Journal of Revenue and Pricing Management (2023) 22:171–180
https://doi.org/10.1057/s41272-022-00404-8

PRACTICE ARTICLE

Personalization @ scale in airlines: combining the power of rich customer data, experiential learning, and revenue management

Alberto Guerrini[1] · Gabriele Ferri[1] · Stefano Rocchi[1] · Marcelo Cirelli[1] · Vicente Piña[1] · Antoine Grieszmann[1]

Received: 28 June 2022 / Accepted: 26 October 2022 / Published online: 11 January 2023

Abstract

Recently, several macro trends have converged to provide airlines new opportunities for one-to-one digital customer engagement and personalization. Airlines have more types and volumes of data available than ever before: shopping-behavior data, data providing context on booking decisions, social media data enriching the information available on travel trends, and more. All of these can play a critical role in defining the right offers and setting the right prices for each shopping request. A plethora of advanced AI and ML techniques have become available on open-source platforms, letting players generate actionable customer insights and leverage vast amounts of existing data. New distribution technology is being deployed to allow airlines to implement real-time retailing capabilities. Consumers have been trained by the likes of Amazon, Netflix, Alibaba, and Starbucks to expect products and services tailored to their individual needs along with superior and engaging content. This paper presents different approaches to price-product personalization that have been tested in airline cases globally. It also explores how the concept of experiential learning is nicely suited to tackling scenarios in which the purchaser is well-identified as well as cases in which not much is known about the visitor except the context of the shopping session.

Keywords Revenue management · Experiential learning · Price personalization · Machine learning · Customer segmentation

Brand loyalty, customer lifetime value, and the power of personalization

Personalized service—it's what every customer wants and perhaps even demands today. Instead, online consumers are overloaded with generic offers of all types, most of which are sent out blindly and are not specific to their wants or needs. Not surprisingly, these shots in the dark have low conversion rates (Gibson 2019).

A lack of information about individual shoppers is responsible for much of this. Personalization, on the other hand, allows companies to present customers with a short list of search results, product recommendations, or service options tailored to their personal needs and preferences so they can more easily determine what is best for them. Retail giant Amazon, for example, sells more than 562 million products. Its machine learning (ML) algorithms proactively display product suggestions personalized to each customer based on everything Amazon knows about their preferences.

Individual consumer expectations are compelling airlines to formulate targeted proposals using the right message in the right channel at the right time, from travel inception to post-travel, at all meaningful touchpoints:

- Inspiration: Providing attractive options to spark desire.
- Shopping: Offering attractive discounts or services.
- Pre-travel: Helping customers with anything they need for their trip.

✉ Stefano Rocchi
Rocchi.Stefano@bcg.com

Alberto Guerrini
guerrini.alberto@bcg.com

Gabriele Ferri
Ferri.Gabriele@bcg.com

Marcelo Cirelli
Cirelli.Marcelo@bcg.com

Vicente Piña
Pina.Vicente@bcg.com

Antoine Grieszmann
Grieszmann.Antoine@bcg.com

[1] Boston Consulting Group, Milan, Italy

- On the ground: Facilitating relevant services before boarding.
- Post-travel: Asking customers to share their positive experiences, listening to feedback, and reengaging.

Personalization is a vehicle to improve the customer experience, allowing individuals to receive more relevant and timely content, improving engagement, and yielding a financial impact by picking up customers' signals (i.e. purchase intent or lack thereof).

A complex problem airlines should tackle when designing personalization solutions is how to harmonize tactical and strategic use cases—striking a balance between the immediate benefit (short-term impact) and the one they can expect from a loyal long-term relationship (customer lifetime value). The loyalty program and its available data are therefore one of the relevant building blocks of personalization.

A comprehensive view of the advantages that personalization can bring to airlines is provided in Fig. 1.

As the figure illustrates, personalization builds on the relationship between customers and airlines to increase sales. Optimizing customer lifetime value (increasing customer loyalty) enhances the profit formula by accounting for how a discount today can translate into more purchases in the future. Reducing distribution costs by attracting more demand to direct channels, assumes personalization can more easily be deployed on channels owned by the airline. Pushing customer segmentation to the next level for revenue management helps because investments in personalization can favor synergies with other revenue-management objectives, such as total revenue management and continuous pricing. Finally, creating a customer data hub that can be used at any point along the customer journey (inspiration, shopping, pre-travel, onboard, post-travel) connects information and sustains initiatives at the individual customer level.

Tangible benefits achieved

From real case experience, we have proved that, by applying advanced analytics for personalization use cases, material customer satisfaction and revenue benefits can be achieved. One key example focused on offering personalized travel destinations with incentive prices. In this case, the airline sent (proactive) targeted offers via email promoting destinations selected according to an affinity model.

The email highlighted six destinations from 16 cities in Europe and North Africa. A control group (20% of the total) received an email offering random destinations. Two training approaches were employed: one with and one without a clustering of destinations (i.e. while training the model for Cologne, a trip to Dusseldorf, in the same cluster, was considered a positive), based on the city attributes (geographical traits as well as attributes like "historical," "nightlife," "culinary," etc.).

