Dynamic Pricing Strategy Using Reinforcement Learning for Indian Airlines

Rupal Tripathi, Dr. Amandeep Kaur
ABV IIITM GWALIOR

***Abstract*—This thesis investigates the use of deep reinforcement learning (RL) techniques to develop a dynamic pricing strategy specifically suited for the Indian airline industry. We use a Markov Decision Process (MDP) framework that incorporates customer patience, which is the willingness to postpone reservations in anticipation of cheaper fares, into the pricing model, taking inspiration from exploration-based models like those of Jo et al. (2024). In order to learn adaptive pricing policies, our RL agents, including Deep Q-Networks (DQN) and Bootstrapped DQN (BDQN), are trained on actual Indian airfare datasets. In addition to addressing the trade-off between fare volatility and customer satisfaction, these models generate more revenue than static pricing heuristics. The study also assesses how fare-setting mechanisms are affected by regulatory frameworks, such as those enforced by the Ministry of Civil Aviation and the Directorate General of Civil Aviation (DGCA). These rules establish the parameters within which dynamic pricing systems must work by guaranteeing openness and equity. The thesis, taken as a whole, provides a strong, data-driven approach to airline revenue management in India and practical advice for industry participants looking to adopt moral, effective and customer-focused dynamic pricing strategies.**

I. Introduction

The more the business environment gets competitive, dynamic pricing strategies settles itself in as it has emerged as a sophisticated tool for improving and maximizing on the revenues while serving the consumers optimally. Such strategies whereby the prices are continually changed based on factors like market demand and competition is one of the most common ones which is currently being practiced is the dynamic pricing strategy. A good example is the airline industry where items on offer and demand have constant elasticity and inventory is perishable. Since customers are not fully rational in their buying behavior, airlines have to make pricing decisions that not only would generate maximum revenues from ticket sales but also take full account of strategic customers who are likely to postpone their bookings until the right price is offered and myopic customers who are likely to book tickets as soon as the prices are reflected on the websites. Over the past few years, the players in the airline industry of India have felt the effects of growth mainly attributable to factors such as expansion in the economy, governmental support, and increase in the class of people considered as middle-class earners. But the corporate world has witnessed a relationship between the price and factors like market demand, seasons, operating cost especially the fuel price, consumer behavior, etc. in the airline's ticket price in India. Although prior literature has considered the topic of dynamic pricing in different contexts, this research will seek to extend its understanding, more potently, to the Indian context, exploring the factors that influence ticket prices in airlines, dynamic pricing methodologies employed by these airlines and the ways in which the behavior of consumers impacts purchasing decisions.

The airline industry in India has grown to unprecedented levels over the last ten years. Currently India is the globe's third largest domestic aviation market, due to growth in disposable income, increase in middle income earners and demand for air transportation. IndiGo, Air India, SpiceJet, Vistara, and GoFirst have tailored their fleets and network to better serve a steadily increasing domestic demand. Government's UDAN scheme targeting the 'any Indian citizen' to fly more often by easing access to aviation has further propped up regional connectivity offered by waiving subsidies to the airlines for the regions. However, the future of the airline companies in India appears to be threatened by various uncertainties; fluctuations in fuel prices, dominant competition and unsteady passenger traffic among them. Dynamic pricing comes in handy for these challenges since it allows the airlines to adapt fares on living factors such as demand, booking, and competitor's fares. Nevertheless, the Indian market is prospective but it demands particular attention to price sensitivity, regulation issues, and its buyers who can be both visionaries and impatient passengers.

Consumer behavior is therefore crucial when the airlines are setting their pricing strategies into place. Due to the wide spectrum of price sensitivity among the consumers in India, the strategic balance to be kept by the airline players is relatively higher because on the one hand, they have to lure the early-booking strategic customers and on the other hand, the myopic customers. Strategic consumers, who keep waiting for some discount, introduce confusion into the demand uncertainty making the pricing process complicated. Inexperienced travelers, on the other hand, are likely to book with little consideration for the overall potential of the prices to change as the flight approaches, meaning that it is critical for airline companies to use the right fare timing strategies to their advantage.

To a certain level, this paper is unique and it has enhanced the understanding of dynamic pricing in the Indian airline industry regarding the nature of strategic and myopic customer behavior, the use of machine learning models and artificial intelligence in revenue management. We will also be focusing on the various measures undertaken by the Indian government for the growth of the aviation industry with regard to pricing and consumers. Through the identification of main factors that influence air ticket pricing and analysis of dynamic pricing strategies, the study will be useful to practitioners and scholars interested in the pricing model improvement in the context of the Indian airline business.

A. Key Concepts

In a highly competitive aviation industry, dynamic pricing has emerged as a vital tool for maximizing revenue and managing perishable inventory like airline seats. Indian airlines increasingly rely on data-driven strategies to respond to fluctuating demand, customer behavior, and market competition. This study focuses on the Indian domestic airline sector, exploring how strategic and myopic consumer behaviors affect pricing decisions. By leveraging reinforcement learning and machine learning models, the research aims to enhance fare optimization while addressing regulatory guidelines and consumer fairness. It also considers the government's role in shaping the industry through initiatives like UDAN and pricing transparency regulations. Before proceeding to unravel the dynamics of dynamic pricing strategies and the consumers' buying behavior in the airline industry, there is the need to provide the meanings and understandings of some concepts that are of paramount importance to this study.

Dynamic Pricing in Airline Industry: Dynamic pricing is the strategy businesses use to set the price of products or services on different factors such as demand, competition, market condition, and time of purchase. This approach proves helpful in revenue maximization since companies can charge their customers the highest prices during demand peaks and offer the lowest prices during demand troughs. For the airline industry, the issue of dynamic pricing is especially sensitive, as that industry relies solely on seat sales as its product is non-storable once the plane takes off, the unsold tickets are of no use. Airlines use dynamic fares to satisfy demand, optimize fares, and charge idle fares without considering daily and weekly market trends.

Strategic vs. Mayopic Consumers: Airlines usually encounter two unique groups of consumers: strategic and myopic consumers. Budget passengers are those people who know about the changes in prices and can afford to balance the current and future prices to buy commodities. They may postpone their purchase if they think prices will reduce nearer the traveling time. These passengers cause demand uncertainty mainly because their purchasing behavior depends on expected price levels, market saturation, and seasonal factors. However, myopic passengers make purchase decisions now and at that ticket price. They are not so bothered about future price drops and are in a position to purchase as long as the price matches the price they are willing to pay.

Role of AI and Machine Learning: AI is defined as the building of complex systems to increase the capability of an entity to carry out operations that previously needed intelligence, which is the mental processes involved in learning, problem-solving solving, or decision-making. Artificial Intelligence (AI) includes an ML, which creates an algorithm that takes the system to learn and enhance the performance by its own experience of the data fed into it rather than being programmed. AI and ML are applied in the airline industry to determine customer behavior, suggest the correct fares, and improve revenue-generating processes. These technologies help airlines make large data sets for faster processing for operational and tactical purposes and to vary fare strategies depending on target markets' conditions and trends.

Application of MDP for Fare Optimization: A Markov Decision Process (MDP) is a model of decision-making that incorporates a stochastic element but where the decision maker chooses actions that determine the transition probabilities. Stochastic programming is generally used in operations research and artificial intelligence to assess pricing strategies under uncertainty, including MDPs. In the context of airline ticket pricing, MDPs can guide airlines to which action would help increase ticket price or reduce it based on current demand, inventory conditions, and customer behavior, thus weighing short-term revenues against long-term profitability.

II. Literature Review

Our work is related to two streams of literature: reinforcement learning (RL) algorithms applied to complex sequential decision-making problems and dynamic pricing with non-myopic customers. Strategic customers and patient customers are the two representative ways that many studies define non-myopic customer models. To understand sellers' pricing strategies, strategic buyers try to postpone purchases until they are as patient as possible (Aviv and Pazgal, 2008; Den Boer, 2015). Conversely, patient consumers stay in the market for a certain amount of time and buy when the price is lower than they are willing to pay (Cao et al., 2015; Liu and Cooper, 2015; Lobel, 2020; Zhang and Jasin, 2022).

