







Article

# 360° Retail Business Analytics by Adopting Hybrid Machine Learning and a Business Intelligence Approach

Abdulmajeed Alqhatani <sup>1,†</sup>, Muhammad Shoaib Ashraf <sup>2,†</sup>, Javed Ferzund <sup>3</sup> , Ahmad Shaf <sup>3,\*</sup> ,  
Hamad Ali Abosaq <sup>4</sup> , Saifur Rahman <sup>5</sup> , Muhammad Irfan <sup>5</sup>  and Samar M. Alqhtani <sup>1</sup> 

<sup>1</sup> Department of Information Systems, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia

<sup>2</sup> Department of Informatics and Systems, School of Systems and Technology, University of Management and Technology Lahore, Lahore 54770, Pakistan

<sup>3</sup> Department of Computer Science, COMSATS University Islamabad, Sahiwal Campus, Sahiwal 57000, Pakistan

<sup>4</sup> Computer Science Department, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia

<sup>5</sup> Electrical Engineering Department, College of Engineering, Najran University Saudi Arabia, Najran 11001, Saudi Arabia

\* Correspondence: ahmadshaf@cuisahiwal.edu.pk

† These authors contributed equally to this work.

**Abstract:** Business owners and managers need strategic information to plan and execute their decisions regarding business operations. They work in a cyclic plan of execution and evaluation. In order to run this cycle smoothly, they need a mechanism that should access the entire business performance. The sole purpose of this study is to assist them through applied research framework-based analysis to obtain effective results. The backbone of the purposed framework is a hybrid mechanism that comprises business intelligence (BI) and machine learning (ML) to support 360-degree organization-wide analysis. BI modeling gives descriptive and diagnostic analysis via interactive reports with quick ad hoc analysis which can be performed by executives and managers. ML modeling predicts the performance and highlights the potential customers, products, and time intervals. The whole mechanism is resource-efficient and automated once it binds with the operational data pipeline and presented results in a highly efficient manner. Data analysis is far more efficient when it is applied to the right data at the right time and presents the insights to the right stakeholders in a friendly, usable environment. The results are beneficial to viewing the past, current, and future performance with self-explanatory graphical interpretation. In the proposed system, a clear performance view is possible by utilizing the sales transaction data. By exploring the hidden patterns of sales facts, the impact of the business dimensions is evaluated and presented on a dynamically filtered dashboard.

**Keywords:** business intelligence; digital revolution; sustainable business model; data warehousing; artificial intelligence; B2C; B2B



**Citation:** Alqhatani, A.; Ashraf, M.S.; Ferzund, J.; Shaf, A.; Abosaq, H.A.; Rahman, S.; Irfan, M.; Alqhtani, S.M. 360° Retail Business Analytics by Adopting Hybrid Machine Learning and a Business Intelligence Approach. *Sustainability* **2022**, *14*, 11942. <https://doi.org/10.3390/su141911942>

Academic Editors: Adam Jabłoński, Marek Jabłoński and Dariusz Zarzecki

Received: 14 August 2022

Accepted: 15 September 2022

Published: 22 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, business intelligence has drastically changed business operations by providing a smart cloud computing-based solution to enhance the working capacity and expand the business on a large scale, such as using knowledge performance indicators (KPIs) to enhance the performance of Nike's company with conversion into analytical reports. One study showed how KPIs are designed smartly to have a direct impact on overall performance management [1]. The core objective of business intelligence services is to view the past performance, diagnose the decision mistakes, and predict and forecast the future demands, benefits, and the best plan based on business facts and figures. By keeping in mind the past, present, and future direction of a business, warehousing business

intelligence (BI) and predictive artificial intelligence (AI) solutions are the best tools for designing flexible and powerful computing applications [2]. The authors of [3] studied how enterprises across the globe are now using big data analytics cloud solutions for their businesses to gain a competitive edge and the strategic planning and development for future demands.

The major challenges in big data analytics and business intelligence are unrefined BI strategies, the right business measures and performance scaling, and the adoption of self-service cross-department analytics [4]. Companies are using customer relationship management (CRM) and the decision support system (DSS) to analyze and develop intelligent and smart business solutions, moving from a product- to customer-centric approach to revise business strategies and high-valued performance metrics adoption [5].

Recently, the customer 360° view term was induced in BI, which comprises the broader view of a business with the help of dynamic filtering and querying responsive dashboards. The process of BI mainly consists of four phases which are business policy and operation framing, data extraction, transformation, and loading, data and machine modeling by applying BI and AI techniques, and at last, presenting the results visually by applying data mining and statistical techniques. Broadly speaking, there are four types of businesses in terms of scale, which are wholesale, business-to-consumer (B2C) retail, business-to-business (B2B) production and manufacturing, direct-to-consumer (D2C) deals and marketing, and facilitators in drop shipping. To acquire the intelligent systems for a specific business, the cost, quality, and access are the core factors in terms of business concerns [6].

To address these issues, software and platforms as services based on cloud computing yield easy access and cost-effective solutions for the end users. The primary concern of business professionals is to lead and expand a business using business intelligence smart applications and big data analytics by exploring why, what, which, where, and how business performance indicators could be evaluated to maintain and increase sales, products, employees, and customers. For effective business solutions, business management faces data management, performance evaluation, and smart business plan issues to smartly manage and lead a business and gain a competitive advantage. With the development of big data analytics and cloud computing solutions, businesses are utilizing their services to reduce costs and provide fast access, quick results, and reliable BI solutions. Customer relationship management, customer experience, departmental business data, quality work in operational systems, and knowledge performance indicators are the major factors driving the best solutions [7].

Data management applications based on database technologies are used in different departments such as administration, accounts, human resources (HR), sales, marketing, and customer care services. These departments produce departmental data such as account details, employee details, product manufacturing, stock and supplier details, transactions, payments, shipment details, customer-related information, and feedback via CRM apps and marketing details. With the advancement of eCommerce businesses, the method of doing business is now very smooth and flexible. Online customers can approach e-business easily and select, order, and track products via social media and eCommerce websites. Business owners know the worth of e-channels and try to facilitate eCommerce websites, mobile apps, and social media. These channels are producing big data as the result of both sides' data management operations. Data entries from administrative operators market products publicly and aggregate customers' information, orders, and reviews. Management with database management apps for day-to-day transactions is not a big issue, but big data analytics is a complex process that requires many resources. For a proactive approach to lead a business, the BI set-up must view different aspects of the past, current, and future performance perspectives.

The actual usage of BI is for utilizing data to support decision making with compelling and interactive data visualization, such as with the Qlik View BI tool [8]. To turn the raw data into knowledge and analytical insights into strategy making, the systematic process is followed with advanced BI tools. At the same time, when designing the data

warehouse, the first knowledge performance indicators describe what to evaluate and the performance scales. In the second step, information source documents are prepared in which abstract detailed information is maintained. The third step matrix documents are prepared to map the information from the source to the target desired analytical attributes using transforming, cleaning, and preprocessing techniques. These documents keep the sources attributes, technical constraints, and target mapping. All these documents are maintained with log files to maintain the versions and for the team's distribution. All source information is gathered via extraction transformation loading (ETL) or an ELT mechanism and placed in the staging area to avoid data loss. These data are then stored in a warehouse scheme for further processing. Mainly, two approaches are used to design and construct a warehouse first: top to bottom, which starts from a complete warehouse and goes to data marts, and the second is bottom to top, which combines data marts with the data warehouse for incremental development. For small types of business, the bottom-to-top approach is usually used. After collecting business requirements, the warehouse schema is designed, which can be of four types: (1) a start schema, (2) snowflake schema, (3) galaxy schema, or (4) clustered galaxy schema.