For each approach, three algorithms—logistic regression, random forest, and gradient boosting machine—were tested. Three KPIs were selected to assess the quality of training:

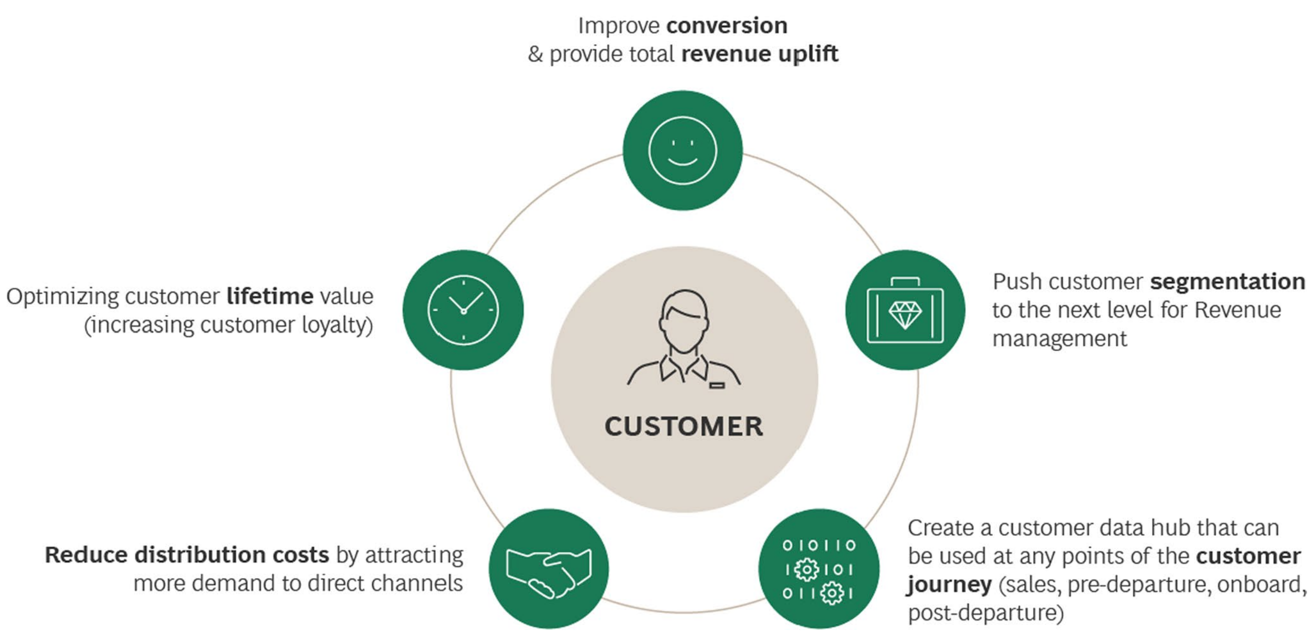

Fig. 1 Benefits of personalization

- Sensitivity = TP/P.
- Lift = Sensitivity of model/sensitivity of random prediction.
- Accuracy = (TP + TN)/(P + N).

TP = Truepositive, TN = True negative, P = Positive, N = Negative

Logistic regression and gradient boosting machine were the best overall models in terms of sensitivity and lift.

In a pilot lasting three months, we measured ~ 5 € incremental revenue per opened email compared to the generic (fixed destination for all) email and a hit rate (correct booking prediction) lift of ~ 2× over the control group receiving the six most popular destinations.

In another airline case, a personalized remarketing tool used advanced analytics and AI to automate offers. It was about sending (reactive) targeted offers to reengage customers who showed interest in a product in an online shopping session but didn't convert. More than 500,000 personalized offers were sent during the pilot. Several types of offers were tested (i.e. price discount, miles bonus, alternative flights). Production data sources used were CRM, loyalty, bookings, inventory, and website traffic. The upside was remarkable: + 15% uplift in website conversion (from 4 to 4.25%) for an estimated annual sales impact of up to $10 million.

Several criteria were used to enhance the likelihood of conversion:

- Attractiveness: Inclusion of conversion incentives (i.e., loyalty points).
- Customer engagement: Personalized header calling customer by name.
- Variety at scale: Over 2800 + ODs tested in just one batch.
- Speed: Acting on a hot lead ~ 24 h after the customers searched.
- Sense of urgency: Offers were valid for a limited period (48 h).
- Exclusivity: URL redirected to specific OD-date previously searched by customer.
- Flexibility: Use of dynamic fields to test multiple offers with the same template.

The ascendance of machine learning in personalization

ML is a branch of artificial intelligence (AI) focused on developing algorithms capable of improving the quality of their outcomes through experience by learning from a potentially broad range of data sources.

Over the past several years, much has been said about how ML can revolutionize the way airlines create offers, set prices, and interact with customers. However, our experience shows that very few players have benefited in tangible ways from the application of such models in their day-to-day business processes. The typical reasons for this are a lack of structured data available to implement these algorithms, challenges attracting and retaining talents with the required skillset, and low levels of confidence from business analysts and company leaders on giving up decision-making power to such algorithms.

We have always taken a pragmatic approach toward the application of ML in airline cases. While we strongly encourage clients to experiment with ML techniques and build the required capabilities, we believe that there are specific use cases where ML can add the most value, like forecasting demand, measuring the price elasticity of demand and willingness to pay, and offering personalized prices, products, and marketing campaigns.