Customers now find it difficult to predict price sequences due to the growing complexity of airline pricing policies, which supports the patient customer model. Liu and Cooper (2015) showed an ideal pricing strategy with decreasing cycles for homogeneous patient customers. A polynomial-time algorithm for calculating the best pricing strategies under varying patient levels was presented by Lobel (2020). However, this is rarely the case in the airline industry, and these models typically assume that customer valuation and patience distributions are known a priori. An online learning and optimization algorithm that does not presume prior knowledge of these distributions was proposed by Zhang and Jasin (2022) to fill this gap; however, their work is still inflexible enough not to account for non-stationary demand and finite inventory constraints. To overcome these constraints, this study loosens three fundamental presumptions: finite inventory, non-stationary demand, and unknown customer-related distributions. Because only a few seats are available for a single flight, inventory-aware pricing strategies are required. Additionally, leisure and business travellers are two distinct customer segments the airline industry usually encounters, with varying arrival patterns and willingness to pay (Bondoux et al., 2020; Wittman and Belobaba, 2018). Relying on fixed or estimated distributions may lead to inconsistent revenue because of the impact of uncontrollable external factors on demand. These relaxations render earlier structural assumptions irrelevant and greatly increase the problem's computational complexity.

Furthermore, learning-based methods are required due to the lack of prior distributional knowledge, which leads to the use of model-free RL algorithms that don't make any assumptions about environmental dynamics (Rana and Oliveira, 2014; Mao and Shen, 2018; Krasheninnikova et al., 2019; Seo et al., 2021; Yang et al., 2022; Dixit and ElSheikh, 2022). Revenue management has long used RL, a tried-and-true solution technique for sequential decision-making problems (Rana and Oliveira, 2014; Pandey et al., 2020; Yang et al., 2022). Gosavi et al. (2002) were the first to apply reinforcement learning to airline revenue management by simulating random arrivals and cancellations across fare classes. A bounded actor-critic algorithm for seat allocation was created by Lawhead and Gosavi (2019), improving upon the computational problems associated with traditional actor-critic techniques. Deep Q-networks (DQN) were used to analyse dynamic airline pricing by Bondoux et al. (2020). Despite these advancements, previous research frequently ignores strategic consumer behaviours and non-stationary arrivals, necessitating effective RL exploration methods. While Yu et al. (2022) and Selim et al. (2022) employed reachability-based exploration techniques, Osband et al. (2016) used multiple networks to address early-phase reward sparsity for complex problems. Hong et al. (2018) and Parker-Holder et al. (2020) supported behavioural diversity, while Lopes et al. (2012) encouraged exploration through empirical learning progress in model-based RL.

Hafez et al. (2019, 2020) introduced intrinsic rewards for increased sample efficiency, while Sekar et al. (2020) suggested latent disagreement for exploratory planning. Tutsoy (2021a, b) introduced adaptive parametric models that use instantaneous rewards to improve policies in the face of real-world uncertainties. These frameworks are especially helpful when external incentives, such as airline pricing, might not provide clear direction. We adopt the model by Osband et al. (2016) and propose reward shaping as a future direction, given our goal of evaluating a new RL algorithm in this domain. Our study fills in three important gaps. First, using realistic airline assumptions, we present a novel Markov Decision Process (MDP) framework for dynamic pricing with patient customers. A novel sequential decision-making formulation that considers historical price sequences is part of this. Second, we evaluate several RL algorithms to determine which works best. As far as we know, this is the first study to address dynamic pricing with patient clients under the specified relaxations. Third, we examine how pricing policies are structured with and without patient customer considerations, providing airlines with helpful information to help them increase revenue through consumer segmentation.

Seong Bae Jo, Gyu M. Lee, and Ilk Yeong Moon (2024) identify how integrating strategic and non-myopic customer behavior affects airline dynamic pricing. Historically, dynamic pricing mechanisms presuppose customers cannot wait for a better offer, which realistic customers do not do anyway. Jo and colleagues offer an MDP framework to model this non-myopic behavior where the 'history of offered prices' is adopted as a state variable. That way, the model can mimic the erratic nature and non-stationary demand prevalent in the airline market. Jo et al. claim that when adopting model-free algorithms and deep exploration-based RL not used in prior works, airlines can create a more realistic pricing model reflecting the unknown distribution of customers' preferences for additional services. This study demonstrates that while making decisions on strategic customers, the price can be high for a short time. Then, low prices bring higher revenues than a continuous daily hike in prices in increased pricing models.

Jin Min Gao, Mei Long Le, and Yuan Fang (2022) also studied the impact of strategic and myopic passengers on dynamic pricing for a related reason. They categorize passengers into two broad categories, high- and low-valuation, depending on passenger evaluation of tickets. Like Jo et al., Gao and his team constructed the dynamic pricing model based on the utility of both the airline and consumer sides. They also use reinforcement learning – specifically, Q-learning – to address the pricing issue in a Markov decision process model. Their evidence states that as the percentage of strategic passengers increases, the airline should employ a more conservative pricing approach. Namely, the price increase should gradually go through corresponding changes or be based on the percentage of high- or low-valuation strategic passengers. An incremental price increase reduces the likelihood of giving in to price-high prices altogether while still getting some revenue from the frame strategic customers. According to the model, the means necessary for generating maximal revenues depend on the composition of the flow of passengers, which requires specific operational tactics and strategies from the airlines (Gao et al., 2022).

In other prior work on dynamic pricing, Kevin R Williams (2020) looked into airlines about the influence of dynamic pricing in the market and how the seating capacity can be allocated to customers with different willingness to pay. Williams introduces a stochastic demand model with characteristics from the revenue management model and pricing models typical for empirical economics. He identifies that dynamic pricing is most applicable in dealing with demand shocks and intertemporal consumer behavior variations. Analyzing the dynamics of dynamic pricing, Williams also described the connection between dynamic pricing and intertemporal price discrimination. Capacity management can be achieved through speed in pricing strategy since consumers' arrival rate and willingness to pay vary. Williams introduces a stochastic demand component and other features from revenue management and practical pricing models widely used in empirical economics. He also realizes that it is most helpful in dealing with demand shocks and temporal cross-section and time series differences in consumers' preferences. Through fares that vary with the demand levels, airlines can give attractive fares to early-booking customers who are less willing to pay while holding back the high fares to late-booking passengers with a higher time utility. According to Williams, dynamic pricing is closely related to intertemporal price discrimination. In this context, airlines can attain a desirable capacity because they can adjust their prices with the rhythm of consumers' arrivals and the prices they are willing to pay. This strategy drives optimal revenues and optimizes seat stock and distribution, with plenty of deep discounts geared to satiating the price-sensitive segments to the exclusion of the high-yielding, time-sensitive consumer (Williams, 2020).

In recent years, boarding fees and seat selection alongside ticket prices have been critical sources of revenue in the air transport industry. Elena Kosonen's (2020) master's thesis examines how these additional services could be priced more effectively using machine learning algorithms. She recruited stakeholders to perform co-creation sessions that led to the creation of a machine-learning model for more apt customer pricing. By understanding customer value through pricing optimization tools, namely pricing controls and internal airline data, this model, proposed by Kosonen, delivers segmented price points for the ancillary services, thus enhancing the organization's revenues and boasting its profit margins. Even though this machine learning model was interrupted due to COVID-19, Kosonen's work shows that data science solutions can bring value for ancillary pricing. Machine learning allows different services, such as extra baggage charges, to be set based on customer profiles and other factors in a real-time fashion. While these ADDITIONAL services also grew into one of the many critical factors determining profits, getting more by applying machine learning for better pricing may also be an appealing option (Kosonen, 2020).

Priester and colleagues' investigation indicates that the PDP strategy could improve airline revenue; however, its application should consider privacy and fairness considerations. This is always the case because it reminds the customers of the personal information they share with the company to be given customized prices. Therefore, the above perspective should guide airlines that are interested in adopting PDP to ensure that pricing policies should not overemphasize the use of customer data while at the same time ensuring fairness (Priester et al., 2020). From the initial attempts to use stochastic demand models to more recent advances in reinforcement learning and machine learning, airlines are better positioned to target prices to meet different consumers and demand changes. According to the research conducted by Jo et al. (2024) as well as Gao et al. (2022), recognizing such a strategic context itself helps airlines improve their selling prices for better profitability; Williams (2020) also reflects that dynamic pricing policy plays a crucial role in optimizing the capacity on the means of transport. However, Priester et al. (2020) note that the given strategies might also raise consumer suspicions about the carriers' motives, which is why the airlines need to perfect the given strategy regarding how consumers receive them to guarantee success.

