

# Chapter 26

## Case Study: AAA Airline



Deepak Agrawal, Hema Sri Sai Kollipara, and Soumithri Mamidipudi

### 1 Introduction

*Steven Thrush*, Chief Revenue Officer of *AAA Airline Corp*, was concerned about his company. The airline industry, buoyed by strong demand and low oil prices, had been on an upswing for the last few years. Rising competition, however, had begun to pressure AAA's operations. Shifting market sentiments and an increasingly complicated market had made travelling to most destinations in the USA dependent for most customers on a number of contrasting factors.

Moreover, the rise of low-cost carriers and online ticket comparison websites had put immense downward pressure on ticket prices, squeezing the margins of companies and forcing them to investigate new avenues of growth in order to maintain their profitability.

Thrush had just returned from a conference focused on the application of data science and analytics in the passenger transport industry. At the conference, researchers and practitioners talked about the rapid advance of big data and its power to understand and predict customer behavior. Thrush grasped that big data

---

**Electronic supplementary material** The online version of this chapter ([https://doi.org/10.1007/978-3-319-68837-4\\_26](https://doi.org/10.1007/978-3-319-68837-4_26)) contains supplementary material, which is available to authorized users.

---

The case study is written by Agrawal, Deepak; Kollipara, Hema Sri; and Mamidipudi, Soumithri under the guidance of Professor Sridhar Seshadri. We would like to thank Coldrena et al. (2003) and Garrow (2010) who are the inspiration to develop this case study.

---

D. Agrawal (✉) · H. S. S. Kollipara · S. Mamidipudi  
Indian School of Business, Hyderabad, Telangana, India  
e-mail: [a.deepak@outlook.com](mailto:a.deepak@outlook.com)

could help his company move toward new models that took into account a dizzying range of factors in order to make better decisions.

When Linda James, the company's head of route planning, approached him to ask about the feasibility of launching a New York–Boston flight, Thrush immediately thought about employing the customer choice models he had heard about in order to understand the proposition. He asked his data team to use the company's database of its customers to understand the question of how well received a new flight from New York to Boston would be. He knew that to answer such a question, the team would also have to investigate many more issues such as what manner of pricing would be most efficient, what type of aircraft would be most efficient, and how best to reach new customers who might not otherwise fly AAA.

Settling on the correct approach to the problem, Thrush knew, would be the best way to deliver the best service possible to customers while maximizing the profit of his company.

## 2 AAA Airline Corp

AAA Airline Corp was founded in 2005, amid a sea change in the travel industry. As Internet penetration grew and price comparison websites became increasingly popular, AAA saw an opportunity for a low-cost carrier to capitalize on the increased customer focus on prices.

Like many carriers founded in the wake of the online boom, AAA's philosophy was to compete purely on price. Instead of focusing on specific regional flights and needs, AAA's philosophy was to identify and fill gaps in the market and in doing so carve out a niche for itself. While most of its flights operated in a hub-and-spoke system out of Boston Logan Airport, the company was not averse to operating point-to-point routes that are the hallmark of low-cost carriers worldwide.

AAA's initial method to identify which routes were profitable relied on a mix of market research and intuition. AAA's original management team consisted mostly of industry veterans hailing from Massachusetts, and they were all well acquainted with the needs of local customers. AAA's in-depth expertise in its initial market helped it survive where many of its rivals failed, prompting it to expand its offering and plan for more ambitious growth.

By 2016, the size of AAA's fleet had risen considerably, prompting Thrush's concern regarding its next steps. AAA's history meant that it had access to a large database of its own customers, which it had so far been using to forecast future demand and traffic patterns. Thrush was keen to know, however, what new tools could be used and datasets found in order to analyze the market and help the company stride into the new era of commercial air travel.

### 3 History and Current Context of the US Airline Industry

The US airline industry had a capacity of more than over 1.1 million available seat miles (accounting for both domestic and international flights) in 2016 and is the largest geography for air travel worldwide. The sector supplies nearly 3500 available seat miles, a measure of carrying capacity, per person in North America, more than double that of the industry in Europe.

The effects of the Airline Deregulation Act of 1978 are still being felt today. Before the Act, American airline companies were strictly constrained by the Civil Aeronautics Board, which was responsible for approving new routes and pricing. The Board could give agreements between carriers anti-trust immunity if it felt it was in the public interest. This resulted in a situation where airlines competed purely on in-flight service and flight timings and frequency.

