Georgia Institute of Technology

Data Analytics in Business - MGT 6203

Group Project - Final Report

“Understanding pricing dynamics from bicycle sales”

Team # 35:

Ahsan Bilal (903962376)

Abdallah Zaid Alkilani (903946663)

Hemil Ray (903957567)

Azib Anees (903965929)

[Project Repository]

Background Information:

A major problem store owners face these days is not understanding how to price their products properly to increase sales across all seasons. There could be two stores in the same city applying the same pricing strategies for the same product type, however one would sell more than the other for no obvious reason, our project aims to answer this problem by providing more insights into what factors could influence increased sales figures.

The analysis of seasonal pricing patterns across different brands, within the same product category, holds immense significance for businesses aiming to optimize their profitability and competitive positioning. By delving into the intricate dynamics of how brands adjust their prices during various seasons, companies gain a strategic advantage in tailoring their pricing strategies to capitalize on consumer behavior and market trends.

Consumer perceptions and responses to price changes within different brands during distinct seasons represent a crucial dimension for businesses seeking to cultivate customer loyalty and satisfaction. By comprehending how consumers react to variations in brand-specific pricing, companies can tailor their marketing messages and promotional campaigns to resonate with consumer preferences. This consumer-centric approach not only bolsters brand loyalty but also provides a foundation for enhanced customer relationships and improved brand perception. Ultimately, the synthesis of these insights contributes to operational efficiency, ensuring businesses allocate resources effectively, optimize product assortments, and make informed decisions that align with both consumer expectations and competitive landscapes.

Business Justification and Objectives:

This project is on the sales of bicycles across three stores in the United States to understand the pricing patterns across different brands of the same product type and the seasonal pricing variations. The objective is to draw informative conclusions on what could influence the future sales of bicycles by uncovering factors influencing these variations and enhancing strategic decision-making for businesses. **This study could help store owners to make the right decisions at the right time to increase sales and better manage their inventory**.

Through primary and a series of supporting research questions, we are attempting to understand pricing dynamics across different stores. The primary research question that we are hoping to answer by the end is:

"How do pricing patterns change across different brands of the same product and what factors influence the observed variations in seasonal pricing among these brands?"

We have identified a set of supporting questions and investigated them. The analysis approach, key findings, interpretations, conclusions, and issues for each hypothesis are presented in this report.

If successful, this analysis can revolutionize how businesses approach seasonal pricing, leading to optimized profit margins and enhanced brand positioning. The monetary impact, if widely implemented, could result in substantial revenue growth, cost savings, and enhanced market competitiveness for businesses adopting optimized pricing strategies.

Dataset Origin and Data Cleaning Process:

The dataset used for the project can be accessed from this **link**. The dataset comprises 2000 rows and 22 columns. Originally, the dataset was in the form of a SQL Server database. The SQL server database is publicy available and can be accessed from this **link.**

To convert the database tables into the comprehensive analytical dataset (ADS) named BikeStore_ADS, we initiated the process from multiple data sources. Each source represented distinct aspects of the bike store's operations, such as order items, orders, customers, stocks, products, staff, brands, categories, and stores.

Following the data loading phase, we merged the data frames into a unified dataset using common columns shared across the data frames. For instance, joins were performed based on shared identifiers like Order_ID or Product_ID to link order items with orders, customers with orders, and products with their respective brands and categories.

Throughout the joining process, we encountered various data cleaning tasks such as handling missing values, standardizing column names, converting data types, and removing duplicates.

Upon completing the data merging and cleaning procedures, we consolidated the transformed data frames into a single analytical dataset named BikeStore_ADS. This unified data set contained comprehensive information about various entities, enabling in-depth analysis of the bike store's operations and performance. The modeling involved feature engineering that includes categorical features of every season and revenue calculation for pricing analysis.

