AVIATION DATA

WHITE PAPER SERIES

JUNE 2019

DATA SCIENCE HYPE OR RIPE FOR AVIATION?

Abstract: As data and associated processing technologies and capabilities are becoming more accessible, there is an opportunity to revisit, rethink, enhance and digitally transform legacy processes. Sales and distribution for example, is a key area where this plays an important role towards more efficiency in the process as well as facilitating the journey towards hyper customer awareness and personalization. Other high potential areas include the passenger experience, comfort and safety. Data sharing concepts have also gained maturity in parallel with gradual cultural changes towards more openness and realization of the benefits of common industry platforms.

IATA

Disclaimer notice

Foreword

Alexandre de Juniac
Director General and CEO,
International Air Transport Association (IATA)

The aviation industry is on the verge of a new era, in which the use of data and the ability to turn data into information and insights will be more crucial than ever. This is reflected in the renewed IATA global strategic priorities, under the pillar, "Control of Data and Digitalization", to facilitate the digitalization of airlines' business models, their standards and associated regulatory environment, and to keep airlines in control of their data and customer relationships.

The ability to process data and the associated capabilities to extract benefits from data, are clearly becoming a key aspect of running a successful airline and equally for other players across the wider value chain. Those able to leverage data are already seeing concrete benefits in areas such as planning, commercial, operations and the customer experience.

IATA is committed to helping our members on this journey through raising awareness and supporting the development of data standards, as well as solutions and guidelines on data management and governance. Ensuring appropriate access and protection of data is also key. Additionally, a strong appetite exists to accelerate the adoption of new data standards, such as the New Distribution Capability (NDC), ONE Order and modern baggage messaging, in order to increase the efficiency of system-to-systems communications and lower the cost of maintaining these.

Having access to data is a prerequisite. It should be followed by building core data science capabilities and acknowledging the strategic value of the ability to process data for such things as network optimization, revenue management, mitigating unplanned maintenance, and optimizing allocation of ground crew by predicting demand and terminal peak times.

There has been a paradigm shift, with companies opening up their data sources and seeing the benefits of sharing data. In parallel, some industry data-sharing concepts have proven to be very effective and valuable to participating users, such as Direct Data Solutions, through which airlines share data to a common platform and reap the benefits of a rich source of data that has strong data management, governance and integrity, meeting the needs and requirements of its data contributors and users.

We are now living in "golden times", as vast amounts of data are available, in conjunction with affordable computing power, facilitating concepts such as dynamic offers, allowing contextualized offers and total revenue optimization. Another great use case is leveraging data to improve the flight experience, making flying more comfortable.

I call aviation the Business of Freedom. It liberates us from the constraints of geography and distance. In doing so, it empowers us to lead better lives, and makes the world a better place. Without aviation, the inter-connected global economy as we know it would not be possible. As we look to an expected doubling in demand for air travel over the next 20 years, our ability to leverage data in all areas of airline activity will be crucial to ensuring the Business of Freedom continues to grow the value it delivers to the global economy.

Acknowledgement

This white paper is the result of a multidisciplinary research, collaboration across different areas and functions within IATA, and the support and contribution from airlines. With special thanks to the team of authors and contributors.

Lead author and editor:

- Houman Goudarzi (Head of Business Intelligence and Industry Engagement, IATA)

Contributing authors:

- Dragos Budeanu (Manager Paperless Operations, IATA)
- Henry Coles (Head of Airline Distribution Standards, IATA)
- Perry Flint (Assistant Director Corporate Communication, IATA)
- Charles de Gheldere (Director of Travel Intelligence, IATA)
- Brent King (Head of Flight Operations Efficiency, IATA)
- Chris Markou (Head of Operational Cost Management, IATA)
- Juan Oliver (Manager Business Intelligence Projects, IATA)
- Andrew Price (Head of Global Baggage Operations, IATA)
- Marie-Claude Simard (Assistant General Counsel, IATA)
- Sebastien Touraine (Head of Dynamic Offer, IATA)

Airline contributing authors:

- Arnaud Capois (Advanced analytics - Customer Data Management, Air France-KLM)
- Charles Girard (Customer Data Officer, Air France-KLM)
- Michael Shores (Director Data Science, United Airlines)

We welcome any feedback (ads_team@iata.org) on this white paper and hope it will raise awareness and trigger discussions to unlock the full potential of data for airlines and the wider value chain.

Table of Contents

1 Introduction to Data Science

The exponential growth of data availability, together with increased accessibility and affordability of hardware and software resources, data processing and visualization of data, generates unprecedented business opportunities in the airline industry. Businesses across aviation, as well as other industries, are embarking on a large-scale Data Science journey, aiming to address new opportunities and enhance efficiencies in various areas.

Data Science, Artificial Intelligence, Machine Learning and Analytics are among the latest data buzzwords. Managed appropriately, they can be very beneficial. While McKinsey has painted a rather gloomy picture about transportation, logistics, travel and tourism when it comes to Artificial Intelligence and being digital frontiers, it does show the great potential if our industry starts moving the digital needle.

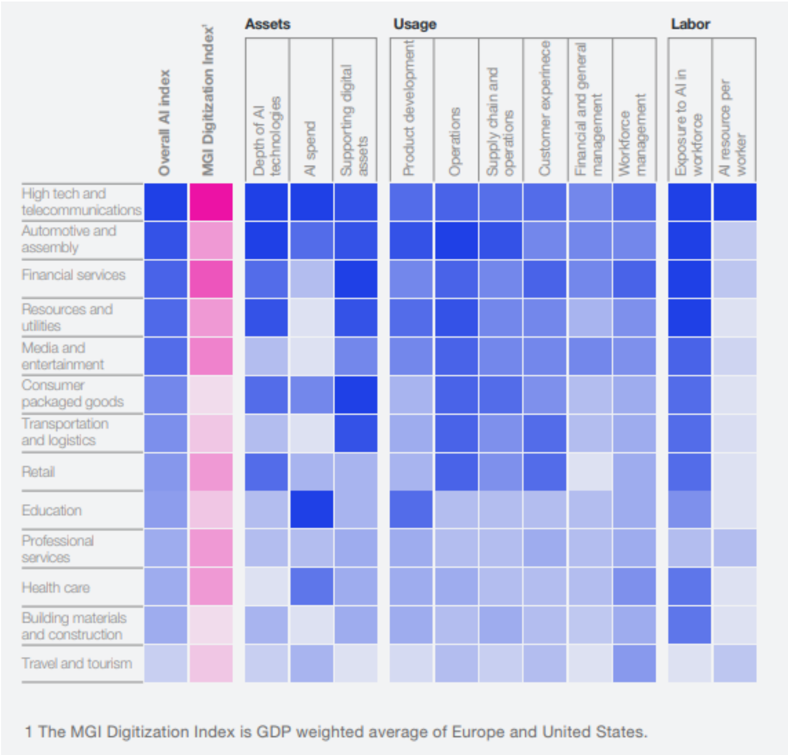


Figure 1: AI adoption when compared with overall digitization (McKinsey Global Institute)[1]

1.1 What is Data Science?

Data Science has a very wide spectrum, is multidisciplinary, and generally refers to the approach of using tools with a set of skills, techniques and knowledge, solving analytically complex problems with data at its core. It is in simple terms how raw data is turned into valuable information, and potentially offering actionable insights to make smarter business decisions.