The dynamic pricing model introduced to KAL by Naman Shukla and colleagues in 2019 was created by Deeper Solutions, and the principle utilized proposed that the rates of routes differ according to supply and demand to achieve greater revenues. Their approach was customer interaction during live booking sessions on the airline's website, with ancillary products as the variable for demand and price optimization. Their model consisted of two key components: A binary classification model for ancillary purchase probability and a revenue optimization model for the best price for the sale of the ancillary product (Shukla et al., 2019). They tried different algorithms, such as logistic mapping and deep neural networks. They realized that their DNN-CL model improved better than traditional pricing systems, mainly in determining the sensitivity of demand for the prices and setting the correct prices. However, some constraints, such as limited and fixed price range and focus on only one accessory during online testing, showed that the extent of applying dynamic pricing in this context was still not fully unleashed (Shukla et al., 2019). Rafael R. Varella, in his paper published in 2017, investigated the effects of new Low-Cost Carriers (LCC). He applied an econometric model to show how mainline airlines changed their tariff strategies to strategies for the new entrant carriers. The research that reviewed more than 96428 price quotes offered by ten key airlines in Brazil identified that incumbents lowered the fares for early bookings to lure price-sensitive travelers primarily targeted by LCCs (Varella et al., 2017). It acknowledged that incumbents raised the number of airfares available through OTAs by 11% and cut basic fares from 3.4% to 9% for bookings one or two months before travel. The change occurred due to high competition in the early time for booking, holding the higher price near the departure date for non-willing-to-haggle clients. This pricing response also promoted competition and reallocated consumer surplus, whereby early bookers benefited most while late bookers were worst affected.

Building on top of dynamic pricing models in the supply function literature, in 2020, Daniel F. Otero and Raha Akhavan-Tabatabaei argued for a stochastic dynamic pricing model that enumerates the inter-arrival time between customer bookings and the likelihood of purchase. It proposed a way to match better the cost of selling a seat at a lower price than its face value or not selling the seat at all. Otero and Akhavan-Tabatabaei's statistical model enhanced prior approaches using phase-type distributions to estimate customer arrival times. As their work led to a more realistic, high-fidelity representation of the data at hand, it is more appropriate for applied/empirical work (Otero & Akhavan-Tabatabaei, 2020). This model was used to solve the problem of where to set the right price because no airline wanted to overprice its fares, and yet, at the same time, no airline wanted to risk leaving seats empty. However, the researchers, as indicated by Otero and Akhavan-Tabatabaei (2020), observed that more research is needed to support the refinement and, thus, the usability of the presented model in field contexts.

Similarly, Ryan J. Lawhead and Abhijit Gosavi presented two algorithms in 2019, Discounter and Averagsolving, solving the Discounted and Andand Average setting, They we; they., airline revenue management. They pointed to deep learning architectures as the key approach, which was attracting much attention among AI researchers. The discounted reward algorithm was tested using minor sample problems where the best policy was known a priori. In contrast, the average reward algorithm was tested on a large-scale testbed using industrial data. These algorithms helped present an understanding of how dynamic pricing models could be further adapted by airlines by implementing reinforcement learning and deep learning approaches. Despite the benefits of these ideas, there is the possibility of further research and development (Lawhead Gosavi, 2019). On the other hand, the successful work of Seong Bae Jo and his collaborators in 2024 introduced the MDP framework to solve non-myopic customer problems—consumers who delay their purchases to enjoy better promotions. Analyzing the nature of unforeseen and non-stationary demand, Jo's research incorporated model-free algorithms into the airline framework. It revealed that if airlines consider patient, non-myopic customers, they could fluctuate between high and low prices to improve their revenue rather than stick to the linear pricing model. It was also revealed that this fluctuating price model provided greater overall sales than the more conventional gradual rise in price (Jo et al., 2024). In addition, the study pointed out that to generate the highest revenue, airlines ought to set very high prices for a long time and set relatively low prices only occasionally if they expect a large proportion of patient customers.

To sum up the present research, it is possible to paint a picture of the way different carriers have implemented dynamic pricing into strategy and examined the strengths and weaknesses of its widespread use throughout the global airline market through the three following studies: It is for this reason that although some recent advancements have been noted in terms of modeling strategic customer behavior and the integration of AI-powered pricing models, there is still much work to be done in continuously unlocking the value of this evolving space. The literature on dynamic pricing in the airline industry has significantly progressed in understanding consumer behavior, especially the differences between strategic and myopic passengers. However, several research gaps remain, including integrating a wider range of customer segments beyond high and low valuations. Future research could examine how behavioral, psychographic, and demographic factors influence pricing sensitivity and preferences. Furthermore, while research looks at how well model-free algorithms and reinforcement learning work for pricing strategies, it usually overlooks how well these models adapt to sudden shifts in the market or outside shocks like pandemics or recessions. Finally, the research could look into how dynamic pricing systems can be designed to react in real time to unanticipated events.

III. Methodology

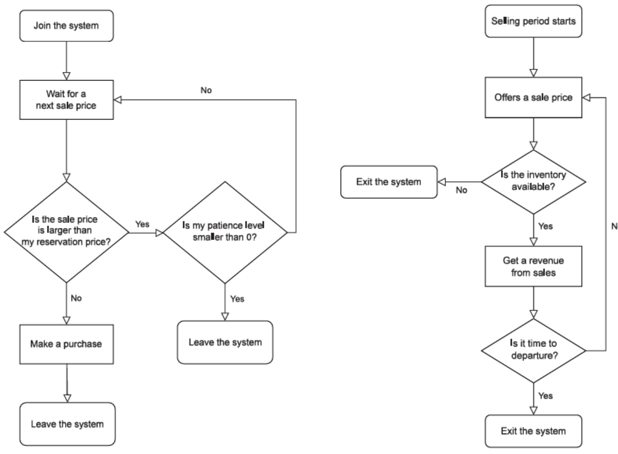
A. Patient customer dynamics

As previously mentioned, our system's customers are patient and possess three attributes that influence their decision-making process: arrival time, reservation cost, and patience level. Customers' maximum willingness to pay is indicated by the reservation price. The customer purchases the product if the seller's sale price is less than or equal to the customer's reservation price. The customers' maximum willingness to wait is indicated by their patience level. Up to $t+1$ purchasing opportunities are available to customers with a patience level of t. In particular, when a customer accesses the system during period k, the sale price and the reservation price from period k are compared to period $k+t$. One makes a purchase and exits the system immediately if they see that the sale price is less than or equal to their reservation price during those times. Otherwise, the longer a person waits to make a purchase, the less patient they become. One exits the system without purchasing if their level of patience declines. Fig. 1(a) displays a flow chart illustrating these patient customer dynamics.

An initial level of patience for a customer who visited during period k is a random variable defined in $\mathcal{T}$, a finite set of non-negative integers with a maximum value of T.A probability distribution with the cumulative density function $F_{k,\tau}(\cdot)$ governs the reservation price for a customer who arrives during period k with patience level τ. The number of customers arriving during period k with patience level τ has a probability density function of $g_{k,\tau}(\cdot)$. We assume that $f_{k,\tau}(\cdot)$, $g_{k,\tau}(\cdot)$, and $F_{k,\tau}(\cdot)$ are unaffected by the seller's price and inventory during each period. All of the study's uncertainties are parametric and related to the client. The number of customers arriving during each period, reservation costs, and customer patience levels are the specific sources of uncertainty.

B. Marcov Decision Process

A popular framework for simulating discrete sequential decision-making problems is the Markov Decision Process (MDP) (Sutton & Barto, 2018). MDPs are frequently used to



(a) Customer (b) Seller

effectively represent problems involving uncertain or stochastic demand, such as inventory management and dynamic pricing (Lawhead & Gosavi, 2019; Ma et al., 2021; Ahiska & King, 2015; He et al., 2023). The four main parts of an MDP are usually states, actions, rewards, and transition probabilities. The state is a representation of the system's present state. This work treats states as fully observable by the decision-maker, in contrast to some previous studies that treat the decision-maker observations independently from the actual system states (Pandey et al., 2020; Shou et al., 2022; Bautista-Montesano et al., 2022). The Markov property is broken if the transition depends on more than just the state and action at that moment. Violating this requirement may harm the quality of the solutions since most algorithms for locating optimal solutions depend on the Bellman optimality equation, which presumes the Markov property. The following section shows how the Markov property can be preserved by structuring the state variables in our dynamic pricing model.