When we observe how airlines currently implement personalized offers, the necessity for an ML-driven personalization approach becomes clear. Traditional methods and processes typically are highly manual, require subjective inputs from the responsible teams, and result in outcomes that frequently don't reach customers with relevant and truly meaningful offers, communications, and prices.

The legacy methods often lead to mass personalization, in which marketing and RM teams design and develop an offer and then look for the best-matching customers to receive it based on a set of criteria and arbitrary business rules. The question asked by airlines is "Who are the best customers for this offer we have just created?" This approach generally leads to a limited number of offers launched due to the effort required in the design phase and, most importantly, low customer engagement, conversion rates, and revenue performance. Since the content and messages aren't personalized at the individual customer level.

A true personalization approach shifts the paradigm to "What is the best offer for this particular customer at this particular touchpoint?". To implement such an approach successfully, three pillars must be in place:

- A *robust customer data model*, integrating all the available data sources that can support the generation of customer insights such as coupons/tickets, ancillaries, frequent flyers, and CRM.
- A *large number of potential offers*, moving from campaigns to "menus" for each customer, resulting in an explosion of product/offer varieties.
- *Human and technology capabilities* to make multiple contacts with individual customers each year, investing in system integration and automation and upskilling talent in both digital and commercial skills.

Given the large number of potential offers (thousands), airlines' vast customer bases (millions), and the higher frequency of contacts required (dozens each year), the possibilities are nearly countless. ML works extremely well in scenarios where millions of decisions must be made in short time frames and high accuracy is expected (Gautam et al. 2021). ML is able to identify patterns from big data and generate insights on customer behaviors and needs that would be nearly impossible to obtain through traditional methods. Most customer propensity models are built using ensemble methods (random forests and gradient boosting) or neural networks framed as a supervised learning problem. When no transaction data, or only restricted historical data, are available, the personalization initiative needs to be framed as a reinforcement learning problem to go through exploration–exploitation cycles on randomly generated offers (customer-product pairings).

Typically, the creation of ML algorithms starts with receiving training data as input. The training data set generally is relevant historical data used to fit the parameters of the specific model—for example, using optimization methods. This step of the process is called model fitting. Afterward, a new data set called test data (that the fitted model has not seen during the training step) is used to provide an unbiased validation of the predictive capabilities of the fitted model. This step is called model validation.

Often, customer profiles could be unevenly represented in the data set. The train/test split has to be designed to deal with this issue, sampling in clusters and applying oversampling and undersampling techniques. Once a model is validated and deemed acceptable for the target application, it is implemented in a real-life context. Eventually, the model may need to be retrained with more recent and updated data to continue performing properly. The frequency of retraining can vary depending on the application: daily, quarterly, or even yearly in some cases.

In other adjacent consumer sectors, personalization leaders such as Amazon, Netflix, and Starbucks have already incorporated ML-driven personalization in their core business processes. In recent years, we have seen an increasingly higher number of airlines investing in the required capabilities to implement ML-driven personalization, and some are already realizing benefits from it.

The right ticket and ancillary price at the right time for everyone

From the airline side, it has always been a challenge to tailor an offer on the basis of individual preferences (what Ben Vinod called "a segment of one" in his 2021a, b paper "The Age of Intelligent Retailing: Personalized Offers in Travel for a Segment of One"). This is due to:

- Incomplete customer data.
- Complexity of the personalization algorithms.
- Limit of the distribution capabilities.

Various approximations have been implemented, starting from reservation booking designator (RBD) segmentation and ticket conditions (fences) that were designed to prevent individuals from buying below their willing to pay (WTP) range and continuing with fare bundles that sweetened the offer by packaging ticket conditions and service items that fit in market segmentation as intended by the airlines. It appears evident that these past approaches were devised because the airlines had a limited knowledge of the consumer base.

Realistically, the use cases modeling the problem need not be as numerous as the purchasers. Moreover, every purchaser has a different need and behavior depending on the context.

Customer profiling and context determination are the building blocks of price personalization. In some scenarios, both elements are known and can be processed by personalization engines. In others, the client cannot be identified yet traveler preferences can be inferred through the purchasing context, e.g., an extra bag for a long stay or seat assignments for a couple.

To set up a solution for a personalized offer, three elements have to be managed efficiently: the identification of all use cases in the scope of the airline's interests, the organization/evolution of a rich and granular client database, and the implementation of a mix of models that are made of ML algorithms compounded by business rule engines. We will reflect on these three topics in the development of the paper.

A comprehensive view on use cases

As noted above, the use cases to model the personalization problem are provided by a high number of combinations. An analysis of the most relevant can be carried out by processing past data and selecting only the combinations that account for a non-negligible share of clients.

For each meaningful combination of customer profile and purchasing context, a proper offering must be devised. This approach would be biased toward customer centricity if the flight revenue condition and projections aren't modeled in the algorithms. Any price offering has to be judged against the flight's revenue target in order to avoid diluting the value from the personalization.

Figure 2 shows an example schema that links use cases to personalized solutions. As seen here, even though ML models make up the backbone of personalization, they need to be compounded by rule-based systems and heuristics to cover for problem complexity.