> *"Data Science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data."*[2]

Data Science can be leveraged to perform descriptive, diagnostic, predictive and prescriptive analysis[3], to respectively obtain insights and foresights on "What happened?", "Why it happened?", "What could happen?", and lastly recommended actions on "What should be done?". The skills required for Data Science are divided across the following three areas[4] which are all equally important:

---

[1] McKinsey Global Institute AI adoption and use survey: Analytics comes of age 2018

[2] Leskovec, Jure; Rajaraman, Anand; Ullman, Jeffrey David. Mining of Massive Datasets (PDF) (Preprint of 3rd ed.). Cambridge University Press. p. 1.

[3] Descriptive, predictive, prescriptive: Transforming asset and facilities management with analytics IBM 2017

[4] The Data Science Venn Diagram, Dr. Conway 2013

- **Computer science, in particular software development:** This is a data scientist's ability to code and develop algorithms.

- **Mathematics, in particular statistics:** These skills should go hand in hand with coding skills to allow a data scientist to understand the data and be able to incorporate and tweak mathematical functionalities within the code. A Data Scientist must be well acquainted with mathematics and statistics to understand the underlying process and logic of complex models applied to a large volume of data. There is a common misconception that Data Science is all about statistics. While statistics is important, it is not the only category of mathematics utilized.

- **Specific business domain expertise:** Understanding the background, context and the domain specific nuances is key in order to develop useful data algorithms. As in any other discipline, to solve a problem, one needs to understand its boundaries, rules, policies and the impact of its dimensions.

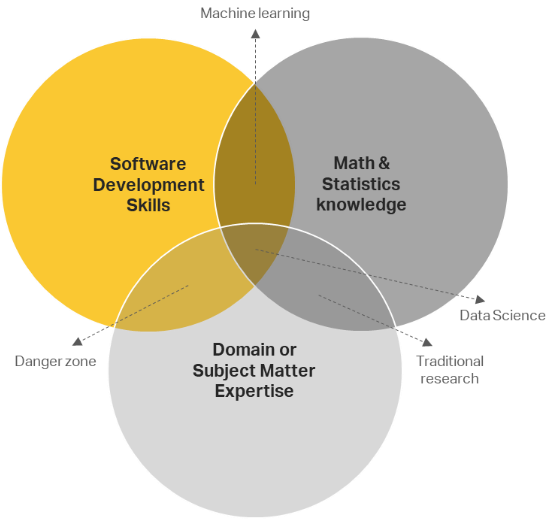


Figure 2: Data Science venn diagram (4)

1.2 Positioning Data Science within the Organization

Data Science plays a key role in many organizations, converting raw data into business insights, supporting leaders in taking data-driven decisions, but also driving product development and business intelligence automation among other use cases. There are many approaches that we have come across and identified. Below are some of the organizational design flavors:

- Data Science as a central support function to all business units.
- Data Science as an integral part of some or all business units.
- Data Science in an IT, Business-Centric or Integrated structure.[5]

While there is no right or wrong on selecting either or any combination of the above approaches to build and setup Data Science capabilities and teams, it is recommended to ensure the team is strongly in touch with the business while avoiding the creation of duplicate capabilities and skills.

[5] How to Structure a Data Science Team: Key Models and Roles to Consider, AltexSoft 2018

2 Aviation areas of potential

Data Science has a very broad impact on all industries and aviation is no exception. There are many areas that could benefit from the use of Data Science capabilities. Our Data Science research and development work spans all aspects of aviation, with many work streams in partnership with airlines, airports, and other travel and technology providers. The following high-level areas of application have been identified as having a high potential of generating significant value and benefits.

Figure 3: Data science areas of application in Aviation

2.1 Data with Dynamic Offers

2.1.1 Tectonic shift in sales and distribution

With the increased adoption of the New Distribution Capability[6] and ONE Order[7] standards, the travel industry is undergoing a major shift, where airlines are constructing offers and orders themselves, with aggregators combining different content to allow for search and comparison of different offers while allowing airlines to customize the offer, and be able to apply changes to the available products and services with a high degree of agility and speed.

Data processing and capabilities to embed and utilize analytics in sales and distribution are critical in order to become strong retailers and be able to offer a unique value proposition moving away from commoditization. Data messaging standards have been undergoing transformation to facilitate the needs of today in terms of the language and interfaces between systems.

2.1.2 Dynamic Offers challenge Revenue Management practices

As a perishable asset, each individual airline's seats require meticulous inventory control. The airline industry pioneered Pricing and Revenue Management techniques as early as the 1980s. Historically, airlines' offer creation has been a disjointed process with static information being filed, to some extent out of the airline's control. It improved over time with the airline's ability to control availability and the fare buckets better. However, airlines still lack the holistic view of improving the complete offer.

With transformation made possible by the New Distribution Capability (NDC) and ONE Order, airlines no longer have to file fares and rules to create and distribute offers in response to shopping requests. Recently, IATA stimulated this debate in the IATA Dynamic Offer Creation white paper[8] by challenging the industry that effective revenue management cannot occur until airlines stop separately managing inventory and price, and start employing the concept of Dynamic Offer. As described in the paper, a Dynamic Offer is provided in real time, in response to a one-time request and can include any types of merchandized products beyond the flight itinerary. This industry trend will bring airlines' sales & distribution and pricing & revenue management teams closer together.

[6] New Distribution Capability (NDC) is a travel industry-supported program, launched by IATA for the development and market adoption of a new, XML-based data transmission standard (IATA NDC webpage).
[7] ONE Order is an industry-led initiative, aiming to modernize the order management processes (IATA ONE Order webpage).
[8] Airline Industry Retailing White Paper: Dynamic Offer Creation, IATA 2018

2.1.3 Leverage data with future Dynamic Offer Engines

Great potential can be unlocked when Dynamic Offer Engines have the ability to provide direct responses to service requests with Offers, especially with data available at the shopping stage. By storing all the offers, especially those not accepted, Dynamic Offer Engines can now better understand the demand based on the requestors' contextual information. This should improve conversion rates by providing the optimal product to the customer.

As the airlines move from an air service provider to retailers of travel, the importance of offering the customer the most relevant offer is increasing. With the advanced capabilities to understand the customer willingness to pay performed by Dynamic Offer Engines, airlines have greater ability to modify dynamically the product attributes of the offer to suit the needs of the consumer. Airlines have now the opportunity to perform total offer management including a holistic view of customer lifetime value, optimization of ancillary products and control over interline or third party content.

In conclusion, by being in control of their NDC offers and by moving away from separate inventory and pricing processes with dynamic offers, airlines have the ability to use vastly more data sources as inputs, and to explore new processing capabilities including Artificial Intelligence[9]. The latter transformation comes with great opportunities to leverage data to gain insights about historical buying trends, correlate with other data sources, e.g. city pair booking behaviors and demographics and move toward personalization by having a comprehensive insights about the customers and be able to act on those.