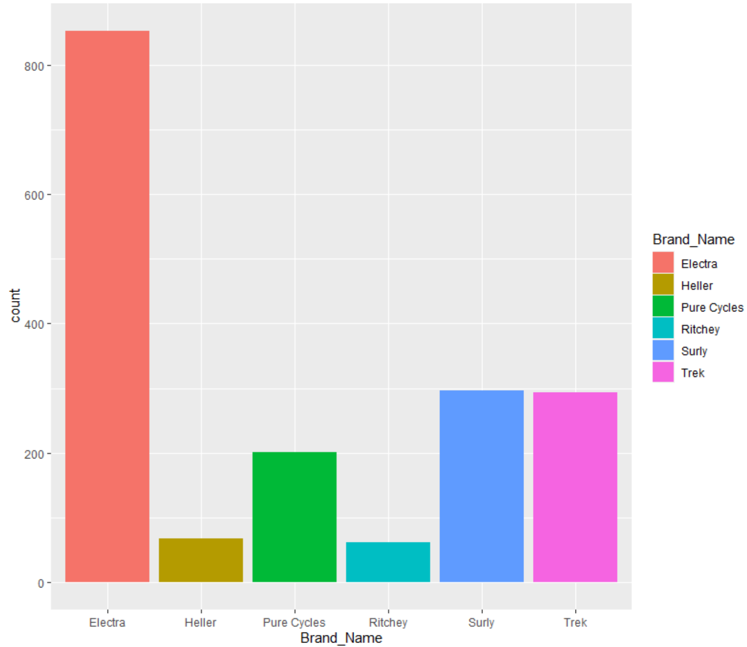
The **key variables** are list price, discount, quantity stocks, brand name, category name, order ID and shipped date. List price and brand name would be dependent variables as we are going to examine pricing dynamics based on independent variables like Order ID and quantity stocks.

Exploratory Data Analysis and Insights:

The table below on the bicycle types that each brand sells, with number of sales in 2016:

	Children Bikes	Comfort Bikes	Cruisers Bikes	Cyclocross Bikes	Electric Bikes	Mountain Bikes
Electra	228	209	416	0	0	0
Heller	0	0	0	0	0	67
Pure Cycles	0	0	201	0	0	0
Ritchey	0	0	0	0	0	62
Surly	0	0	0	152	0	145
Trek	0	0	0	0	73	220

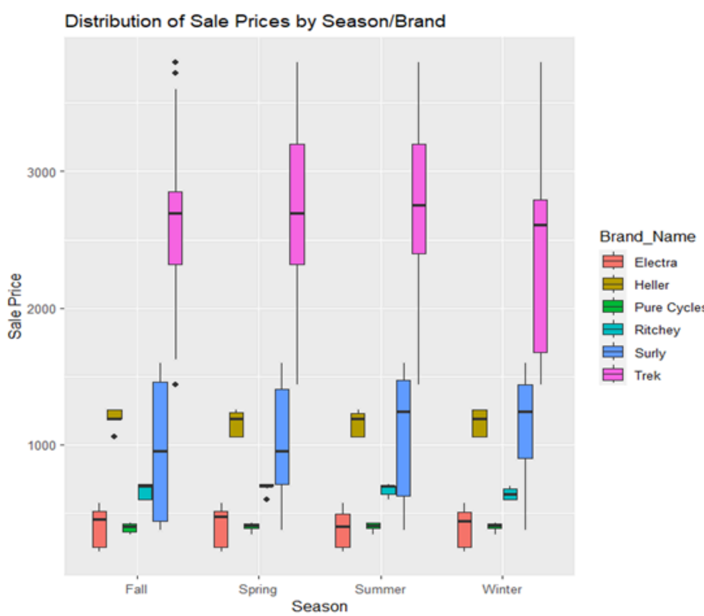
Below plot on number of sales across brands to assess if we have sufficient data points:



To understand the distribution of prices and brands across different seasons, there are two points that could be looked at:

- Comparing how brand prices differ within one season.
- Comparing how one brand prices change across all seasons.

