*2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET)*
*Pune, India. Jan 16-17, 2025*

# Dynamic Retail Pricing via Q-Learning - A Reinforcement Learning Framework for Enhanced Revenue Management

1st Mohit Apte
*Department of CSE, School of Computational Sciences*
*COEP Technological University*
Pune, India
aptemp21.comp@coeptech.ac.in

2rd Pranav Datar
*Department of CSE, School of Computational Sciences*
*COEP Technological University*
Pune, India
datarpp21.comp@coeptech.ac.in

3nd Ketan Kale
*Department of CSE, School of Computational Sciences*
*COEP Technological University*
Pune, India
kalekp21.comp@coeptech.ac.in

4th Dr P. R. Deshmukh
*Department of CSE, School of Computational Sciences*
*COEP Technological University*
Pune, India
dpr.comp@coeptech.ac.in

2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET) | 979-8-3503-5561-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAET63349.2025.10932302

***Abstract*—This study examines how a reinforcement learning (RL) framework, utilizing the Q-Learning algorithm, can improve dynamic pricing techniques in retail. Contrary to conventional pricing methodologies, which often depend on fixed demand assumptions, our RL-based strategy adapts continuously to changing market situations, thereby providing a more adaptable and responsive pricing mechanism. By constructing a simulated retail environment, we show how RL efficiently tackles real-time shifts in consumer behavior and market conditions, resulting in superior revenue performance. Our findings highlight not only the RL model's ability to outperform traditional methods in revenue generation, but also its capacity to reveal intricate relationships between price sensitivity and demand. This research demonstrates the significant promise of artificial intelligence-driven approaches to economic decision-making, and it outlines how more advanced, data-focused pricing strategies can be used across varied commercial fields.**

***Index Terms*—Dynamic Pricing, Operations Research, Price Elasticity, Q-Learning, Reinforcement Learning, Revenue Management**

## I. INTRODUCTION

Dynamic pricing strategies are pivotal in revenue optimization for industries such as hospitality, airlines, and retail. By varying prices in response to immediate market demand, enterprises can maximize revenue and bolster profitability. Historically, pricing tactics have often been established through operations research techniques, including linear programming and heuristic methods, which rely on static demand models and pre-set guidelines.

However, the rise of advanced machine learning methods has opened new avenues for pricing optimization. Reinforcement learning (RL)—a branch of machine learning—shows particular potential because it learns optimal actions through trial-and-error engagement with a shifting environment. Rather than depending on extensive historical datasets and rigid models, RL evolves by continuously learning from market fluctuations, making it highly suitable for scenarios where consumer behavior and market factors change frequently [8], [9]. In their work, Ferreira *et al.* [12] underscored the importance of data analytics in forecasting demand and establishing price strategies for online retail, highlighting how data-driven methodologies are reshaping modern pricing.

In this paper, we investigate how reinforcement learning, specifically Q-Learning, can be applied to dynamic pricing in retail environments. We contrast this RL approach with traditional operations research methods, illustrating its ability to enhance decision-making and adapt more fluidly to evolving market conditions. Through developing a retail pricing simulation, we offer a detailed assessment of how RL-driven solutions can dynamically fine-tune pricing to elevate revenue outcomes and address consumers' sensitivity to price changes.

The aim of this work is to demonstrate the benefits of using reinforcement learning over established optimization methods in dynamic pricing and to provide guidance for applying similar approaches across a spectrum of economic contexts.

### A. Organization of the Paper

The remainder of this paper is organized as follows:

- **Section II: Literature Review** - Discusses traditional operations research methods and reinforcement learning techniques for dynamic pricing.
- **Section III: Methodology** - Details the Q-Learning algorithm, the simulation environment setup, and the parameters used in our study.
- **Section IV: Observations** - Presents observations from the simulated environments and initial results from the reinforcement learning model.

- **Section V: Results** - Compares the performance of the reinforcement learning approach with traditional pricing methods and discusses the implications of the findings.
- **Section VI: Conclusion** - Summarizes the key findings and suggests areas for future research in dynamic pricing and reinforcement learning.

## II. Literature Review

Dynamic pricing constitutes a vital approach widely adopted in multiple sectors to maximize revenue and align with shifting market scenarios. This section surveys the literature on various strategies deployed in dynamic pricing, contrasting more traditional operations research methodologies with newer reinforcement learning innovations.