Traditional ML techniques can solve one or two problems *at a time*, not optimize a full customer journey with the millions of permutations available.

The datamart of personalization

Data is the lifeblood of personalization. One of the biggest challenges regarding data is that many systems that hold and process

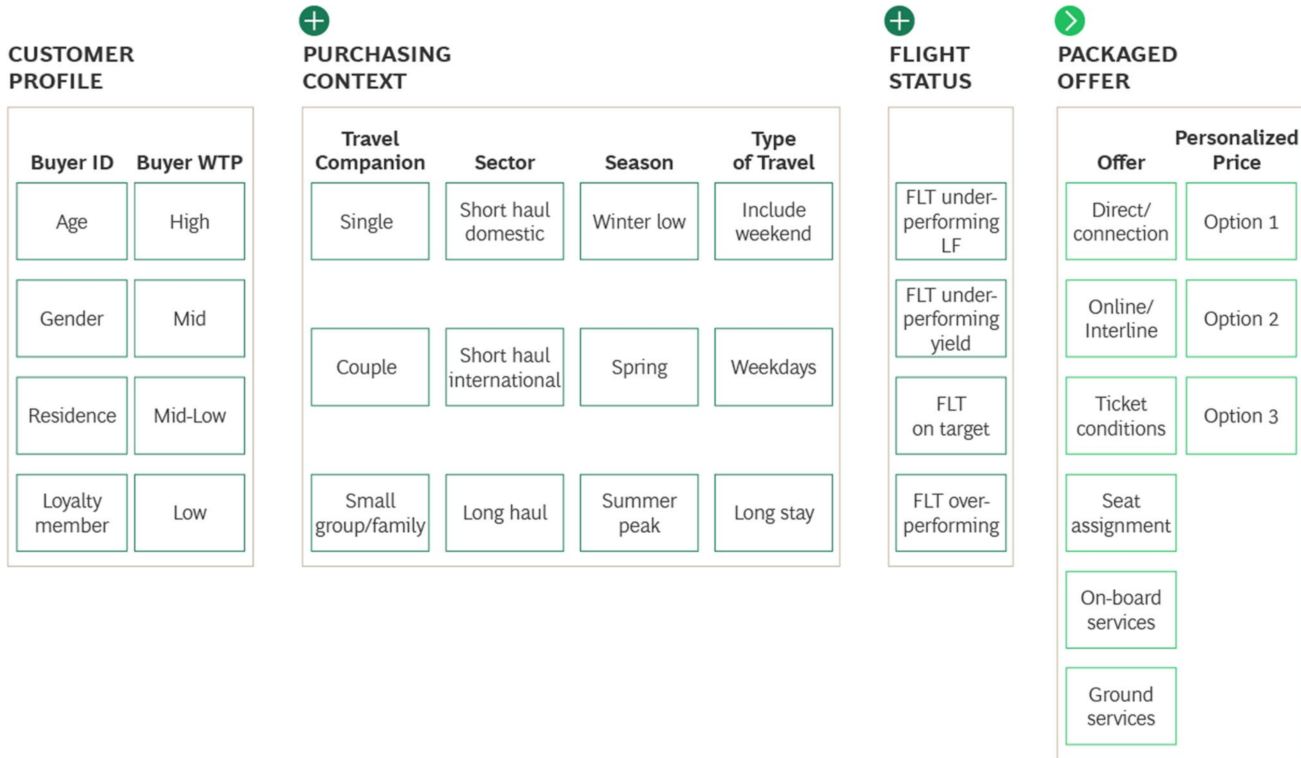

Fig. 2 Use cases building personalized solutions

vital data for personalization have been developed to fulfill specific functions. These systems have been created in the past as verticals and store data in transactional formats not favoring a holistic view of the customer and the customer journey.

From BCG experience, typical shortcomings found in data are shown in Fig. 3.

We know from experience that any personalization initiative, to be successful, must rely on a rich set of

DATA SOURCES	CORE DATA	TYPICAL SHORTCOMING VS. PERSONALIZATION
Website	App collects user data at various touchpoints (booking, check-in, changes)	System often is not designed to track individual customer behavior (e.g., type of searches and look-to-book ratio)
CRM	System is log-based. Logs are opened at a specific customer request, resolved, and closed	System doesn't store customer history of interactions along with their needs and preferences
Shopping data	Very rich transactional data containing booking history (PNRs) from reservation system	PNR record doesn't hold data connected to website utilization Data not processed regularly to feed customer data hub
Loyalty	Very rich data about loyalty program subscribers. Connected with PNR database and CRM	
Marketing cloud (e.g., social media)	Social media data & contact details to sustain marketing initiatives	This stand-alone data is not integrated with other sources and is also complex to process

Fig. 3 Data shortcomings

different data. Functional systems must be adapted to collect and store more data. A big pool of data has to be created with volumes not common to other airline services and functions. It is good practice not to connect the personalization models directly with the various data sources, but rather to create an intermediate customer data hub properly designed to support personalization. The customer data hub can sustain many different use cases, from precision marketing to remarketing to personalized pricing.