Legacy carriers—airlines founded before deregulation—are able to offer better service and benefits such as loyalty schemes as a result of the environment in which they operated at their founding. Airlines such as these tend to have larger planes and operate in a hub-and-spoke system that means that their flights are largely based out of a single airport.

After the industry was deregulated, airlines became free to decide what routes to fly and what prices to offer. New low-cost carriers like AAA entered the market and shifted the paradigm by which companies in the industry functioned, forcing full-service airlines to adapt. Since 1978, more than 100 airline carriers have filed for bankruptcy,<sup>1</sup> underscoring the tumultuous nature of the industry.

The proliferation of the Internet was no less disruptive to the airline and travel industries. Customers were more able than ever to compare flights, and their ability to discriminate between a multitude of choices at the tap of a key left companies the world over scrambling to keep up. This meant that companies such as AAA were forced to use ever more complicated models in their attempts to understand and predict customer demand while at the same time keeping track of their costs.

Thrush knew that AAA's spoke-and-hub system helped to keep costs low and enable the airline to fly a large number of passengers. However, he was also aware hub airports were especially hard-hit by the increase in the number of passengers using them, meaning that pressure on his staff and his operations was mounting daily. The industry's domestic load factor, the fraction of available seats that were sold, had risen to 85% in 2016 from 66% in 1995.<sup>2</sup> Domestic ASMs rose 29% in the same period to 794,282. However, the sizes and capacities of hub airports had not risen in line with this explosive growth in passengers due to property, environmental, and financial constraints.

---

<sup>1</sup><http://airlines.org/dataset/u-s-bankruptcies-and-services-cessations/> (accessed on Jul 21, 2018).

<sup>2</sup><http://web.mit.edu/airlinedata/www/2016%2012%20Month%20Documents/Traffic%20and%20Capacity/Domestic/Domestic%20Load%20Factor%20.htm>, accessed on Jul 15, 2017.

The airline industry had so far tackled the problem of being able to supply its customers with the flights they needed by looking to strategic alliances and code sharing deals. Airlines that were part of the same alliance agreed to pool their resources by agreeing to be located in the same terminals in hub airports, operating flights under the banner of more than one carrier, and offering privileges to members of fellow members' loyalty programs. By doing so, companies ensured that they did not have to operate and fly every route their customers demanded.

"We need to consider whether it makes sense to abandon our spoke-and-hub system. Our rivals that use point-to-point routes are eating into demand, and I'm sure passengers are noticing the kind of queues that are building up in the larger airports," James told Thrush.

## 4 Industry Data Sources

The airline industry uses three main types of data to interpret the environment in which it operates—demand data, such as booking and ticketing; supply data, such as schedules; and operational data, such as delays, cancellations, and check-ins. Thrush found that data scientists used these databases to uncover traveler preferences and understand their behavior.

The demand data in the industry comes from booking and ticketing databases, and detail a plethora of factors that affect customers while booking flights, and take into account exactly what information is available to customers at the time of their purchase. Supply data is usually accessible so that customers are able to identify flights, but the industry's main sources are schedules and guides provided by the Official Airline Guide (OAG). These guides collate information including origin, destination, trip length, and mileage for individual flights globally.<sup>3</sup> Data regarding the operational status of flights is usually available freely, though it is often not granular. AAA, like its competitors, kept detailed records of operational data in order to catch patterns of inefficiency. In addition, The US Department of Transport maintains a databank that consists of 10% of all flown tickets in the country.<sup>4</sup> The databank provides detailed ticketing, itinerary, and travel information and is freely available for research purpose.

---

<sup>3</sup><https://www.oag.com/airline-schedules>, accessed on Jul 15, 2017.

<sup>4</sup>[https://www.transtats.bts.gov/Tables.asp?DB\\_ID=125&DB\\_Name=Airline%20Origin%20and%20Destination%20Survey%20%28DB1B%29&DB\\_Short\\_Name=Origin%20and%20Destination%20Survey](https://www.transtats.bts.gov/Tables.asp?DB_ID=125&DB_Name=Airline%20Origin%20and%20Destination%20Survey%20%28DB1B%29&DB_Short_Name=Origin%20and%20Destination%20Survey) (accessed on Jun 24, 2018).

## 5 Meeting with John Heavens, Consultant

Thrush met with John Heavens, a data scientist and airline travel consultant, to inquire further about the possibility of using advanced data models in order to understand and forecast customer behavior.