To formulate the airline dynamic pricing problem described at the beginning of Section 3 as an MDP with finite states and actions, we consider an airline as a decision-maker. The behavior of the airline can be presented in Fig. 1(b). The action space contains possible prices that the airline can offer to customers. Because the airline's objective in this paper is to maximize the total revenue over the finite selling periods, we set the revenue gained in each period as a reward. In this study, we define the discount factor of the MDP as one. Before defining our state variables, we define some notations:

We model the airline pricing problem as a Markov Decision Process (MDP), following Jo et al. (2024). A MDP is defined by states, actions, transition dynamics, and a reward function.

- q_t : Number of remaining seats in period t
- l_t : Remaining time to departure in period t
- W : Maximum patience level that customers can have
- s_t : State of the system in period t
- a_t : Price offered by the decision-maker in period t
- $\mathcal{S}$: Finite set of states
- $\mathcal{A}$: Finite set of actions
- t_k : Group of customers who arrive in period t with patience level k

Initially, we determine the likelihood that the quantity of seats sold during time period t will be i.

For $t' \le t \le t' + k$, a customer in $\mathcal{G}_{t'}^k$ behaves as follows:

If the customer's reservation price is less than every price in the set $\{a_{t'}, a_{t'+1}, \dots, a_{t-1}\}$, then the customer does not make a purchase until period t.

Furthermore, the customer makes a purchase during period t if their reservation price is greater than or equal to the offered price in that period, i.e., the lowest price in the set $\{a_{t'}, a_{t'+1}, \dots, a_t\}$.

Based on these facts, we can calculate the likelihood that a customer in $\mathcal{G}_{t'}^k$ will make a purchase during period t, where $t' \le t \le t' + k$, as follows:

$$p_t^1 = \left(F_{t'}^k\left(\min\{a_{t'}, \dots, a_{t-1}\}\right) - F_{t'}^k(a_t)\right)^+$$

Here, $F_{t'}^k(\cdot)$ denotes the cumulative distribution function (CDF) of reservation prices for customers in $\mathcal{G}_{t'}^k$, and $(\cdot)^+$ represents the positive part function, i.e., $\max\{0, \cdot\}$. If the number of customers arriving during time t' with a patience level k is $n_{t'}^k$, then the likelihood that $l_{t'}^k$ seats are sold by customers in $\mathcal{G}_{t'}^k$ during period t is given by the binomial probability:

$$\Pr\left(X_{t',t}^k = l_{t'}^k\right) = \binom{n_{t'}^k}{l_{t'}^k}(p_{t',t}^k)^{l_{t'}^k}(1 - p_{t',t}^k)^{n_{t'}^k - l_{t'}^k}$$

Using this, we can now compute the probability that exactly i seats are sold during period t, given the historical price path $H_t = \{a_{t-W}, \dots, a_{t-1}\}$.

- $\bar{P}_t^i$: The probability that i seats are sold during period t, given $H_t = \{a_{t-W}, \dots, a_{t-1}\}$.
- $X_{t',t}^k$: A binomial random variable with parameters $n_{t'}^k$ and $p_{t',t}^k$, representing the number of purchases by customers in $\mathcal{G}_{t'}^k$ during period t.
- $\mathbf{n}_t$: A vector representing the number of customers in each group $\mathcal{G}_{t'}^k$ for $t - W \le t' \le t$ and $t - t' \le k \le W$.
- $\mathbf{i}_t$: A vector representing the number of seats sold to customers in each group $\mathcal{G}_{t'}^k$ for $t - W \le t' \le t$ and $t - t' \le k \le W$.
- N_W^i: The set of all vectors of W non-negative integers that sum to i.

The likelihood that i seats will be sold during period t is computed based on the fact that customers who arrived during the periods $t - W, \dots, t$ may still purchase in period t. This is expressed as:

$$\bar{P}_t^i = \sum_{\mathbf{n}_t \ge \mathbf{i}_t} \left[\left(\prod_{l=0}^{W} \prod_{u=0}^{l} d_t^k(n_{t-W+l}^{W-u}) \right) \times \left(\sum_{\mathbf{i}_t \in N_W^i} \prod_{l=0}^{W} \prod_{u=0}^{l} \Pr\left(X_{t-W+l,t}^{W-u} = i_{t-W+l}^{W-u}\right) \right) \right] \tag{1}$$

where:

- $\bar{P}_t^i$: Probability that exactly i seats are sold during period t
- $\mathbf{n}_t$: Vector of customer group sizes $n_{t'}^k$ for all relevant t' and k
- $\mathbf{i}_t$: Vector of corresponding seat sales from customer groups
- $d_t^k(n)$: Probability mass or weight function (e.g., probability distribution over customer arrivals)
- $X_{t',t}^k$: Binomial random variable for group $\mathcal{G}_{t'}^k$
- $N_{\hat{W}}^i$: Set of all vectors of $\hat{W}$ non-negative integers summing to i
- $\hat{W} = \frac{(W+1)(W+2)}{2}$: Total number of (t', k) group combinations over the window W

Equations (1) and (2) show that $\bar{P}_t^i$ depends on the price history $\{a_{t-W}, \dots, a_{t-1}\}$, which represents the prices chosen by the decision-maker during the periods $t-W$ through $t-1$.

This implies that when customers exhibit patience, the expected number of seats sold in the upcoming period may be significantly misestimated if the decision-maker fails to consider actions taken in the previous W periods.

Recall that the system state s_t was defined solely using q_t (remaining seats) and l_t (remaining time to departure). In this setup, the system fails to satisfy the *Markov property*, because $\bar{P}_t^i$ is not consistent for identical s_t and a_t when the history of prior actions differs.

Specifically, the transition probability

$$\Pr(y_{t+1} \mid s_t, a_t)$$

cannot be accurately determined without knowing $\bar{P}_t^i$, where $y_{t+1} = (q_t - i, \, l_t - 1)$. This dependence on historical prices introduces a form of memory into the system, violating the assumption that future states depend only on the current state and action. Although invariability is still employed to calculate the transition probability, the consistency is ensured by the fixed set $\{a_{t-W}, \dots, a_{t-1}\}$ in s_t for the same $\bar{P}_t^i$ and a_t.

It is important to note that Equations (1) and (2) are standard equations that capture the stochastic dynamics of the problem under study. However, deriving state variables from these equations can offer a novel perspective for the literature that applies model-free algorithms and formulates the airline revenue management problem as a Markov Decision Process (MDP).

In the context of this study, the airline must decide on a range of prices that may be offered to passengers prior to departure. We refer to these as *pre-defined prices*. One of the agent's actions is to select a price from this predetermined list and offer it to the clients. The action space $\mathcal{A}$ can be defined as $\{p_1, \dots, p_A\}$, where each p_i represents a pre-defined price. The action space consists of real numbers that are positive.

The state space is denoted by $\mathcal{S} = (q, l, H) \in [0, Q] \times [0, T] \times \mathcal{H}_W$, where H is a vector of size W that, as previously mentioned, contains previously offered prices. Here, Q and T are positive integers representing the total number of seats and the selling horizon, respectively.

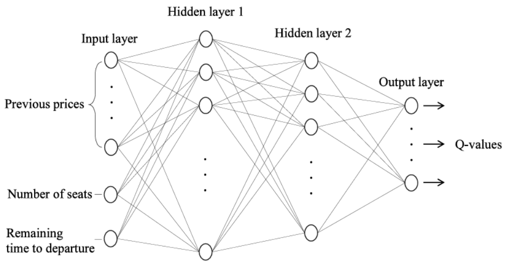
The function $r(s, a)$ represents the instantaneous reward for taking action a in state s. In this study, when the agent offers price a based on state s, the reward $r(s, a)$ is defined as the immediate revenue from the sale.

Transition probabilities can be computed using Equation (2), which captures the dynamics of the system. Equation (2) can be used to compute the transition probabilities. Our objective is to determine the best pricing strategy π that optimizes the anticipated total revenue over a limited selling horizon T:

$$\max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{T-1} r(s_t, a_t) \right] \tag{3}$$

With the solution methods described in the following section, the agent can learn a pricing policy that is nearly optimal, as it can observe the reward immediately after taking action in a given state. One of the key advantages of Deep Reinforcement Learning (DRL) is its flexibility in solving high-dimensional state-space problems (Gosavii et al., 2002). The dimension of the state space is $W+2$, as the state y_t consists of q_t, l_t, and the historical price vector $(a_{t-W}, \dots, a_{t-1})$. Since many actual customers in the airline industry are aware that ticket prices are subject to change, their maximum patience level may be high, which raises the state space's dimension. In RL, it may be unmanageable in a high-dimensional state space if value functions or policy functions are not approximated. Thus, the agent can learn policies for difficult real-world problems by using approximators such as neural networks (Gosavii et al., 2002; Yang et al., 2022).

