

Sales Forecasting and Visualization

¹K R Kaushik Kumar, ²Jyothsna R, ³Sindhu K

^[1,2]Undergraduate Students of Department of Information Science and Engineering, BMSCE, Bangalore

³Assistant Professor of Department of Information Science and Engineering, BMSCE, Bangalore

Abstract - Sales forecasting plays a crucial role in predicting future sales trends and facilitating informed decision-making in business settings. This paper presents a comprehensive approach to sales forecasting by leveraging the SARIMA (Seasonal Autoregressive Integrated Moving Average) model, renowned for its ability to capture seasonality and trend patterns in time series data. The study further incorporates the powerful suite of tools provided by the Google ecosystem, including Google Drive, Google Sheets, Google Colab, Google Data Studio, and Google Cloud Console, for efficient data storage, collaborative work, and visualization. By utilizing these platforms, the research enhances the accuracy and accessibility of sales forecasts, enabling businesses to optimize sales strategies and improve inventory management. The findings of this study contribute to empowering organizations with valuable insights for making informed decisions in dynamic market environments.

Index Terms - Business analytics, Sales Forecasting, Statistical methods, time series forecasting, Deep learning, Machine Learning, profit prediction, SARIMA.

I. INTRODUCTION

Business Analytics is a rapidly evolving field that aims to harness the power of data and analytics to drive decision-making and enhance organizational performance. In a comprehensive overview by [2], the field of Business Analytics is defined, and a research agenda is identified. The study emphasizes the importance of understanding the various aspects of Business Analytics and its potential impact on organizations. Building on this research agenda, in [3] propose a study to investigate the impact of Business Analytics on decision-making processes within organizations. Their research highlights the need to explore how the adoption and utilization of Business Analytics tools and techniques can transform decision-making processes and contribute to organizational success.

Furthermore, [4] delve into the intersection of Business Analytics and Data Science, shedding light on the similarities, differences, and potential synergies between these two fields. Their study emphasizes the complementary nature of Business Analytics and Data Science and how organizations can leverage both disciplines to gain valuable insights and drive strategic decision-making. Additionally, [5] conduct a study that examines the factors influencing the adoption and implementation of Business Analytics solutions within organizations. Their research explores key factors such as organizational culture, data quality, and leadership support that play a crucial role in the successful deployment of Business Analytics initiatives.

Business Analytics encompasses various types of analyses that organizations employ to extract insights from data for informed decision-making. Sales forecasting, one such analysis, involves analyzing historical sales data and market trends to predict future sales volumes and patterns. It aids businesses in estimating future demand, planning inventory levels, and devising effective sales and marketing strategies. Revenue forecasting focuses on predicting an organization's future revenue streams by analyzing historical financial data, market conditions, customer behavior, and other relevant factors. This assists in financial planning, investment evaluation, and strategic decision-making. Financial forecasting, on the other hand, utilizes historical financial data and statistical models to predict an organization's future financial performance, aiding in budget planning and assessing the financial feasibility of strategic initiatives.

Customer Analytics leverages customer data to gain insights into behavior, preferences, and buying patterns, enabling organizations to enhance customer experience through segmentation, personalization, and optimized acquisition and retention strategies. Marketing Analytics evaluates the effectiveness of marketing campaigns, measures return on investment (ROI), and optimizes strategies by analyzing data from various marketing channels. Supply Chain Analytics optimizes supply chain operations, reduces costs, and improves performance through data analysis of procurement, inventory, logistics, management, and demand forecasting. Risk Analytics identifies and manages potential risks related to operations, finance, compliance, and cybersecurity. Operational Analytics improves efficiency and performance by analyzing data from various operational processes to identify bottlenecks, optimize resource allocation, and streamline workflows. Together, these types of Business Analytics provide organizations with valuable insights and decision support to drive strategic planning, operational excellence, and financial success.

Collectively, these studies contribute to the understanding of Business Analytics as a dynamic and impactful field. They emphasize the need for further research to explore the implications, challenges, and opportunities associated with the adoption and utilization of Business Analytics within organizations. By leveraging data-driven approaches and analytics tools, businesses can unlock valuable insights, drive innovation, and gain a competitive advantage in today's data-driven landscape. This paper specifically focuses on "Sales Forecasting" and aims to provide an in-depth analysis of this critical aspect of Business Analytics.