<u>Note:</u> For below visuals in this section, *Sale Price = bike's list price - the discount rate*



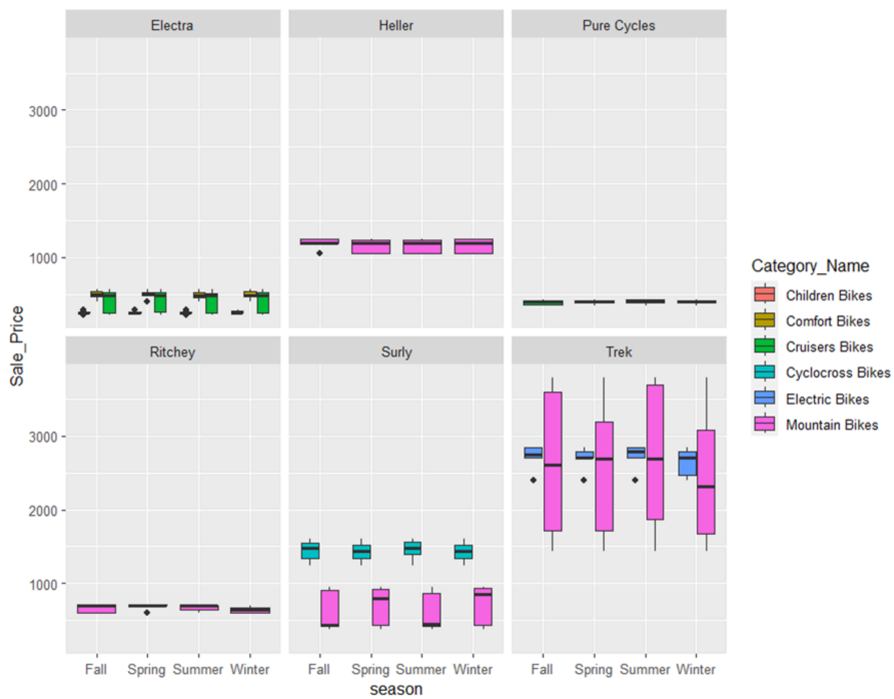
Boxplot showing distribution of sale price by brand per season

Trek brand had the most expensive sale prices, which dropped relatively the most in winter. One possible conclusion is that this brand only sells Electric & Mountain bikes, which could not be very popular during the winter season. Therefore, it might consider offering higher discounts in Winter.

Surly, the brand that sells Cyclocross & Mountain bikes has the second highest sale prices across all seasons. During Fall they dropped their prices the most, which could be to compete with the brand Trek that was selling more Mountain bikes. Therefore, Surly would have tried to drop their prices to increase sales.

Heller, a brand that sells only Mountain bikes, had relatively similar sale prices across all seasons, and their sales were the least compared to the brands that sell Mountain bikes, Trek & Surly. One possible reason is that they did not apply similar discounting techniques across seasons and ended up with lower total sales.

We can conclude that applying different discount strategies across and within seasons might have an influence in having higher sales figures for these brands.