Figure 4 shows a real case example of a customer data hub. All the data sources available at this airline had been created from different functional needs, and the integration of a customer-centered data hub proved inefficient: Only 12% of distinct clients could be clearly identified and enriched with features on travel habits created from historical purchases (has traveled with kids, has traveled with a companion, type of repeat customer, never/occasionally/often bundles ancillaries, etc.).

The low matches between PNRs and client accounts suggest that airlines don't design their data infrastructure around the client but rather around the function for which each application was created (e.g., shopping, reservation, check-in, loyalty, CRM). A transformation toward a comprehensive IT infrastructure focused on the client will have to be enacted by airlines that aspire to a personalized approach.

For situations where customer information is limited, we use rule-based segmentation on PNR attributes. This approach allows the identification of specific customer groups or macro-segments, e.g., families with children or business travelers (see Fig. 5), and some of their preferences. The method consists of defining and combining different criteria with attributes that are easy to identify during the booking process, even with limited information about the client. This includes travel cabin, children, number of people traveling, day of the week of departure and return, and more. Once a macro-segment is identified, a price or offer relevant to them can be proposed.

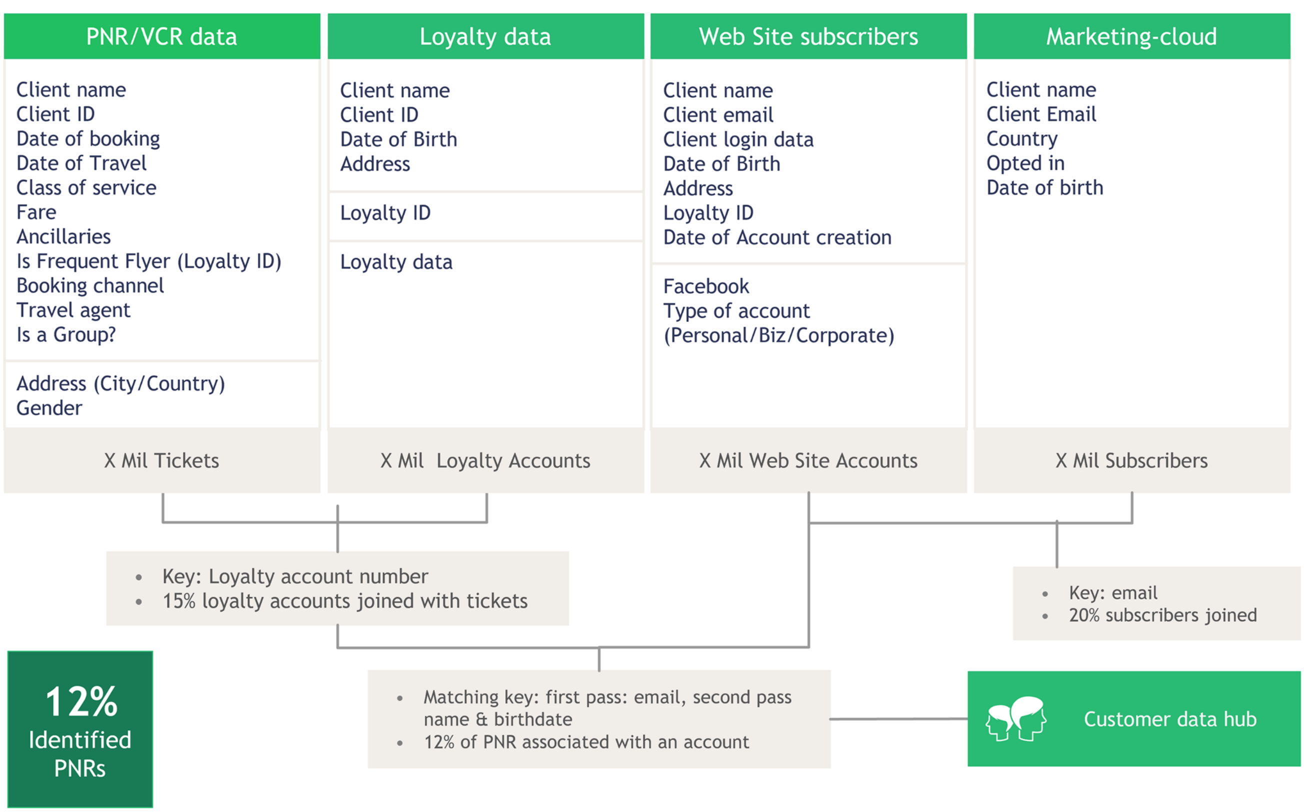

Fig. 4 An airline data hub

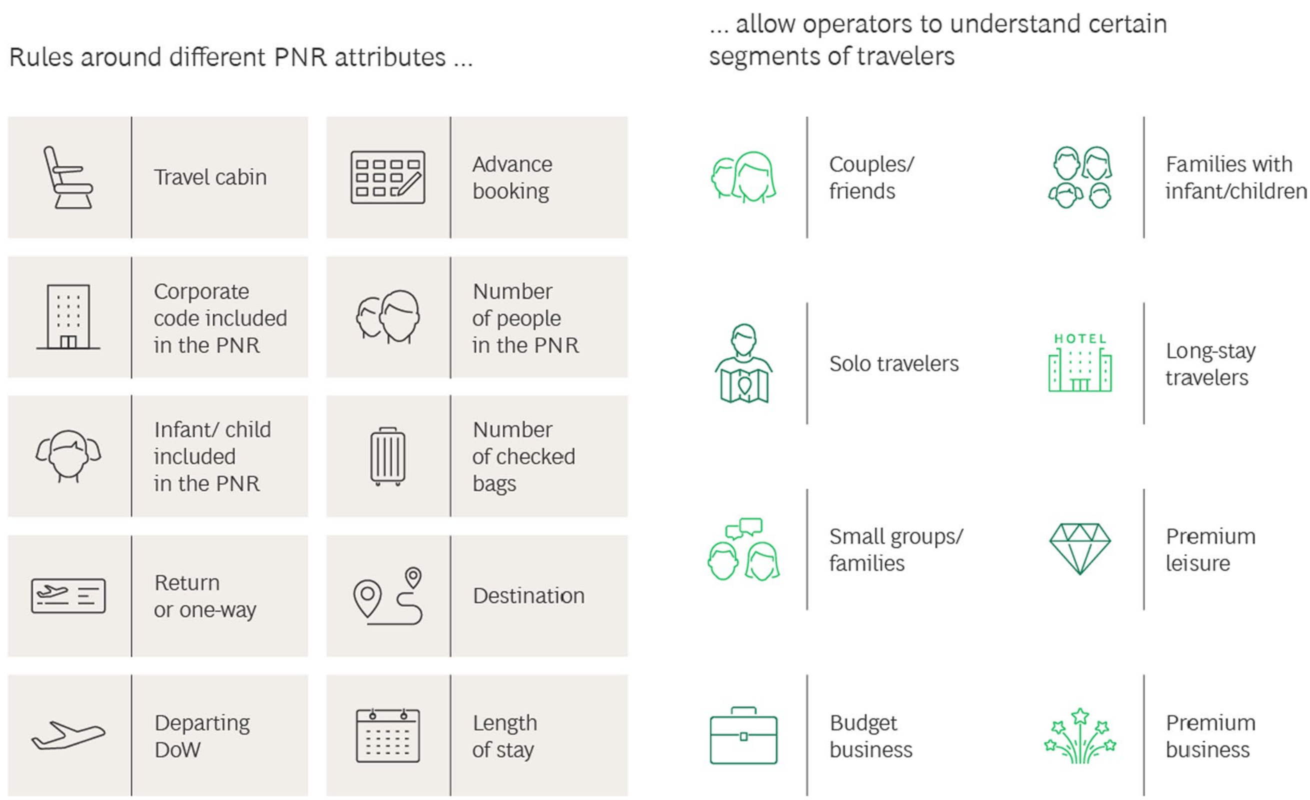

Fig. 5 Travelers' rule-based segmentation

Solutions for personalized pricing

Solving a personalized pricing problem, as hinted above, requires a customer description and segmentation that goes beyond the traditional airline segmenting methodologies. This is because it is assumed that the number of features to be used are in higher number and typology: The broader the set of data sources that can be connected for the same customer, the more accurate the profiling and offering will be.