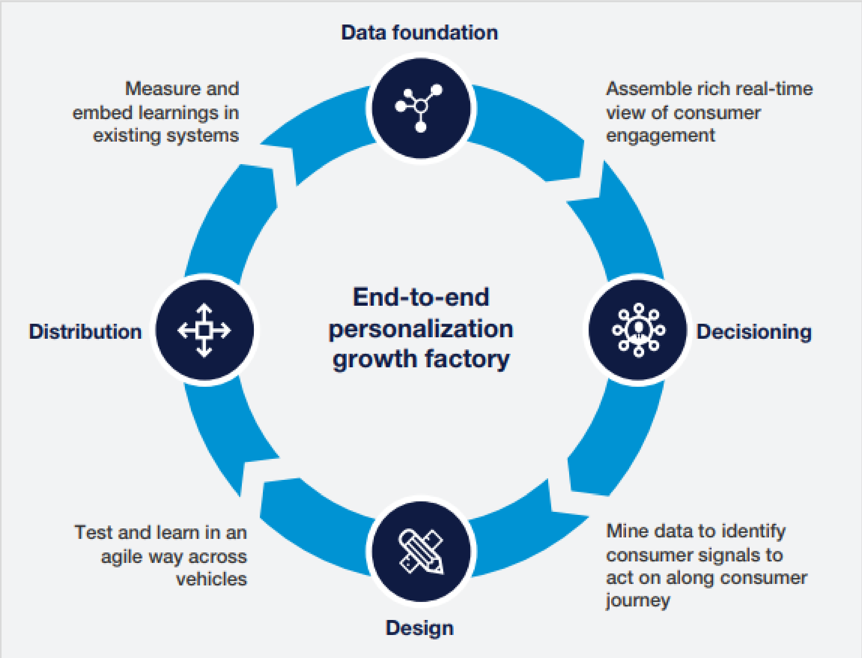


Figure 5: Building deeper 1-to-1 relationships with consumers at scale (McKinsey Global Institute)

2.2 Industry data sharing platforms

Data sharing has proven to be an effective way to make strong improvements in leveraging not only a user's own data, but also a wider rich data set provided by other participants and shared across an industry data sharing platform. These kinds of platforms require strong governance and a neutral custodian to ensure data provided by participants is protected, e.g. through anonymization, aggregation, and averaging of data fields as per the governing framework of the platform.
To support effective decision making, efficient planning and operations, airlines need to be and remain in control of their data. This will enable them to unlock the benefits provided by new technology and digital capabilities, ensuring procompetitive usage which promotes new entry and innovation.

Data sharing platforms have been proven to support the industry ambitions described above. Airlines have already started such a platform for their ticketing data through the IATA Direct Data Solutions (DDS) platform[10], where airlines share data to a common platform and reap the benefits of a rich source of data that has strong data management, governance and integrity, thereby meeting the needs and requirements of its data contributors and users. Building on the success of DDS, the industry

[9] Airline Industry Retailing White Paper: Dynamic Offer Creation, IATA 2018
[10] IATA Direct Data Solutions (DDS): An industry-sponsored program that provides the travel industry with timely, accurate and cost-effective access to global airline market data.

could extend the concept to other types of data, with data-sharing platforms to aggregate aviation industry data from airlines and other sources, e.g. airports, agents, and regulators. This could enable actionable insights for the participating users, based on industry governed and managed data.

Among the many data initiatives in IATA, an Aviation Data Hub is a new industry data sharing concept that is being investigated, to allow "data sharing, without sharing data". The Aviation Data Hub would consume raw data contributed by its users, in return, giving back to users, predictive and prescriptive analytics services through APIs[11], without exposing any of the raw data that has been contributed by users, rather using the contributed data to train the analytics algorithm.

At a high level the concept would involve:

- Individual users contributing data according to a pre-agreed data model;
- All contributed data sets to be combined;
- The combined data set is used to develop analytics algorithms (e.g. Machine Learning);
- The data and the algorithm itself are not accessible to users;
- Users can use an API to get predictive analytics that is leveraging the algorithm.

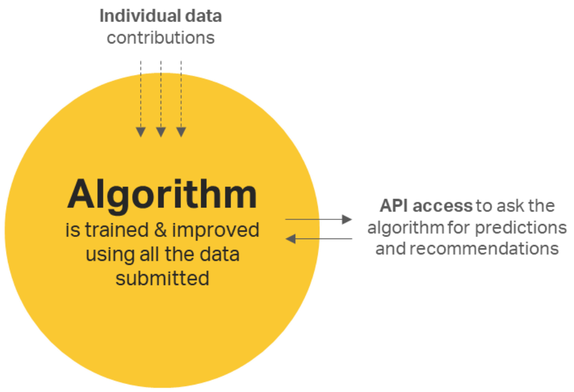


Figure 4: Aviation Data Hub concept

2.3 What is the role of data in network planning?

Airline network planning involves determining where to fly, when, and with which aircraft. To maximize efficiency, airlines manage the dual objectives of maximizing network revenue opportunities while minimizing network cost. Often these objectives are difficult to meet at the same time. As airlines work to maximize the return on capital-intensive aircraft and ground assets, networks become increasingly complex. This complexity increases the number of variables impacting cost and revenue. These variables must be successfully balanced when building and executing a network plan. New data sources offer significant opportunities in many network planning processes. New data capabilities such as artificial intelligence also offer new mechanisms to optimize networks in the design phase, and to maximize efficiency during operations.

2.3.1 Moving beyond the rear-view mirror: Forecasting demand and estimating market share

In an increasingly deregulated global aviation market, airlines have more freedom than ever in how they deploy capacity to meet demand. One of the most fundamental processes of network planning is demand forecasting. This involves estimating the number of passengers willing to travel from one place to another, and how this demand will materialize across all available transport options.

Historically, network planning activity involved using historical travel patterns to predict future demand. This approach has an inherent limitation in that historical travel patterns are constrained by existing travel options. For brand new nonstop markets, airlines often resort to using existing markets with similar size or characteristics as a starting point, often with limited success.

In today's world the volume of data that is available on consumer behavior in virtually every market segmentation is extensive, and under-utilized. Next generation models to predict unconstrained demand will use many variables to estimate potential demand, from abandoned searches on online travel agents and metasearch sites, to the perceived value of hotel and activity options gathered from travel review websites. Social media chatter will provide indications of potential demand in the leisure sector, just as professional networking site data and data around recruitment and procurement activities may provide predictors of emerging corporate markets. As social media and the digital economy transform the way people behave and interact, so too are market responses to new direct flights and increased capacity changing. Artificial intelligence will become a key enabler (Figure 5) to continuous calibration of demand models, ultimately predicting which new markets will survive, and how and when social media amplification will provide a halo effect to the announcement and operations of new services. These are all factors which airlines today typically observe after the fact, and which are not integrated into decision making models.