Boxplot showing distribution of brand sale price by category per season

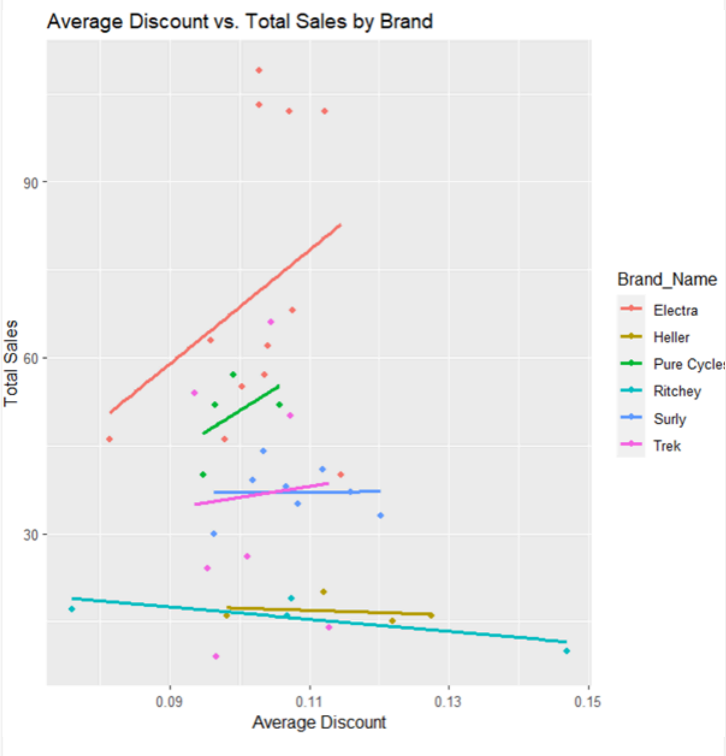
Trek's sale prices distribution for the Mountain bikes category varies greatly and are more expensive compared to other brands selling Mountain bikes. Since *Trek* sells the most of them, this could indicate that they use intelligent pricing/discounting methods that results in higher sales despite being more expensive.

Electra and *Pure Cycles* both sell Cruisers bikes, but *Electra* sold more than double the amount. This could be due to the varied selling prices across each season, creating the perception for customers that they are getting a better deal if they buy from this brand.

For other brands since the distribution of sale prices is not as high, it is difficult to draw further conclusions based on that.

Overall, from this initial analysis, applying different discount strategies across and within seasons seems to have a positive influence for higher sales figures, encouraging customers to buy more from such brands.

The plot below shows the relationship between Discount Rate and Sales Figures for each brand, based on the average discount and total number of sales for each brand/category/season:



- Electra brand appears to have a positive relationship between discount and total sales, sales figures increased with the increase of discount rate.
- The same could be noticed for the brand Pure Cycles, however since the discount range is minor, where the scale of the x-axis should be taken into consideration, showing there is a chance the increase in sales is seasonally driven.
- In general, from this plot, it was observed that higher discounts did not necessarily increase the sales figures for most of the brands.

It is interesting to see how for many brands the discounts given did not necessarily improve the sales figures, this could be tied to other reasons, for example seasons and store geographical location, that could have also impacted sales figures. **Further analysis would be needed before drawing final conclusions.**

Overview of Hypotheses:

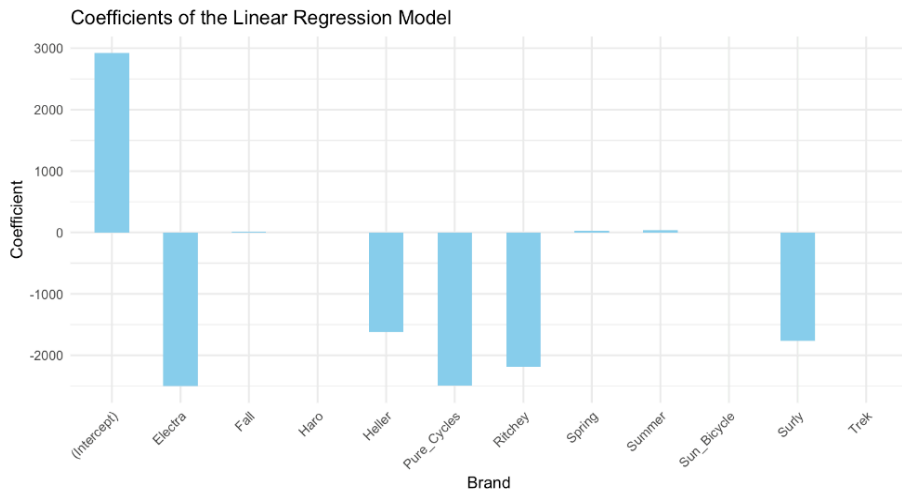
In this section, we will provide an overview of the hypotheses/supporting questions we have identified. Answering these questions will give us insights about patterns in our data that will link well with our problem statement and can be justified from a business perspective.