The proposed system aims to improve sales forecasting by integrating the SARIMA model into the Google ecosystem using various tools such as Google Drive, Google Sheets, Google Data Studio and Google Cloud Console. The process begins by securely storing a set of data in Google Drive, providing centralized data management and easy access. Data preprocessing, including operations such as data cleaning and feature design, takes place in Google Sheets. The core of the system is based on the SARIMA model, which stores sales data and seasonality to create accurate forecasts. To visualize the results and get an overview, the system takes the forecasts to Google Data Studio and creates interactive dashboards and reports. The Google Cloud Console manages the entire ecosystem, ensuring seamless integration and unified resource management. Overall, this integrated system aims to improve the accuracy and usability of

sales forecasts, enhance informed decision-making and promote business success.

II. LITERATURE SURVEY

Sales forecasting is a crucial aspect of business analytics that involves analyzing historical sales data and market trends to predict future sales volumes and patterns. By examining past sales performance and considering factors such as seasonality, economic conditions, and consumer behavior, organizations can estimate future demand and plan their sales strategies accordingly. Sales forecasting enables businesses to make informed decisions regarding production levels, inventory management, pricing, and marketing efforts. It provides valuable insights into potential sales growth, identifies opportunities for expansion or optimization, and helps businesses align their resources to meet customer demands effectively. With accurate sales forecasts, organizations can improve operational efficiency, optimize revenue generation, and stay competitive in the market.

The paper [1] presents a study on the use of the SARIMA model for forecasting daily sales in the food retail industry. The study demonstrates that the SARIMA model is able to capture the temporal dependencies and seasonal patterns in the data, resulting in highly accurate sales forecasts. In a comparative analysis conducted in [2], different machine learning models were evaluated for forecasting seasonal items sales. The study found that the SARIMA model was the most accurate model, followed by the Prophet model and the ARIMA model. Similarly, [3] focuses on sales forecasting for Amazon using time series modeling. The study highlights the effectiveness of the SARIMA model in capturing seasonal patterns and accurately forecasting sales on Amazon's platform.

In the general trade industry, the SARIMA model was utilized for sales forecasting in [4]. The study demonstrates the SARIMA model's ability to capture temporal dependencies and seasonal patterns in the sales data, resulting in highly accurate sales forecasts. The paper [6] delves into sales forecasting using neural networks, emphasizing their ability to capture complex sales patterns and provide accurate predictions. Their research showcases the potential of neural networks in enhancing sales forecasting accuracy. In a similar vein, [7] focuses on sales forecasting for retail stores, employing machine learning techniques. Their study investigates the application of machine learning models to analyze historical sales data and predict future sales accurately. The research conducted by Krishna et al. sheds light on the optimization of sales forecasting in the retail industry.

The paper [8] contributes to the field with research on machine learning models for sales time series forecasting. By leveraging a range of machine learning algorithms, Pavlyshenko aims to capture underlying patterns and dynamics in sales data, leading to improved accuracy in sales forecasts. The paper [9] proposes intelligent sales prediction using machine learning techniques. Their research integrates advanced algorithms such as support vector machines and random forests to analyze sales patterns and make accurate predictions. The study highlights the effectiveness of machine learning in enhancing sales forecasting. Moreover, [10] provides a comprehensive review of trends in machine learning applied to demand and sales forecasting. Their study offers insights into advancements and emerging techniques in machine learning for sales forecasting, providing a holistic view of the field. Additionally, [11] presents a case study on the use of supervised machine learning for B2B sales forecasting, specifically focusing on spare parts sales at an after-sales service provider. Their research highlights the practical application and benefits of supervised machine learning techniques in sales forecasting.

Furthermore, [12] conducts a case study applying Long Short-Term Memory (LSTM) neural networks to sales forecasting in the retail industry. Their research demonstrates the efficacy of LSTM networks in capturing sequential dependencies and improving sales forecasting accuracy. Lastly, [13] discusses various sales forecasting techniques. Their research provides an overview of different methods used in sales forecasting, enabling a better understanding of the available techniques and their applicability in different contexts.

By incorporating the findings from these studies, businesses can enhance their sales forecasting capabilities and reap numerous benefits. Accurate sales forecasts provide organizations with a competitive edge by enabling them to optimize their operational efficiency. With reliable predictions of future sales volumes and patterns, businesses can effectively plan their production levels, align their inventory management strategies, and streamline their supply chain operations. This, in turn, minimizes the risk of overstocking or stockouts, leading to improved cost-effectiveness and resource allocation. Furthermore, accurate sales forecasts empower businesses to make informed pricing decisions. By anticipating future demand, organizations can adjust their pricing strategies to maximize profitability. They can identify price elasticity trends and optimize their pricing structures to capture the highest possible revenue while remaining competitive in the market.