[11] API: Application programming interface

Where industries will put practical AI to work

Ranking of AI impact by its potential to free up time, enhance quality, and enhance personalization

Ranking	Industry	High-potential use cases
1	Healthcare	• Supporting diagnosis by detecting variations in patient data • Early identification of potential pandemics • Imaging diagnostics
1	Automotive	• Autonomous fleets for ride sharing • Semi-autonomous features such as driver assist • Engine monitoring and predictive, autonomous maintenance
3	Financial services	• Personalized financial planning • Fraud detection and anti-money laundering • Automation of customer operations
4	Transportation and logistics	• Autonomous trucking and delivery • Traffic control and reduced congestion • Enhanced security
5	Technology, media, and telecommunications	• Media archiving, search, and recommendations • Customized content creation • Personalized marketing and advertising
6	Retail and consumer	• Personalized design and production • Anticipating customer demand • Inventory and delivery management
7	Energy	• Smart metering • More efficient grid operation and storage • Predictive infrastructure maintenance
8	Manufacturing	• Enhanced monitoring and auto-correction of processes • Supply chain and production optimization • On-demand production

Figure 5: AI use cases across different industries (PwC)[12]

Airlines have also historically used relatively blunt tools to forecast organic market growth. These typically involve a reference to an indicator of economic activity such as gross domestic product, with some adjustment for market stimulation. Better access to a wider range of data sets will also allow such forecasts to be more accurate, and for possible reactions to external shocks to be better understood. For example, indicators such as the 'baked bean index' tracking baked bean sales in supermarkets are often reported to have provided a better indication of short-term drops in disposable income than other more established macro-economic indicators. Such data could be invaluable in predicting demand changes on leisure routes. Similar indicators may be valuable in other segments. Access to larger data sets and the use of Data Science techniques allow thousands of such indicators to be integrated into decision modelling and calibrated continuously to determine how useful they may be in predicting market changes.

Another fundamental process in network planning is the modelling of choice options, and the estimation of market share. This typically involves mapping out every available means of travel, e.g. every possible connecting itinerary offered by competitors in the market, and then attempting to quantify how every individual traveler may value each alternative. The limitation of this approach has typically centered around a lack of data. In traditional 'Quality of Service' (QSI) models, itinerary attributes focus on schedule, number of stops, aircraft type and price. Customer characteristics such as gender, age and income are also typically ignored as data is not available.

With better access to anonymized data from multiple sources, there is now almost no limit to the number of seemingly trivial factors of quality which can, and do, subtly change perceived value for every different type of customer. This could include for example social media chatter influencing reputation, affinity with co-branded credit cards, previous experiences with the brand or product, travel to the airport, experience in the specific terminal, and the availability of car sharing on arrival. Airlines that pursue new data sources and new processing capacity to constantly evolve and re-calibrate models will see improvement in market share modelling, leading to meeting customer demand more accurately.

2.3.2 The end of scheduling seasons?

Airlines typically build schedules around two scheduling seasons, i.e. northern hemisphere summer season, and northern hemisphere winter season. In many ways this approach of a relatively fixed published schedule predates computer technology and harkens to a time when schedules were published in printed guides. There are of course many operational aspects to air transport that require careful coordination of intended departures, such as the coordination of airport capacity, and planning around resource requirements. In reality, many airlines respond to changes in demand and operational considerations after a schedule is published. This often leads to many changes to scheduled departures occurring up until a flight departs. Because

[12] 2018 AI predictions: 8 insights to shape business strategy, PwC 2018

these changes occur within an eco-system which has not been designed around agility, this often causes inconvenience: to passengers who are frequently advised their flight has been rescheduled, to travel agents who are managing a passenger's hotel and ground arrangements, to interline partners who may unexpectedly lose or gain passengers. As new data processing capacity allows more seamless exchanges of data and more real-time resource planning activities, it is conceivable that the nature of published schedules will change, and that something more dynamic than seasonal schedules will become the practice.

2.3.3 Fleet Assignment, and aircraft routing: closing the feedback loops

Decisions about fleet often occur as part of strategic assessments of market entry and exit, and capacity changes. There is also a more tactical component to fleet assignment, as airlines attempt to optimize the assignment of different fleet types onto different scheduled departures. Building routings of actual aircraft across the network (into specified 'lines of flying') then occur to maximize revenue and minimize cost.

These processes both involve considering many competing variables. Changing passenger demand on different departures across different times of day is balanced against operational constraints of aircraft performance and efficiency, and ground resource availability. Sophisticated airlines typically use advanced models that consider as many variables as possible to solve fleet assignment and aircraft routing problems. Variables include such things as anticipated passenger loads, crew limitations and maintenance requirements. Many airlines approach operational planning in a linear way. Fleet assignment processes drive a published schedule, the schedule drives ground and air crew planning and rostering, and rostering drives costs associated with wages and trip expenses such as hotels, meals and allowances. There is often some limited feedback in these linear models, with several schedule iterations attempting to drive a reduction in crew overnights, for example. There is significant opportunity in bringing more data and better processing capability into these models. The future of operational planning will involve predictive and intelligent consideration of all anticipated operational cost implementations being more cohesively considered at the time of network planning, schedule creation and aircraft routing, and more real-time adjustment of lines of flying to respond to actual and predicted changes.

The need for efficiency and asset utilization is balanced against the drive towards on-time performance (OTP), which becomes more critical as reporting and disclosure requirements increase. The use of more data to predict and address OTP issues (from likely ramp delays, to the profile of passengers travelling, and likely traffic delays at the city or departure) could improve these metrics and reduce cost.

There are also opportunities to use Data Science and predictive analytics to influence the constraints in the model and create a more compact feedback loop. Airlines struggle today with closing the loop between changes on the revenue or cost side, and actions taken on the network. Typically, while revenue trends may be monitored in relatively real-time within revenue management systems, the interaction between revenue and cost is not tracked outside of route profitability processes. As a financial function, this often aligns with financial reporting periods which may be calendar months, and financial reporting timeframes which may see data produced for a reporting period several days or weeks after the end of that period. Just as new data processing capability could see a huge increase in the number of inputs into network planning models, thereby improving timeliness and accuracy, so too could route profitability processes be improved. Smart contracts and distributed ledger platforms bring the opportunity to close the loop between airlines and suppliers, bringing procure-to-pay processes into real time, and eliminating manual reconciliation and reporting within finance functions. This would see cost data being another real-time feed into network planning models and allow real time route profitability. Artificial intelligence could also be used to better calibrate cost allocation models driving route profitability, to ensure network cost and revenue impacts are more accurately modelled when looking at individual network changes.

2.3.4 What does the future hold?

As airlines and solution providers build capability in the area of Data Science and artificial intelligence, it is important to consider network optimization challenges as a valuable use case. New data sources are increasingly available from a variety of sources that may be useful in estimating demand, shaping customer choice models, and better optimizing schedules, fleet assignment and aircraft routing. As data processing capacity improves within the industry, continuous testing of new data sources and data processes against existing processes becomes possible. New data sources can calibrate existing models and may also begin to replace them as the environment changes.

2.4 Irregular Operations: The business challenge

An irregular operation occurs whenever an unplanned event causes a flight not to depart or arrive as it was scheduled, which then interrupts the travel plans of passengers. Airlines try to manage irregular operations to minimize customer inconvenience. As with many airline commercial activities, this involves a delicate balance, and the assessment of a vast number of variables in any individual situation.

2.4.1 The crystal ball – what if we knew when things would go wrong?

New data sources and data processing capabilities could provide airlines and other stakeholders within the ecosystem, with better intelligence to predict when and how disruptions might occur. Historical data on delays and cancellations assessed with reference to airports, gate assignments, and specific aircraft could be further calibrated with new data sources such as specific passenger profiles, total airport capacity and social media chatter. Using wider data sources may also improve forecasts around potential industrial disputes and weather incidents. Such intelligence could then become an input into processes such as scheduling, aircraft assignment and resource planning to better manage the impacts of disruptions.