A first step in segmenting travelers can be achieved by using queries and rule-based segmentation of different attributes of a PNR, such as travel cabin, number of travelers in the PNR, and infant/child included in the PNR. Airlines are able to tell for a portion of their travelers whether they are business travelers, small groups, families with children, and so on.

In terms of personalized offerings, we see three different scenarios to be resolved. The offline scenario is also known as precision marketing or remarketing, depending on whether the proposal sent to the client is *proactive* (the client is selected based on a promo initiative launched by the airline) or *reactive* (the promo offering is triggered by a previous shopping session from a client who didn't convert). The other two scenarios are defined as online: The personalized offer is created dynamically by the online shopping engine when the customer connects to search, either as a logged-in client (exactly identified) or as an unidentified visitor where the model can only rely on contextual elements to formulate an offering.

The model process flow describing how the three problems are tackled is depicted in Fig. 6.

The core of the approach stands in the feature enrichment models. These models can be structured differently based on the quality of available data (e.g., length of historical series or meaningful number of repeat purchases at customer and context level).

The first class of models are descriptive. Conceptually, they are set up to assess how many profiles can be identified as statistically meaningful in terms of customer descriptor combinations and to validate how many clusters can be identified within the ticket purchase context dimensions (e.g., season, time of departure, length of stay, including weekend) or travel descriptors.

The next step is to match customer descriptor clusters and travel descriptor clusters to make an actionable set on which to base the personalized offer. For example, a profile could be made of customers older than 40 who

