An Analysis of Passengers' Satisfaction on US Airlines

Daniela Johnson
Computer and Information Science Department
Fordham University, NY USA
djohnson122@fordham.edu
CISC6080 Capstone in Data Science

***Abstract*— There has already been a notable increase in flights this year. Passenger satisfaction is important for airlines because it can increase profits and customer loyalty. This project aims to find which features are most important to passenger satisfaction through forward feature selection, linear regression, and random forests. We will also train the model to predict passenger satisfaction using KNN and the decision tree model to see which model performs better to unseen data so that US airlines can collect data on their passengers and predict satisfaction themselves.**

Keywords*— *feature importance, passenger satisfaction, correlation coefficient, rank

I. INTRODUCTION

Measuring passenger satisfaction can help airlines adjust their flights to boost customer satisfaction and increase purchases. I examine the correlation between information about the customer and the many varying services airlines provide and whether each passenger is 'satisfied' or 'neutral or dissatisfied' with their flight. These factors are included before, during, and after take-off. Over the twenty-two features that will be analyzed, there are three different types of variables: categorical, infinite numerical, and discrete numerical on a 1-5 scale, with 0 being inapplicable. I hope to find which features highly correlate with airline passenger satisfaction using forward feature selection, linear regression, random forests, and customer satisfaction prediction using KNN and decision trees.

II. BACKGROUND INFORMATION

This dataset was updated four years ago by its author TJ Klein, who collected it before COVID-19. During COVID-19, cleanliness, seat comfort, and legroom were critical features when it came to passenger satisfaction. People would have wanted to socially distance themselves from each other and with seat comfort and plenty of leg room that would give the passengers more space. During the pandemic, according to a recent study published in 2023, "passenger satisfaction reached record highs," which surpassed pre-pandemic levels [5]. This, most likely, was due to less crowded flights, which led to emptier seats and more room and comfort for the passengers on board. Since this dataset is from before COVID-19, other factors may be more relevant. In 2013, a survey was conducted where "passengers who were greeted by airline staff with a smile reported satisfaction scores between 105 and 211 points higher than those who were not warmly greeted" and "passengers who use in-flight wi-fi are 39 points happier than those who do not use it" [6]. Information before COVID-19 in this situation could be more beneficial than information during the pandemic because, although it was more recent, many people were afraid to travel. This year (2024), it is predicted that there will be 40.1 million flights, which has steadily increased since 2004 [10]. There was a drop in 2020 due to the COVID-19 pandemic, but air travel has since recovered. Therefore, there are more similarities between the present day and before the pandemic, despite a record-breaking flight increase this year [12].

Some of the data set features listed in Table 1 need additional explanation. We have one feature called "inflight services" and another called "on-board service." Inflight services include pilot briefings, voice communications, progress reports, weather advisories, etc. [3]. On-board services, conversely, include the care the flight attendants provide, pillows, earphones, and blankets being provided, etc. Another feature that may need some clarification is "Online Checkin." Online check-in tends to allow the passenger to choose a seat before they get to the airport (if they did not pay for one), skip lines (if they do not need to check their bags), and it can, at times, be cheaper [7]. Some people wait to check-in in person at the airport, but online check-in is done before arrival.

III. EXPERIMENTS & FUNCTIONS

This project aims to help U.S. airlines boost their sales by improving customer satisfaction. Using Python and the Scikit learn libraries, I code and obtain an output that tells me which features play the largest role in passenger satisfaction and train models to predict passenger satisfaction.

a) Dataset Details and Preprocessing

The training and testing datasets came from Kaggle, which had 103,905 records on the training dataset and 25,977 on the testing dataset. The 22 features are shown in Table 1. The features with "scale" in the "type of variable" column are rated based on a 1-5 scale (one being the lowest, five being the highest), where 0 means the service is inapplicable. For example, for the feature WIFI, if a passenger were to record a 0, they did not use WIFI on the flight and did not want to. It is important to note that the departure and arrival delay features are in minutes. As for missing values, there were only missing values in the arrival delay feature. In the training data, there were 310 missing values, and in the testing data, there were 83 values. These values were such a small percentage in our overall dataset (0.298% in the training data and 0.320% in the testing data) that I deleted the data in these rows. There was still a sufficient amount of records from which to obtain results. The dataset is also acceptably balanced, with 43.3% of passengers reporting being satisfied with their flights and 56.7% reporting being neutral or dissatisfied. Since forward feature selection, linear regression, and KNN require numerical input, I used Label Encoder from the Scikit preprocessing library to convert all categorical data to numerical. Although random forests and decision trees can work with categorical data, converting the data to numerical can obtain optimal performance.