Each design has its pros and cons. Usually, the choice of design is based on storage and time constraints. BI is the systematic process that holds the data and business needs and implements a data collection mechanism, ensuring data quality and implementing data marts and data lakes with ETL and ELT to put data into the warehouse [9]. The BI warehousing provides a descriptive framework to view the business performance in terms of business dimensions in a detailed manner using data warehousing, which involves extraction, transformation, and loading to collect relevant business data from operational systems and external sources and preprocessing and data modeling under a relational schema to make them usable for relational online analytical processing (ROLAP) and multidimensional online analytical processing (MOLAP), which filter the data using business dimensions such as time, geo-graphs, products and HR, and a data visualization process in which ad hoc reporting is performed to view and explore the business performance as well as the impact of business features. On the other hand, machine learning (ML) allows for supporting the decision makers with helpful trained models which capture the data trends, acting as a magic box to answer the queries with probability. ML offers multiple analytics (personalized recommendations, sales, demand forecasting, and customer behavior prediction) in eCommerce [10].

ML offers capturing data trends such as customer segmentation, sales forecasting, and product recommendation systems. For predictive analysis to see the future based on past and current business data, strategy makers set up various questions. In [11], the authors performed predictive analysis to predict the sales in weekly collected data.

In [12], the authors proposed an ensemble approach by combining S-ARIMA, vector auto regressor, and long short-term memory (LSTM) for demand forecasting of monthly product distributor orders with some external features such as weather, campaign, and holidays. In [13], the author implemented a model to manage chain stores via BI modeling and presented a dashboard to report on the managers and strategy makers for effective decision making. The proposed model considers the KPIs and divides the dashboard reports for different business layers to support the supply chain and, ultimately, the impact on sales. For practical customer and sales data usage in a retail business, the author of [14] developed a framework under the BI tool and quickly communicated with a chat bot using point of sale (POS) transitional data. The author [15] studied the impact of business intelligence and analytics for retailing intelligent data insights. They used 25 global retailers' online databases and web source datasets and analyzed the retailer activities in different phases.

The authors of [16] presented a profit-based forecasting model for shoe multi-seasonal retail data. They evaluated ML models with 10 parts via cross-validation and real-time demand values and used the mean absolute percentage error. In retail chains, the authors of [17] applied a forecasting (special calendar days as features) framework on a centralized

food bakery to predict the product facility in 100 stores on a daily basis. In [18], the authors evaluated the customer engagement data for the telecom sector in terms of social demography and services used by the customer. Segmenting in similar customer traits and then predicting the expected behavior were the main concerns in their study. In [19], the authors applied a data mining approach on an electronic sales dataset (mixed B2B and B2C) for a complete online business process from customer analysis (by customer segmentation via k-means clustering), product analysis (by association mining via the Apriori method), and predicting customer behavior outcomes (by the decision tree method).

The authors implemented an ML model on time series data of B2C Amazon quarterly sales data in 2019 in [20]. After data transformation, three models were applied (Holt winters exponential smoothing, automotive regression integrated moving average (ARIMA), and ANN auto regression), and they compared the accuracy of the three models by evaluating the mean absolute percentage error (MAPE) and root mean squared error (RMSE), as well as other metrics for the implemented models. The seasonal ARIMA was best in the projection of revenue.

In [21], a comparison of two powerful ML algorithms (XGBoost and LSTM) to predict the sales for retail products is performed. The authors used online analytical processing (OLAP) methods to analyze the sales measures on a multi-dimensional data model which contained sales transactions, customers, and item information [22]. The outcome of their analysis was a web dashboard which showed the sales trends for the given dimensions.

In [23], the author implemented a model to manage chain stores via BI modeling and presented a dashboard to report to the managers and strategy makers for effective decision making. In [24], a data warehouse monitoring framework is presented in various applications for effective results and integration with management pipelines. In [25], the authors identified the main KPIs, such as the net income, rate of investment, details, equity, and gross margin, as well as their implementation with machine learning and presented the results on a dashboard.

As discussed above, the intelligent BI and AI mixture for a business provides a smoother platform for dealing with day-to-day operational management and making better decisions for a business. This research primarily focuses on business to consumers, in which the retail and wholesale types target business subdomains. This research will be conducted on utility product industries to add valuable business insights. The following are the core research questions that are going to be addressed in this article:

- What is the sales performance in terms of the date and time dimension?
- What is the sales performance in terms of the product selection dimension?
- How can business knowledge communicate with a central visual data platform for a business stakeholder to view the performance, operations, and strategic planning?
- What is customer behavior in business, and what are the reasons for these insights?
- How many business dimensions are dependent and impact on sales performance?
- How can we make more accurate analytical models to find impactful business insights?
- Which business are dimensional features highly impactful to driving a business strategy?

Retail business sales performance can be evaluated using BI techniques for efficient sales views in various business dimensions and effective strategies. A light and flexible warehouse schema needs to be designed by keeping in mind the business type, analytical needs, and transnational data produced. The main concerns of a retail business are to evaluate the sales so that the appropriate action can be taken. Therefore, the proposed framework is based on a hybrid cloud solution which mainly consists of warehousing business intelligence, which provides statistical analysis about the past and current views of a business and predictive artificial intelligence to forecast the business plans. The dataset used in this study was taken from the UCI repository, having 541,909 data entries of transactions by different customers from 2010 to 2011. By utilizing this data, BI techniques were applied to fulfill the analytical research objectives. Compared with existing techniques, this will provide a lighter weight, less errors, and a centralized and comprehensive view of

a business at a low cost and with fast access and dynamic business analysis. The leading business intelligence solutions are the cloud-based data pipelines with deep data analytics as well as data visualization.

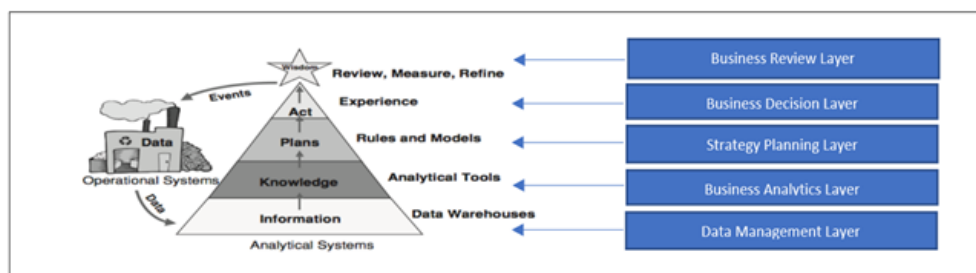
The rest of the paper is formatted as follows. Section 2 describes the system model, while the results are discussed in Section 3. In Section 4, the conclusions and future works are discussed.

## 2. Materials and Methods

The business intelligence techniques under cloud computing and big data processing are the core ones to drive the insights from data. In recent years, various cloud services have offered making decision support systems from data collection to dynamic dashboards by Watson IBM, Qlik View, Microsoft Power BI, Tableau, Oracle NetSuite, Amazon Redshift, and Google Data Studio. The proposed methodology is based on a customer 360° view with four dimensions and an advanced business analytical dashboard to drive smart business insights. The core components are data sources, ETL, data consumers, a processing module, and an analytical dashboard. The figure below demonstrates the smart application framework, consisting of the following:

- Business data sources;
- The process of extraction, transformation, and loading;
- Relational warehouse and machine learning data modeling;
- Descriptive BI analytics;
- Diagnostic performance modeling;
- Prescriptive planning modeling;
- Predictive AI analytics;
- A dynamic central dashboard for ad hoc queries and a customer 360° view.

Finding the business insights from data is a systematic process which contains multiple layers and processes to drive the actionable results for business strategic development. Data are the core components, which are collected from operational management and point-of-sale systems, CRM systems, as well as social media marketing channels, which are combined into a centralized data repository. Figure 1 depicts the layers of the smart analytical system.