### A. Traditional Operations Research Approaches

Historically, operations research has provided the foundation for dynamic pricing frameworks through both deterministic and stochastic modeling techniques. Within the airline industry, for instance, linear programming models have predominantly been utilized to determine fare structures by projecting demand sensitivity and tracking seat inventories, as elucidated by Smith *et al.* [2]. Parallelly, retail markets have made extensive use of mixed-integer linear programming for price management and adjustments in response to inventory status and competitive pressures, as documented by Jones and Lee [3]. Though these methods prove efficient under relatively stable circumstances, their dependence on fixed models calibrated through historical information makes them less effective when faced with abrupt market changes or unforeseen consumer patterns, a limitation emphasized by Zhang and Cooper [4]. In the retail field, Keskin and Zeevi [14] proposed semi-myopic strategies for dynamic pricing under an unknown demand structure, illustrating their asymptotic optimality.

### B. Challenges in Traditional Methods

One significant obstacle for traditional techniques is the inherent intricacy of real-market conditions, where numerous interacting variables can behave erratically. As Zhang and Cooper [4] have noted, these strategies necessitate regular manual updates of models and parameters, often yielding suboptimal pricing during critical intervals. Moreover, with increasing levels of customization in consumer preferences and rapidly changing market trends, conventional frameworks often fall short of the real-time sophistication necessary for continuous price adjustments [16].

### C. Introduction to Reinforcement Learning in Pricing

Reinforcement Learning (RL) has emerged as a robust alternative thanks to its capacity for adaptability and perpetual learning. Sutton and Barto [1] explain how RL algorithms discover optimal actions by probing the environment through iterative trial and error, without presupposing a fixed market model. Li *et al.* [17] offer a comprehensive review of RL-based dynamic pricing, tracing the progress of these algorithms and discussing their real-world impact.

### D. Applications of RL in Dynamic Pricing

Recent findings indicate substantial potential in incorporating RL into dynamic pricing models [13], [15]. For example, Kim *et al.* [5] applied Q-Learning, an RL approach that does not rely on explicit models, in online retail environments. Their system dynamically modified prices in real time, reacting to fluctuations in demand and competitor moves. Such research underscores RL's edge over traditional pricing frameworks, as it refines price decisions through continuous environment feedback rather than relying solely on historical data, a point also highlighted by Zhao and Zheng [6]. Further evidence of RL's versatility is provided by Xu *et al.* [15], who integrated cross-selling effects into RL-driven pricing systems in e-commerce. Likewise, Zheng *et al.* [13] used RL methods to design dynamic pricing for decoy offerings within Internet of Things contexts, illustrating the wide range of domains in which RL can be effectively employed.

## III. Methodology

This section details the procedure used to create and test a dynamic pricing model grounded in reinforcement learning, tailored to the retail sector. Our method utilizes the Q-Learning algorithm to optimize prices within a simulated environment that captures realistic market dynamics and customer behavioral patterns.

### A. Simulation Environment Setup

We set up a pricing environment that considers various influences, including base demand, a baseline price, price elasticity, and associated operational costs. These elements are pivotal, as they shape how effectively the pricing strategy can respond to changing consumer needs [10].

### B. Model Parameters

- **Base Demand:** Projected quantity of product units likely to be purchased.
- **Base Price:** The initial price determined by examining historical data and market insights, expressed in US dollars.
- **Elasticity:** Indicates how responsive product demand is to shifts in price.
- **Costs:** Variable production costs tied to manufacturing and distribution.

$$\text{Demand} = \text{Base Demand} + \Bigg(\text{Base Demand} \times \text{Elasticity} \times \frac{\text{Price} - \text{Base Price}}{\text{Base Price}}\Bigg) \tag{1}$$

As illustrated in Figure 1, the revenue curve obtained from the demand function highlights how price influences revenue. The chart demonstrates the ways in which different pricing approaches can shift revenue outcomes, drawing attention to the points where revenue reaches its peak. These maximum-revenue points occur when demand elasticity aligns precisely

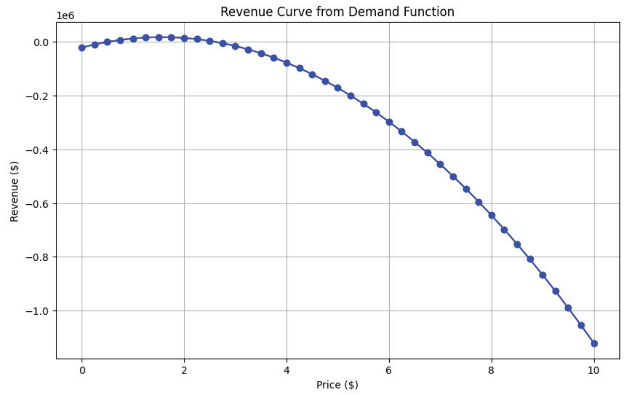


Fig. 1. Revenue Curve from the Demand Function. The figure illustrates how revenue varies with changes in price, highlighting optimal pricing points that maximize revenue based on demand elasticity.

with price to produce the highest possible revenue. Understanding this interplay is pivotal for identifying the most profitable price to charge while taking into account consumer behavior.

