# Survival Analysis in Online Retail

***Case Background***[1]

The proliferation of web 2.0 technologies since the early years of the millennium resulted in a new type of business commonly referred to as online retail. In a typical online retail business, potential shoppers visit a website that lists a set of products. These potential shoppers can browse the products, read online reviews and other information related to them, and decide whether they want to make a purchase. With the success of Amazon, eBay, Alibaba, and Etsy, many other players entered the online retail business. In addition, many brick-and-mortar businesses such as Walmart, Costco, Kroger, and Ikea also decided to improve their online presence by creating effective and easy-to-use online marketplaces, making this sector even more competitive. With the elevated competition, online retailers have tried a variety of strategies to create more value for online shoppers. Some tried unleashing the web 2.0's true potential by asking previous shoppers to review their purchases. They would then process these reviews to create a set of reputation scores (star ratings) for each product they list. Such information, which signals the quality of products, reduces the information asymmetry for online shoppers, helping them make their purchase decisions faster and easier. Over time, other online retailers also offered similar reputation systems for their products, eliminating the competitive advantage due to offering a reputation system.

[1] INTERNAL WORKING DRAFT. NOT FOR DISTRIBUTION OUTSIDE MCINTIRE. This case was prepared by Reza Mousavi, Assistant Professor of Commerce.

Furthermore, big players such as Amazon found a new competitive advantage. By consistently decreasing the click-to-door time (the number of days/ hours from purchase to delivery), Amazon outpaced its competition and grew into one of the most successful firms in U.S. Strategic decisions such as decreasing click-to-ship time, taking control of the last-mile deliveries, and optimizing warehouse locations helped Amazon to decrease its click-to-door time from almost 6 days in December 2015 to 3 days in March 2018.


**Figure 1. Amazon's Click-to-door Time Between December 2015 and March 2018**

Although in 2022 Amazon's click-to-door time was between 1 and 2 days, many other retailers competed with Amazon by leveraging their physical presence. Firms such as Walmart, Costco, and even tech firms such as Apple, started taking orders online and preparing them for customer pick-up at their numerous stores, effectively eliminating any significant time between ordering and delivering (pick-up). This meant that Amazon's original competitive advantage due to the decrease in click-to-door time became less competitive as other firms found ways to deliver their products to their customers faster.

The brief review of online retail sector provided above reveals a non-stop competition for attracting more customers to sell more products. Major players in online retail have found effective ways to make purchase decisions easier and deliveries faster. Many of these efforts required long-term, strategic commitments by firm leaders. Adding more warehouses, solving the last mile problem, or optimizing warehouse processes to decrease click-to-ship time are all examples of long-term and costly commitments.

On the other hand, many online retailers have found a much more flexible and much less costly way to create a competitive advantage: dynamic pricing. Dynamic pricing is a type of pricing strategy where instead of using a fixed price an adjustable price is used. The dynamic pricing model's fundamental premise is to provide the same product to various customer segments at various prices. As depicted in Figure 2, dynamic pricing results in an increase in demand and therefore more sales. Given that online retailers can easily adjust the content of their websites, they can instantly change their prices. For instance, in a dynamic pricing scheme, an online retailer can easily offer discounts to nudge online shoppers to make a purchase. With the digital transformation being underway during the past few years, online retailers can set their prices using automated systems. Advances in artificial intelligence have also fueled the use of dynamic pricing. Advanced machine learning models can be trained to learn when, for what product-shopper pairs, and how much discount/ price adjustment should be applied.

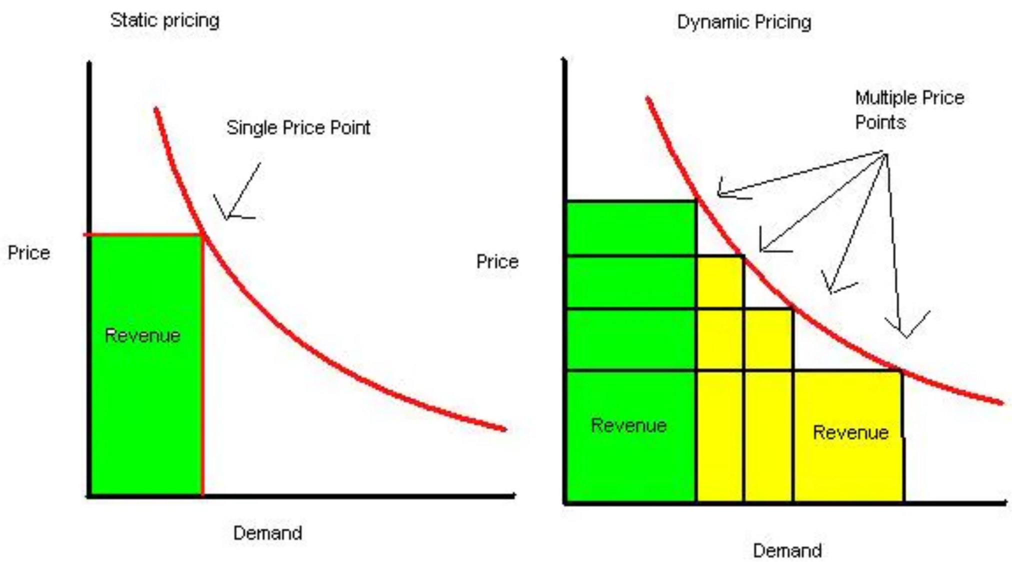


**Figure 2. Dynamic Pricing vs. Static Pricing**[2]

Although dynamic pricing was originally designed to alter the prices for different customer segments, it can be used to alter the prices for products that don't sell within a specific period of time. That is, dynamic pricing can be used to adjust the prices of products based on the expected time of sale. An online retailer can predict whether a product is likely to sell or not within a specific period. If the product is unlikely to sell, the retailer can offer incentives (such as discounts) to sell it.