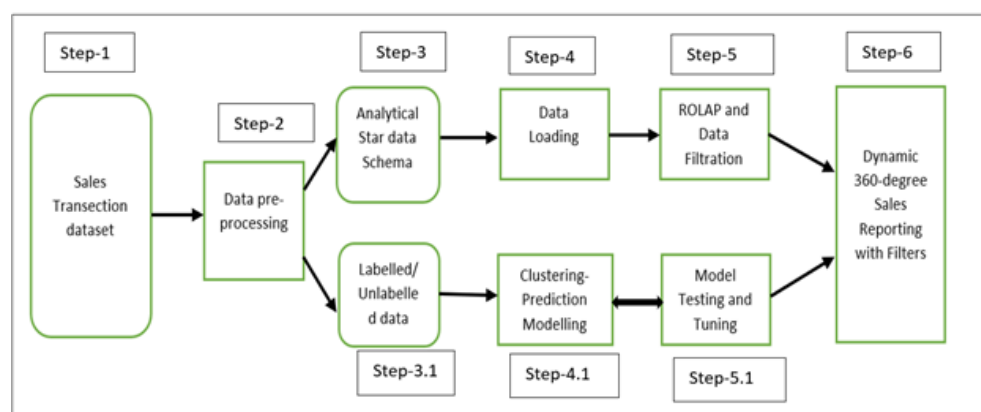
**Figure 1.** The business operation and analytical system mapper for understanding the actual business needs to feed into analytical system development.

Data sources are diverse and may produce different types of data stored in file formats, such as how department management system data are stored in SQL and Excel or CSV files which contain customer data, HR data, store data, financial data, sales, and marketing data from local sources. Customer data can also be collected from social media and mobile apps, which are used to engage and retain customers for a long time. These data are used in the warehousing and predictive data modeling after ETL processing. The output of this process is the refined form of data repositories, which are further used in data modeling for the descriptive diagnostic business intelligence (DD-BI) phase and PP-AI phase. This DD-BI phase uses statistical techniques to explore the data relations and interpret the results in the form of graphs. The PP-AI phase uses machine learning predictive and diagnostic techniques to predict the future outcomes with time, customer, and business demographics



graphs. The final phase is the dashboard analytical phase, which interprets the results in the form of dynamic visuals with the help of BI cloud solutions. The dashboard will give the smart insights of a business for each department as well for the business owners.

The process flow for the proposed research framework is given in Figure 2, showing the abstract flow of the analytical techniques. First, the sales data was collected from the UCI repository, and the initial data was preprocessed to clean it and remove duplicates and irrelevant data entries. For the comprehensive sales analysis, we derived some date and time features from the invoice dates. After that, we constructed a schema design and the ML's needed data. This involved two main components: one was machine learning, and the second was warehousing. The work for these two components was performed in parallel. For the DD-BI phase, one subprocess was compiled. With the help of power BI, the complete development was performed for data loading, relational design, and interactive data reporting. Similarly, for the predictive forecasting machine learning (FP-ML) phase, a second subprocess was compiled with Anaconda Jupyter Notebook to prepare the trained test datasets and the ML modeling and evaluation. The results were compiled on the dashboard. In the next subsections, the complete dataset description and all working components are described.

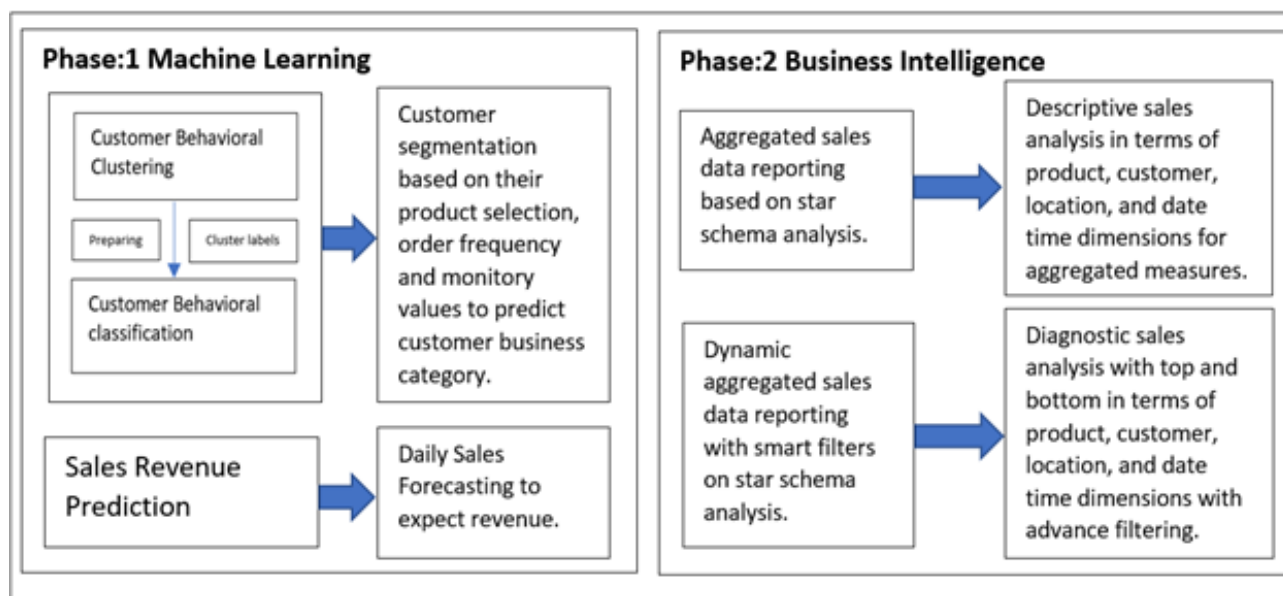


**Figure 2.** Sales analysis process flow to find the performance insights using warehousing and machine learning algorithms.

The proposed research model has two phases which are based on business intelligence data warehousing and machine learning solutions. To fulfill the research objectives, how these phases will sort out the issues and provide a comprehensive assistive solution for better and more friendly insights is shown in the technical flow diagram.

After applying this model, the final analytical report outcomes address the business objectives with graphical interpretation. In the figure, we can see the analytical solutions of hidden business insights.

A deeper view of the research framework in terms of development and its core techniques are discussed here. As the proposed methodology consists of BI and AI development with deep analytics, using big data analytics and web applications enabled us to present the quick results and effective business analytics for business professionals. The proposed research design provides the BI solution for a business to view the performance and employ friendly and accurate assistance to make future decisions in terms of the product pipeline, customer relationship management, and core business dimensions for strategy development to lead and manage a business in quick, productive, and efficient ways, as shown in Figure 3.



**Figure 3.** Proposed model's ML-BI technical flow for retail data analysis in 360 degrees to answer business analytical questions.

### 2.1. Designing Knowledge Performance Indicators

The data analytics process begins from business process framing and conversion into the smart metrics needed in performance evaluation, decision making, and strategy building. Designing KPIs is not a straightforward process, as it needs a thorough understanding of a business's set-up, sales channels, entire environment, and policies. As the dataset used in this study was taken from the UCI repository, having eight lakhs of data entries of transaction by different customers from 2009 to 2011, by utilizing this data, business intelligence techniques were applied to fulfill the analytical research objectives.

Being a business stakeholder, one has to be concerned about business performance and expansion plans to compete in a market. This research focuses on business performance indicators and customer relationship management by targeting the retail business level. The following are the core business questions which explain the problem statement regarding this research proposal:

- Which business information (product, management, and customer features) does one need to decide the better plans for a business?
- How can business performance indicators be used to carefully choose and make an impact on a strategy with smart intelligent dashboards to take a competitive edge in the market?
- Can one make smart and intelligent business decisions without visual business insights which are based on facts and figures?
- Is there any proper mechanism for capturing a customer's dynamic behavior which is necessary for a strategy planning?
- How can operational systems link with cloud BI solutions to make live contact with business events and performance for control management?

Business strategy and planning development, which are addressed in this proposal, are designed for B2C businesses. For business managers and owners, a proactive approach leads one's business on a high scale. Therefore, some problems are mentioned in retail and wholesale businesses in different departments to find the operational and planning efficiency and effectiveness.