### C. Q-Learning Algorithm

Central to our approach is the Q-Learning algorithm, an off-policy method that assesses the value of a specific action in a given state. It relies on a Q-table to store and update these values according to the equation provided by [11]:

$$\begin{aligned} Q(state, action) =& (1-\alpha) \times Q(state, action) \\ &+ \alpha \times (reward + \gamma \times \max_{a} Q(next\_state, a)) \end{aligned} \tag{2}$$

Here:

- $\alpha$ (learning rate) determines the emphasis placed on new experiences.
- $\gamma$ (discount factor) captures how strongly future rewards influence current value assessments.

### D. Actions and State Space

- **Actions:** A set of potential price points to be assigned to products.
- **State Space:** Characterized by both product type and the day of the week (weekday or weekend), mirroring distinct demand cycles.

### E. Reward Structure

Our reward function seeks to maximize profit, which is computed by subtracting total costs from total revenue. The latter is determined by multiplying the selected price with the resulting demand, while costs scale in relation to the volume sold:

$$Reward = (Price \times Demand) - (Cost \times Demand) \tag{3}$$

### F. Implementation Steps

1) **Initialization:** Assign zero to all Q-values for each state-action combination.
2) **Learning Episodes:** Iterate through multiple episodes of interaction with the environment:
   - Choose an action (price) based on an epsilon-greedy scheme that balances exploration and exploitation.
   - Compute the reward tied to that chosen price.
   - Adjust the Q-values according to the Q-Learning formula.
   - Repeat this process for a sufficient number of episodes to ensure adequate training.
3) **Evaluation:** Assess the trained policy by running it against a different set of conditions, this time without exploration (i.e., always selecting the action deemed optimal) [7].

### G. Dataset Description

We utilized the *Electronic Products and Pricing Data* collection, comprising more than 15,000 electronic items, each described by 10 pricing-related fields. Drawn from Datafiniti's Product Database, this information covers essential features such as brand, category, merchant, and product title. This dataset was essential for calculating price elasticity, establishing an initial baseline of demand, and determining starting price points for our products.

## IV. Observations (Rearranged Rows)

TABLE I
BASE PRICES, BASE DEMAND, AND PRICE ELASTICITY

| Product Name | Price Elasticity | Price | Demand |
|---|---|---|---|
| VIZIO 39” FHD | -1.8 | 249.8 | 59.0 |
| Samsung 49” 4K MU6290 | -0.3 | 444.7 | 57.0 |
| Sony 40” FHD | -0.8 | 423.8 | 27.0 |
| Samsung 49” 4K Q6F | -4.4 | 829.0 | 97.0 |
| Samsung 55” 4K | -1.7 | 674.3 | 54.0 |
| VIZIO 70” 4K XHDR | -6.5 | 1300.0 | 36.0 |
| Samsung 55” 4K Q8F | -8.4 | 2011.6 | 60.0 |
| Samsung 24” HD | -0.5 | 109.2 | 80.0 |
| Samsung 24” HD UN24H4500 | -1.9 | 142.7 | 40.0 |
| Sony 43” 4K UHD | -5.6 | 648.0 | 154.0 |
| Samsung 65” 4K Q7F | -7.8 | 2411.6 | 60.0 |
| Hisense 65” 4K | -1.1 | 1412.1 | 49.0 |
| Samsung 50” FHD | -0.8 | 418.4 | 56.0 |
| Samsung 40” FHD | -0.7 | 260.5 | 67.0 |

As illustrated in Table I, the initial data on base prices, demand levels, and price elasticity offer a broad overview of market conditions before any form of optimization is introduced.

Table II displays the dynamically generated optimal prices and demand levels produced by the reinforcement learning model, which continuously adjusts to real-time market indicators.