	Feature Name	Type of Variable
1	Gender	string
2	Loyal or Disloyal?	string
3	Age	integer
4	Personal or Business trip?	string
5	Business, Eco, Eco Plus	string
6	Flight Distance	
7	WIFI	scale
8	Departure & Arrival Time	scale
9	Online Booking	scale
10	Gate Location	scale
11	Food & Drink	scale
12	Online Checkin	scale
13	Seat Comfort	scale
14	Entertainment	scale
15	On-board Service	scale
16	Leg Room	scale
17	Baggage Handling	scale
18	Checkin Service	scale
19	Inflight Services	scale
20	Cleanliness	scale
21	Departure Delay	integer
22	Arrival Delay	integer

Table 1. Features

b) Feature Correlation

The correlation matrix will show us which features greatly impact each other and our target variable of passenger satisfaction. This is significant because if two features are closely related, and if the airline were to improve one feature, the other feature could possibly also improve or worsen. It is important to note whether the two features correlate positively or negatively. If two features are highly related, they could be combined as one feature, but their correlation must be close to 1 or -1. Instead of combining two features that indicate multicollinearity, you could also drop one to reduce redundancy and improve stability to simplify the model.

c) Forward Feature Selection

Forward feature selection determines which features improve the model's performance. Cross-validation is used to assess the model, which will show us how well the model generalizes to unseen data. The features selected first will have the most crucial effect on passenger satisfaction, and those selected later in the process do not have as large of an impact, but they still contribute to the model.

d) Linear Regression and Random Forest Feature Importance

The linear regression coefficients and magnitudes from passenger satisfaction are analyzed to determine feature importance. Feature importance will also be analyzed through the random forest model, and the importance of the feature will be calculated by how much it reduces impurity. A variety of models is always necessary because a limitation of linear regression is that it assumes a linear relationship between the features and the target variable, so it may not capture complex relationships, whereas random forests can.

e) Evaluating KNN and Decision Trees on the Testing Data

Finally, KNN is implemented to train and evaluate the model on the testing data. It is important to run the model using different values of k because it can lead us to different conclusions about the data. A lower k value can lead to overfitting, so the model would perform well on the training data but could perform poorly on unseen data. A higher k provides a more generalized model, but if it gets too high, it may not capture important patterns, leading to underfitting. A better accuracy with a lower k would imply the data is relatively clean and not noisy. The method of cross-validation will be used to choose the optimal k. I implemented the models and evaluated the accuracy using Confusion Matrices and Classification Reports from the Scikit learn library.

Another model that was implemented was Decision Trees. Decision trees use feature selection to repeatedly split the data by selecting the best feature. The tree-building process will continue until all 22 features have been selected. Similarly to how KNN was evaluated, the model will perform on the testing data and output the Confusion Matrix and Classification Report. Then, pruning will be applied to avoid overfitting the model. The features that do not improve accuracy or add little predictive power will be removed. Again, the new pruned decision tree model will be evaluated on the testing data using the Confusion Matrix and Classification Report. Finally, we will compare the performance of KNN and decision trees on our test data.

IV. RESULTS

a) Feature Correlation

It is difficult to assess some of the correlations involving features rated on the 1-5 scale because of the option of passengers being able to answer with a 0. It is important for the data that passengers had that option because if not, our data would lead us to draw false conclusions. For example, hypothetically, for “flight distance” and “seat comfort,” if people with shorter distances rate seat comfort higher than others, that would lead to a negative correlation between the two features. If passengers, however, with a long flight distance put 0 because they do not have an opinion, this would skew and weaken our correlation. The rows with 0s in any of the columns where the feature is rated on the scale could be removed, and the correlation matrix could be recreated. However, these 0s are still very important to our dataset, which puts us in a difficult position. We must draw inferences based on the matrix in Figure 1, knowing that the 0 value skews the relation between some features.