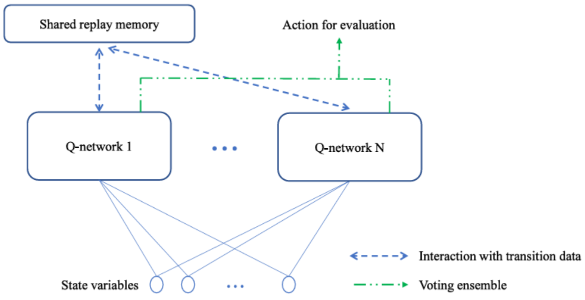
1) Deep Q-Network: The value-based RL literature has used the parameterised Q function to solve large-scale problems. The Q-function is estimated using fewer parameters rather than tracking Q-values for every state–action pair. As opposed to state-action pairs. After that, they are updated frequently in order to get close to the ideal Q-function. The deep Q-network (DQN), which was proposed by Mnih et al. (2013), is the most widely used value-based algorithm that uses a parameterised Q-function. They approximated Q-functions for learning optimistic control policies for a few Atari 2600 environments using convolutional neural networks. For numerical experiments, we employ Mnih et al. (2015)'s DQN algorithm. Appendix A contains the algorithm's detailed process.



(a) DQN

Fig. 1. DQN

2) Bootstrapped DQN (BDQN):: To achieve functional improvements for more difficult environments, some sophisticated RL algorithms based on DQN have been investigated (Van Hasselt et al., 2016; Schaul et al., 2015; Wang et al., 2016). The exploration problem is one of those complex problems. Even though larger rewards can be offered in other states, RL agents can easily become stuck in states that are close to the initial state in an episodic environment when relatively small rewards are offered. When RL agents employ inefficient exploration techniques, such as greedy policies, it takes many episodes to determine the best action.



(b) BDQN

Fig. 2. B-DQN

The same exploration problem applies to the dynamic pricing problem of airlines in this study. An RL agent learns suboptimally if it focuses on optimizing rewards from customers who arrive earlier with lower reservation prices because the percentage of customers with higher reservation prices rises over time and the number of seats is limited. In fact, we found that the agent learned pricing policies that focused on the early times of the selling horizon, leading to early episode termination, when we performed numerical experiments using DQN.

As a result, RL algorithms with better exploration techniques would be needed. One of the RL algorithms that has been shown to be successful for episodic exploration in solving difficult problems is Bootstrapped DQN (BDQN), which was proposed by Osband et al. (2016). Appendix B contains the detailed process of the algorithm. We present the findings of numerical experiments in the following section, which suggest that BDQN is also more efficient in the dynamic pricing environment of airlines. A shared network architecture was proposed by Osband et al. (2016), in which a network directly connected to input data is shared by several Q-networks. To cut down on computational expenses, it learns a feature representation of the input data. However, since our input data is not a two-dimensional image frame, we do not use the architecture. In this scenario, the state variables may be overly shortened by the shared network, which would leave the BDQN agent with inadequate environmental information. As a result, we use Q-networks that directly receive state variables as input when designing the BDQN framework. Like the DQN architecture we employed, each Q-network is built separately. The network architecture of DQN is depicted in Fig. 2(a), while the architecture of BDQN, which is made up of several networks with structures identical to those in Fig. 2(a), is depicted in Fig. 2(b). Five networks with the same structure as DQN are used in BDQN. The hyperparameter values described above are chosen to ensure that the RL algorithms produce sufficient results, which are shown in the following subsection. Python 3 with an Intel Core i5-9400F and 16 GB of RAM is used for all experiments.

C. Comparison between formulations in presence of patience customers

In this subsection, we develop two distinct MDP formulations. The state s_t is the only element that distinguishes different decision-making contexts. When a Markov Decision Process (MDP) includes a sequence of past actions $(a_{t-h}, \ldots, a_{t-1})$ as part of its state s_t, it is referred to as an *MDP with action history*. On the other hand, if the state s_t comprises only the current system state and time, (x_t, t), without incorporating previous actions, it is referred to as an *MDP without action history*. Therefore, when patient customers are in the system, MDP without action history does not satisfy the Markov property. Even when the Markov property is met, there is no theoretical guarantee that DQN and BDQN will converge or perform better. However, based on the numerical experiment results, we found that the MDP with action history had higher revenue than the one without. DQN doesn't create nearly ideal policies. Nonetheless, DQN pushes the agent to investigate less-than-ideal policies during the training episodes. The difference between DQN's revenue and the upper bound may be narrow if the agent chooses the best course of action with probability one during the evaluation episodes and eliminates exploration.

graphicx

[a] MDP states include only current state variables.
[b] MDP states include past action sequences.

DQN doesn't create nearly ideal policies. Nonetheless, DQN pushes the agent to investigate less-than-ideal policies during the training episodes. The difference between DQN's revenue and the upper bound may be narrow if the agent chooses the best course of action with probability one during the evaluation

TABLE I
AVERAGE REVENUE OF EVALUATION EPISODES OVER FIVE RANDOM SEEDS.

Case	**Customer**	**T**	**Algorithm**	[c]Average Revenue
(No History[a] / With History[b])				
1	Patient	20	DQN	66.31 / 71.24
2	Patient	20	BDQN	65.66 / 72.68
3	Patient	40	DQN	129.82 / 141.32
4	Patient	40	BDQN	131.07 / 144.45
5	Myopic	40	DQN	119.69 / 119.13
6	Myopic	40	BDQN	119.94 / 120.04

[a] MDP states include only current state variables.
[b] MDP states include past action sequences.

episodes and eliminates exploration. The outcomes of DQN and BDQN in evaluation episodes are shown in Table 1. The mean of the average revenue with action history over five random seeds is indicated by the first value in the last column, while the second value is for entities without action history. Cases 1, 2, 3, and 4 relate to different experimental configurations. Regardless of RL algorithms, the average revenue in the evaluation episodes rises when the action history is incorporated into the state, just like in the training episodes. Furthermore, the optimal expected revenues with infinite inventory in Cases 2 and 4 are 74.45 and 149.8, respectively, and policies derived from DQN and BDQN produce revenue near the upper bounds. These numerical findings suggest that the DRL algorithms can identify nearly optimal policies even in dynamic pricing environments where perfect information is not available.

D. Comparing Pricing Algorithms for Inadequate Inventory and Non-Stationary Demand

In this subsection, when demand is non-stationary and inventory is inadequate, we illustrate the pricing algorithms using the MDP with action history. One of the most used customer segmentation techniques in the airline industry is the division of customers into leisure and business customers (Li et al., 2014; Bondoux et al., 2020; Varella et al., 2017). In the preceding section, every component of the MDP formulation is defined in the same way, and five runs of 50,000 training episodes are also carried out. According to earlier research, business and leisure clients are presumed to enter the Poisson process (Flapper et al., 2012; Yousuk & Luong, 2013; Yu et al., 2019).

As time passes, the arrival rate of leisure customers falls linearly, and the opposite is true for business customers. Although uniform distributions generate reservation prices for business and leisure customers, business customers typically pay higher average prices than leisure customers. It is anticipated that twice as many customers will enter the system as there are seats. We compare each algorithm's performance for a few examples to confirm which is suitable.

IV. NUMERICAL TEST AND EXPERIMENTATION

The results of numerical experiments are presented in this section. To simulate dynamic pricing, we created a synthetic environment. In the simulations illustrated in Fig. 1, the seller and the buyers make their choices. The efficacy of the suggested MDP formulation with patient clients present is described in the first subsection. The performance of algorithms for scenarios with non-stationary demand and insufficient inventory is compared in the second subsection. We examine pricing policy structures in the final subsection.

TABLE II
COMPUTATIONAL TIME FOR EACH ALGORITHM.

Figure	T	W	Algorithm	Time (min)[a]
3*Fig. 4(a)	3*50	3*11	DQN	128.76
			BDQN	263.20
			Modified OLD	147.49
3*Fig. 4(b)	3*50	3*29	DQN	156.74
			BDQN	314.11
			Modified OLD	166.46
3*Fig. 4(c)	3*100	3*11	DQN	273.86
			BDQN	568.80
			Modified OLD	289.19
3*Fig. 4(d)	3*100	3*29	DQN	318.13
			BDQN	723.36
			Modified OLD	200.65

[a] Average computational time to complete 50,000 training episodes.