2.4.2 Seamless recovery – balancing customer and cost

When things do go wrong, it is important that a recovery solution is formulated for every passenger who is impacted. In many airlines, these recoveries are built manually. Even where automation is used, re-accommodation often occurs on the basis of relatively simple robotics that attempt to find available capacity for each passenger. These robotics often use static business rules that don't consider all cost items, and most importantly they don't consider or anticipate customer value beyond broad-brush aspects such as frequent flier program tier, or cabin class. In many cases, existing automation does not take into account opportunities that are available to mitigate the impact of disruptions, e.g. by making tactical operational changes such as cancellations, aircraft swaps, gate allocation swaps, or retiming other scheduled flights.

In all scenarios, every possible recovery solution involves a complex and highly connected set of variables around available network capacity, aircraft and crew limitations, available airport capacity, and the costs associated with hotel accommodation, meals and passenger compensation. As with many aspects of operational planning, many airlines take a linear approach to these considerations, designing a recovery for passengers and then considering impacts on downstream processes. Better use of all operational data and real time cost analysis is critical for ensuring optimal decisions are made. Artificial intelligence may be valuable in learning of the outcomes of different recovery options. It is conceivable that systems could consider large data sets involving factors such as cost implications, competitor reactions, news media and customer experiences shared on social media. These systems could learn how to manage recoveries more effectively through predictive and prescriptive analytics.

2.4.3 Bringing the customer to the center of recovery

Better use of data allows all parameters impacting cost and customer preference to be integrated into recovery planning in real time. The customer preference aspect is perhaps the most important. Better use of customer data can allow customers preferences to be better considered in a re-accommodation. For example, a customer who has just experienced a disruption on their most recent journey with the airline may be a candidate for special handling. A passenger known to be travelling for business may prefer a night in a hotel after a late arrival, providing that their connecting flight still gets them to their destination for the start of the work day the next day. A leisure customer may prefer a one-way rental car hire to complete a journey rather than an overnight stay – an outcome which may also be more cost effective for the airline.

Many of these attributes can be better predicted through integration of more customer data. What is most effective, however is when these considerations are actively integrated into the planning for each re-accommodation. Today, the extent to which airlines involve the customer in re-accommodation is often limited to push-notifications when a recovery has been arranged. Obtaining customer input on their willingness to accept compensation in exchange for an inferior experience is typically limited to overboarding situations at a busy gate, where cash offers may be made for volunteers who are happy to take a later flight. As more and more customer interactions occur directly within mobile devices, customers can be more actively involved in service recovery.

For example, it could be imagined that when a flight departs late leading to known misconnections downstream, passengers could be invited to interact with the airline either through the onboard entertainment system or using their own devices. These passengers could indicate their preference for an immediate re-accommodation, or their willingness to break their journey or accept a different itinerary – perhaps in exchange for frequent flier points or miles, for an upgrade in cabin of service on a later flight, or even cash compensation. Where customer preference becomes an active input into the recovery planning, there is a far higher probability that customers will have a more positive overall experience. Active engagement with customers, in particular in the cabin environment, also eliminates resource requirements for ground staff, and eliminates customers waiting for information or re-accommodating options at the airport or with call centers. Finally, assessing customer preference more directly also provides data which can be better used when planning for and managing disruptions in the future.

Brand damage and negative press are also an increasingly prevalent risk for airlines, in an age of social media. Using social media chatter as an input into disruption recovery allows airlines to identify issues where a proactive response is required, but also to identify customer or customer segments where negative experience may involve a higher risk of social media amplification.

2.4.4 Network economics – using data together

In mass disruption events, many carriers may be impacted at the same time. This may involve weather disruption, where a whole airport, or even several airports are closed, or industrial disputes impacting air navigation service providers. In these situations, there are significant opportunities for airlines to better collaborate with data sharing to more effectively match disrupted passengers with available capacity, and to support better automation. There is also opportunity to use network economics to improve the sharing of information to predict and respond to disruption. In Europe and Unites States this occurs to a limited extent with the Collaborative Decision-Making process (CDM) between EUROCONTROL and European carriers, and the Federal Aviation Administration with carriers in the United States. CDM could be extended to more markets and expanded in scope to allow more data sharing where this adds value to the entire industry.

Data Science and Artificial Intelligence offer significant opportunities for the way irregular operations are managed. New data sources are increasingly available to better anticipate customer preferences for recovery options, and to better incorporate all available solutions and their associated costs into decision-making processes. Above all, better integration of the customer into these models, and more agile information exchange with passengers leveraging mobile devices and inflight connectivity is critical to managing customer experience, and reputational risk.

2.5 Aircraft Health Monitoring

Big data is said to have become the lifeblood of aircraft maintenance decisions. There is no questioning of the essential role played by data in the design and manufacturing of an aircraft. The basis of all aviation engineering solutions governing an aircraft type birth is data and, moreover so for new generation aircraft, it's big data. Endless hours of design, simulation and testing of individual parts and systems, up to the complete aircraft, are churned on ground and in flight by computers under close engineering expertise.

Thus it should come as no surprise that data is becoming part of the after-birth destiny of the aircraft: the post "design and production" decades of expected operational life in the airlines' fleet. Aircraft are made to fly and, while doing so, to provide a sustainable business for the airlines operating them and a great experience to their passengers. To achieve that, airlines must address the aircraft necessary maintenance, repair and overhaul (MRO) required for preserving the aircraft airworthiness throughout its operational life. The optimal answers to what to do and when to do it are heavily dependent on the individual operation logged by the given tail number. For everything to happen optimally, you must have a smart aircraft to tell you what and when, and you must be a wise listener when the aircraft talks to you. There is great opportunity to process all these available data sources and build algorithms for prediction and recommendations to support aircraft operation.

Health monitoring is a rather common place concept for contemporary human beings. We're used to monitoring and tracking our blood pressure, temperature, weight, cholesterol, and other vital signs. Monitoring the health of an aircraft is similar; we just have different metrics to track: flight hours (for getting a sense of aircraft's "used age"), flight cycles (for knowing the demands put on the aircraft structure by aircraft landings and airframe pressurization cycles), engine cycles (i.e. engine start – takeoff power use – engine shutdown), temperatures, pressures, fluid levels, vibrations, wear (e.g. of tire treads), free-play (e.g. in the attachment of the actuating rod to a flight control surface), voltage – frequency – current (e.g. of the output from an electrical generator) etc.

2.5.1 What is AHM?

Aircraft Health Monitoring (AHM) is a set of activities enabling the timely collection and analysis of aircraft systems' functional data to optimize the aircraft maintenance, ensuring a maximum aircraft technical availability for a safe and efficient flight operation. AHM should be understood as an end-to-end chain of activities where hardware and software empower expert analysis and decision-making (possibly in a Human – Artificial Intelligence combination) to occur in real- or almost real-time. This requires a sum of aircraft capabilities, to be designed and qualified by the OEM[13] for the aircraft type, and airline specific processes and procedures to operationalize such capabilities. Both the on-board and on-ground segments must come together to enable AHM.