The correlation matrix in Figure 1 may be overwhelming, but the colors are the most important. The closer the shade of the box is to dark red (positive correlation) or dark blue (negative correlation), the more highly correlated the two features on the x-axis and y-axis are. Let me draw your attention to the features “WIFI” and “Online Booking.” They have a positive correlation of 0.72. We can assume that those who booked online and found it very easy are technically savvy and found it easy to use the WIFI. There is a slight correlation between “WIFI” and “age,” which is 0.018, and with “online booking” and “age,” which is 0.025. I compare these features with age because teenagers and young adults tend to be more technically savvy than those who are older. In 2014, it was found that “kids aged 6-7 achieved an average DQ [Digital Quotient] score of 96, and the average score stays above 100 until the age of 35 when it begins to drop off,” followed by “a particularly steep fall from the age of 60” [digital trends]. With the growing technology, it can be assumed that these DQ levels are all proportionally higher today. Young children are not booking flights, which could affect the correlation value between “age” and “online booking.” Still, young adults should generally have an easier time booking flights online than older ones. Children would answer 0, while it can be predicted that young adults would answer a 4 or 5 and older adults would answer a 1 or 2. With children answering 0, the correlation is skewed to be lower and positive because calculating correlation assesses the strength and direction of a linear relationship. When it comes to “WIFI,” it is a difficult feature to discuss because the WIFI feature talks about how well the WIFI connection is on the flight and not who is using it. However, those who answered 0 on the dataset tend to be older because they did not use the service on the flight.

Another feature to look closely at would be “departure/arrival time convenient.” In Figure 1 it is labeled as “timing” for spacing reasons. As mentioned, those who find online booking easier are most likely technically savvy and around 18-35 years old. The “departure/arrival time convenient” and “online booking” features have a correlation of 0.44, which is relatively high. This means that passengers who had an easier time booking their flight online are more satisfied with their flights' departure and arrival times. This would make sense because when booking a flight, one factor that affects which flight you decide to purchase is the arrival and departure times. None of the features correlate close enough to 1 or -1 to combine them or drop one.