Hypothesis #1: Brand-Specific Seasonal Pricing

- Do different brands of the same product exhibit distinct seasonal pricing patterns?
- How does the magnitude of price changes vary among competing brands during different seasons?

To examine the pricing dynamics of different brands for the same product, we employed a linear regression analysis. This involved treating each brand as a categorical variable within the regression model, allowing us to discern the distinct pricing effects associated with individual brands.

The linear regression model employed in this analysis facilitated the estimation of coefficients for each brand variable, elucidating the significant differences in pricing strategies among the brands under investigation. Notably, brands such as Electra, Heller, Pure Cycles, Ritchey, and Surly exhibited substantial negative estimates, indicating lower average prices compared to the reference brand. These findings highlight the distinct pricing dynamics adopted by each brand, potentially reflecting differences in brand positioning, target market, or competitive strategies. Additionally, the presence of singularities for certain brands underscores potential challenges in estimating their specific pricing effects within the model. Despite these challenges, the regression analysis provided valuable insights into the brand-specific pricing behaviors within the market, shedding light on the competitive landscape and consumer preferences.



Bar graph depicting the coefficients of the linear model (List Price ~ Brand)

The visualization displays the coefficients obtained from the linear regression model. Each bar represents the coefficient associated with a specific brand.

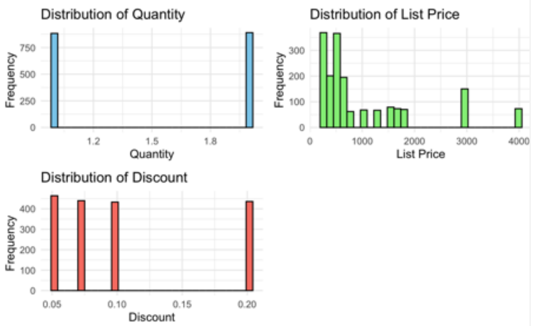
Interpreting the plot, positive coefficients suggest an increase in the predicted list price for the associated brand, while negative coefficients indicate a decrease. The magnitude of the coefficient represents the strength of the effect: larger coefficients signify a greater impact on the predicted list price.

Brands with taller bars have a stronger effect on the predicted list price, whereas those with shorter bars have a weaker influence.

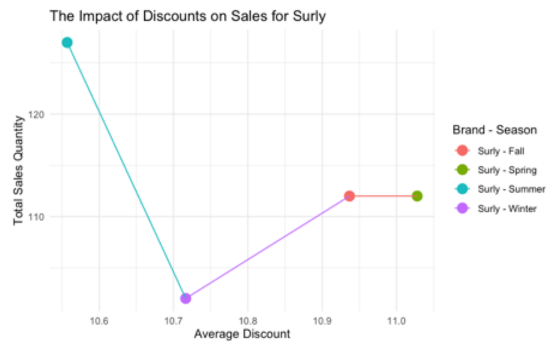
Hypothesis #2: Brand Strategies in Response to Seasons

- What strategies do individual brands employ to adjust pricing in response to seasonal changes?
- Are there brands that consistently adopt higher or lower prices during specific seasons?

In this section, we plan to study the dynamics between Discount offered vs the Impact on the Total amount of Sales. We track the total amount of sales by their seasons to identify trends across individual brands. This method also allows us to compare different brands across different seasons. We first created a season subset - Winter, Spring, Summer, Fall - that allows us to track the seasons. We start with a summary of our data and an example that showcases the impact of discount on total sales quantity on a specific brand by season.



Descriptive Stats Visualized



Sales vs. Discount for Surly across seasons

We first started with a linear regression model that grouped Quantity by Brand Name, then created a Linear Model where bikes sold are regressed against individual brands and a separate model by individual Categories. We found no statistical significance and rejected the null hypothesis as the P-values were greater than 0.05. Accounting for examples such as the Sales Vs. Discount for Surly across Seasons, we noticed a visual trend across all brands. Offering an average discount on a particular season resulted in a contrasting sales behavior for brands. We moved forward to investigate any statistical significance with a Linear Model accounting for each Brand and Season combination.