Sales forecasting also plays a pivotal role in guiding marketing efforts. By understanding future sales trends, businesses can tailor their marketing campaigns to target specific customer segments and optimize their promotional activities. This ensures that marketing resources are effectively allocated, resulting in higher return on investment and improved customer acquisition and retention. Moreover, sales forecasting provides valuable insights into potential sales growth and expansion opportunities. By identifying emerging market trends and consumer preferences, businesses can proactively adapt their strategies and capitalize on new market segments. This enables them to stay ahead of competitors and explore untapped markets, ultimately fostering business growth and profitability.

In summary, the collective findings from these studies on sales forecasting contribute to a deeper understanding of the field and offer practical insights for businesses. By leveraging advanced techniques such as neural networks and machine learning models, organizations can improve the accuracy of their sales forecasts, leading to better decision-making, optimized resource allocation, enhanced operational efficiency, and ultimately, improved revenue generation and competitiveness in the market.

III. METHODOLOGY

In this research, the Sales dataset was obtained from the eforexcel.com website [21], consisting of approximately 1 million sales records spanning 46 years. To ensure data quality, preprocessing techniques were applied, and descriptive statistics and data visualization were performed. The primary forecasting method used was the SARIMA model, known for its effectiveness in capturing seasonal patterns in sales data. The Google ecosystem, including tools like Google Drive, Sheets, Colab, Data Studio, and Cloud Console, was utilized for data storage, manipulation, visualization, and computational resources. By following this methodology as shown in Fig. 1, the research aimed to provide accurate sales forecasts and contribute to the understanding of sales forecasting techniques in real-world business scenarios.

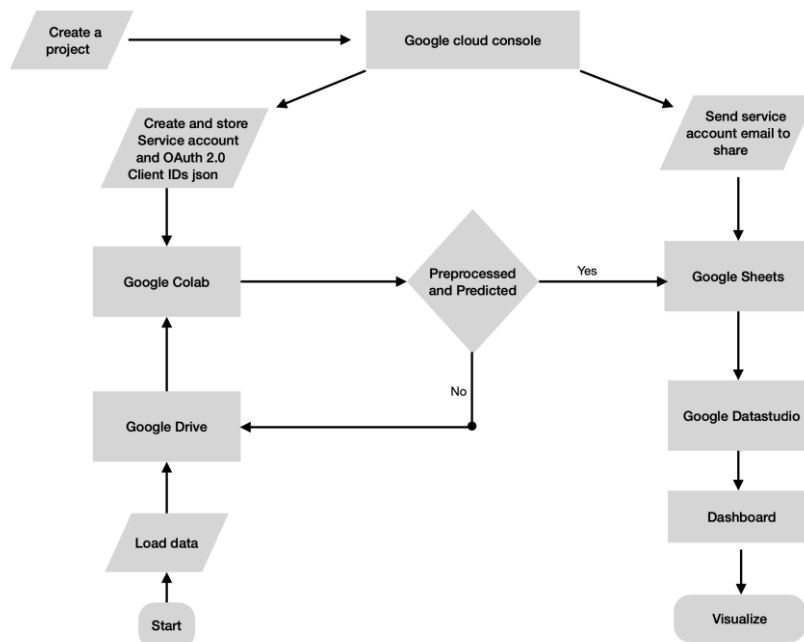


Figure 1 : System Architecture

1) Project Setup and Connection Establishment:

A project was created in the Google Cloud Console to run the proposed method. This project serves as a central hub for managing data analysis and visualization tasks. Once the project is complete, the next step is to create connections between the various components of the Google ecosystem. Connections are made between Google Cloud Console, Google Colab, Google Drive, Google Sheets and Google Data Studio. This ensures seamless data flow and communication between these platforms, facilitating the efficient execution of data analysis and visualization tasks. A service account is created in the Google Cloud Console as part of the installation process. A service account is a special Google account that represents a project, not an individual user. It allows project resources to be accessed and managed programmatically.