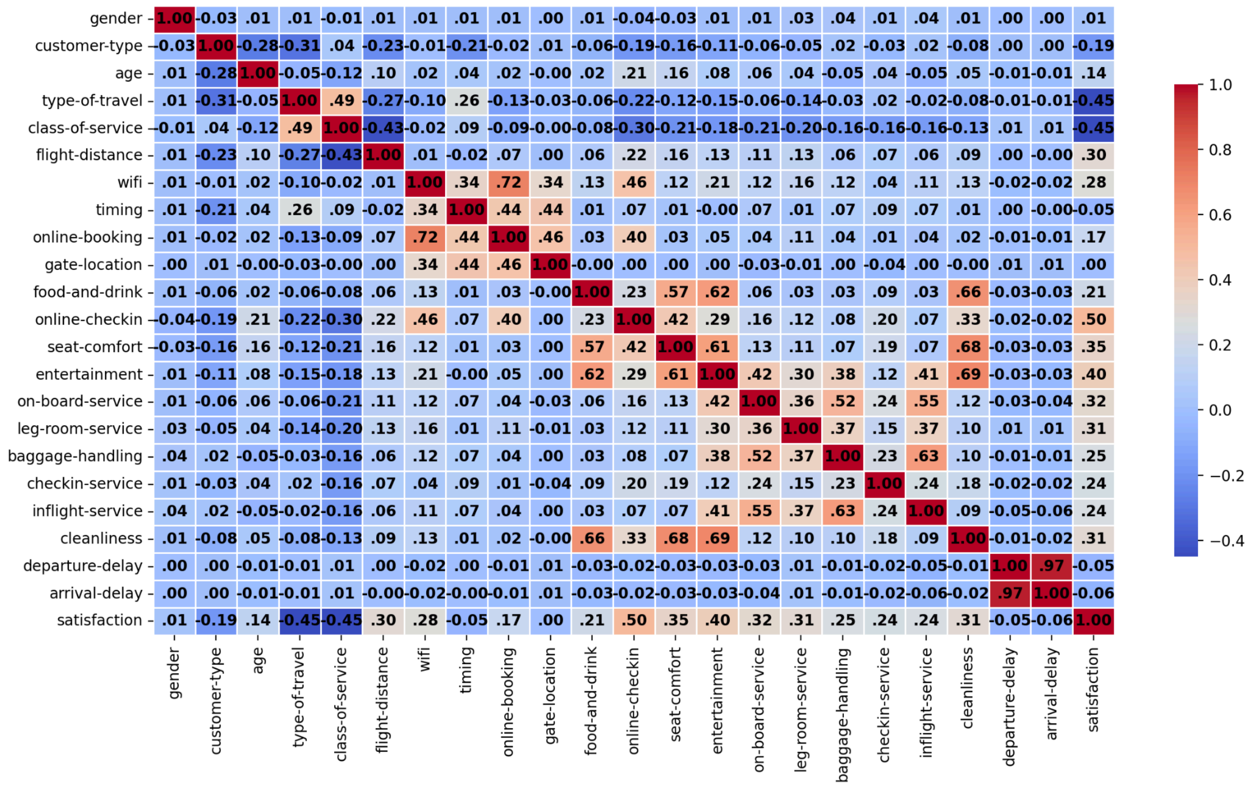

Fig. 1. Correlation Matrix

b) Forward Feature Selection

Using the forward feature selection method, it can be seen in Table 2 that “online check-in” had the largest impact on passenger satisfaction. The benefits of online check-in include avoiding the additional costs of checking in at the airport, choosing a seat, and less time at the airport since the passenger already has their boarding pass; it also secures the passenger’s spot on the flight in case it is overbooked since they already have their seat [8]. With these benefits, it is a no-brainer that this would be so heavily correlated with passenger satisfaction.

The second most important feature is “Personal or Business trip.” If someone is going somewhere for vacation, which is a personal trip, they are usually very excited. In contrast, if they are going somewhere to work, they may not be looking forward to it. If the trip is personal and business, we can assume most people would answer with a personal trip if they were looking forward to it, or they would still consider it a business trip if they were dreading it, proving the heavy correlation to passenger satisfaction to hold true. Personal trips remind people of vacations, and the word “business” does not have the same exciting connotation the word “vacation” does to it.

The third most important feature is “WIFI.” Nowadays, people are glued to screens, so it is very important to a passenger to know how well the Wi-Fi works or if there is any Wi-Fi at all. In 2023, it was found that the average screen time across various devices in the US was 7 hours and 3 minutes [11]. It is hard for most people to sit on a plane, no matter the distance (wifi and flight distance only correlate 0.0071), without access to the internet because it provides entertainment and allows people to be productive. Dr. Mark, a Professor from the University of California, states that “the internet and digital devices have affected our ability to focus,” and our attention spans are shrinking [13]. People no longer have the patience to wait for something to load due to a bad Wi-Fi connection. Someone traveling for business may be extra frustrated if they are flying to a conference where they need to finish their presentation during the flight and the internet connection is slow. The features “WIFI” and “personal or business trip” have a correlation of -0.11, as can be seen in Table 2. The negative does not matter in this case since personal or business trip is a categorical, discrete variable. Whether there is wifi or not and how well it works could be very important to someone on a business trip. But overall, it is very important to passenger satisfaction for entertainment and productivity.

Rank	Feature Name	Test Accuracy
1	Online Checkin	78.77%
2	Personal or Business trip?	84.35%
3	WIFI?	84.76%
4	Entertainment	84.57%
5	Cleanliness	84.58%
6	Departure & Arrival time	84.54%
7	Gate Location	84.42%
8	Loyal or Disloyal?	84.61%
9	On-board Service	86.17%
10	Checkin service	86.70%
11	Buiness, Eco, Eco Plus	87.02%
12	Leg Room	87.27%
13	Inflight Services	87.39%
14	Arrival Delay (mins)	87.46%
15	Baggage Handling	87.52%
16	Flight Distance	87.52%
17	Food & Drink	87.52%
18	Gender	87.51%
19	Age	87.52%
20	Seat Comfort	87.55%
21	Online Booking	87.55%
22	Departure Delay (mins)	87.56%

Table 2. Forward Feature Selection

c) *Linear Regression and Random Forest Feature Importance*

In Figure 2, it is displayed that linear regression and random forest obtained very different results when it came to feature importance. This is as expected, though, because linear regression cannot capture complex relationships like random forests. The lower the point on the line for the feature, the more important the feature is, and the higher the point for the feature, the less important it is.