Heavens told Thrush that the industry's old time-series/probabilistic models had become too outdated. Multinomial logit decision-choice models were the industry's mainstay tools in understanding consumer demand. These models broke itineraries down by assigning utility values to each flight and attempting to determine which factors were most valuable to customers. By observing the factors that affected customer choices for each origin–destination pair, Thrush would be able to predict with confidence where customers were looking to travel next.

However, Heavens also gave Thrush a third option. “Even the decision-choice models are becoming old, and we’re moving in new directions now,” he said. The consultant pointed out that the industry's MNL models were essentially linear in nature, and were not able to deal with factors that were correlated. In addition, their rigid need for data input meant that they could not predict the demand for new routes and new markets.

Instead, Heavens pointed to groundbreaking artificial intelligence research as the vanguard of an array of new technological tools that could be used to predict future demand. Techniques such as random forests, gradient-boosting machines, and artificial neural networks were able to produce better out-of-sample results without sacrificing in-sample goodness-of-fit. While these techniques lacked the readability and simplicity of MNL models, they were ultimately more efficient.

## 6 Meeting with Veerabhadram, the Data Scientist

After being presented with the models, Thrush knew he had a difficult decision to make. Moving to new methods of analysis had clear advantages, yet the significant investment in time and effort needed to be justified. Training and hiring employees and conducting ongoing analysis would be a drain on the company's resources.

Thrush looked to Hari Veerabhadram, the newest member of his team, to explain to him exactly which models are best suited to understand customer preferences. Hari knew that he had to explain how the models worked. He started thinking about which variables in the models he would use and what would be the most important. He knew that it would be crucial to explain why particular variables were the most important and which model was better at predicting customer preferences. Management always like visual proof of analysis. Thus, he felt that he would need to explain and compare the models both statistically (mean squared error, percentage variance explained by model, etc.) and through visualization methods (predicted vs. actual fit, training vs. validation results, etc.).

**Table 26.1** Itinerary features (data description)

S. no.	Variable name	Description
1	ODPair_ID	Origin–destination pair identifier
2	Itinerary_ID	Itinerary identifier within O–D pair
3	Airline	Identifier for airline (A, B, . . . , H, and all others) AAA Airline can be identified as Airline = “A”
4	Aircraft_Type	Type of aircraft in itinerary—propeller, jet, mainline
5	DoW	Day of the week that itinerary flies
6	Service_Level	Level of service of itinerary—non-stop, direct, single connect, double connect
7	Best_Service_Level	Best service level available for itinerary in the O–D Pair Non-stop > Direct > Single connect > double connect
8	Dep_Time	Time of departure at origin (HH:MM)
9	POS_Presence	Point of sale presence of airline at origin and destination airports. This measures the number of flights departing out of airports
10	Code_Share	Whether or not the itinerary is operated as a code-share with another carrier
11	Mileage	Distance travelled by the itinerary (converted into ratio)
12	Fare	Average fare of the itinerary (converted into ratio)
13	Pass_Cnt	Number of passengers who chose itinerary
14	No_of_itinaries	Number of itineraries available for O-D pair
15	Pass_Cnt_ODPair	Number of passengers in O-D Pair (Market Size)
16	Time_Bucket	Hourly time-of-day buckets corresponding to the departure time of the first leg of the itinerary 5 = (4 a.m.–5:59 p.m.); 6 = 6–6:59; . . . ; 22 = 10 p.m.–11:59 p.m. No departures from 12 midnight to 4 a.m.
17	Airline dummies	Whether the itinerary is with Airline XX
18	Service Level dummies	Whether the itinerary is having specific service level
19	Time bucket dummies	Whether the departure time of the itinerary falls into time bucket XX
20	Aircraft type dummies	Dummy variables for different aircraft type for each the itinerary

Thrush told Veerabhadram to start by looking at the itineraries in which AAA was lagging behind its competitors. Were there any patterns? Which markets did AAA find itself out of the top three? Did AAA need to do anything different? Veerabhadram knew that he would have to recommend changes to AAA’s operations and began thinking about how to use the data at his disposal.