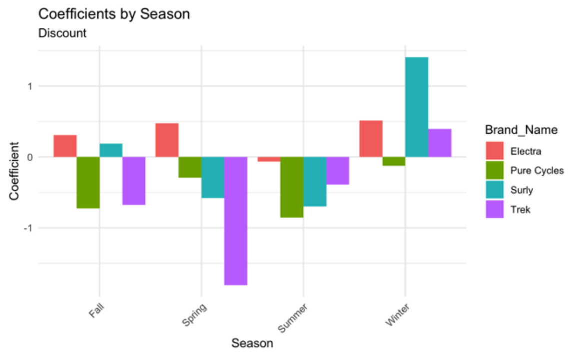


Image: Coefficient by Season (Discount)

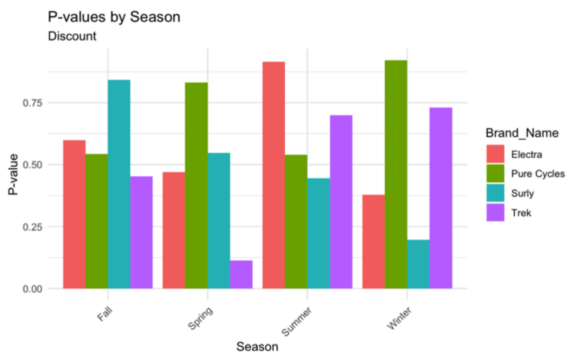


Image: P-Value by Season (Discount)

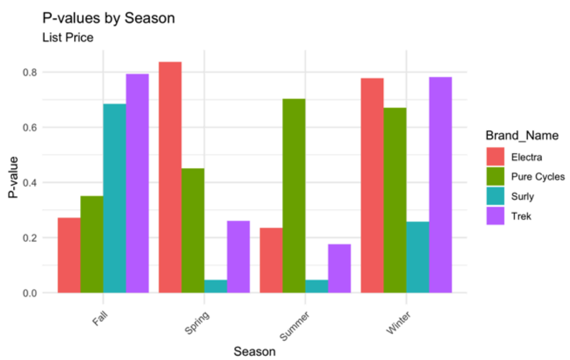


Image: P-Value by Season (List Price)

- The analysis of P-Values for brand names across their respective seasons revealed a critical insight: none of the observed P-Values meet the established threshold of 0.05 for statistical significance. We also cannot reject the null hypothesis, concluding that discounts lack statistically significant impact on bicycle sales across brands and seasons.
- Despite visual trends suggesting potential associations between discounts and sales performance, the regression models provide compelling evidence that suggest otherwise. The discrepancy shows us the importance of relying on statistical backing over visual interpretation alone. Visual examples, such as the "Sales vs Discount for Surly across Seasons," while visually appealing, do not have a statistical backing.

The lack of statistical significance in the relationship between discounts and bicycle sales across different brands and seasons could be attributed to several factors. These include a limited sample size, leading to reduced statistical power due to the data being confined to just one year. Additionally, variations in sales driven by customer preferences and other unmeasured factors may overshadow the impact of discounts. Furthermore, seasonal variations, where customers naturally show a higher inclination to purchase bicycles during certain seasons, could also play a significant role, potentially masking the effects of discounts offered.

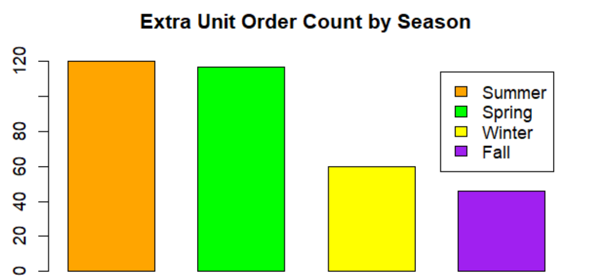
Hypothesis #3: Consumer Perceptions and Brand Choices