To formalize the business problems into detailed analytical objective points, knowledge performance indicators are the best way to describe them easily. The ultimate objective of this study is to answer these KPIs by exploring the sales data and applying ML-BI techniques. These KPIs are dependent on given business information, as the less-studied

data contains sales transnational data. Therefore, the following KPIs are the defined goals which will be met at the end of analytical experimentation and data result reporting:

- What is the customer behavior (similarities in buying) in product selection?
- What will be the expected category of customers in product selection?
- What are the total sales measures (revenue, sold quantity, customers, products, orders, and canceled quantity)?
- How much are total dimension values (location, customers, products, date time)?
- What is the total sales revenue for location-wise order placement (maximum and minimum revenue locations)?
- What is the relation of products with date time (frequently in date hierarchy)?
- What is the relation of products with customers (location wise products orders)?
- What is the relation of products with customers and date time (smart data filters to view measures)?
- What is total sales revenue customer-wise (frequent and infrequent customers)?
- What is total sales revenue product-wise (frequently and infrequently sold products)?
- What is the total sales revenue year-wise (maximum and minimum revenue time frames)?
- What are the prices of products with the maximum revenue generation?
- What are the sales and quantity distribution among all orders?
- What are the sales and quantity distribution among weekends and weekdays?
- How many orders were placed year- and month-wise?
- What is the average of the sales revenue in terms of top customers, products, country, and months?
- Which are the top and bottom customers in terms of maximum revenue?
- Which are the top and bottom customers in terms of sold quantity?
- Which are the top and bottom customers in terms of order booking?
- Which are the top and bottom customers in terms of maximum orders quarterly and monthly?
- Which are the top and bottom customers with the maximum revenue generation location-wise?
- Which are the top and bottom customers having the maximum product orders location-wise?
- Which are the top customers for every month with the maximum revenue generation?
- What are the similar customers in product ordering (segmentation and category prediction)?
- What is the relation of customers with date time (frequently in date hierarchy)?
- What is the relation of customers with order products (location-wise products order and features)?
- Which are the top and bottom products in terms of sales revenue generation?
- Which are the top and bottom products in terms of sold quantity?
- Which are the top and bottom products in terms of order booking?
- Which are the top and bottom quarters and months in terms of sales revenue?
- Which are the top and bottom quarters and months in terms of sold quantity?
- Which are the top and bottom quarters and months in terms of order booking?
- What is the relation of quarters and months with customer engagement (frequently in date hierarchy)?
- What is the relation of quarters and months with order products (location-wise products order and features)?
- What are expected sales revenue in near future based on daily sales?
- How much was contributed to the total sales revenue by the top and bottom customers with the maximum sales revenue?
- Which are the top and bottom products with the maximum sales location-wise?
- Which are the top and bottom products with the maximum sold quantities location-wise?



- Which are the top and bottom products with the maximum sales month-wise?
- What is the difference between the quantity sold, canceled quantity, and total sales semester-wise?
- What is the difference between the quantity sold, canceled quantity, and total sales quarter-wise?
- What is the difference between the quantity sold, canceled quantity, and total sales month-wise?
- What is the difference between the quantity sold, canceled quantity, and total sales in terms of the week of a month?
- What is the difference between the quantity sold, canceled quantity, and total sales in a month by day?
- Which are the order's quantity and sales revenue in terms of the maximum sales revenue generation?
- Which are the product's quantity and sales revenue in terms of the maximum sales revenue generation?
- What is expected of one month's sales revenue minimum, maximum, and trend lines?
- What are the expected sales revenue minimum, maximum, and trend lines over two months?
- What is the sales revenue distribution among locations having orders returned?
- What is the sales revenue distribution among product prices having orders returned?
- Which is the top sales revenue generated by a product's sales revenue with orders returned?
- Which is the top sold product's revenue with orders returned?
- Which is the most frequent product revenue when orders are returned?
- What is the relation between the sold quantity and total sales for the highest-revenue products?
- What is the relation between the sold quantity canceled quantity for the highest-revenue products?
- What is the relation between the sold quantity and total sales for the most sold products?
- What is the relation between the sold quantity and canceled quantity for the most sold products?
- How many orders are canceled and returned (order, quantity, price)?
- What is the relation of invoice orders with customers location and products?
- What are the observation and frequency of invoices in date time hierarchy?

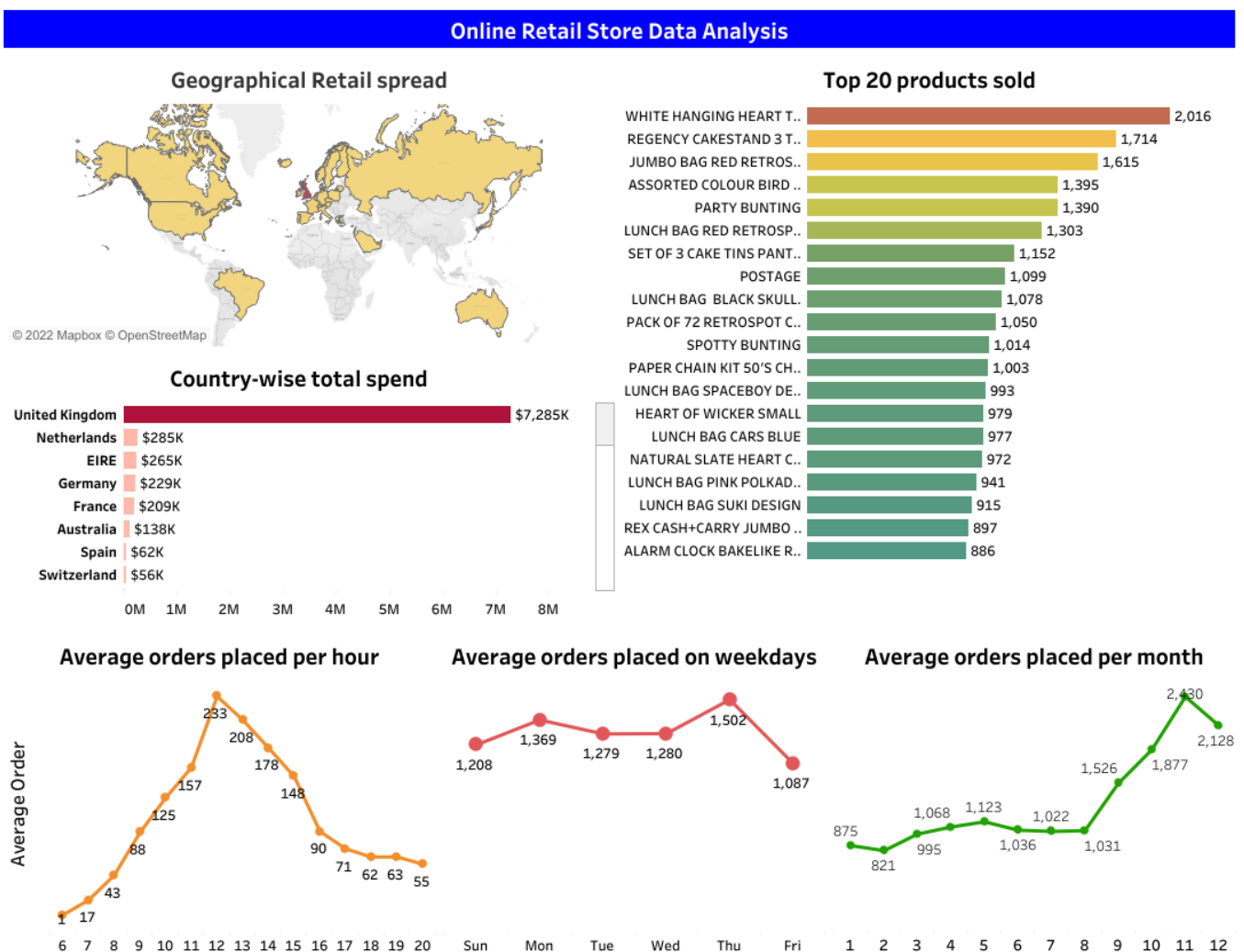
These need to be used to deeply analyze the relationships of relational sales attributes to find the hidden insights so that improvements can be made in weak areas. For the hidden picture of business data, this is the best problem-framing technique for mapping the research objectives. In addition, these KPIs will be useful in the evaluation of the proposed analytical techniques.