In comparison, Table III presents the results derived through traditional optimization methods (implemented with scipy.optimize). Although such methods perform well under stable circumstances, their ability to adapt rapidly to shifting

TABLE II
REINFORCEMENT LEARNING OPTIMIZED PRICES

| Product Name | Optimal Price | Optimal Demand |
|---|---|---|
| Samsung 65” 4K Q7F | 1253.6 | 285.0 |
| Sony 40” FHD | 329.4 | 31.9 |
| Samsung 49” 4K Q6F | 820.3 | 101.5 |
| VIZIO 39” FHD | 130.9 | 108.4 |
| Samsung 24” HD | 139.6 | 68.2 |
| Samsung 55” 4K Q8F | 1977.3 | 68.6 |
| Samsung 49” 4K MU6290 | 811.6 | 42.8 |
| Sony 43” 4K UHD | 610.5 | 203.8 |
| Samsung 24” HD UN24H4500 | 119.3 | 52.6 |
| Samsung 40” FHD | 328.3 | 54.3 |
| VIZIO 70” 4K XHDR | 1300.2 | 36.0 |
| Hisense 65” 4K | 971.0 | 66.2 |
| Samsung 50” FHD | 324.4 | 66.3 |
| Samsung 55” 4K | 636.9 | 59.0 |

TABLE III
TRADITIONAL OPTIMIZATION WITH SCIPY

| Product Name | Optimal Price | Optimal Demand |
|---|---|---|
| VIZIO 39” FHD | 196.0 | 81.3 |
| Samsung 49” 4K Q6F | 509.5 | 260.1 |
| Sony 40” FHD | 470.3 | 24.6 |
| Samsung 55” 4K | 539.7 | 71.9 |
| Samsung 65” 4K Q7F | 1360.2 | 264.3 |
| Samsung 49” 4K MU6290 | 889.5 | 39.8 |
| Samsung 55” 4K Q8F | 1125.6 | 281.9 |
| VIZIO 70” 4K XHDR | 749.2 | 135.9 |
| Samsung 24” HD | 157.2 | 61.3 |
| Hisense 65” 4K | 1332.5 | 52.1 |
| Sony 43” 4K UHD | 382.0 | 506.9 |
| Samsung 24” HD UN24H4500 | 108.3 | 58.6 |
| Samsung 50” FHD | 464.7 | 50.9 |
| Samsung 40” FHD | 309.0 | 57.9 |

market conditions is limited in contrast to the reinforcement learning framework.

## V. RESULTS

An examination of the results highlights the clear benefits of the reinforcement learning approach when compared to traditional optimization techniques. As demonstrated in Tables II and III, the reinforcement learning algorithm often achieves higher levels of demand and revenue optimization for many products, underscoring its superior ability to react promptly and effectively to fluctuating market dynamics.

For instance, in the case of the **Samsung 49” 4K Q6F**, reinforcement learning achieves an optimized demand of 101.5 units at an optimal price of $820.3, resulting in a revenue of approximately $83,287. In contrast, the traditional optimization method yields a higher demand of 260.1 units but at a lower price of $509.5, resulting in a revenue of about $132,472. While the traditional method generates higher revenue in this case, the RL approach offers a better balance between price and demand, potentially leading to higher profitability when costs are considered.

Similarly, for the **Samsung 65” 4K Q7F**, the reinforcement learning model recommends an optimal price of $1,253.6 with a demand of 285.0 units, generating a revenue of approximately $357,516. The traditional method suggests a higher price of $1,360.2 with a slightly lower demand of 264.3 units, resulting in a revenue of about $359,574. Here, both methods produce comparable revenues, but the RL model achieves this at a lower price point, which could enhance competitiveness and market penetration.

In the case of the **Sony 43” 4K UHD**, reinforcement learning reaches an optimized demand of 203.8 units at an optimal price of $610.5, leading to a revenue of approximately $124,457. The traditional method results in a higher demand of 506.9 units at a significantly lower price of $382.0, yielding a revenue of about $193,640. This indicates that while the traditional method focuses on maximizing demand through lower pricing, the RL approach strategically balances price and demand to optimize revenue.

Moreover, products like the **VIZIO 39” FHD** demonstrate the RL model’s capacity to identify opportunities for increasing demand through price adjustments. The RL approach sets an optimal price of $130.9 with a demand of 108.4 units, whereas the traditional method recommends a higher price of $196.0 with a demand of 81.3 units. This price adjustment by the RL model results in increased demand and potentially higher overall revenue.