The following is an explanation of the DQN algorithm's processQ-network, parameterized with ϕ, approximates the Q-function. The stochastic gradient descent method is used to It aims to find Q by minimizing the mean squared error between the parameterized Q-function Q_θ and the optimal Q-function Q^*. Since Q^* is unknown, unlike in supervised learning scenarios, gradients of the mean squared error are computed using the estimated target value, which in the t-th update is:

$$\phi_{k+1} = \phi_k - \alpha \Bigg[\sum_{j \in \mathcal{B}} \Big(Q_{\phi_k}(s_j, a_j) - r_j + \gamma \max_{a' \in \mathcal{A}} Q_{\phi_k}(s'_j, a') \Big) \nabla_{\phi_k} Q_{\phi_k}(s_j, a_j) \Bigg] \quad (2)$$

It denotes a minibatch selected from the replay memory $\mathcal{D}$ that stores samples that the agent gained from previous interactions with the environment. Due to the moving target in Eq. (A.1), the stability of the SGD method can be degraded. To alleviate this problem, Mnih et al. (2015) used an additional target Q-network, denoted by Q_{θ^-}, in the update rule. The target value in Eq. (A.1) is substituted with the target network Q_{θ^-}, and it copies the original Q-network every τ updates.

Algorithm 1: DQN Algorithm

For exploration, the BDQN algorithm makes use of several Q-networks. The transition tuples are saved in the shared replay memory, and one network is chosen randomly to perform steps in a training episode. Then, like in DQN, a minibatch of transitions is selected randomly from the replay memory. While some Q-networks update their weights using it, others do not. An agent may attempt several suboptimal actions for

Algorithm 1: DQN Algorithm

```
Algorithm 1 DQN algorithm
Initialize replay memory $D$
Initialize parameters $\phi$ for Q-network
Set $\phi^- = \phi$
for episode $= 1, \ldots,$ C do
    Initialize state $s_0$
    Set $t = 0$
    while $s_t$ is not the terminal state do
        Select a random action $a_t$ with probability $\epsilon$, otherwise $a_t = \arg\max_a Q_\phi(s_t, a)$
        Implement action $a_t$ and get reward $r_t$ and next state $s_{t+1}$
        Store transition data $(s_t, a_t, r_t, s_{t+1})$ in replay memory $D$
        Randomly select a minibatch of transitions $(s_j, a_j, r_j, s_{j+1})$ from $D$
        Set target $y_j = \begin{cases} r_j & \text{if } s_{j+1} \text{ is the terminal state} \\ r_j + \max_{a'} Q_{\phi^-}(s_{j+1}, a') & \text{otherwise} \end{cases}$
        Update $\phi$ using the SGD method according to Eq. (A.1) on $y_j$
        Set $t = t + 1$
    end while
    Set $\phi^- = \phi$ every $E$ episodes
end for
```

```
Algorithm 2 BDQN algorithm
Initialize replay memory $D$
for $i = 1$ to $N$ do
    Initialize parameters $\phi_i$ for Q-network $i$
    Set $\phi_i^- = \phi_i$
end for
for episode $= 1, \ldots,$ C do
    Select $k \sim \text{Uniform}(1, \ldots, N)$
    Initialize state $s_0$
    Set $t = 0$
    while $s_t$ is not the terminal state do
        Set $a_t = \arg\max_a Q_{\phi_k}(s_t, a)$
        Implement action $a_t$ and get reward $r_t$ and next state $s_{t+1}$
        for $i = 1$ to $N$ do
            Generate a bootstrap mask $u_i \sim \text{Bernoulli}(\frac{1}{2})$ for Q-network $i$
        end for
        Set vector of bootstrap masks $\hat{u}_t = (u_1, \ldots, u_N)$
        Store transition data $(s_t, a_t, r_t, s_{t+1}, \hat{u}_t)$ in replay memory $D$
        Randomly select a minibatch of transitions $(s_j, a_j, r_j, s_{j+1}, \hat{u}_j)$ from $D$
        for $i = 1$ to $N$ do
            Set target $y_j = \begin{cases} r_j & \text{if } s_{j+1} \text{ is the terminal state} \\ r_j + \max_{a'} Q_{\phi_i^-}(s_{j+1}, a') & \text{otherwise} \end{cases}$
            if $i$-th element of $\hat{u}_j$ is equal to 1 then
                Update $\phi_i$ using the SGD method in Eq. (A.1) on $y_j$
            else
                do not update
            end if
        end for
        Set $t = t + 1$
    end while
    Update target networks every $E$ episodes
end for
```

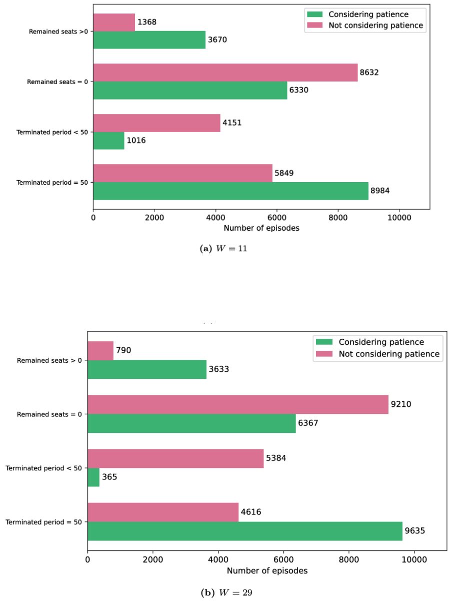
several time steps due to the randomness in choosing and updating multiple Q-networks. Osband et al. (2016) demonstrated through numerical experiments that the BDQN agent required a significantly smaller number of training episodes than the DQN agent to escape from suboptimal policies.

A. Real-World Validation: Indian Airline Pricing Data

We examined a dataset of 10,683 domestic flights in India, including details like ticket prices, carrier type, route, and number of stops, to confirm the applicability of our findings. The data showed significant trends that matched the outcomes of our simulation. The average cost of nonstop flights was 5,024, one-stop flights were 10,594, and two-stop flights were 12,716. The prices of flights with more stopovers were significantly higher. Legacy carriers like Jet Airways and Air India had much higher median fares exceeding 9,000, while low-cost airlines like SpiceJet and IndiGo had median prices between 3,800 and 5,000. For example, the Bengaluru-Delhi route had fares ranging from 3,257 to 25,913, indicating a notable price disparity within the same routes.

V. Key Findings

Airlines modify their pricing strategy to alternate between high and low prices over the selling horizon when they identify the presence of patient customers. The conventional monotonic price increase over time contrasts with this dynamic pricing strategy. The idea is based on revenue optimization: strategically timed low prices lower the risk of unsold inventory close to departure. In contrast, high prices target customers with a high willingness to pay. Consequently, this strategy raises the percentage of episodes that end with unsold seats and those that end at the point of departure—nevertheless, overall revenue increases despite increased unsold inventory. Interestingly, the average difference between high and low price points widens, and the frequency of high price charges increases as customer patience (W) rises.



(a) $W = 11$

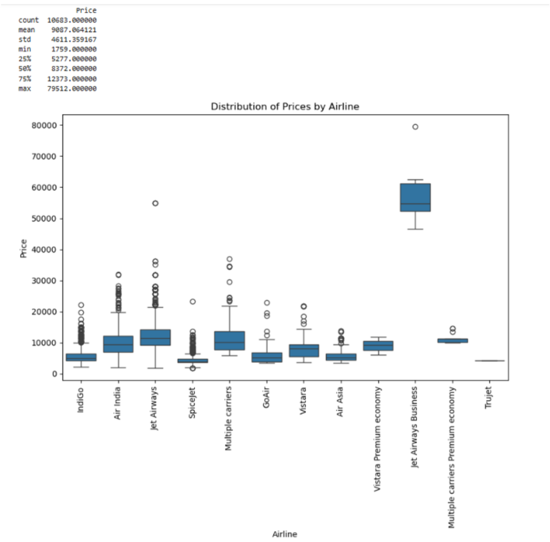
(b) $W = 29$

Fig. 3. Number of episodes grouped by terminated period and the number of remaining seats.

This implies that by reserving expensive slots for extremely patient, high-reservation-price passengers, airlines are getting more value out of them. As a result, fewer inexpensive slots are offered. The observed pattern lends credence to the claim that dynamic, non-stationary systems involving patient and strategic customer behavior are better suited for non-monotonic, alternating pricing.

