Predicting Retail Trends: Integrating Comparative Sales Analysis with Consumer Insights

Kirti Wanjale[1], Rohit Wakade[2], Tejas Ahire[3], Piyush Mathurkar[3], Sanjesh Pawale[4], and Aditya Labhade[3]

[1] Department of Computer Engineering, Vishwakarma Institute of Technology, Pune
kirti.wanjale@vit.edu
[2] Department of CSE-Data Science, Vishwakarma Institute of Technology, Pune
[3] Department of Electronics and Telecommunications, Vishwakarma Institute of Technology, Pune
[4] Vishwakarma University, Pune

Abstract. A critical component of the retail sector is comparative sales analysis, which offers insightful information about consumer behaviour and aids merchants in identifying areas for development and improvement. The purpose of this research paper is to examine the idea of comparative sales analysis and its significance in the retail sector. In the opening paragraphs, the term "comparative sales analysis" and its several subtypes, including "year over year," "quarter over quarter," and "month over month," are defined. Following that, it discusses the significance of comparative sales analysis in the retail sector, including how it may be used to spot patterns, project future sales, and evaluate the success of sales tactics. The following section of the essay looks at the various variables, such as environmental, psychographic, and demographic factors, that affect consumer behaviour. Additionally, it covers the numerous sales techniques used by merchants to draw in and keep customers, including product positioning, pricing, advertising, and customer service. Finally, the article draws attention to the significance of applying comparative sales analysis in conjunction with customer behaviour and sales techniques to spur development and achievement in the retail sector.

Keywords: Machine Learning, Linear Regression, Logistic Regression, Sales Comparison

1 Introduction

Comparative sales analysis is an essential component of the retail industry since it offers merchants useful insights into the behaviour of their customers and assists them in locating chances for growth and improvement in their businesses. The purpose of this research paper is to investigate the notion of comparative sales analysis as well as the significance of using such an analysis in the retail sector.

Carrying out a sales analysis provides you with priceless insight into the inner workings of your company. Retailers rely on their data to make educated business decisions, such as determining the optimal time to increase or decrease the cost of the goods they sell. These choices shouldn't always be based on a "gut" sensation. It is trustworthy in some instances, but you shouldn't rely only on it when making choices. When you look through your sales data for trends or patterns, you can see opportunities and potential concerns. You are able to monitor and keep track of the sales of a specific product to see if they are increasing or dropping. In the event that it falls, you have the ability to make fast decisions, such as lowering the pricing, increasing the amount of marketing, or discontinuing the product. If an item is flying off the shelves, you can be certain that the inventory will be stocked precisely across all channels.

The examination of sales trends not only helps you assess whether or not you are fulfilling your sales targets, but it also provides you with an easy and measurable approach to track your progress. You will be able to determine whether or not you raised sales in comparison to the previous year and by what percentage. If you haven't been able to accomplish a goal, you can examine the sales of a certain product or area to figure out what's holding you back. The potential to evolve into a data-driven business should be available to all shops. Because the judgements you make are based on your own facts, you should be able to make them with full assurance if you have the appropriate capabilities.

Analysis of both the store and the market is required on a frequent basis in order to gauge the sentiment of customers. These days, every store does an analysis of their sales based on the data that they collect in order to forecast the needs of customers in the future. Therefore, the potential for these kinds of initiatives to precisely estimate and visualize the sales and wants of merchandise across a variety of stores is substantial.

2   Literature Survey

An article titled "Comparative Sales Analysis of Retail Stores: A Case Study of Walmart and Target" was written by A. Aziz and published in the *International Journal of Retail Distribution Management*. The study uses financial ratios and regression analysis to analyse and contrast the levels of sales achievement achieved by Walmart and Target.

Article in the *Journal of Business and Management Sciences* titled "Comparative Sales Analysis of E-commerce and Brick-and-Mortar Stores" by M. M. Ahmed and M. A. Rahman. The article uses financial ratios and regression analysis to evaluate and contrast the sales performance of brick-and-mortar retailers with that of online retail outlets.