Veerabhadram started exploring the available data (refer to Table 26.1 for details) sourced from the company’s internal database and third-party data providers. He noticed that the sales data is at the itinerary level for multiple airlines, but not at customer or transactional level. The data consists of 1885 distinct origin–destination (O-D) pairs with all possible itineraries on respective routes (O-D pair). Each row is an itinerary for a given O-D pair and provides additional information such as

the identity of the airline, how many passengers chose that itinerary, what was the offered aircraft type, departure day and time, service-level, best service level available on that route, mileage, average fare, etc. For example, say, the O-D pair “5” represents the New York to Los Angeles route, and Airline = “A” represents all the itineraries offered by AAA Airline. Pick Itinerary ID “3.” This itinerary offers a Small Propeller service on the route as a single connect option departing at 7 a.m. from New York. Single connect is the best service possible on this route across all the airlines serving that OD pair.

The basic summary statistics (Table 26.2) helped Veerabhadram to understand the variability in the data.<sup>5</sup> He observed that AAA Airline is one of the top performing airlines connecting significant number of cities through single and double connect itineraries. It also means that even a minor change in route can have

**Table 26.2** Summary statistics

Basic summary statistics					
<i>Variable name</i>	<i>Distinct Obs</i>	<i>Mean</i>	<i>Median</i>	<i>Minimum</i>	<i>Maximum</i>
ODPair_ID	1885	NA	NA	1	1885
Itinary_ID	82,892	78	63	2	311
POS_Presence	NA	10	7	0	100
Code_Share	NA	NA	0	0	1
Mileage	NA	112	109	100	153
Fare	NA	99	100	10	471
Pass_Cnt	NA	1.7	0	0	534
Market size using Pass_Cnt	NA	212	52	1	3828
Airlines’ market share					
<i>Variable name</i>	<i>Total itineraries</i>		<i>Proportion (%)</i>		
Airline A	18,234		22		
Airline B	11,907		14		
Airline C	4463		5		
Airline D	12,593		15		
Airline E	13,849		16		
Airline F	3185		4		
Airline G	12,226		15		
Airline H	5231		6		
Airline others	1564		2		
Flight options offered by all the airlines					
<i>Variable name</i>	<i>Total itineraries</i>		<i>Proportion (%)</i>		
Non-stop	450		1		
Direct	454		1		
Single connect	35,155		42		
Double connect	46,833		57		

<sup>5</sup>Refer to the website to download the data (csv and excel version).

a major impact on scheduling and resource allocation. Veerabhadram wondered whether more analysis and better visualization might help represent the current situation in a meaningful way to decision makers.

“Changing the timing of our flights poses serious challenges, Steven. We’re better off keeping them the same so that we don’t confuse customers and complicate our logistics,” James said. Now, Veerabhadram started thinking of how to improve AAA’s performance without changing the schedule.

Thrush’s second priority was understanding which itineraries have gaps that could be filled by AAA. “Our business was built on flying between places that nobody realized needed more flights. Where do you think we need to go next?” he asked Veerabhadram.

Thrush told him that AAA’s fleet currently consists of 120 propellers, 40 jets, and 100 mainline aircraft, all of which were in service. Veerabhadram decided he would pick the five best unserved routes to recommend to Thrush. He also started to consider which routes AAA would have to discontinue in order to start new flights.

At the end of these discussions, Veerabhadram’s notebook was full! He decided to start with the following questions:

- Fit a customer choice model to the given data and analyze whether the features, model complexity, and size of data were adequate for answering the business questions.
- Identify variables of importance and make sure they made sound business sense. He thought he would try MNL and other techniques.
- Identify the currently unserved routes and recommend the five best unserved routes where the airline can introduce a new itinerary to increase revenue.
- Identify the O-D pairs where AAA Airline is not among the top three in terms of market share and then:
  - Analyze the offerings of the top three airlines in these O-D pairs.
  - Recommend changes in the itinerary features to improve the market share.
  - Suggest possible routes the airline could drop and explain why.

## Electronic Supplementary Material

All the datasets, code, and other material referred in this section are available in [www.allaboutanalytics.net](http://www.allaboutanalytics.net).

- Text 26.1: Airline Instruction manual.docx
- Data 26.1: AAA\_Airline\_dummy.csv
- Data 26.2: AAA\_Airline\_Template.xlsx

## References

- Coldrena, G. M., Koppelmana, F. S., Kasturirangana, K., & Mukherjee, A. (2003). Modeling aggregate air-travel itinerary shares—Logit model development at a major US airline. *Journal of Air Transport Management*, 9, 361–369.
- Garrow, L. A. (2010). *Discrete choice modelling and air travel demand: Theory and applications*. New York: Routledge.