A. Price Distribution for Tickets on Different Airlines

A detailed analysis of the ticket price distribution across Indian airlines reveals significant variability. The mean ticket price is approximately 9,087, while the median is 8,372, indicating a moderately right-skewed distribution. According to the box plot, prices range from around 1,759 to 79,512. This wide dispersion, along with a high standard deviation of 4,611, underscores the diverse pricing strategies adopted by different carriers. Full-service airlines such as Jet Airways (Business Class) charge premium fares, whereas budget airlines like IndiGo consistently offer lower prices. Numerous outliers—often associated with last-minute bookings, peak travel seasons, or seat class upgrades—highlight pricing anomalies. These observations suggest that static pricing models are inadequate for capturing the complex and dynamic nature of airline fare structures.



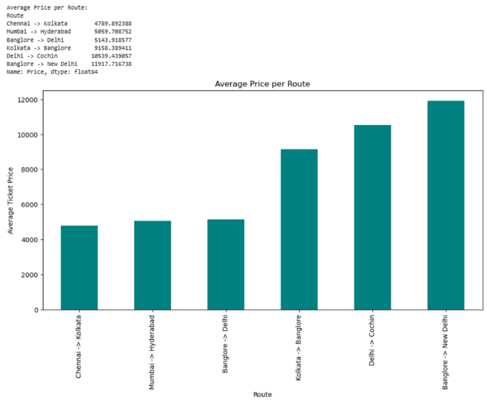
```
             Price
count  10683.000000
mean    9087.064121
std     4611.359167
min     1759.000000
25%     5277.000000
50%     8372.000000
75%    12373.000000
max    79512.000000
```

Instead, the data emphasizes the necessity of dynamic pricing strategies that consider factors like service class, demand spikes, customer type, and time of purchase. Additionally, the dispersion suggests significant space for optimization using cutting-edge revenue management systems, which could more effectively match pricing to passenger willingness to pay and market demand. Recognizing these patterns supports the potential efficacy of dynamic pricing algorithms based on machine learning in practical settings.

B. Price Differences Based on Route

Significant differences influenced by various market factors can be seen when comparing the average ticket prices for various domestic routes. Flights from Bangalore to New Delhi, for instance, routinely have the highest average fares, often surpassing 12,000. In contrast, flights from Chennai to Kolkata are substantially less expensive, with average fares of less than 5,000. Demand elasticity, market competition, flight distance, and route popularity are the reasons behind these disparities. While routes with lots of options show downward price pressures from airline competition, heavily trafficked routes with little competition typically command higher prices.

Furthermore, higher demand for business travel is frequently reflected in metropolitan connections, which helps to justify premium pricing. When creating airline pricing policies, the analysis emphasizes route-level optimization. Route-specific features, such as seasonal patterns, local demand curves, and customer segmentation, must be taken into account by dynamic pricing models. In addition to improving the accuracy of price-setting algorithms, route-based knowledge guarantees that airlines maintain their competitiveness while optimizingoptimizing profits. Airlines can customize their pricing strategies for every flight path by incorporating route data into reinforcement learning models such as BDQN.

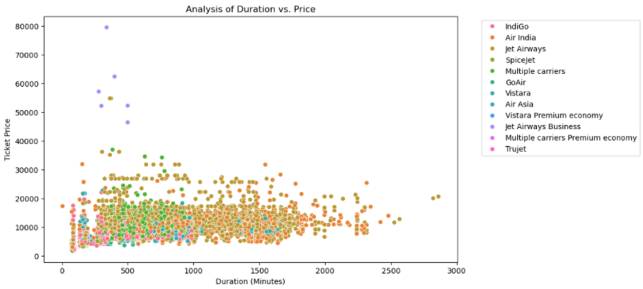


```
Average Price per Route:
Route
Chennai -> Kolkata         4789.892388
Mumbai -> Hyderabad        5059.708752
Banglore -> Delhi          5143.918577
Kolkata -> Banglore        9158.389411
Delhi -> Cochin           10539.439057
Banglore -> New Delhi     11917.716738
Name: Price, dtype: float64
```

C. Flight Duration vs. Cost of Ticket

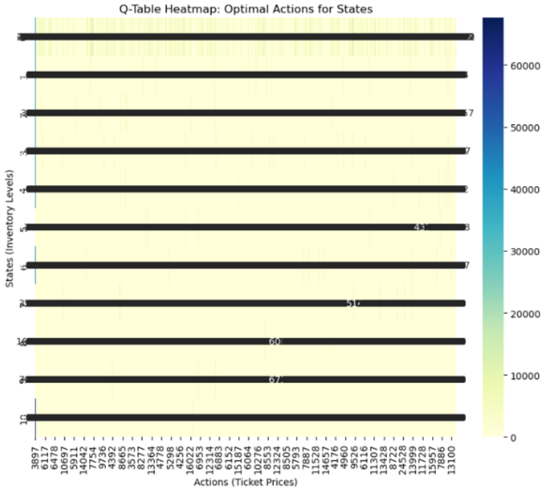
A general trend of rising costs with longer travel times can be seen in the relationship between flight duration and ticket price. A scatter plot, however, shows considerable variability within this pattern upon closer inspection. Even though longer flights are usually more expensive, many short-duration flights are also costly, which is a result of other factors like the popularity of the destination, the time it takes to book, and the class of travel. Outliers are particularly noticeable on long-haul flights, indicating that demand spikes, strategic pricing, and stopovers impact these costs and travel time. This intricacy shows that conventional models considering duration or distance cannot explain pricing decisions.

More accuracy is provided by multi-variable methods that consider behavioral and temporal factors. These findings support the need for dynamic pricing schemes considering demand elasticity, booking time, passenger segmentation, and travel time. Airlines can maximize revenue from various customer segments by matching price to value perception across flight durations.



According to this distribution, most travelers plan their trips balanced. From a pricing standpoint, the substantial presence of strategic customers calls for applying dynamic pricing models that predict customer behavior over time. Pricing optimization gains a temporal dimension because these customers are more likely to act on anticipated fare trends or wait for better offers.

On the other hand, smaller Myopic segment frequently accepts higher prices in exchange for immediacy and makes impulsive or early reservations. To set unique pricing strategies that can dynamically adjust based on real-time demand signals and passenger profiles, airlines must thoroughly understand this behavioral segmentation. That maximizes revenue and reduce inventory risk.



Closely followed by strategic passengers, this indicates a high demand for responsive, intelligent pricing systems that incentivize patience and information gathering. The fact that myopic passengers continue to be the smallest group suggests that there is little room for profit through early high pricing.

According to insights, airlines must adjust their prices according to time and in response to anticipated passenger reactions. For example, offering high prices at the outset might not turn off Neutral or Myopic customers. Still, it might cause Strategic customers to put off purchases, lowering early revenue. Dynamic pricing mechanisms, particularly those powered by reinforcement learning such as Q-learning or BDQN, can instantly adjust such behavior patterns. To maintain profitability for all passenger types, segment-aware algorithms test and modify prices over time, matching prices to perceived value and behavioral patterns.

VI. Conclusion

With an emphasis on passenger behavior, particularly in non-myopic (strategic) contexts, this study investigated using a deep exploration-based reinforcement learning (RL) framework in dynamic pricing for the Indian airline industry. By accommodating strategic and myopic customer types, the suggested BDQN (Bootstrapped Deep Q-Network) model showed enhanced pricing policies, leading to increased revenue and better seat utilization. BDQN successfully learned to alternate high and low prices over time by modeling a scenario with 18 available seats and different price sets. Real-world fare data from routes like Bangalore to Delhi, where prices range from 3,257 to 25,913, was in line with this approach. By offering discounts to early buyers and higher prices later to maximize revenue from less price-sensitive travelers, the model took advantage of the wide range of willingness to pay.

Although there are obvious revenue benefits to RL-based dynamic pricing, there are also issues with customer perception. Abrupt fare changes may affect passenger satisfaction and trust in a price-sensitive market like India. Airlines may need to smooth out price adjustments or limit abrupt increases to balance fairness and profitability. The long-term sustainability of customer loyalty and goodwill requires this consideration. The study has limitations despite its success. The model assumes a discrete and comparatively small set of prices. Value-based approaches like BDQN might not work well in real-world settings where pricing can be more precise. Policy-based DRL algorithms that are more appropriate for continuous action spaces may be investigated in future research. Furthermore, outside factors like holidays, the weather, and rival prices are not considered by our model. These elements can be added to the Markov Decision Process (MDP) by enlarging the state space to improve responsiveness and realism.