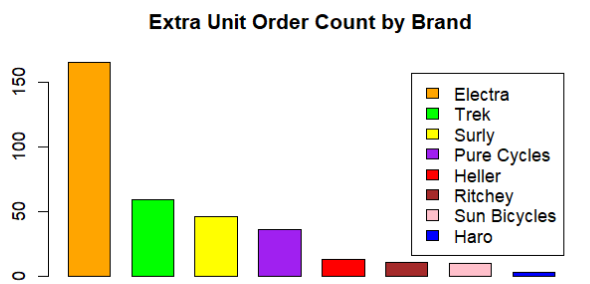
- How do consumers respond to price changes within different brands during different seasons?
- Are there seasonal trends in consumer preferences for specific brands within the same product category?

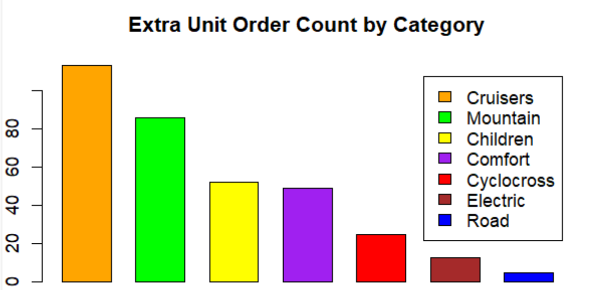
After performing the necessary data cleaning that included date formatting, removing NA values and filtering to return 2016 values, the approach was to first identify existing and potentially new variables that will help answer the question.

The existing relevant variables from the dataset were List_Price, Discount, Order_Date and Brand_Name.

Through data manipulation, new variables such as Discounted_Price, Revenue, Seasons and some more were created. Before diving into the modeling phase, some descriptive visuals were created that describes consumer behavior based on certain factors.







- In the three bar plots to the left, the y-axis shows the count or frequencies of the number of customers who ordered an extra unit of bicycle. The column 'Quantity' has two values, either 1 or 2. Hence, customers who order 2 units of bicycles are classified as people ordering an extra unit compared to just one.
- Most customers purchase an additional unit of bicycle during the summer season in 2016.
- Most customers purchase an additional unit of bicycle for the Electra Brand in 2016.
- Most customers purchase an additional unit of bicycle for the Cruiser Bikes product category in 2016.

Adjusted R-squared Model Comparisons and Evaluations:

Hypothesis	All seasons	Seasons			
		Spring	Winter	Summer	Fall
#1	0.8470	0.8398	0.8340	0.8609	0.8492
#2	0.8228	0.8207	0.8159	0.8539	0.8407
#3	0.8126	0.8119	0.8119	0.8121	0.8131

In hypothesis #1, the models' performances vary slightly, with Winter exhibiting the lowest adjusted R^2 value of 0.834 and Summer showing the highest at 0.8609. These adjusted R^2 values indicate the proportion of variance in list prices explained by the predictor variables while adjusting for the number of predictors in each model. Interestingly, despite slight fluctuations, all seasons display strong explanatory power, with adjusted R^2 values ranging from 0.834 to 0.8609. This suggests that while certain seasons may marginally outperform others in terms of predictive accuracy, the impact of weather conditions on list prices remains consistently significant throughout the year. Such findings underscore the importance of considering seasonal variations when modeling pricing dynamics in retail contexts.

Since we are testing consumer behavior in hypothesis #3, the appropriate linear models were built using Revenue as the response variable. Since *Revenue = (List_Price – Discount) * Quantity,* the quantity variable provides evidence for consumer preferences as it represents the units of bicycles consumers order. For each individual season's model, we have considered each season's order as a binary predictor (EX: Ordered in Summer? Y/N). The adjusted R^2 values are fairly uniform. However, the values being close to 1 are mainly because of the predictive power of the price variable. Category and Brands do not have any significant coefficients in each of the models.

Conclusion:

Based on the exploratory analysis, visualization and modeling processes, we can make the following conclusions pertaining to the three main hypotheses/research questions.

