passenger Dataset has provided valuable insights that airlines and other aviation industry businesses can use to enhance their customer experience. Our study has identified certain key factors that affect passenger satisfaction, including online boarding, class, and type of travel, while also highlighting certain features that do not significantly impact satisfaction such as gate location and gender. By leveraging machine learning techniques like Decision Trees we have generated accurate and coherent results that can be used to improve the overall customer experience and drive growth for airlines and related businesses. Relevant businesses can also use these insights to prioritize the areas that are lacking in their services and develop strategies that can boost customer loyalty and satisfaction. Overall, this study can help stay ahead of the competition by providing valuable insights that help them understand their customers and deliver services that exceed their expectations.