The line graph in Figure 2 could be misleading. There are only 3 features that have the same exact rank in both models: “Business, Eco, Eco Plus,” “Gate location,” and “Arrival Delay.” Although the two lines intersect multiple times, such as between “seat comfort” and “gender,” it does not imply that the models match rank at those points. Random forests ranked “Seat Comfort” as the 6th most important feature and “gender” as the 22nd most important feature. Because these features were rated 15th and 16th using linear regression, the line representing random forests had to cross the linear regression line. Essentially, the intersections could be similarities between the way the two models ranked the features or large differences, and to determine this, you must look very closely at the graph to see exactly where the points are.

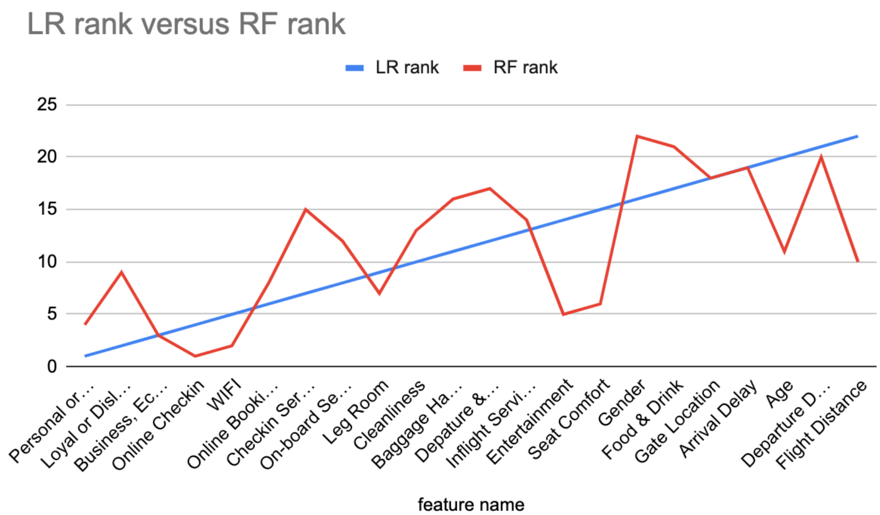

Fig. 2 Comparing Linear Regression & Random Forests' ranks of features

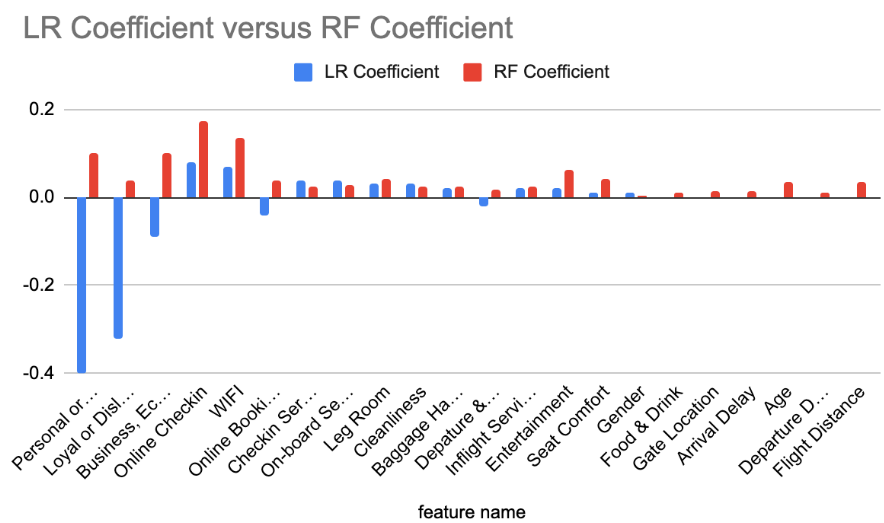

Fig. 3 Comparing Linear Regression & Random Forests' Correlations to Passenger Satisfaction