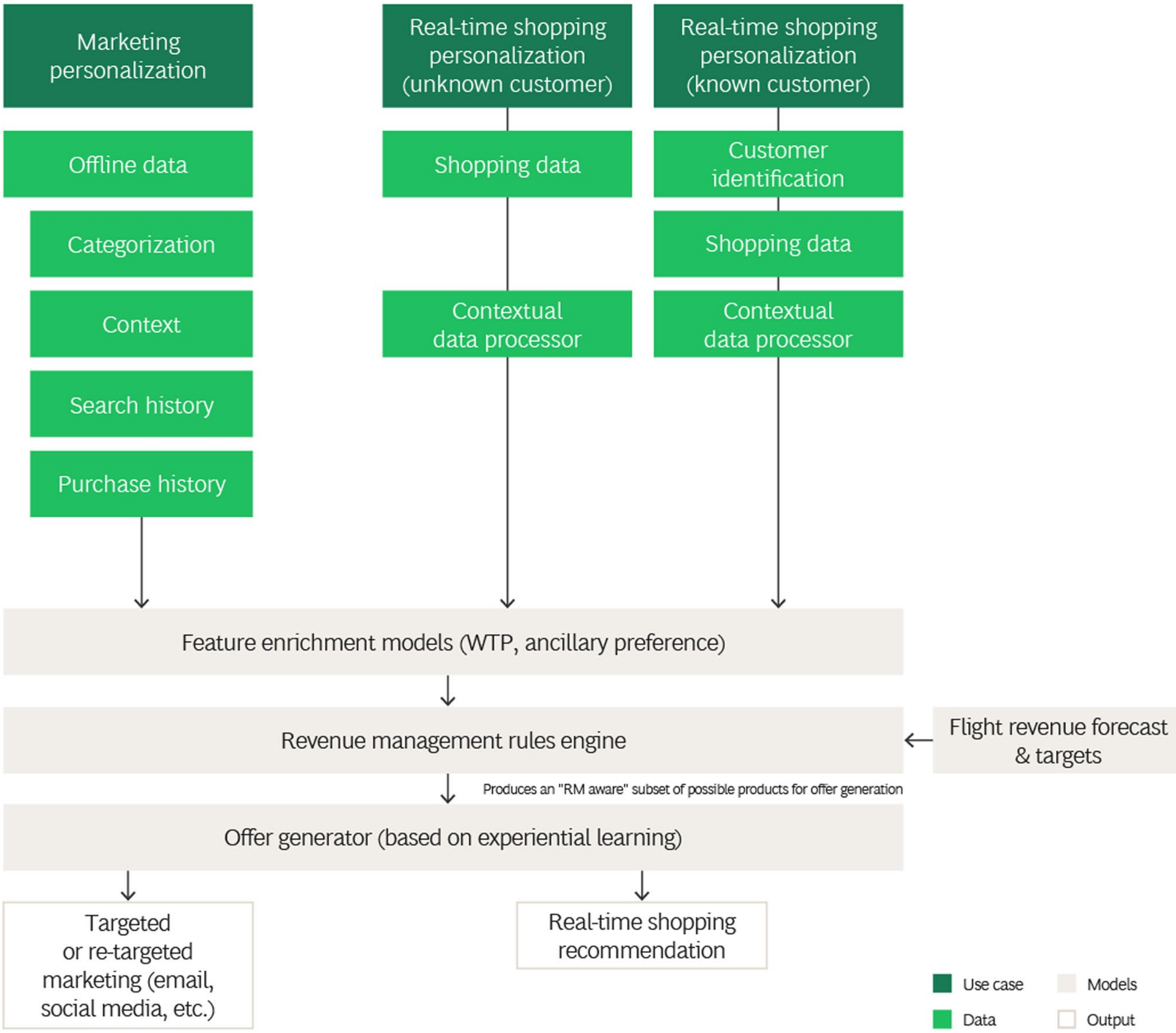

Fig. 6 Scenario process flow

are not loyalty members and have traveled multiple trips on the same route and never purchased an ancillary item. The travel profile can be round-trip/low-season/mid-fare tickets. This intersection represents a significant share of the total customer database.

Then the problem itself, assigning an ad-hoc price/service to the identified customer set, can be treated by a heuristic or rule-based system. In the case illustrated above, for instance, the customer type could be offered a fare discount only if they hold a top-fare family ticket, while an ancillary offering for a free or discounted flight shouldn't be considered.

The model described has the advantage of being pretty simple and clear enough to allow a good supervision of results. On the other hand, it may not always be applicable due to data limitation—a considerable part of the customer data couldn't be classified because some clusters are low-populated or the number of variables in the meaningful clusters is too low to propose meaningful personalized pricing.

A more powerful approach, enabling segments of one, can instead be based on an ML model (framed as a probabilistic classification problem) set to estimate the likelihood that a given customer would purchase a certain product in a specific price range in a particular context and/or that an ancillary item could be of interest.

These models are trained on historical data to estimate the customer WTP (i.e., the maximum price range the customer will be willing to pay for a specific product). They can also help simulate how the WTP varies by keeping customer features fixed and making product features vary (e.g., high/low season, day of week, leisure/business destination). Consider the following profile: Italian customer, age 30–39, corporate account. This profile has the highest probability to buy in the top price percentile range in October on the Rome to Venice route, but the probability to buy in the same price class in August is much less.

Another approach is used when historical data are not reliable and rich enough. For a new offer, the personalization model has no a priori information about customers' WTP and preferences. However, the customer can be identified and tracked for future purchases. In this scenario, the ML model will be trained based on a cold start experiential learning method.

One approach is to assume all new offers have the same uniform probability of conversion. In other words, they're selected at random. As we get data from the random offer trials, the model can learn those that have better response rates for a given customer set. Note that collecting response data is more important than having the best model. This means starting with a simple model should give faster results than waiting for a sophisticated model.