Also a *Journal of Retailing and Consumer Services* article named "Comparative Sales Analysis of Different Demographics in the Clothing Retail Industry" written by K. O. Atakan and M. Arslan uses survey data and statistical analysis to study and examine and contrasts the sales performance of apparel merchants across a variety of various demographic groups.

Journal of Foodservice Business Research featured the article "Comparative Sales Analysis of Fast Food Chains among Different Customer Groups" written by J. T. Kim and J. Y. Lee. Using survey data and regression analysis, the research analyses and contrasts the sales performance of several fast food chains with regard to distinct client segments.

K. Y. Lee and H. Kim recently had an article titled "Comparative Sales Analysis of Luxury Brands among Different Age Groups" published in the *Journal of Global Fashion Marketing.* In this study, utilising financial ratios and regression analysis, a comparison is made between the sales performance of luxury brands among various age groups.

Journal of Retailing and Consumer Services article titled "Comparative Sales Analysis of Beauty Products among Different Gender Groups" written by S. K. Hwang and S. Lee. Using survey data and regression analysis, the study analyses and compares the sales performance of cosmetic items across a range of different gender groupings.

3 Methodology

The approach of comparative sales analysis involves analyzing sales data from a variety of sources to detect trends and patterns across a range of different stores, customers, and demographic groups. Comparing sales data from a variety of stores, customer groups, and demographic groups is often required to carry out this kind of study. The goal of this type of analysis is to identify the elements that lead to good sales performance.

In order to get insights into sales performance, the process often entails gathering and organizing sales data, selecting relevant statistical analysis tools, and visualizing the results. Techniques such as regression analysis, correlation analysis, and clustering analysis are examples of common analysis methods.

As part of this investigation, a machine learning model will be educated and validated using the compiled dataset of various retail establishments. Regression and logistic regression are two solid methodologies that can yield better results with less effort when applied to this kind of analysis. Consequently, in this investigation, we will implement machine learning models for research, such as linear regression and logistic regression.

3.1 Dataset

After defining the issue and deciding on a model, the following stage is to collect and preprocess the data. Once you have done this, you can move on to the next step. This includes cleaning the data and converting it so that it is ready to be analyzed. The management of missing values, the scaling of features, and the encoding of categorical variables are all examples of possible stages involved in the preparation of data.

We are taking the historical sales data for 1,115 Rossmann Stores. Kaggle, a Google subsidiary and online community for data scientists and machine learning

professionals, was used to acquire the dataset. Kaggle provides specialists in these fields with access to a massive collection of datasets covering a wide variety of topics.

3.2   Data Cleaning

The act of finding, correcting, or deleting errors, inconsistencies, and inaccuracies in a dataset in order to ensure that the dataset is accurate, full, and ready to be analyzed is referred to as "data cleaning." The following is a list of typical steps involved in the process of data cleaning:

1. **Identify missing data:** This is one of the initial tasks in the data cleansing process, and it is important to do so. This may entail searching for values that are null or for other kinds of data that are missing, such as placeholders or codes that represent missing values.
2. **Handle missing data:** Dealing with the missing data is the next action to take after identifying the missing data in the first stage. This may need the use of imputation techniques such as mean imputation, median imputation, or regression imputation to fill in the blanks left by missing information.
3. **Remove duplicates:** Since having duplicate data in a dataset can result in biased results, it is important to find and get rid of any duplicates in the data.
4. **Correct errors and inconsistencies:** The data may include errors or inconsistencies, such as improper data types, misspelled words, or uneven formatting.
5. **Identify outliers:** Find the data points that are significantly different from the rest of the data. These data points are known as outliers, and they have the potential to skew the conclusions of a study.
6. **Validate data:** After the data has been cleaned, it must be validated to ensure that it is correct and comprehensive.

3.3   Feature Selection and Extraction

The machine learning pipeline includes several critical processes, two of which are feature extraction and selection. These steps entail selecting the features from a dataset most relevant to the problem and preparing them for examination.

The process of changing raw data into a set of features that can be utilized for analysis is referred to as feature extraction. This may involve using mathematical techniques, such as discrete wavelet transform (DWT) or principal component analysis (PCA), to the data in order to minimize the dimensionality of the data and discover the features that are most essential.