The difference between Figures 2 and 3 is that Figure 2 is the features graphed by their rank in order of importance, and Figure 3 is the correlation coefficient for each feature. In Figure 3, the features "Personal or Business trip" and "Loyal or Disloyal" show a large difference between linear regression correlation coefficient to passenger satisfaction versus random forest.

Both of these features are discrete and categorical, so the direction of the coefficient does not matter.

After comparing our results in this section with our results from the forward feature selection method, random forests generally agree with the importance of features with the forward feature selection method, while linear regression does not. Therefore, random forests and forward feature selection capture feature importance more accurately for our dataset, meaning the features have significant non-linear relationships with passenger satisfaction. Had they had a linear relationship, the results from the forward feature selection method would have matched those from the linear regression method.

Two features that would have been thought to have had a heavy correlation with passenger satisfaction would have been "Arrival Delay" and "Departure Delay." Using the forward feature selection method, they were ranked 14th and 22nd, and using random forests, they were ranked 19th and 20th. From a study done this year (2024), although delays have a large impact on customer satisfaction, "passengers are more inclined to judge airlines based on their satisfaction with the in-flight service provided by the flight attendants and aircraft facilities" [2]. Proper communication with flight delays is key because passengers may develop a negative view of the airline. If the airline could already compensate a passenger for a flight without them having to fight for it, they could avoid the passenger developing a negative view of the airline. By compensation, I mean providing travelers with alternative transportation, meals and refreshments as they wait, complimentary upgrades, or even a refund [1]. It is impossible for airlines to avoid delays related to weather conditions, but if they know how to deal with them, they can avoid reducing passenger satisfaction. Onboard and inflight services are "the drivers of airline success," especially regarding delays [2]. It is important that airlines have plans for even the things they cannot control to maximize passenger satisfaction.

d) Evaluating KNN and Decision Trees on the Testing Data

Using different values of k for KNN is key to finding the optimal k value and drawing conclusions about the data. I tested k = 1,5,10,25,50. The greatest obtained accuracy was 73.7% when k = 5. This was calculated in the Classification Report. A low k value captures the finer details and variations in the dataset. This could lead to overfitting, which means the data would not perform well to unseen data, but when k=5, the model performed the best on the testing data, which is the unseen data. When performing cross-validation, the model also had a very low standard deviation score of 0.0035, proving that the model performs reliably across different subsets of data. In Table 3, we have 2,786 false positives and 3,774 false negatives.

		Actual Values	
		Positive	Nefative
Predicted Values	Positive	11742	2786
	Negative	3774	7591

Table 3 KNN Model (k=5) — Confusion Matrix

The Decision Tree model demonstrated significantly superior performance. The accuracy before pruning was 94.7%; after that, it was 94.6%. The only feature not used in the pruned decision tree was "gender." The original importance of this feature was 0.001854, so it did not contribute much to the model performance before pruning. The feature was excluded from the pruned decision tree to reduce the model's complexity. Now, it would rely more heavily on the remaining features, which have larger impacts on model performance. A study done in 2018 reported that "there are no significant differences between female and male passengers'" importance of service features on airlines [4]. They reported having the same three most important criteria: "cabin features and responsiveness, employee competency and booking & reservation, and in-flight services" [4]. Therefore, "gender" does not seem to be a necessary feature for predicting passenger satisfaction.