In our experience, it takes some time (four to five months) to take the model to a satisfactory level of predictability and stability (see Fig. 7).

Some methodologies balance exploration and exploitation during the testing/learning phase. As an example, Thompson sampling supports a model of the reward probabilities. In a simplified frame, the reward could be conversion or no conversion (i.e., offered price or ancillary item is purchased or not). When a random variable has only two possible outcomes, its probabilistic behavior can be described by the *binomial distribution*. In the cold start approach, we have no notion of the initial probability of conversion, so we can assume the probability distribution of profit for the offer is a flat line. Proceeding with the test, the parameters of the binomial increase and shape the distribution to be either large and low (low probability of conversion) or narrow and high (high probability of conversion).

Thompson sampling (Chapelle and Li 2011) helps select the offer with the highest returned profit, but it also accounts for frequency of testing for a specific offer. In this way, an offer that currently has a low estimated mean profit but has been tested fewer times than an offer with a higher estimated mean can return a larger sample value and therefore become the selected offer at this time step.

The last component of the architecture is the revenue management rules engine. It satisfies a fundamental necessity in the personalized engine: ensuring that all elements in the logic (such as customer lifetime value, WTP, preferences, and flight performance necessities) are orchestrated and any discounts are offered only if the cost-opportunity equation is well-balanced. As an example, we can imagine an identified client has a WTP below the current public price offered, is not a loyalty member, and has a low propensity toward ancillaries. In this case, a ticket discount will be offered only if the flight in point is currently tracking on the target yield curve but is underperforming on volumes.

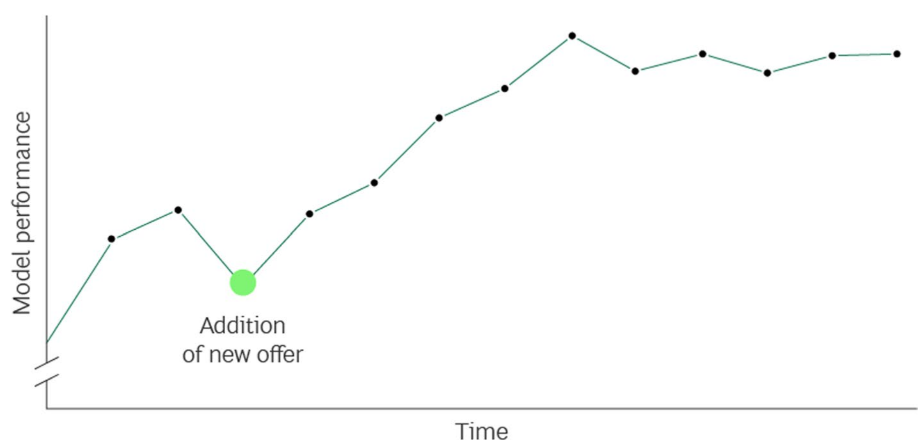

Fig. 7 Model performance over time

How should it evolve?

Having seen many airlines' setup, we think the journey to mature personalization is still long. Customer data, analyst capabilities, ML models, and distribution technology must advance in an organic fashion that we call a bionic approach (Guerrini et al. 2020). Models and technology must support more challenging problems, but human capabilities need to be evolved to be on top of more complex architecture to run.

References

Chapelle, Oliver and Lihong Li. 2011. *An Empirical Evaluation of Thompson Sampling*. NIPS.

Gautam, Nitin, Shriguru Nayak, and Sergey Shebalov. 2021. Machine learning approach to market behavior estimation with applications in revenue management. *Journal of Revenue and Pricing Management* 20 (3): 344–350.

Gibson, Rebecca. 2019. Respectful hyperpersonalization. *Compass* magazine. https://compassmag.3ds.com/respectful-hyperpersonalization/. Accessed 30 Sep 2022.

Guerrini, Alberto, Gabriele Ferri, Tom McCaleb, Matt Beckett, Rohit Ramesh, John Elder, and Carlo Pacifico. 2020. Bionic revenue management in travel and tourism. https://www.bcg.com/publications/2020/bionic-revenue-management-travel-tourism. Accessed 25 Sep 2022.

Vinod, Ben. 2021a. Artificial Intelligence in travel. *Journal of Revenue and Pricing Management* 20 (3): 368–375.

Vinod, Ben. 2021b. The age of intelligent retailing: personalized offers in travel for a segment of one. *Journal of Revenue and Pricing Management* 20 (4): 473–479.

Alberto Guerrini is a managing director and senior partner of Boston Consulting Group and leads the firm's travel and tourism sector in Italy, Greece, and Turkey.

Gabriele Ferri is a managing director and partner of Boston Consulting Group and a core member of the firm's travel and tourism practice.

Stefano Rocchi is an associate director for the Boston Consulting Group's travel and tourism sector and a core member of the firm's Center for Revenue Management in Travel.

Marcelo Cirelli is a project leader of Boston Consulting Group and a core member of the firm's Center for Revenue Management in Travel.

Vicente Piña is a project leader of Boston Consulting Group and a core member of the firm's Center for Revenue Management in Travel.

Antoine Grieszmann is a project leader of Boston Consulting Group and a core member of the firm's Center for Revenue Management in Travel.