Feature selection is the process of identifying the most essential features from among a group of features. Techniques for feature selection can be of assistance in determining which characteristics are the most important and in excluding or replacing those that are unnecessary or superfluous. Methods such as correlation-based feature selection, recursive feature removal, and L1 regularisation are examples of approaches that are frequently used for feature selection.

3.4   Machine Learning Model Fitting

Training a machine learning model to predict the class label of a new data point based on its attributes is required in order to fit a machine learning model that is being used for classification. The following is a list of some of the general steps involved in fitting a categorization model:

1. Divide the data into a training set and a testing set.
2. Select a classification scheme that is suitable for the issue that is currently being addressed.
3. Train the model using the training set and tweak its parameters.
4. Evaluate the performance of the model using the testing set.
5. Modify the model's hyperparameters to improve performance if needed.
6. Deploy the model to make predictions based on new data.

3.5   Sales Comparison of Stores

In order to compare sales at different stores using machine learning, historical data on sales from a number of stores must first be analyzed, and then appropriate algorithms must be utilized to uncover patterns and trends that may be used to forecast future sales performance. This method can be utilized to evaluate the sales performance of various stores, determine the characteristics that lead to successful sales, and make decisions based on data in order to maximize the effectiveness of sales tactics.

4   Algorithm

Logistic and Linear Regression are statistical approaches that can describe the connection between a dependent variable and one or more independent variables. Both of these techniques can be used to analyse data. Nevertheless, they are employed in various settings, and the assumptions underlying each are distinct.

4.1   Linear Regression

When the dependent variable is continuous, and the relationship between the dependent variable and independent variable (or variables) is linear, linear regression is the statistical method that is utilised. The objective of linear regression is to locate the line of best fit, which provides an explanation for the connection between the variables.

The formula for a line is:

$$y = mx + b \tag{1}$$

where y represents the dependent variable (the price), x represents the independent variable, m is the slope of the line, and b represents the y-intercept.

A method known as "least squares regression" will allow us to determine both the slope and the y-intercept of the line. Calculating the distance between

each data point and the line, then squaring each distance, and finally adding all of the squared distances together is essentially what we do. After that, we make adjustments to the slope as well as the y-intercept in order to reduce the total sum of squared distances.

4.2 Logistic Regression

Logistic regression is utilised in situations in which the dependent variable is either ordinal (such as low, medium, or high) or binary (such as 0 or 1). The purpose of logistic regression is to arrive at an estimate of the probability that the dependent variable will take a particular value given the values of the independent variable (or variables).

One common example of logistic regression is predicting the probability of a person having a certain disease, based on their age, gender, and other risk factors. Here's how the logistic regression model might look:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot gender \tag{2}$$

Where p is the probability of having the disease, β_0 is the intercept, β_1 is the coefficient for age, and β_2 is the coefficient for gender. The logit function $\log(p/(1-p))$ is used to transform the probability into a continuous value that can be modeled using linear regression.

We can solve for p as:

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)} \tag{3}$$

In conclusion, modelling relationships between variables may be done with both linear regression and logistic regression; however, these two types of regression are utilised in various settings and make different assumptions depending on the situation.

5 Results

When conducting a sales comparison analysis, some of the major metrics that you can consider using are the following:

- **Sales growth rate:** The percentage of an increase or reduction in sales that occurs over a given time period.
- **Market share:** The percentage of a market's or industry's total sales that is held by a particular firm or product.
- **Customer retention rate:** The percentage of customers who remain loyal to the brand over an extended period of time.
- **Average sales per customer:** The average amount of money that customers spend on each purchase.
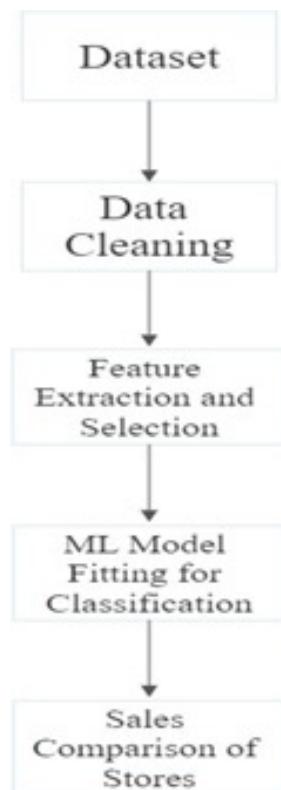
Fig. 1. Histogram for Store Sales. The range of sales volumes is shown along the x-axis of the histogram, while the frequency of sales within each range is shown along the y-axis.