In addition, OAuth 2.0 client credentials are generated in the Google Cloud Console. These client identifiers are used to authenticate the project and establish secure connections with other Google services. The OAuth 2.0 protocol provides a standardized way for the project to access and interact with Google APIs. After the service account and OAuth 2.0 client credentials are created, the JSON files associated with those accounts are retrieved. These JSON files contain important credentials and configuration information. JSON files are stored in Google Colab, the environment of choice for data analysis and modeling.

If you follow these steps, the project will be configured correctly and the necessary connections will be made between Google Cloud Console, Google Colab, Google Drive, Google Sheets and Google Data Studio. It enables seamless integration of these platforms and leverages their features for data analysis and visualization.

2) Share Email from OAuth Client:

To establish a connection between the OAuth client and the Google Sheets where the data is stored, the email associated with the OAuth client created in the Google Cloud Console must be obtained. This email serves as a unique identifier for the customer and is necessary to access the necessary resources. Access to the OAuth client settings or information in the Google Cloud Console is required to download the email.

The email address is usually in the format customername@projectname.iam.gserviceaccount.com. Once received, the email can be shared in a Google Sheets document. Sharing an email with the appropriate permissions in Google Sheets ensures that the OAuth client has the necessary permissions to interact with the data stored in the spreadsheet. In the "Share" settings of the Google Sheets document, you can navigate and enter the desired email address. Here you can set the desired permissions, such as view-only permissions, edit permissions or custom permission levels according to project requirements. By sharing the email with the appropriate access permissions, the OAuth client is authorized to read, write, and modify the data in the Google Sheets document. This enables seamless integration between the OAuth client and the spreadsheet, allowing for efficient data retrieval, updates, and synchronization between different components of the project.

3) Data Storage in Google Drive:

For efficient information management and ease of use, Google Drive acts as a centralized archive to store project information. Google Drive offers a secure and reliable cloud-based storage solution that allows you to store and organize data in a structured way. When the data is ready to be saved, folders and subfolders can be created in Google Drive to organize the data according to project requirements. This hierarchical organization helps maintain a logical structure and allows for easy navigation and retrieval of information when needed. Files containing data can be stored in any directory. These files can be in different formats (eg CSV, Excel or other compatible format) depending on the nature of the data. By organizing the data warehouse in this way, it ensures that the data is well organized and readily available for analysis, processing and other activities throughout the life of the project.

Google Drive also offers features such as search capabilities, version control, and collaboration features that further improve data management. Platform features can be used to efficiently find specific files, track changes, and collaborate with team members or project stakeholders. Overall, using Google Drive as a data storage solution provides a secure, scalable and user-friendly platform for storing and organizing project data. This simplifies data management, improves collaboration and ensures that data is readily available for further analysis, processing and integration into other parts of the project.

4) Data Fetching to Google Colab:

To perform analysis and model training, data must be downloaded from Google Drive to Google Colab. Using data stored in Google Drive requires the drive to be connected to the Colab environment. By installing Google Drive, a link is established between Google Colab and Google Drive, allowing direct access to and processing of data in Colab notebooks. This connection ensures seamless integration between the two platforms, enabling efficient work with data. Once the drive is successfully installed, the necessary data can be downloaded from Google Drive to the Colab environment. This can be done by specifying a file path or using the appropriate APIs provided by Google Colab and Google Drive. Taking data brings it into Colab notebooks, making it readily available for analysis, processing and model training.

Once the data is retrieved, the preprocessing process can begin. In preprocessing, raw data is converted into a format suitable for modeling. This step may include tasks such as cleaning the data, handling missing values, scaling attributes, and performing other necessary transformations to ensure that the data is suitable for analysis. By preprocessing data in Google Colab, it prepares other modeling tasks. This step is crucial to ensure the quality and integrity of the data and to optimize its suitability for the specific modeling techniques intended to be used. In short, it can be said that bringing data from Google Drive to Google Colab allows you to use and process data in the Colab environment. Installing the drive and preprocessing the data creates a seamless workflow that enables analysis and modeling using data obtained from Google Colab.

5) Data Storage and Automatic Update in Google Sheets:

To ensure efficient data management and smooth updates, a separate table is created in Google Sheets, where data and predictions generated by models are stored and pre-processed. This spreadsheet serves as a centralized repository for project information and provides a structured and organized format. The link between Google Colab and Google Sheets uses APIs and libraries available to enable communication between the two platforms. Using these tools, updating the spreadsheet with new forecasts can be automated.