In Tables 4 and 5, the decision tree model before pruning had a larger number of false positives, 706, and the model after pruning had a larger number of false negatives, 791. Our results for both models are very similar in our classification reports and confusion matrices. The pruned decision tree is better for predicting passenger satisfaction because it avoids overfitting, reduces complexity, and is more generalized, so it would perform better with unseen data. This model is easier to interpret.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	13822	706
	Negative	656	10709

Table 4 Decision Tree Model (before Pruning) — Confusion Matrix

		Actual Values	
		Positive	Negative
Predicted Values	Positive	13919	609
	Negative	791	10574

Table 5 Decision Tree Model (After Pruning) — Confusion Matrix

V. CONCLUSION

The dataset used for these analyses did not state which airline each individual traveled on. However, all US airlines can benefit from these results because they should improve the features that were highly correlated with passenger satisfaction. These features include online check-in, inflight wifi, and inflight entertainment. The type of travel (personal or business) also greatly impacts passenger satisfaction, but there is not much the airline can do about that. If there are empty seats in business class, those traveling for business should be upgraded first. This could improve their satisfaction rate. If a specific airline would like to collect data specifically from their passengers to predict passenger satisfaction, they should use the pruned decision tree model because it has very high accuracy.

VI. REFERENCES

Bhanushali, Harsh. “Impact of Flight Delays.” ClaimFlights, 17 Oct. 2023, claimflights.com/impact-of-flight-delays/.

Cen Song a, et al. “The Adverse Impact of Flight Delays on Passenger Satisfaction: An Innovative Prediction Model Utilizing Wide & Deep Learning.” *Journal of Air Transport Management*, Pergamon, 26 Oct. 2023, www.sciencedirect.com/science/article/pii/S0969699723001540.

“Chapter 4. In‑Flight Services.” *General*, www.faa.gov/air_traffic/publications/atpubs/fs_html/chap4_section_1.html#:~:text=Section%201.,General&text=In%2Dflight%20services%20are%20those,Pilot%20briefings. Accessed 4 Aug. 2024.

Feride Bahar Kurtulmuşoğlu a, et al. “Does Gender Matter? Considering Gender of Service in the Airline Industry.” *Journal of Air Transport Management*, Pergamon, 7 May 2018, www.sciencedirect.com/science/article/abs/pii/S0969699717304994.

Francisco Pereira a, et al. “The Impact of the COVID-19 Pandemic on Airlines’ Passenger Satisfaction.” Journal of Air Transport Management, Pergamon, 7 June 2023, www.sciencedirect.com/science/article/pii/S0969699723000844#:~:text=Interestingly%20passenger%20satisfaction%20reached%20record,and%20convenience%20of%20air%20travel.

Haq, Husna. “What Makes Air Passengers Happy?” BBC News, BBC, 25 Feb. 2022, www.bbc.com/travel/article/20130517-what-makes-air-passengers-happy.

How Does Online Check-in Work? | Travelsupermarket, www.travelsupermarket.com/en-gb/blog/travel-advice/how-does-online-check-in-work/. Accessed 4 Aug. 2024.

Mariiadolgopolova. “Online Check-in: Advantages and Tips.” Flightright, 29 June 2023, www.flightright.com/blog/online-check-in-advantages-and-tips.

Nield, David. “Teens Top the Tech-Savvy Chart While Adults Lag Behind.” *Digital Trends*, Digital Trends, 10 Aug. 2014, www.digitaltrends.com/computing/teens-top-tech-savvy-chart-adults-lag-behind/#:~:text=With%20the%20average%20DQ%20score,for%20those%20aged%20over%2075.&text=Kids%20aged%206%2D7%20achieved,from%20the%20age%20of%2060.

Published by Statista Research Department, and Jun 28. “Airline Industry Worldwide - Number of Flights 2024.” *Statista*, 28 June 2024, www.statista.com/statistics/564769/airline-industry-number-of-flights/.

“Revealing Average Screen Time Statistics for 2024.” *Backlinko*, 11 Mar. 2024, backlinko.com/screen-time-statistics.

“Summer Travel 2024: Record Air Travel, Fewer Cancellations, Better Airports, More Passenger Protections.” *U.S. Department of Transportation*, www.transportation.gov/briefing-room/summer-travel-2024-record-air-travel-fewer-cancellations-better-airports-more. Accessed 4 Aug. 2024.

“Why Our Attention Spans Are Shrinking, with Gloria Mark, Phd.” *American Psychological Association*, American Psychological Association, www.apa.org/news/podcasts/speaking-of-psychology/attention-spans. Accessed 4 Aug. 2024.