2.5.2 The AHM Current Circumstances

There were gradual advances in AHM over time; they reside with areas like aircraft jet engine operation, where the regulation governing airline ETOPS (extended twin engine operations) flights, is requiring an Engine Condition Monitoring (ECM) Programme, or helicopter operations based on Health and Usage Monitoring Systems (HUMS).

---

[13] An original equipment manufacturer (OEM) is a company that produces parts and equipment that may be marketed by another manufacturer.

The industry standards are evolving towards establishing a robust level of AHM enabled Condition Based Maintenance (CBM) which could become dominant in the realm of aircraft maintenance programs. The efforts of standard-setting bodies like SAE Aerospace Council and Maintenance Programs Industry Group (MPIG) are steps in the right direction but much still needs to be accomplished.

This is where a mature Data Science implementation for aircraft systems modelling fueled with big data is beginning to enable AHM's major benefit: accurate prognostics of maintenance need. OEMs are the primary source of the robust aircraft systems modelling. OEMs own the design and production details of the aircraft. Airlines, however, know the operational behavior of the aircraft and its components. Entities which develop sensors, write algorithms, and employ state-of-the-art technologies are essential in data generation, capture, data transmission from the aircraft as well as relevant analysis to be used for various maintenance activities.

The steps forward are always conditioned by Regulators granting "maintenance credits" to the use of AHM: this is the condition enabling the use towards compliance with the aircraft Instructions for Continuing Airworthiness (ICAs).

2.5.3 The "Must Have" for Answering YES to the AHM Implementation Question

While AHM is a potential game changer for the airline operation, the AHM implementation is definitely not a single stakeholder journey: you need the Operator, the OEM, the Maintenance provider and the Regulator to converge for the AHM implementation to happen. Of course, data scientists, IT specialists and experienced business experts need open and continuous communication channels to achieve maximum levels of efficiency.

Although the Cost – Benefit analysis is the airline prerequisite of proceeding with the AHM implementation business case, we will consider that a "for AHM" decision has already been made and we will mention hereby a few considerations governing the technical and regulatory feasibility which must be catered to in order to implement AHM:

- Certificated aircraft capability (with hardware and software assurance) for collection of technical and operational performance data/information ;
- Secure data transmission with data fidelity and quality assurance;
- Data analysis using robust Data Science with accurate modeling of monitored system/aircraft and relevant expertise to generate actionable information;
- Maintenance alerting and resolution assurance based on the actionable information;
- Processes and procedures with time-bound outcomes for an AHM uninterrupted flow with fall-back assurance mechanism to address exceptions with limited or nil operational impact;
- Qualification and training of personnel relevant to their roles and responsibilities in the AHM loop.

All of the above require various levels of regulator's vetting. The airline internalization or the allocation of different parts to airline external entities (be it the OEM or a Service Provider) is to be decided by each operator on a case-by-case basis considering its own capabilities, resources and business model.

2.6 In Flight Data Driven Turbulence Mitigation

Turbulence is an irregular motion of the air that ranges from a few annoying bumps to severe throws of an airplane during flight. Turbulence is associated with for example: fronts, wind shear, and thunderstorms. It is the most unpredictable weather phenomenon affecting flight operations. It is a leading cause of injuries to passengers and flight crews. In addition to safety, turbulence impacts many aspects of the airline business, including customer service, aircraft maintenance, legal, insurance, sales and marketing. As such, damages from turbulence incidents amount to millions of dollars in direct and indirect costs each year. Furthermore, current research[14] indicates that turbulence may be significantly increasing in severity and frequency in the coming decades due to climate change.

Even though most airlines across the globe continue to use traditional tools, e.g. PIREPS[15] and forecasts such as SIGMETs[16] and Significant Weather Charts, to mitigate the impact of turbulence on their operation, these subjective tools do not provide enough detail regarding the location and severity of turbulence. More efficient approaches are being sought by airlines to manage turbulence and meet both safety and efficiency demands of operations globally. Recent technical advancements now enable aircraft to accurately measure the turbulence state of the atmosphere around an aircraft in flight. The data generated by these aircraft is invaluable from a situational awareness perspective. Given the limitations associated with traditional tools, many airlines are adopting a data-driven turbulence mitigation strategy by using real-time, objective aircraft-sensed turbulence reports and forecasts using this data. This new approach is proving to reduce turbulence injuries and costs.

To maximize the benefits of the afore-mentioned reports and build a global turbulence data coverage, IATA has developed a turbulence sharing platform[17] to consolidate, standardize and enable access to worldwide real-time objective turbulence data collected from multiple airlines around the globe. The primary purpose of the Turbulence Aware system (currently at pilot phase) is to provide airline pilots and airline operation center personnel with real-time, very detailed turbulence awareness and support a global industry shift towards data-driven turbulence mitigation.

2.6.1 Industry-wide Objective Turbulence Data Sharing

The collection and reporting of automated, objective in situ turbulence data is exponentially increasing. Several airlines currently employ turbulence reporting products. While helpful, often they do not look at turbulence beyond their own airline. As flight schedules vary throughout the day this leaves pilots and dispatchers with gaps in coverage making it more difficult to plan and fly around areas of turbulence.

There is strength in numbers and with many airlines contributing their turbulence data to a shared database the industry will be able to see beyond their own routes, viewing objective turbulence reports from other air carriers. Additionally, we will see that there is a "critical-mass" that can be obtained by eliminating turbulent blind spots when we have many airlines contributing data to a global system[18].

2.7 Data in Baggage Operations

Baggage operations is critical from an operational efficiency perspective, and it concerns a fundamental part of the passenger journey, delivering the passenger's bag back to them on arrival at their airport of destination. In order to achieve this aim, airlines and airports have worked together through various stages of baggage handling development. Baggage handling development is not a linear process though, and thus not all airports have the same level of system complexity and layers of controls. The basic process however remains the same:

- Generate a tag with a 10-digit license plate that identifies the bag;
- Attach the tag to the bag;
- Send the bag into a system of conveyors that delivers the bag to a security screening area;
- Ensure the bag is safe to travel;
- Send the bag through more conveyors to an area where all the bags for a particular flight are grouped;
- Load the bag into a cart or container and deliver to the aircraft;
- Throughout this process the bag is identified by the barcode on the label attached to it.

There is then a second level of handling, more or less the same in every airport, where the bag is taken from one flight and placed on another. This involves the following steps:

- Unload the bags from the arriving flight;
- Move the bags either directly to the next flight or:
    - Place the bag onto a series of conveyors;
    - Identify the bag from its license plate;
    - If needed, ensure that the bag is safe to travel;

14 Increased light, moderate, and severe clear-air turbulence in response to climate change, Paul D. Williams, 2017

15 A pilot report or PIREP is a report of actual weather conditions encountered by an aircraft in flight. Wikipedia

16 SIGMET, or Significant Meteorological Information AIM 7-1-6, is a weather advisory that contains meteorological information concerning the safety of all aircraft. Wikipedia

17 The IATA Turbulence Aware Platform, IATA Press release no. 73

18 It is understood that more than one global system could co-exist and that airlines remain free to contribute their own data to any desired service provider.