Once connected, the integration is configured to automatically update a particular spreadsheet in Google Sheets when new forecasts are created in Google Colab. This ensures that forecasts are quickly displayed in a spreadsheet without manual intervention. Implementing automatic updates simplifies the process and eliminates the possibility of errors or delays in manual transmission of forecasts. This integration improves overall workflow efficiency and reliability by providing real-time access to up-to-date forecasts in the Google Sheets environment.

In conclusion, creating a separate spreadsheet in Google Sheets and connecting to Google Colab facilitates seamless data storage and automatic updates. This integration ensures that the spreadsheet contains the latest predictions generated by the models, providing a complete and accurate representation of project results.

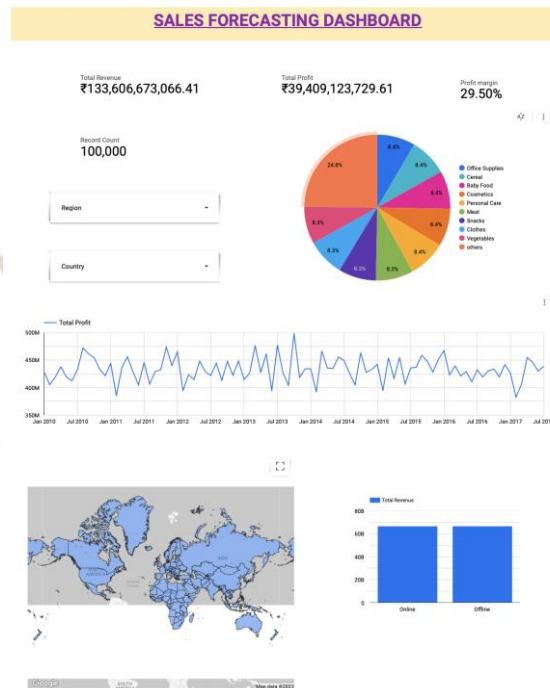
6) Preprocessing :

Data preprocessing is a critical step in preparing the sales data for analysis and model training. Given the large volume and unstructured nature of the sales data, it required several preprocessing techniques to clean, transform, and enhance the data for further processing. The following techniques were applied to the dataset:

- A. **Handling Missing Values:** If missing values are detected in the dataset, they can be handled using the SimpleImputer class from the scikit-learn library, which provides a more flexible and comprehensive approach for imputing missing values. The SimpleImputer class allows us to specify different imputation strategies, such as mean, median, mode, or constant, depending on the nature of the data.
- B. **Outlier Detection and Treatment:** Outliers, if present in the dataset, can be detected using suitable methods such as statistical measures or visualization techniques. Depending on the analysis requirements and the significance of outliers, various approaches can be considered, including removal of outliers, transformation techniques, or using robust statistical models.
- C. **Data Validation and Quality Checks:** To ensure the integrity and quality of the dataset, various data validation and quality checks can be performed. This involves examining the data for inconsistencies, duplicate records, incorrect data types, or any other anomalies that could affect the reliability of the analysis.
- D. **Feature Selection and Extraction:** Feature selection and extraction techniques can be employed to identify the most relevant and informative features for the analysis. This involves analyzing the relationship between the features and the target variable, and selecting a subset of features that contribute significantly to the analysis. Techniques such as correlation analysis, statistical tests, or

domain knowledge can be considered for feature selection. In this context, the target variables can be Total Profit alone for prediction or both Total Profit and Total Revenue for prediction.

- E. Data Transformation: Data transformation techniques are used to normalize or standardize the data, handle skewed distributions, or remove scale differences among features. Common techniques include normalization, standardization, logarithmic transformation, or power transformation.
- F. Encoding Categorical Variables: Categorical variables need to be encoded into numerical representations for analysis. This can be



done using techniques such as one-hot encoding, label encoding, or ordinal encoding.

Figure 2 : Dashboard for Preprocessed Data

7) Data Visualization in Google Data Studio:

Data visualization plays a vital role in effectively communicating insights and trends from the preprocessed data. After the data is preprocessed in Google Colab and stored in a Google Sheets spreadsheet, it can be accessed and utilized for further analysis and visualization in Google Data Studio. Google Data Studio offers a wide range of visualization options to create interactive and visually appealing dashboards based on the preprocessed data.