## 2.2. Dataset Description

Data are the hub of analysis, and they can be in any file format, local or global source of data, and type of data such as text, sound, or visuals. The textual type of data is mainly used for analysis specifically in the banking sector and business industries. These data are produced everywhere with the help of web and mobile applications from the client and administration sides to complete the needed tasks. In this research, online retail sales transactional data will be used to conduct research experiments for B2C and B2B businesses. Most of the transactions were made by a UK-based, registered, non-store internet retailer between 1 December 2010 and 9 December 2011. In this multinational data collection process, 541,909 entries were collected from the following countries, given as percentages; the United Kingdom (88.9%), Germany (2.3%), France (2.1%), the EIRE (1.8%), Spain (0.6%), the Netherlands (0.6%), Belgium (0.5%), Switzerland (0.5%), Portugal (0.4%),

and Australia (0.3%). The following data columns were used in the data collection phase and also described in Figure 4:

- InvoiceNo: The invoice number is a nominal, six-digit integral number issued to each transaction specifically. If this code begins with the letter “c”, then a cancellation has occurred.
- StockCode: The product (item) code is a nominal, five-digit integral number issued specifically to each individual product.
- Description: The product (item) name, which is nominal.
- Quantity: The number of each product (item) in a single transaction, which is numeric.
- InvoiceDate: The invoice date and time is numeric and shows when each transaction was created, both in terms of day and time.
- UnitPrice: The unit price is the numeric product price per unit in GBP.
- CustomerID: The customer number is a nominal, five-digit integral number uniquely assigned to each customer.
- Country: The country name is the nominal name of the country where each customer resides.

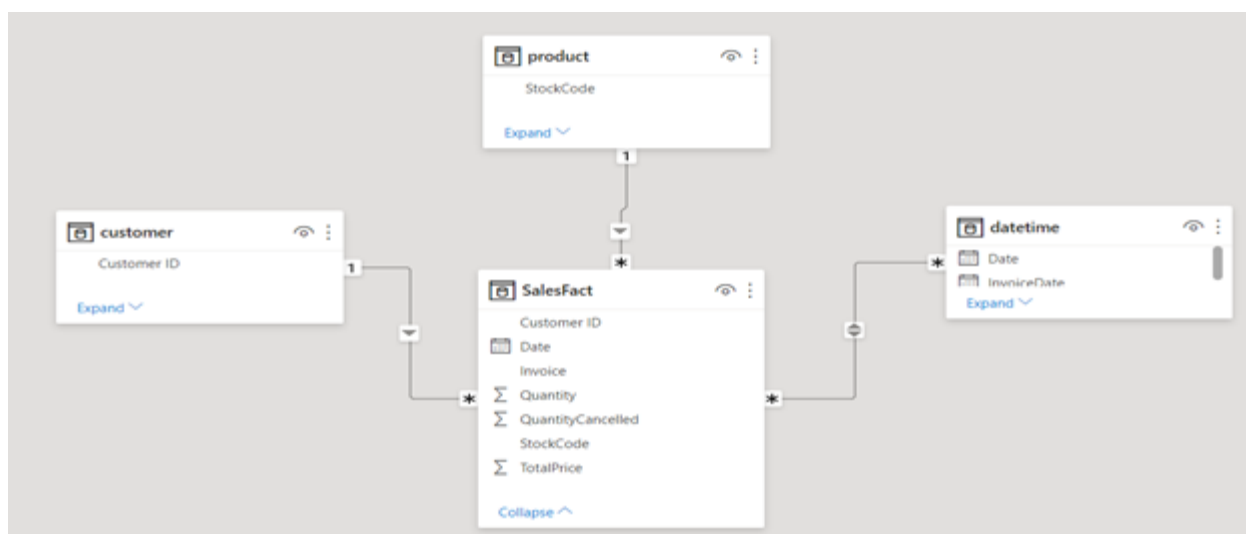


**Figure 4.** Data analysis of online store.

For the detailed dimensional analysis provided, the features were not enough, so we derived some more features from the date and time column to conduct overall fact measurements and paired-fact measurement analysis.

### 2.3. BI Data Modeling

Business intelligence-based application demands structured warehousing data models, which are used for the development of data marts, data lakes, and complete warehouses. Raw data are collected from multiple sources, which consist of different file formats such as SQL from operational systems, Excel, and Word reports from management reporting systems. In this study, online retail data are used to prepare a data model with the help of Python scripting to first prepare dimensional data in different Excel sheets and then prepare a warehousing schema with power BI. As shown in Figure 5, the proposed schema data model contained four dimensions and one fact table to measure the revenue, sold quantity, and number of orders.



**Figure 5.** Data warehousing analytical schema to support ad hoc queries and data reporting.

This is the dimensional star schema model which contains three business dimensions—customers, product, and date and time—and one fact sales table connected with all dimensions. These are extremely helpful for ad hoc analysis and how these dimensions are having an impact on sales growth.

### 2.4. ML Data Modeling

Machine learning-based applications demand featured data models which are divided into training and testing data to train the ML model using supervised or unsupervised algorithms according to the data attributes and business goals. Usually, Excel, CSV, TSV, and SQL data files are used, from which features are extracted and set as dependent or independent variables for predicting forecasting. In this study, after collection and preprocessing, we prepared the data for unsupervised learning to create similar bought product- and customer-based clusters on unlabeled (product category and customer category) data. Therefore, we first arranged the data for products sold and customer orders. To prepare the data for forecasting, we just picked two features from the given data, which were the invoice date and total sales.

Similarly, for customer classification, the customer-driven features were based on product selection and their participation in sales transactions. In the first segment, the products totaled five buckets, as in order transactions, and they were labeled as customer category buckets, which totaled 11, as well the customer minimum, maximum, mean, and sum values contributed in their sales orders. This customer oriented-data will be used in the FP-ML phase for customer classification.

#### 2.4.1. Descriptive Diagnostic Business Intelligence

The deeper analysis provides relevant and accurate insights about the data, which drives actionable results. DD-BI is the first parallel phase and highlights the past and current states of the data by exploration, detection, and association techniques. Experimenting with clustering and association techniques to explore the data by segmentation and dimensionality reduction highlights the customers, products, deviated sales, demand behavior, and aggregated business impact. This phase consists of descriptive and diagnostic analysis, which will be performed on BI data models. To describe the patterns in the data, initial visualization will be performed to know the associations, correlations, and deviations to find and identify the positive and negative factors.