- Send the bag to a place where all the bags for a particular flight are grouped;
- Load the bag onto a cart or into a container and deliver to the aircraft.

Either of these sets of processes present just two challenges. Identifying the bag and knowing what to do with it. Both of these challenges are data problems, not baggage problems.

2.7.1 The impact of data on baggage operations

Having looked at the key challenges there are a number of areas where sophisticated data processing can play a role in improving baggage operations.

Optical Artificial Intelligence

As mentioned earlier, the barcode used in baggage is not a robust identification technology. It would be possible to build image-based identification that can interpret a barcode even when part of the code is obscured through creases and folds in the paper. This should be a relatively simple task, as barcodes are made from straight lines.

There is also a further possibility that is currently being investigated by airlines, and that is the use of optical artificial intelligence on the bag, rather than on any label. This allows a bag to be uniquely identified without tagging, and trials have uniquely identified up to 4 million bags. While airline IT experts talk about having a bag without a label, the baggage operations experts are rightly concerned with this approach, as tag-less bags are the ones that are most difficult to return to the passenger. However, the technology has several other interesting applications. Optical systems can be used to identify the dimensions of a bag, which is very useful for cabin baggage monitoring. The optical systems can also identify damage to a bag and the type and color of the bag, which makes the processing of baggage claims easier.

Machine learning for baggage probabilities

A mathematician would say that for any bag there is a probability that it will be loaded correctly. Working out these probabilities is complex, but machine learning and artificial intelligence offer a way to manage the data necessary to calculate the probabilities without actually knowing all the parameters of the equation. By giving every bag a probability of being loaded then the bags that fall below a certain threshold can be identified and proactively handled to have them make their connections. With sufficient data, this decision can be moved towards the start of the baggage journey.

A further use of probabilities can be made if they are applied to the passenger. All airlines build a certain no-show rate into their booking profiles, but knowing which passengers are less likely to make a flight is quite difficult. Today we have proactive offloads when a passenger fails to make a specific milestone on their progress through the airport. The main milestone is security, where a failure to pass a certain number of minutes before the flight results in offloading the passenger and the bag. There are other factors that can influence a passenger making their flight, including the length of time spent at the airport, access to lounges, the shopping offers and other attractions in the airport and the size of the airport. Discovering the probability that a passenger will travel would allow bags belonging to passengers with a low probability of travel to be loaded at the end of boarding, thus reducing the time spent searching for an offloading baggage.

Route analysis and profitability

One of the goals in the IATA Baggage Vision is to unlink the passenger and baggage journeys. The aim is to allow a passenger greater freedom for how their bag travels, for instance by allowing collection from home and delivery to destination. If we can do that, then the question arises why the passenger and baggage journey have to be linked any longer. Breaking this link would allow airlines to send baggage by the most efficient route to the final destination, even if this were on a different carrier's aircraft. This would allow routes to be selected based on the probability of loss, the value of cargo space, excess capacity or delivery window availability, rather than having the bag follow the passenger route, which is chosen on either a cost or time basis.

These are just three examples of how baggage data can be used to transform the baggage business, and while each requires much further investigation and effort before they are realized, they are likely contenders for implementation in the next few years. Bringing in data scientists to the world of baggage creates a rich new playground for ideas, and should be encouraged.

3 Conclusion

The journey of data and Data Science proliferation starts with gaining access to internal and external data sources. In some cases the data could be generated, but there is no mechanism in place to capture it. In other cases data is being generated but not stored, either as a conscious decision or owing to technical restrictions or challenges.

Raising awareness on the value of data to ensure appropriate management and governance is critical. The journey continues with ensuring the input is reliable, i.e. measuring and enhancing the quality of data, identifying and addressing discrepancies with other data sources.

While digital and data frontiers in other industries such as high tech and telecommunication and the automotive industries have come to the conclusion that adopting open data sharing cultures has significant benefits, there are still challenges related to data sharing frameworks, standards, and governance. Aviation is a surprisingly harmonized industry when it comes to data standards and data sharing platforms. There are currently more than 600 data exchange standards across the aviation industry, with initiatives[19] aiming at creating even more alignment across existing standards and newly generated ones.

Respected thought leaders and futurists such as Yuval Noah Harari[20] have put the ability to process data as the main capability of the companies of the future, above any other capability. Success of companies will depend strongly on their ability to consume data, build appropriate analytics models and algorithms, and finally integrate and embed these in their processes.

Many airlines and other players across the aviation value chain have a long way to go to become data processing experts. As the aviation industry enters this new era, the "golden age" of data, those that rethink their company DNA and include data and digital in the core of their strategy and follow through towards integration in business process, will have a significant competitive advantage.

Figure 6: 5 key aspects in proliferation of data and Data Science

19 IATA Airline Industry Data Model

20 Yuval Noah Harari is an Israeli historian and a professor in the Department of History at the Hebrew University of Jerusalem. He is the author of the international bestsellers Sapiens: A Brief History of Humankind (2014), Homo Deus: A Brief History of Tomorrow (2016), and 21 Lessons for the 21st Century (2018). His writings examine free will, consciousness, and intelligence.

4 Data Science Aviation Case Studies

There are many activities across and airlines and the wider aviation to leverage data. This section is a compilation of case studies provided by individual airlines.

4.1 Future of Data Science at Air France-KLM

Strengthened by several successful projects, such as demand forecast models or ancillary recommendations, the Data Science at Air France KLM is taking a growing importance with applications in all domains. Internal pioneer in the implementation of machine learning models, the customer data management is designing a new way of working, capitalizing on the creation of the 360° view of the customer. Analyzing customer data as a sequence of events, sticking to the customer multi-channel journey, will help to be always more customer centric. This purpose requires to break silos and enhance collaborative work between data scientists, analysts and business experts.

The overarching aim of customer journey analytics is to be able to propose the best action at the right time for each customer. This very complex use-case might be taken step by step compelling ancillary recommenders and various new predictive models. The main steps are claim prediction, commercial campaigns targeting and pro-active information targeting, especially in case of disruption.

More than predictive, the Data Science that Air France KLM is building is intended to be prescriptive, improve robustness of ROI measurement, enable its automation, and propose recommendations based on machine learning that should help the test and learn approach for new products and services.

In parallel with the improvements done on customer personalization, many efforts are now dedicated to operational excellence. Preventing disruptions is key to offering the best experience to our customers. Chaining effects of delays, delay forecasts, flight time and rolling time forecast with root cause analysis should help to smooth the customer journey and reduce costs. Regarding other operational questions, failure prediction for maintenance is a hot Data Science topic for Air France-KLM.

The employee journey is also fruitful for new use-cases. Absenteeism root cause analysis and prediction and improved mobility proposals might be the next Data Science challenges for human resources.