In Google Data Studio, stakeholders can leverage various types of charts, graphs, and visual elements to present the insights derived from the preprocessed data in a meaningful and understandable way. These interactive dashboards enable users to explore the data, uncover patterns, and gain valuable insights. The dashboards can be customized and tailored to suit the specific needs and preferences of the stakeholders.

One of the key advantages of using Google Data Studio is its ability to provide real-time updates. Any modifications or updates made to the preprocessed data in the Google Sheets spreadsheet automatically reflect in the visualizations within Google Data Studio. This ensures that the dashboards are always up to date and accurately represent the latest insights, allowing stakeholders to make data-driven decisions based on the most current information.

Furthermore, Google Data Studio offers interactive features such as filters, slicers, and drill-down capabilities. These features allow users to interact with the preprocessed data as shown in Fig. 2, filter specific subsets, and dive into detailed views. By interacting with the visualizations, stakeholders can gain a comprehensive understanding of the preprocessed data, explore different perspectives, and uncover deeper insights.

By leveraging the capabilities of Google Data Studio for data visualization, stakeholders can effectively communicate the insights derived from the preprocessed data. The combination of preprocessing techniques, storing the data in Google Sheets, and visualizing it in Google Data Studio empowers stakeholders to make data-driven decisions and gain valuable insights from the preprocessed data. Overall, the integration of preprocessing, data storage in Google Sheets, and visualization in Google Data Studio creates a comprehensive and efficient workflow for analyzing and communicating the preprocessed data. The seamless connection between these platforms enables stakeholders to effectively utilize the preprocessed data, explore visual representations, and derive actionable insights for informed decision-making.

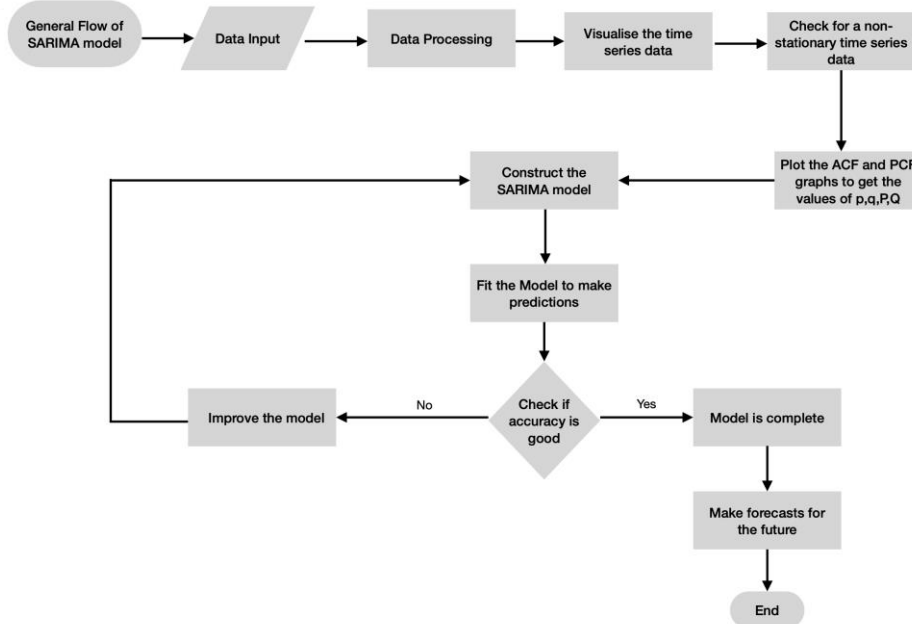
8) SARIMA Model Building and Prediction:

- A. Data Preprocessing: The initial step in the code involves data preprocessing. The sales dataset is loaded, and the data is processed by grouping it based on different factors such as country, region, item type, and order date [16,17]. This grouping allows for the aggregation of relevant information required for analysis. Specifically, the total profit and total revenue values are calculated by summing the corresponding values for each group [16,17]. This results in a new dataset that contains the aggregated values, which will be used for further analysis.
- B. SARIMA Model Building: The code proceeds with building the SARIMA (Seasonal Autoregressive Integrated Moving Average) model. SARIMA is a popular time series forecasting model that incorporates autoregressive (AR) and moving average (MA) components, along with differencing to handle non-stationary data [16,19]. The specific parameters for the SARIMA model are provided by the authors of the research paper, and these parameters are incorporated into the code (References 1, 4). These parameters include the number of autoregressive terms (p), the degree of differencing (d), the number of moving average terms (q),

$$SARIMA \left(\underbrace{p, d, q}_{non-seasonal} \right) \left(\underbrace{P, D, Q}_m \right)_s$$

as well as the seasonal components (P, D, Q) and the length of the seasonal period (s) [16,19].