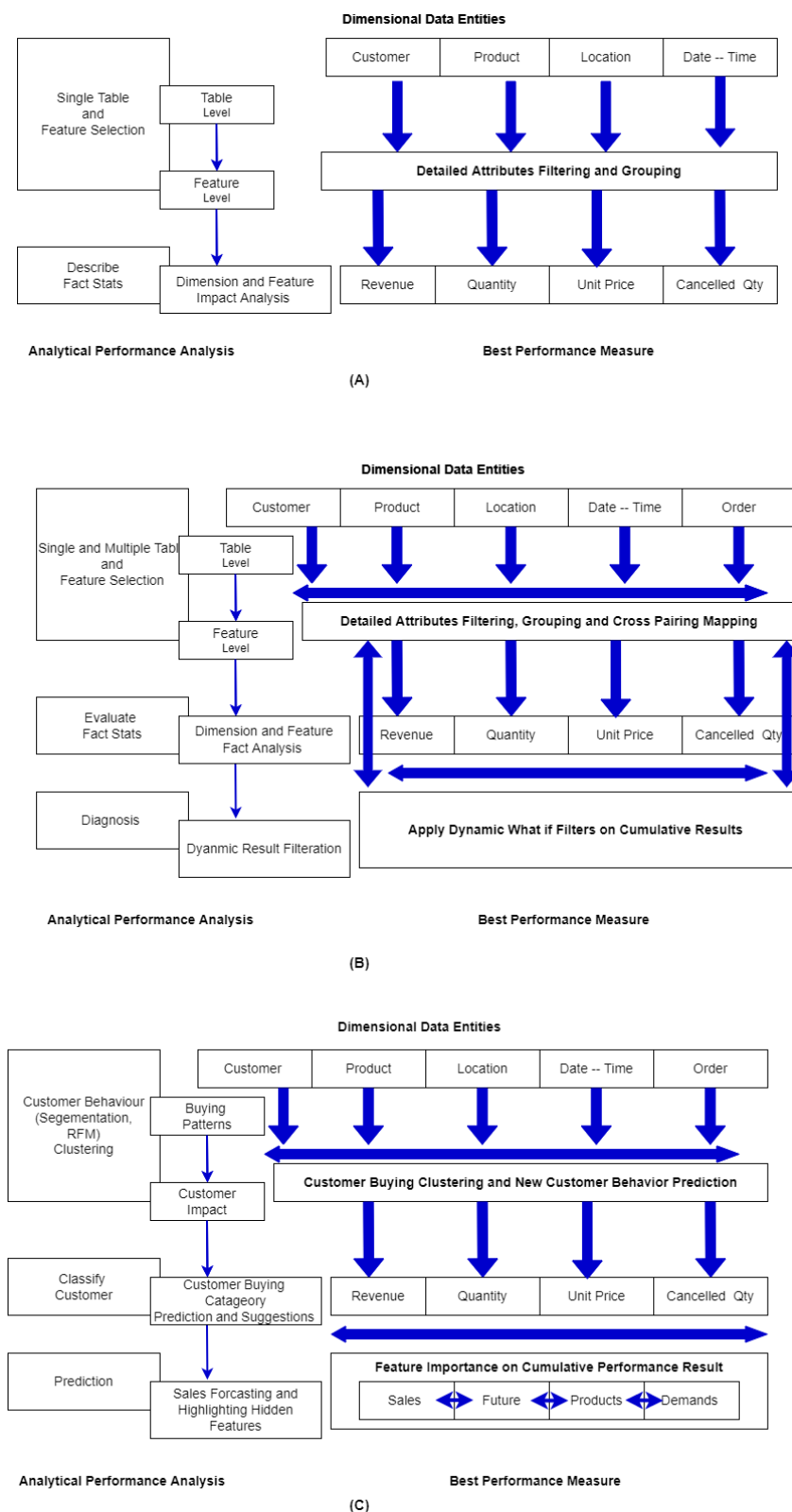
Here, we describe the basic static data measures to overview the data values. The total unique customers, products, orders, canceled orders, quantity sold, and total revenue show the basic descriptive analysis. To diagnose the sales and their dependency on the product, customer, location, and date and time, multiple types of filters were implemented to view the measures in 360°. These analytical reports were on the completely self-descriptive properties implemented with power BI. Stakeholders can easily find the sales with on-demand business questions and design the strategy based on these stats. This is the flow which shows how descriptive and diagnostic analysis will be performed to find the sales KPIs.

#### 2.4.2. Predictive Forecasting Machine Learning

Proactiveness is the key to attain success in the corporate world while managing any enterprise business to gain a competitive edge in the industry. FP-ML is the second parallel phase to view and decide in advance on the future outcomes which will be faced. This involves first segmenting the ordered products and customer segments based on product clusters with K-mean and principle component analysis. Clustering will be used to prepare the customer-oriented data that will lead to classifying the customers into various categories regarding sales contribution. This mechanism is explained in the ML data modeling section. Forecasting analysis will be performed using ML data models to forecast the performance, behavior, engagement, and demands related to the business domain. To view the future demands, a predictive ML technique by ensemble for ARIMA, forecasting models with Python scripting, and a power BI automated tool are used to determine what suitable action will be best in future to lead the business from the front in terms of decision support systems.

This flow shows how the patterned mining of products in sales orders and customer behavior prediction will be performed with K-means clustering and classifying customer categories with multiple multi-class enabled algorithms. To predict the sales in near future, ARIMA will be applied to find the sales trends.

This proposed methodology contains a BI-ML hybrid framework for effective sales analytics for online retail businesses. As our research objects mainly concern sales facts, which are needed for sales managers and owners to facilitate customers directly, this is dual process uncovers the sales insights from the past to the future as well as user-friendly interpretations. In the next section, the complete implementation is given, which shows both phases and each step of experimentation. As explained in Figure 6, the abstract view of the proposed model which uses BI and ML effectively to view the retail sales yields the KPIs with the help of interactive charts on retail data. In Figure 6C, the analytical BI model and ML model data are shown. Similarly, Figure 6A,B shows how dimensional OLAP analysis will be performed for the descriptive and diagnostic models. The results of these two phases are presented as reports. These reports are highly beneficial in performance analysis, strategy making, and inductive decisions. The real output from this analytical model provides a self-service BI solution for effective sales analysis for performance monitoring.



**Figure 6.** (A) Warehousing descriptive data analysis to know the past overall business performance. (B) Warehousing diagnostic data analysis to know the past business performance with dynamic filtration and feature importance. (C) Machine learning predictive data analysis to predict the future business performance with customer segments and sales forecasting.





	recency	frequency	monetary	r_quartile	f_quartile	m_quartile	RFM_Score
Customer ID							
18102.0	0	145	580987.04	1	1	1	111
17450.0	8	51	244784.25	1	1	1	111
13694.0	3	143	195640.69	1	1	1	111
17511.0	2	60	172132.87	1	1	1	111
16684.0	4	55	147142.77	1	1	1	111
15061.0	3	127	126389.02	1	1	1	111
17949.0	1	118	117314.08	1	1	1	111
15311.0	0	208	114966.42	1	1	1	111
13089.0	2	203	113416.91	1	1	1	111
12931.0	21	57	92347.34	1	1	1	111

Figure 9. Customer recency, frequency, and monitoring values after RFM analysis and computing.

The forecasting results were better, forecasting a two-month sales forecast as shown in Figures 10–15. The sales daily trends, ARIMA modeling, fitted model stats, and forecasting values are displayed. Due to some shortcomings in data length, the results were not up to the mark but showed a clear view of the sales for the next two months in the results of daily seasonality.

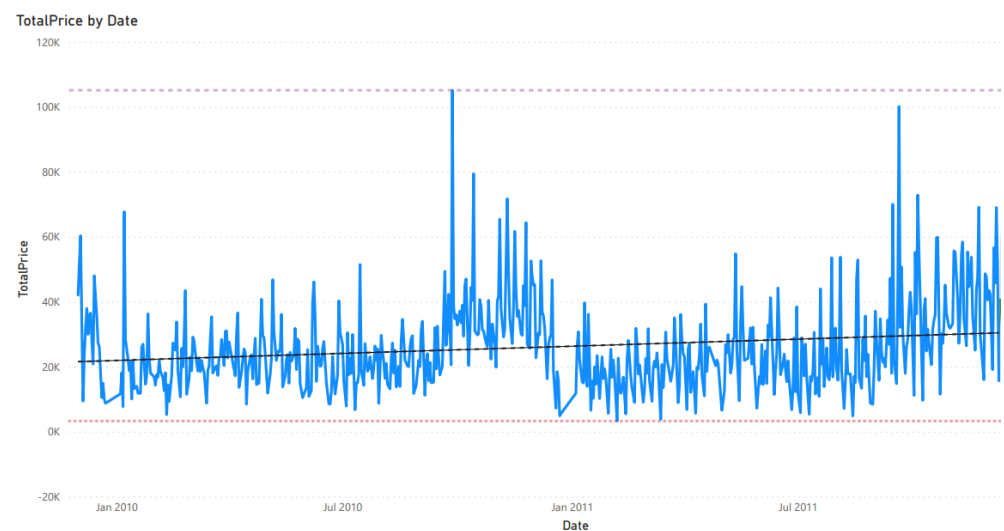


Figure 10. Sales values with day-to-day transactions by invoice date.