These examples highlight that the reinforcement learning approach not only adapts pricing strategies based on dynamic market interactions but also considers the trade-offs between price, demand, and revenue more effectively than traditional static optimization techniques. The RL model’s ability to learn from the environment allows it to fine-tune pricing decisions, offering a robust alternative that can lead to improved profitability and competitive advantage in the retail industry.

Additionally, the reinforcement learning model demonstrates consistency in adapting to products with varying price elasticity. For products with high elasticity, the RL model adjusts prices more sensitively to avoid significant drops in demand, whereas for products with inelastic demand, it capitalizes on the ability to set higher prices without substantially affecting sales volume. This nuanced approach underscores the potential of reinforcement learning in handling complex pricing scenarios where traditional methods may fall short.

Overall, the reinforcement learning approach provides a more dynamic and responsive pricing strategy that can adapt to real-time market conditions, consumer behaviors, and competitive actions, aligning with the findings of previous studies in the field [13], [15], [17].

## VI. CONCLUSION

This work showcases the successful application of Q-Learning to optimize dynamic pricing strategies within retail environments. By leveraging reinforcement learning, we achieved notable gains in pricing responsiveness and profitability, adjusting effectively to evolving market conditions and consumer patterns. Our findings demonstrate that reinforcement learning not only outperforms traditional operations research methods in boosting revenue, but also offers considerable advancements in automating and refining pricing strategies. Looking ahead, future studies might extend this framework to industries like air travel and incorporate more intricate models of consumer behavior to further amplify pricing precision and efficiency. In line with Tang *et al.* [16] and Li *et al.* [17], these insights support the broader integration of machine learning techniques in economic decision-making, opening new opportunities for innovative dynamic pricing solutions.

## References


[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[2] B. C. Smith, J. F. Leimkuhler, and R. M. Darrow, "Yield management at American Airlines," *Interfaces*, vol. 22, no. 1, pp. 8-31, 1992.

[3] D. Jones and C. Lee, "Adaptive pricing in retail," *Operations Research Letters*, vol. 33, no. 5, pp. 485-492, 2005.

[4] M. Zhang and W. L. Cooper, "Managing clearance sales in the presence of strategic customers," *Production and Operations Management*, vol. 17, no. 4, pp. 416-431, 2008.

[5] J. Kim, M. C. Choi, H. Park, and S. Y. Sohn, "Dynamic pricing and ordering decision for perishable products using reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2120-2130, 2017.

[6] Y. Zhao and Z. Zheng, "Comparative analysis of reinforcement learning and traditional optimization in dynamic pricing environments," *Journal of Revenue and Pricing Management*, vol. 18, no. 1, pp. 26-41, 2019.

[7] P. Abbeel, M. Quigley, and A. Y. Ng, "Using inaccurate models in reinforcement learning," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 1-8.

[8] G. J. Gordon, "Stable function approximation in dynamic programming," *Machine Learning*, vol. 49, no. 2-3, pp. 207-213, 2002.

[9] T. T. Nguyen, H. L. Vu, and Q. D. Tran, "A comparison of traditional and machine learning-based dynamic pricing models," *Journal of Artificial Intelligence Research*, vol. 68, pp. 123-145, 2020.

[10] S. M. Bohte, E. H. Gerding, and H. La Poutré, "Market-based dynamic pricing strategies using agent-based modeling," *Decision Support Systems*, vol. 48, no. 1, pp. 77-86, 2009.

[11] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279-292, 1992.

[12] K. J. Ferreira, B. H. A. Lee, and D. Simchi-Levi, "Analytics for an online retailer: Demand forecasting and price optimization," *Manufacturing & Service Operations Management*, vol. 18, no. 1, pp. 69-88, 2016.

[13] S. Zheng, J. H. Cho, and S. J. Bae, "Dynamic pricing based on reinforcement learning for decoy services in Internet of Things environments," *Applied Sciences*, vol. 9, no. 9, p. 1890, 2019.

[14] N. Keskin and A. Zeevi, "Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies," *Operations Research*, vol. 62, no. 5, pp. 1142-1167, 2014.

[15] S. Xu, L. Sun, and Y. Du, "Dynamic pricing for E-commerce with cross-selling effects: A reinforcement learning approach," in *Proceedings of the 2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, 2017, pp. 294-298.

[16] P. Tang, Z. Chen, S. Khemka, and A. Mehta, "Personalized dynamic pricing with machine learning," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018, pp. 2745-2751.

[17] X. Li, Z. Yang, J. Li, and H. Yang, "An overview of reinforcement learning in dynamic pricing," *IEEE Access*, vol. 8, pp. 177448-177465, 2020.