- C. Stationarity Check: Before training the SARIMA model, it is crucial to check the stationarity of the time series data (References 1, 4). Stationarity implies that the statistical properties of the data, such as mean and variance, remain constant over time. The code performs a stationarity check using the Augmented Dickey-Fuller (ADF) test. The ADF test provides a p-value, and if the p-value is greater than 0.05, it suggests that the data is non-stationary. In such cases, differencing is applied to the data to make it stationary [16,19]. Differencing involves subtracting the previous observation from the current observation to remove trends or seasonality,



ensuring that the data meets the stationarity assumption required for SARIMA modeling.

Figure 3 : SARIMA Model

- D. SARIMA Model Training: Once the data has been preprocessed and made stationary if necessary, the code proceeds with training the SARIMA models. Separate models are built and trained for total profit and total revenue using the specified SARIMA parameters [16,19]. The order and seasonal order have been taken as (2,1,4) and (0,1,1,7) from the paper[1]. The parameters for the training process involves fitting the SARIMA model to the preprocessed data, which estimates the model's coefficients and other parameters. This allows the model to capture the patterns and relationships present in the data, enabling it to make accurate predictions.
- E. Forecasting: After training the SARIMA models, the code utilizes them to make predictions for the next 12 months [16]. This is accomplished using the get_forecast() function, which takes the trained SARIMA model as input and generates the forecasted values for total profit and total revenue. The function utilizes the model's parameters and the historical data to produce predictions for the future time periods [19]. The forecasted values represent the expected future values for the respective time series, providing insights into the future trends and patterns in total profit and total revenue.
- F. Storage of Predictions: The predictions for total profit and total revenue are stored in a dictionary structure for easy access and analysis. The dictionary is organized with keys representing the unique combinations of country, region, and item type [19]. For each combination, the corresponding values in the dictionary contain the predicted values for total profit and total revenue for each month. This storage mechanism ensures that the predictions are organized and can be efficiently retrieved for further analysis or visualization [16].

G. Printing Predictions: The code concludes by printing the predictions for all combinations of country, region, and item type [16]. It iterates over the stored predictions in the dictionary and displays the country, region, item type, and the predicted total profit and total revenue values for each month. This provides a clear and concise representation of the forecasted values, allowing stakeholders to examine and interpret the predicted outcomes easily.

By following these steps and utilizing the SARIMA model with the specified parameters from the research paper, the code effectively preprocesses the data, builds and trains the SARIMA models, performs forecasting, and stores the predictions [18]. The predictions can serve as valuable inputs for decision-making, resource allocation, and strategic planning in the context of sales and business operations [16]. They provide insights into the expected future trends and patterns in total profit and total revenue for various combinations of country, region, and item type, supporting informed decision-making processes [19].

9) Automatic Update of Predictions in Google Sheets:

When new forecasts are created in Google Colab, there are two ways to automatically update the forecast table in Google Sheets: scripts and scheduled execution. A script refers to the execution of a program or script that is executed in Google Colab. This script is responsible for creating forecasts and updating the forecast table in Google Sheets. To achieve this, Google Colab and Google Sheets must be connected using appropriate APIs or libraries, such as the Google Sheets API or Python's gspread library. The script can be scheduled to run manually, or it can be triggered by an event, such as the completion of a forecasting process.

Scheduled execution involves defining a schedule or time-based trigger to run the forecasting process automatically. This can be done using various available tools or services such as Google Cloud Scheduler or external cron jobs. By defining a schedule, the forecasting process, including creating new forecasts and updating the forecast table in Google Sheets, can be done at regular intervals without manual intervention. The purpose of both methods is to make it easy to update the forecast table in Google Sheets, every time new forecasts are created. Enabling a script or scheduled execution ensures that forecasts automatically appear in the spreadsheet and provide up-to-date information without manual updating. The choice between the two methods depends on the specific requirements, preferences and tools available in the environment.