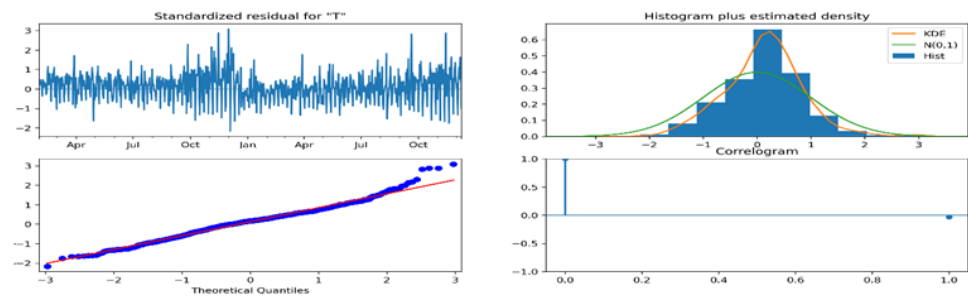
```

===== SARIMAX Results =====
Dep. Variable: TotalPrice      No. Observations: 374
Model: SARIMAX(1, 0, 1)x(1, 1, 1, 30)    Log Likelihood: -3545.487
Date: Sat, 01 May 2021      AIC: 7100.973
Time: 16:22:17      BIC: 7119.688
Sample: 12-01-2010      HQIC: 7108.453
Covariance Type: opg

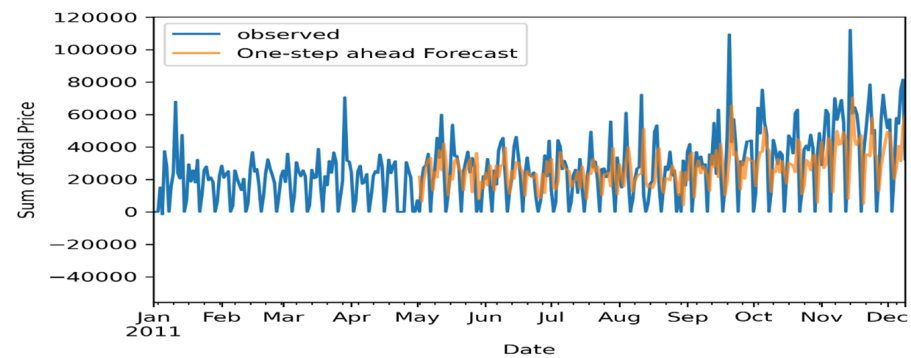
=====
coef      std err      z      P>|z|      [0.025      0.975]
-----
ar.L1      0.2633      0.222      1.185      0.236      -0.172      0.699
ma.L1      0.2300      0.207      1.111      0.266      -0.176      0.636
ar.S.L30   -0.2132      0.121     -1.765      0.078      -0.450      0.024
ma.S.L30   -0.7000      0.084     -8.335      0.000      -0.865     -0.535
sigma2     6.538e+08   5.52e-11   1.18e+19   0.000     6.54e+08   6.54e+08
=====
Ljung-Box (L1) (Q): 2.32      Jarque-Bera (JB): 10.78
Prob(Q): 0.13      Prob(JB): 0.00
Heteroskedasticity (H): 3.08      Skew: 0.04
Prob(H) (two-sided): 0.00      Kurtosis: 3.91
=====

```

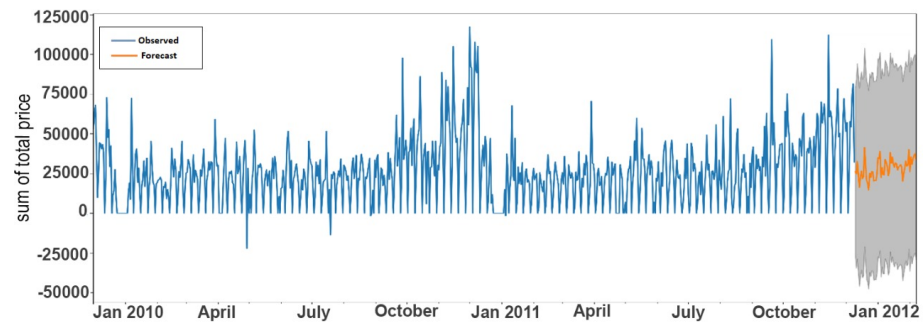
Figure 11. Sales forecasting behavior summaries.



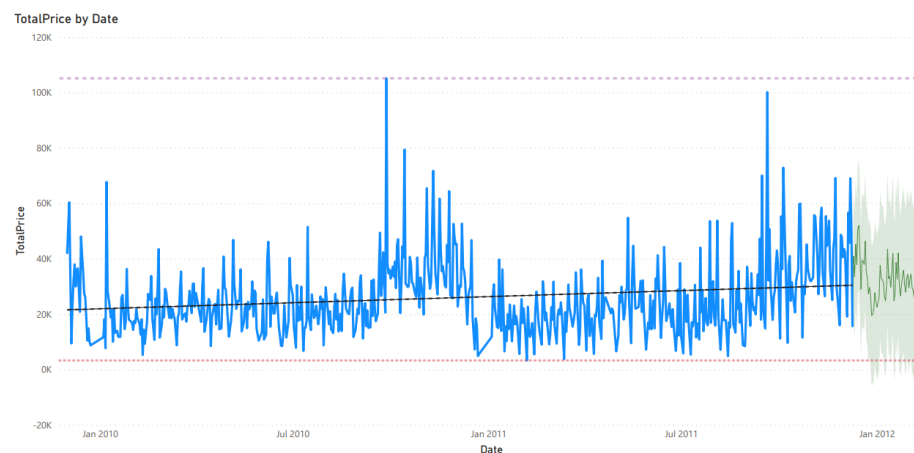
**Figure 12.** Sales forecasting model with summary visuals to show the model's training fit.



**Figure 13.** Sales forecasting model validation of existing dates to show the model's training fit.



**Figure 14.** ARIMA mean sales forecasting of two months to forecast future sales



**Figure 15.** Automatic forecast using power BI sales forecasting (with minimum, trend, and maximum values) for two months to forecast future sales.

The overall results were compiled on an interactive dashboard to show the data insights statically as well dynamic reporting with multiple ad hoc filters. Here, we only



The most important dimension in this study was the time data, which gave more rich features (derived from the date) and had multiple scenarios for the evaluation of the facts. Using smart filters, the 360-degree analysis was possible for all of the data. We similarly extended the generic sales performance for order characteristics such as how many orders were placed within a specific price range of products, a comparison of revenue on the weekends and working days, and multiple options in the date and time hierarchy in terms of years, semesters, quarters, months, and weeks.

The order analysis for products and customers showed the quantity of products sold and sales revenue with the order characteristics. With smart filters, viewing the data with multiple factors was very informative and helpful to answering the short- and long-term performance-monitoring questions. The cancellation order analysis for products and customers was summarized. Quality results are totally dependent on the given data and applied analytical framework, as discussed above regarding the results for ML and BI. All the statistical graphs showed the hidden insights as well the analytical framework evaluation. Every section of the charts was cross-filtering enabled, which showed the business insights in a detailed manner. The interactive reports were summarized on a sales dashboard for specific dimensions and are shareable with associated sales persons and stakeholders.