That being said, dynamic pricing has its own challenges: on the one hand, offering unnecessary discounts for products that sell without any discounts would cut the profit. On the other hand, not offering discounts for products that may not sell well without any discounts would cut the revenue. To address this issue, online retailers have relied on advanced techniques in data science and machine learning to *predict* which products need discounts and when such discounts

[2] Source: https://www.priceintelligently.com/blog/dynamic-pricing-strategy

should be offered to online shoppers. Many online retailers also use dynamic pricing during special sales events. Such sales events are typically offered in a limited period of time and provide shoppers either with additional discounts or other perks. During these events, retailers offer incentives to sell their current inventory to have liquidity and space for restocking new products. Hence, dynamic pricing is a very effective tool to adjust the prices and nudge shoppers to buy before the end of the online sales event.

***Dynamic Pricing at ORX:***

In this case study, we want to help an online retailer, which we refer to as ORX, to use dynamic pricing during an online sales event. ORX offers a variety of products such as smartphones, TVs, computers, and kitchen appliances. ORX automatically captures and records online shoppers' clickstream data, products they view, date and time they make their decisions (either exit the product page or purchase the product) as well as a variety of information about the products listed. ORX typically uses monthly sales events to sell its existing products and restock for the next sales event. This business model helps ORX to be responsive to the seasonal trends. In addition, selling and restocking on a monthly basis helps ORX to quickly adapt when new consumer or technology trends emerge. Although being agile and responsive to market trends are great benefits of this business model, ORX needs to sell its items during the sales events. Otherwise, ORX would not have sufficient liquidity or the capacity to restock new items for the next event. To address this issue (to make sure items are sold within the sales event), ORX started using a dynamic pricing strategy to discount unsold items towards the end of each sales event. That policy, however, meant that ORX should wait until the last few days of the sales event, identify unsold items, and then offer discounts for those items. Over time, ORX identified two main issues with this dynamic pricing strategy:

1- Some of the customers who became familiar with this strategy learned how to strategically postpone their purchases until the last few days of the sales event to benefit from the potential discounts. This resulted in two additional problems: a) ORX needed to discount on more items than necessary, b) since more items were sold towards the end of the sales event, the volume of processing and shipping of the items became very imbalanced (low volumes during early days and high volume towards the end of the sales event). Consequently, the imbalanced volume resulted in either under-utilization of the resources during the early days of the event or over-utilization of the resources towards the end of the event.

2- For the new shoppers who were not familiar with the discount policy, they were more likely to see the original price rather than the discounted price because for most of the days during the sales event the items were still not discounted. For instance, a potential customer who would have purchased the item if it was discounted 5% visits ORX's website on day 10 of the sales event. Since the item is not discounted yet, the shopper will not purchase the item. Therefore, ORX misses selling the products to these shoppers.

To address these issues and offer the discounts intelligently, ORX is sharing its data with you. ORX is asking you to use their data and predict which items will not sell within 30 days so that they could offer the discounts on the early days of the sales event rather than waiting until the very end of the sales event.

In this case study, we only work with a subset of the data both in terms of the features and observations. Our dataset is based on data collected during an online sales event which started on October 1st, 2019 and lasted until October 31st, 2019. Although ORX sells a variety of items, we only focus on smartphone sales. Smartphones typically have the same fixed price across different online retailers. Therefore, even a marginal discount (e.g., 5% discount rate) may result in a significant boost in sales. That is why in this case we only focus on this type of product. Table 1 provides the names of the columns in the data along with their descriptions.

In this phase of the project, ORX is willing to only discount 10% of the listed items. Therefore, they are asking you to select the top 10% of the items that would be unsold at the end of the sales event if they are not discounted.

**Table 1. The Data Fields and Their Descriptions**

| **Column Name** | **Description** |
|---|---|
| id | Unique product identifier |
| product_id | Product subcategory identifier |
| brand | Product brand |
| price | The listed price including processing and handling fees |
| color | Product color |
| purchased | Whether the product was purchased or not (0: not purchased, 1: purchased) |
| days_on_market | The number of days the product was listed on the website. If the product was never purchased, the value for days_on_market will be 30. Otherwise, the value for days_on_market refers to the number of days until the item was sold. |

***Your Task:***

In your mod teams, use survival analysis to predict which products are likely to remain unsold after 30 days (i.e., on day 31):

1- Train multiple survival/ hazard models and evaluate them.

2- Select your best model.

3- Apply your best model to the Kaggle data.

4- Export the predicted scores and submit the exported file ("to_kaggle.csv") to Kaggle: https://www.kaggle.com/t/1fd732518dbd43539f115bf559ce5883

5- Find the top 10% of items in Kaggle data to discount. Export the file ("sample_top17.csv") and email it to mousavi@virginia.edu before 5 PM Eastern time on Tuesday 21st.

***Evaluation Criteria:***

Your submissions will be graded based on the following two criteria:

1- Your Kaggle score.

2- Percent of items in your "sample_top17.csv" that did not sell within 30 days.

Your final score is the average of the two scores above.

***Discussion Questions:***

Please be ready to discuss the followings items in class:

1- What are the benefits of dynamic pricing in online retail?

2- What are the main disadvantages of dynamic pricing in online retail?

3- Did you make any changes to the data pre-processing steps? What changes?

4- What was your best model? What other models you tried?

5- What are the top 2 most important features in your final model?

6- How can the predictive performance of the model affect dynamic pricing and sales?

7- How can the predictive performance of the model impact online retailers' competitive advantage?