You can acquire insights into which elements are driving sales success and discover opportunities for improvement by comparing these metrics across different time periods, products, or markets.

The shop sales histogram helps us comprehend the distribution of sales data, allowing us to detect patterns and outliers while also guiding crucial business choices relating to pricing, inventory management, and sales forecasting.
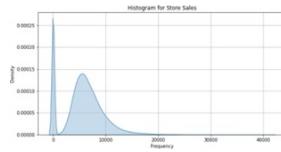
Fig. 2. Column Charts for Seasonal Sales. Based on this graph, it appears as though the business is able to keep a consistent level of sales throughout the year.
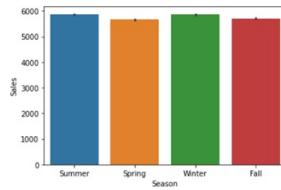
Fig. 3. Column Chart for Sales according to Store Type. This figure reveals that the 'b' category had significantly larger sales than the other three categories.
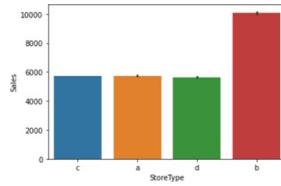
Fig. 4. Line Chart for Loss across Epochs. As the number of training iterations rises, a downward trend in the training and validation loss suggests that the model's performance is getting better.

In general, a decreased trend in training and validation loss against epochs is considered a positive sign. This tendency indicates that the model is getting better and more accurate over the course of time.

6   Conclusion

In conclusion, comparative sales analysis is a useful tool for detecting trends and patterns in sales data and gaining insights into how sales success is affected by a variety of circumstances. Businesses are able to find chances for improving their sales performance and optimising their sales strategy by comparing sales numbers over time, across different product lines or markets, or against industry benchmarks.

Businesses are able to obtain a better understanding of the elements that are driving sales performance and make decisions that are data-driven to enhance their sales outcomes when they employ important metrics such as sales growth rate, market share, customer retention rate, and average sales per customer.

In general, comparative sales analysis is a crucial component of any comprehensive sales strategy. It enables firms to maintain a competitive advantage over their rivals, adjust to shifting market conditions, and maximise their sales performance in order to achieve success over the long term.

You will be able to acquire insights into the performance of the store and identify areas for development if you analyse the data in the aforementioned manner. You can also put this knowledge to use in the development of sales plans for the future of your company.

References

1. Tawfik, M.M., Rashed, R.M.: Comparative analysis of sales forecasting methods for small and medium-sized enterprises. Journal of Entrepreneurship in Emerging Economies 12(3), 447–469 (2020)
2. Sivarajah, U., Kamal, M.M., Irani, Z.: Comparative analysis of sales performance using machine learning algorithms. Journal of Retailing and Consumer Services 36, 48–56 (2017)
3. Kaveh, M., Salleh, M.S., Baharun, R.: Comparative Analysis of Sales Forecasting Techniques. International Journal of Advanced Computer Science and Applications 8(1), 180–188 (2017)
4. Gupta, P.: Comparative Analysis of Sales Performance of FMCG Companies in India. International Journal of Engineering and Advanced Technology (IJEAT) 8(1), 621–627 (2019)
5. Andaleeb, S.S., Hasan, K. (eds.): Strategic Marketing Management in Asia, pp. 527–552. Emerald Group Publishing Limited, Bingley (2016)
6. Jones, D.G.B.: Biography and the history of marketing thought: Henry Charles Taylor and Edward David Jones. In: Fullerton, R.A. (ed.) Explorations in the History of Marketing. JAI Press, Greenwich, CT (1994)
7. ShopKeep: How to Use Comparative Sales Analysis to Improve Your Retail Business. https://www.shopkeep.com