#### 4. Conclusions

Business intelligence is a growing technology that uses data and computational analytical techniques to find business insights. Business owners and management are facing data management, customer relationship management, customer experience, departmental business data, operational system quality, and KPI issues to smartly manage their businesses and gain competitive advantages. The proposed framework is based on a hybrid solution that mainly consists of warehousing and ML empowered with BI data reporting, which provides statistical, ranking, and dimensional analysis of the past and current views of a business and a predictive view to forecasting a business's performance. These analytical sales reports are user-friendly, easily shareable across departments, and very useful in performance monitoring and for designing the best strategies. Users can utilize smart filters to obtain fully dynamic and versatile business insights that can truly answer analytical questions. This is purely a self-service business intelligence model implementation with ML and effective data reporting. The overall process is very efficient from data loading to processing, schema design, pattern computation, and data reporting, as well sharing with stakeholders to view the performance quickly. The smart framework comprises the sales metrics with respect to different dimensions that are very helpful for designing business strategies. Compared with existing techniques, this will provide a lightweight, centralized, and comprehensive view of a business with less errors, a low cost, and fast access with dynamic business analysis.

#### *Future Work*

As the dataset under consideration belongs to online retail, which has customer-based features, no shipping or payment information was given in the detailed product information, and any supplier information was not provided. The proposed approach will give more interesting and needed insights if the data are enriched with more detailed information. This study can be further extended with more inter-department and external source data such as product reviews and demands to evaluate the overall growth, dependent areas, and future goals and strategies.

**Author Contributions:** Conceptualization, M.S.A.; formal analysis, H.A.A., S.R. and M.I.; investigation, A.A., J.F., A.S., S.R. and M.I.; methodology, M.S.A., J.F. and A.S.; project administration, A.A., S.M.A., S.R. and M.I.; resources, A.A., H.A.A., S.R., S.M.A. and M.I.; software, M.S.A. and A.S.; supervision, H.A.A.; validation, J.F. and S.M.A.; visualization, A.A. and H.A.A.; writing—original draft, M.S.A. and J.F.; writing—review and editing, A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Deanship of Scientific Research at Najran University and the Kingdom of Saudi Arabia for funding this work under the research group funding program, grant code number NU/RG/SERC/11/3.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data and code presented in this study are available on <http://www.kaswa.ai/>.

**Acknowledgments:** The authors acknowledge the support from the Deanship of Scientific Research at Najran University and the Kingdom of Saudi Arabia for funding this work under the research group funding program, grant code number NU/RG/SERC/11/3.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Nica, I.; Chiriță, N.; Ionescu, Ș. Using of KPIs and Dashboard in the analysis of Nike company's performance management. *Theor. Appl. Econ. Asoc. Gen. Econ. Rom.—AGER* **2021**, *28*, 61–84.
2. Sang, G.; Xu, L.; de Vrieze, P.T. Implementing a Business Intelligence System for small and medium-sized enterprises. In Proceedings of the SQM 2016: 24th International Software Quality Management Conference, Bournemouth, UK, 21–22 March 2016.
3. Ukhalkar, P.K.; Phursule, D.R.N.; Gadekar, D.D.P.; Sable, D.N.P. Business Intelligence and Analytics: Challenges and Opportunities. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 2669–2676.
4. Tripathi, A.; Bagga, T. Leading Business Intelligence (BI) Solutions and Market Trends. In Proceedings of the International Conference on Innovative Computing Communications (ICICC), Delhi, India, 20–22 February 2020.
5. Gupta, S.; Ramachandran, D. Emerging Market Retail: Transitioning from a ProductCentric to a Customer-Centric Approach. *J. Retail.* **2021**, *97*, 597–620. [CrossRef]
6. Mueller, C. Usage of Business Intelligence Solutions within the Mergers and Acquisitions Process. In Proceedings of the 14th IWKM—International Workshop on Knowledge Management, Bratislava, Slovakia, 7–8 November 2019.
7. Jin, D.H.; Kim, H.J. Integrated understanding of big data, big data analysis, and business intelligence: A case study of logistics. *Sustainability* **2018**, *10*, 3778. [CrossRef]
8. du Plessis, D.P. A data warehouse model for quicker and less expensive implementation. In Proceedings of the 2nd International Conference on Intelligent and Innovative Computing Applications, Plaine Magnien, Mauritius, 24–25 September 2020; pp. 1–9.
9. ElMalah, K.; Nasr, M. Cloud business intelligence. *Int. J. Adv. Netw. Appl.* **2019**, *10*, 4120–4124. [CrossRef]
10. Simanjuntak, M.; Putri, N.E.; Yulianti, L.N.; Sabri, M.F. Enhancing customer retention using customer relationship management approach in car loan business. *Cogent Bus. Manag.* **2020**, *7*, 1738200. [CrossRef]
11. Ragulan, B.; Subash, R. Designing a Data Warehouse System for Sales and Distribution Company. *Big Data Min. Anal.* **2021**, *1*, 1–6.
12. Sohrabpour, V.; Oghazi, P.; Toorajipour, R.; Nazarpour, A. Export sales forecasting using artificial intelligence. *Technol. Forecast. Soc. Chang.* **2021**, *163*, 120480. [CrossRef]
13. Lahbi, H. The Power of Business Intelligence on the Decision-Making Process at Linköping University A Case Study. 2018. Available online: <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-153451> (accessed on 13 August 2022)
14. Kalaiarasan, T.R.; Anandkumar, V.; Nivetha, R.; Mathumitha, B. Data Analytics for Retail Industry Using QLIK. *J. Xi'an Univ. Archit. Technol.* **2020**, *12*, 1276–1279.
15. Moorthi, K.; Dhiman, G.; Arulprakash, P.; Suresh, C.; Srihari, K. A survey on impact of data analytics techniques in E-commerce. *Mater. Today Proc.* **2021**, in press. [CrossRef]
16. Karb, T.; Kühl, N.; Hirt, R.; Glivici-Cotruta, V. A network-based transfer learning approach to improve sales forecasting of new products. *arXiv* **2020**, arXiv:2005.06978.
17. Kaipov, I.; Nedzved, A. Sales forecasting of goods in shoe retail. *Cent. Eur. Res. J.* **2020**, *6*, 10–17.
18. Huber, J.; Stuckenschmidt, H. Daily retail demand forecasting using machine learning with emphasis on calendric special days. *Int. J. Forecast.* **2020**, *36*, 1420–1438. [CrossRef]
19. Abdi, F.; Abolmakarem, S. Customer Behavior Mining Framework (CBMF) using clustering and classification techniques. *J. Ind. Eng. Int.* **2019**, *15*, 1–18. [CrossRef]
20. Exenberger, E.; Bucko, J. Analysis of online consumer behavior-Design of CRISPDm process model. *AGRIS On-Line Pap. Econ. Inform.* **2020**, *10*, 13–22. [CrossRef]
21. Singh, B.; Kumar, P.; Sharma, N.; Sharma, K.P. Sales forecast for amazon sales with time series modeling. In Proceedings of the 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, 3–5 January 2020; IEEE: New York, NY, USA, 2020; pp. 38–43.
22. Swami, D.; Shah, A.D.; Ray, S.K. Predicting Future Sales of Retail Products using Machine Learning. *arXiv* **2020**, arXiv:2008.07779.
23. Rino, R. Implementation of Business Intelligence in Data Superstore Sales with Online Analytical Processing Method. *bit-Tech* **2020**, *3*, 44–50.

- 
24. Sureddy, M.R.; Yallamula, P. A Framework for Monitoring Data Warehousing Applications. *Int. Res. J. Eng. Technol.* **2020**, *7*, 7023–7029.
  25. Bhatiasevi, V.; Naglis, M. Elucidating the determinants of business intelligence adoption and organizational performance. *Inf. Dev.* **2020**, *36*, 78–96. [[CrossRef](#)]