Furthermore, our simulations used limited types, even though the RL agent doesn't rely on known probability distributions. Future studies should examine a range of reservation prices, patience, and passenger arrival distributions to prevent overfitting. There are significant ramifications for the Indian airline sector. While considering the peculiarities of each market, adaptive dynamic pricing models adapted to changing passenger behaviors can improve revenue management and seat occupancy. RL-based pricing provides a data-driven strategy to successfully satisfy customer preferences, which range from value-driven to convenience-oriented.

In conclusion, BDQN offers a solid foundation for dynamic airline pricing in non-myopic scenarios; however, practical implementation will require improving realism, considering outside influences, and guaranteeing pricing equity. The study offers a strong starting point and new directions for further investigation into responsive, intelligent pricing systems in the airline industry.

VII. Recommendation and Implications

The study's conclusions lead to the following suggestions and ramifications for Indian regulators and airlines:

1) To effectively respond to real-time demand variations in the Indian market, airlines should implement dynamic pricing tools based on reinforcement learning, like DQN or BDQN. Through ongoing learning and policy improvement, these models can dynamically modify fares in response to local demand spikes, regional events, and holidays, maximizing revenue.
2) Pricing models should include customer segmentation. Airlines can adjust their pricing strategies by classify-

ing passengers according to their price sensitivity and patience levels. One way to maximize seat occupancy and revenue across various customer types is to reserve premium last-minute pricing for business travelers and offer discounted fares in advance for leisure travelers.

3) Fairness and transparency in pricing procedures must be given top priority. Airlines should put protections in place, like limits on fare increases between updates, and give customers accurate and consistent pricing information. Visible price ranges or fare calendars are two tools that can improve customer trust and lessen annoyance brought on by abrupt price changes.
4) Airlines should monitor consumer opinions and comments about their pricing strategies. Airlines should modify their tactics if travelers show discontent or uncertainty regarding fare structures. Pricing models are guaranteed to align with consumer expectations through ongoing feedback gathering, perhaps through surveys or complaint analysis.
5) Guidelines on dynamic pricing practices ought to be supplied by regulatory organizations like the Ministry of Civil Aviation and the DGCA. To safeguard consumer interests while permitting pricing flexibility, policies include mandatory disclosure of pricing logic, prohibitions on exorbitant last-minute pricing, and fair practice audits.
6) It is crucial to inform customers about dynamic pricing. Airlines should create resources like mobile apps, fare alerts, or tutorials to assist passengers in understanding fare patterns. Passengers are more likely to accept dynamic pricing as a normal market aspect and are less likely to feel misled when informed.
7) These suggestions seek to strike a balance between airline revenue targets, customer satisfaction, and regulatory monitoring. In the cutthroat Indian aviation market, implementing fairness, transparency, and reinforcement learning-based intelligent pricing strategies can boost profitability and passenger trust.

References

[1] Jo, S., Lee, G. M., & Moon, I. (2024). Airline dynamic pricing with patient customers using deep exploration-based reinforcement learning. *Engineering Applications of Artificial Intelligence*, 133, 108073.

[2] Ahiska, S.S., & King, R.E. (2015). Inventory policy characterisation methodologies for a single–product recoverable manufacturing system. *European Journal of Industrial Engineering*, 9(2), 222–243.

[3] Ahmadi, M., & Shavandi, H. (2014). Joint pricing and rationing in a production system with two demand classes. *European Journal of Industrial Engineering*, 8(6), 836–860.

[4] Aviv, Y., & Pazgal, A. (2008). Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management*, 10(3), 339–359.

[5] Bautista-Montesano, R., Galluzzi, R., Ruan, K., Fu, Y., & Di, X. (2022). Autonomous navigation at unsignalized intersections: A coupled reinforcement learning and model predictive control approach. *Transportation Research Part C*, 139, 103662.

[6] Bondoux, N., Nguyen, A.Q., Fiig, T., & Acuna-Agost, R. (2020). Reinforcement learning applied to airline revenue management. *Journal of Revenue and Pricing Management*, 19(5), 332–348.

[7] Cao, P., Fan, M., & Liu, K. (2015). Optimal dynamic pricing problem considering patient and impatient customers' purchasing behaviour. *International Journal of Production Research*, 53(22), 6719–6735.

[8] Caro, F., & Gallien, J. (2012). Clearance pricing optimization for a fast-fashion retailer. *Operations Research*, 60(6), 1404–1422.

[9] Den Boer, A.V. (2015). Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*, 20(1), 1–18.

[10] Dixit, A., & ElSheikh, A.H. (2022). Stochastic optimal well control in subsurface reservoirs using reinforcement learning. *Engineering Applications of Artificial Intelligence*, 114, 105106.

[11] Flapper, S., Gayon, J.-P., & Vercraene, S. (2012). Control of a production–inventory system with returns under imperfect advance return information. *European Journal of Operational Research*, 218(2), 392–400.

[12] Gosavi, A., Bandla, N., & Das, T.K. (2002). A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking. *IIE Transactions*, 34(9), 729–742.

[13] Hafez, M.B., Weber, C., Kerzel, M., & Wermter, S. (2019). Efficient intrinsically motivated robotic grasping with learning-adaptive imagination in latent space. In *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob)*, IEEE, pp. 1–7.

[14] Lawhead, R.J., & Gosavi, A. (2019). A bounded actor–critic reinforcement learning algorithm applied to airline revenue management. *Engineering Applications of Artificial Intelligence*, 82, 252–262.

[15] Otero, D.F., & Akhavan-Tabatabaei, R. (2015). A stochastic dynamic pricing model for the multiclass problems in the airline industry. *European Journal of Operational Research*, 242, 188–200.

[16] Victor, V. (2020). An experimental research on the consumer response towards online personalised pricing strategies: A comparative study between Indian and Malaysian online consumers. M.Sc. thesis, Szent István University, Gödöllő, Hungary.

[17] Victor, V., Thoppan, J., Jeyakumar Nathan, R., & Farkas, M. (2018). Factors influencing consumer behavior and prospective purchase decisions in a dynamic pricing environment—An exploratory factor analysis approach. *Social Sciences*, 7(9), 153.

[18] Gao, J., Le, M., & Fang, Y. (2022). Dynamic air ticket pricing using reinforcement learning method. *RAIRO-Operations Research*, 56, 2475–2493.

[19] Lu, R., Hong, S.H., & Zhang, X. (2018). A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. *Applied Energy*, 220, 220–230.

[20] Raju, C.V.L., Narahari, Y., & Ravikumar, K. (2006). Learning dynamic prices in electronic retail markets with customer segmentation. *Annals of Operations Research*, 143, 59–75.

[21] Rana, R., & Oliveira, F.S. (2015). Dynamic pricing policies for interdependent perishable products or services using reinforcement learning. *Expert Systems with Applications*, 42, 426–436.

[22] Collins, L., & Thomas, L. (2013). Learning competitive dynamic airline pricing under different customer models. *Journal of Revenue and Pricing Management*, 12, 416–430.

[23] Cao, J., Liu, Z., & Wu, Y. (2020). Learning dynamic pricing rules for flight tickets. In *Knowledge Science, Engineering and Management*, Springer.

[24] Liu, Q., & Zhang, D. (2013). Dynamic pricing competition with strategic customers under vertical product differentiation. *Management Science*, 59, 84–101.

[25] Levin, Y., McGill, J., & Nediak, M. (2009). Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *INFORMS Journal on Computing*, 55, 32–46.

[26] Bertsaks, D. (2000). *Dynamic Programming and Optimal Control*. Athena Scientific.

[27] Bobbio, A., Horváth, A., Scarpa, A., & Telek, M. (2003). Acyclic discrete phase type distributions: Properties and a parameter estimation algorithm. *Performance Evaluation*, 54(1), 1–32.

[28] Škare, V., & Gospić, D. (2015). [Title not provided]. *Vol. 63/No. 4*, 515–528. [Please update with correct title if known].

[29] Fiig, T., Isler, K., Hopperstad, C., & Belobaba, P. (2010). Optimization of mixed fare structures: Theory and applications. *Journal of Revenue and Pricing Management*, 9, 152–170.

[30] Williams, K.R. (2020). Dynamic airline pricing and seat availability. *Cowles Foundation for Research in Economics, Yale University*, Working Paper 208281.