Case study prepared and written by:

Arnaud Capois
Advanced analytics - Customer Data Management, Air France-KLM

Charles Girard
Customer Data Officer, Air France-KLM

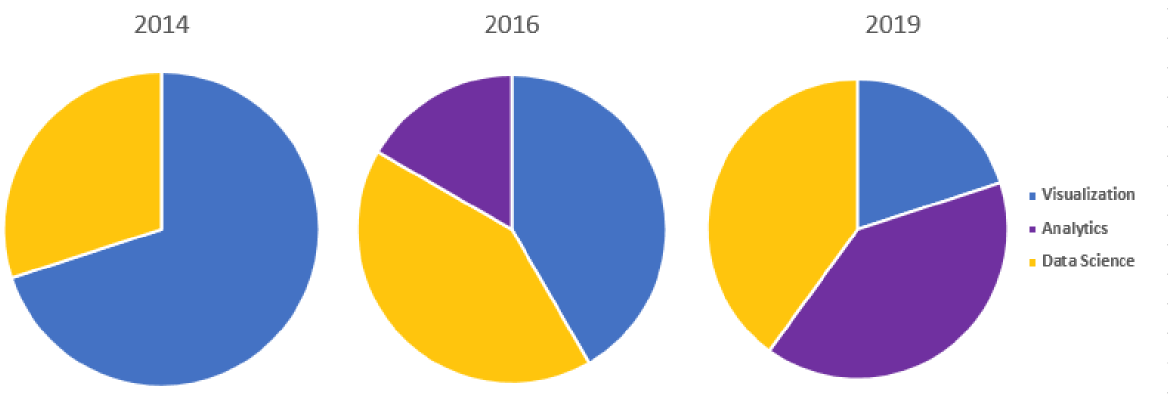
4.2 Data Science's Role at United Airlines

The airline industry generates a tremendous amount of data, ranging from ticket sales to maintenance logs. Deriving insights from data is part of what helps airlines like United become more efficient and dependable. For this reason, United embarked on a journey to ingrain Data Science into the enterprise. As with all journeys, it has had its stumbles and its successes. Five years in, United has established a centralized Data Science and Analytics team, whose evolution we present here.

In 2014, United was shifting attention from managing the many complexities of the merger with Continental Airlines to creating a better customer experience. United long had an Enterprise Data Warehouse, which allowed business units to store and analyze their respective data. United also had a well-established Continuous Improvement & Enterprise Optimization team: a team with major initiatives like network planning, revenue management forecasting and operational optimization. However, many teams remained without analytical capabilities and inter-departmental knowledge of data.

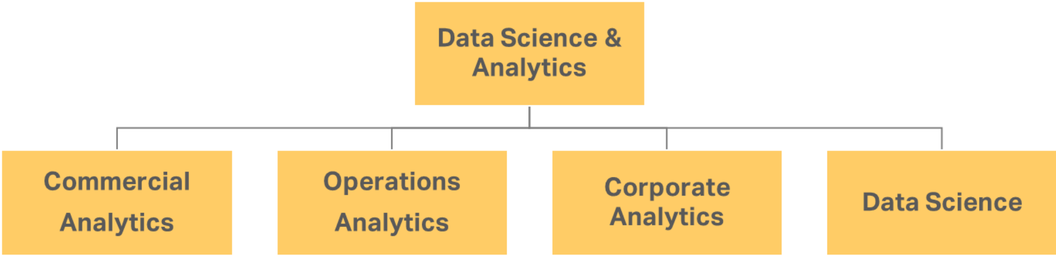
United created a new group to solve this problem. It was then known as Customer Data and Optimization and the goals were clear: show the value of and democratize access to customer data. In those early days, the Customer Data and Optimization team focused on developing capabilities with initiatives ranging from selecting an enterprise data visualization tool to developing relationships with new partners.

The team's first major analytical success was the "Customer Signal Hub". Leveraging the already-robust Enterprise Data Warehouse, the team first created customer-level attributes that could be easily reused across myriad projects. Customer Signal Hub allowed United to change the way it dealt with customer data. No longer did business units have to spend their time writing custom queries that could lead to inconsistent definitions of key customer metrics. Instead, there was now a robust, well-recognized source of truth. Customer Signal Hub had enabled business units to focus their efforts on the strengths: understanding their business, developing strategic initiatives and executing on new projects.



During the following years, the Customer Data and Optimization team continued to build on the success of Customer Signal Hub. The team developed machine learning models to predict the likelihood of customer behaviors, such as the propensity to buy an Economy Plus seat. Again, by building models for the enterprise to consume, the team was able to shift some departments' behavior from use of complex business rules to a simple metric whose performance could be easily monitored over time. These models were so successful that they were (and remain) embedded into a number of automated programs at United, including United's real time system for ancillary products.

By 2017, the team's accomplishments were spreading across the organization. Partnering with Operations teams, the team launched a variety of applications to get customer information into the hands of frontline coworkers who regularly interact with United's customers. Flight Attendants and Customer Service Representatives were able to more quickly understand their customers and their recent experiences with United. Steps like this have helped United to operate more efficiently and improve customer experiences. The team then launched its first machine learning model in Operations, which was embedded in some of the applications used by Customer Service Representatives to board flights. Of course, in deploying a predictive model, the team learned some important lessons, like the importance of explaining to users how to use and interpret the tool. Lessons like these are now core to the team's strategy of model deployment.



Fast forward to 2019, and the team's analytical and Data Science skills have deepened considerably. Renamed Data Science & Analytics, the team works across the organization, with team members dedicated to partnering with specific business units. This organization has allowed the Data Science & Analytics team to have better business context and effectively deliver meaningful analyses or tools. Additionally, a separate Data Engineering group creates better support and enables data consumption. These groups have helped to educate other groups on what Data Science is and identify potential projects for the dedicated Data Science team. For example, the Operations Analytics team, working closely with the Technical Operations groups created both a dashboard and alerting system to identify certain situations that could indicate an impending issue with an airplane part. And, Operations Analytics members have helped the Data Science team to use Natural Language Processing algorithms to automatically categorize mechanics' logs into particular ATA chapters. Data Science & Analytics and Data Engineering are now working with almost every group throughout United, meeting other departments at their technical level.
In the years to come, United sees tremendous appetite and opportunity for Data Science to deliver improvements both within United and across the industry. With Data Science moving so quickly, it is difficult to say exactly what will be the next biggest change that Data Science can bring to the airline industry. However, United is already seeing big steps in video, image and analytics; along with the introduction of biometrics to certain steps in the customer journey.

Case study prepared and written by:

Michael Shores
Director Data Science, United Airlines

4.3 Delta Air Lines' Success Story with Turbulence Data

At Delta Air Lines the biggest driver for these efforts is safety. Firstly, objective and timely reports provided to the flight crews on a tablet device give a much more objective and comprehensive view of the atmosphere. Secondly, a new turbulence model which updates every 15 minutes called the GTG NowCast developed by NCAR has EDR inputs to improve the accuracy of the model. The provision of both observations and an improved model has been labeled a game changer by many pilots who use the observations to validate the model. As the volume and geographic distribution of EDR inputs increases, the model will become even more accurate.

Injuries and financial liability avoided due to use of real-time turbulence data
Based on a PIREP, ATC relayed to a Delta aircraft that smooth ride conditions were ahead. The Captain chose to rely on his Flight Weather Viewer tool utilizing real-time, continually updated objective turbulence data and decided to secure the cabin. The aircraft subsequently flew through moderate and severe turbulence. Some loose objects were tossed about during the encounter, however no injuries to the passengers or cabin crew occurred.

Case study prepared and written by:

Delta Air Lines