- It is not guaranteed that providing discounts on prices without considering when to apply them over the year would increase sales. And there are other factors that could have an impact on the sales figures as well like geographical location, seasons and more.
- It was noticed how the brands that applied different discount strategies across and within seasons seem to have influenced higher sales figures, in comparison to other brands that gave about the same discount ratio regardless of the season/time of the year.
- But despite visual trends suggesting potential associations between discounts and sales performance, the regression models provided a different statistical answer.
- The lack of statistical significance in the relationship between discounts and bicycle sales across different brands and seasons could be attributed to several factors such as sample size, consumer preferences or seasonal variations.
- From the descriptive analysis, there are trends in consumer preferences for brands and product categories of bicycles based on changes in seasons and pricing.
- There are distinct differences in consumers purchasing different quantities in different seasons, and even when you break it down by brands and product categories. However, we cannot conlude that changing of the seasons causes consumers to change their purchasing behavior.
- The differences in quantities ordered based on categories can be attributed to random real-life factors such as convenience, usage purposes and other internal consumer attributes.
- The major difference in quantities ordered based on brand can be attributed to brand loyalty.

Literature Survey/References:

1. Neubert, M. (2022). A systematic literature review of dynamic pricing strategies. *International Business Research*, *15*(4), 1-17.
2. Baldwin, G. (2019, March 8). *How retail seasonality is changing*. Pricing Tool for Retailers and Brands: Unlock Profitability with Omnia. https://www.omniaretail.com/blog/how-retail-seasonality-is-changing

Reference 1: A Systematic Literature Review of Dynamic Pricing Strategies

This publication examines the use of dynamic pricing models and its impact on consumer behavior and perception, where it synthesizes data across multiple research streams, including factors moderating the impact of dynamic pricing on customer behavior, strategic purchasing behavior in response to dynamic pricing, and the effect of dynamic pricing on customer perception of fairness. Providing further evidence of how bicycle brands that dynamically change their prices within and across seasons may have higher sales figures.

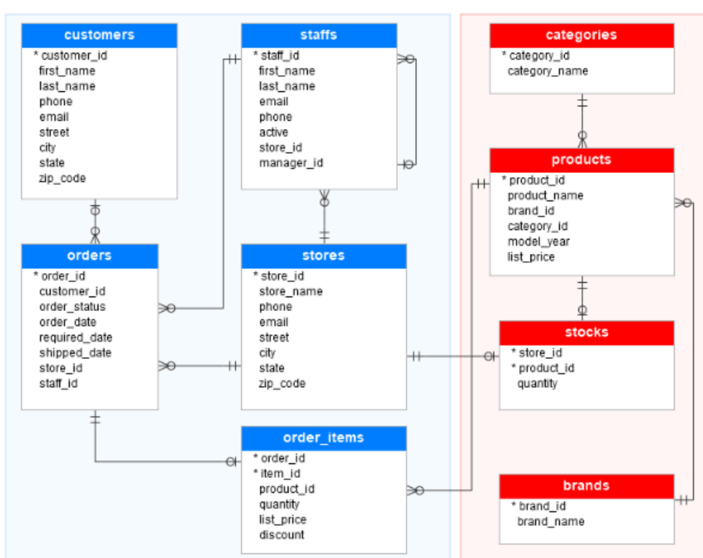
Reference 2: How Retail Seasonality is Changing

This article explores the yearly seasonal influence on retail and examines how stores can adjust their strategies to match these changes:

- Not all seasonality is the same, and it's important to illuminate the different drivers of consumer spending.
- Seeing seasonality as two separate categories (holiday-driven seasonal shopping and climate-driven seasonal shopping), will help understand the sales figures and allow stores to optimize for following years.

This could help this project findings in understanding what drives would influence the yearly sales figures.

Appendix:



Entity Relationship Diagram (ERD) for original dataset.
Source: Bike Store Relational Database | SQL ~ kaggle.com