10) Automatic Update of Visualizations in Google Data Studio:

After successfully automatically updating the forecast table in Google Sheets, the next step is to ensure that these updates are reflected in Google Data Studio visualizations and dashboards. Thus, dynamic and up-to-date visual representations can be obtained from the data. Article [15] provides a good example of how Google Data Studio can be used to visualize time series data. In this study, the authors used Google Data Studio to create a dashboard illustrating the number of tourist visits to Bali over time. The dashboard included various visualizations such as line charts, bar charts, and pie charts. Using these visualizations, the authors were able to track trends in tourist visits over time and identify the factors behind these trends.

To automatically update visualizations in Google Data Studio based on changes to the forecast table, follow these steps:

- A. Connect Google Data Studio to the predictions spreadsheet: Create a connection between Google Data Studio and a forecast table in Google Sheets. This connection allows you to import spreadsheet data into Google Data Studio reports.
- B. Create visualizations: Utilize the data from the predictions spreadsheet to create visualizations and charts in Google Data Studio. Select appropriate visualization types that effectively represent and communicate the insights.
- C. Configure automatic update settings: Google Data Studio provides an automatic update feature that allows you to specify the refresh rate or schedule for updating the data in your visualizations. This feature ensures that the visualizations reflect the latest data from the predictions spreadsheet.
- D. Test and verify the setup: After configuring the automatic update settings, validate that the visualizations in Google Data Studio automatically update based on changes in the predictions spreadsheet. Make changes to the spreadsheet and observe if the visualizations reflect those changes in real-time.

By utilizing the automatic refresh feature of Google Data Studio, a smooth and synchronized flow of data between the forecast table and visualizations can be maintained. All updates to the forecast table are automatically reflected in the visualizations without manual intervention or updating. This automation streamlines the process and ensures that visualizations and dashboards always show the latest predictions and insights. It can be used to make informed decisions and share current information with the public. Regular monitoring of the automatic update is recommended to ensure its continued effectiveness and to troubleshoot issues.

11) Iterative Analysis and Visualization:

After successfully completing the previous step of setting up the automatic update of the predictions spreadsheet in Google Sheets, the next phase involves iterative analysis and visualization using Google Colab, Google Sheets, and Google Data Studio. This iterative process is essential for continuously improving the accuracy and performance of predictions and staying up to date with the latest insights. Iterative analysis and visualization involve the following steps:

- Continuous data analysis and visualization: Utilize the capabilities of Google Colab, Google Sheets, and Google Data Studio to regularly analyze and visualize the data. This includes exploring patterns, trends, and relationships within the data and identifying any anomalies or outliers. By consistently engaging in data analysis and visualization, one can gain a deeper understanding of the data and uncover valuable insights.
- Iteration on model training, evaluation, and prediction: Enhance the accuracy and performance of predictions through an iterative approach. This iterative process allows for continuous improvement in the accuracy and reliability of predictions.

- Monitoring of automatic updates: Follow the automatic updates of Google Sheets and Google Data Studio regularly to stay up to date with the latest predictions and statistics. Monitoring automatic updates ensures you're using the latest information and can make informed decisions.

By following this formalized iterative analysis and visualization process, individuals can refine their models, gain deeper insights from the data, and improve the accuracy of predictions over time. The integrated workflow of Google Colab, Google Sheets, and Google

| Country | Region | Item Type | Month | Profit | Revenue |
|-------------|-----------------|-----------|-------|-------------|-------------|
| Afghanistan | Middle East and | Baby Food | 1 | 320691.6843 | 854015.8247 |
| Afghanistan | Middle East and | Baby Food | 2 | 403201.2792 | 1073740.96 |
| Afghanistan | Middle East and | Baby Food | 3 | 435406.8517 | 1159507.747 |
| Afghanistan | Middle East and | Baby Food | 4 | 201184.7356 | 535763.6199 |
| Afghanistan | Middle East and | Baby Food | 5 | 430460.9918 | 1146339.969 |
| Afghanistan | Middle East and | Baby Food | 6 | 260862.2065 | 694689.5457 |
| Afghanistan | Middle East and | Baby Food | 7 | 367539.8262 | 978778.5408 |
| Afghanistan | Middle East and | Baby Food | 8 | 470505.3421 | 1252979.134 |
| Afghanistan | Middle East and | Baby Food | 9 | 681309.9033 | 1814362.571 |
| Afghanistan | Middle East and | Baby Food | 10 | 617234.5089 | 1643729.207 |
| Afghanistan | Middle East and | Baby Food | 11 | 306806.8876 | 817049.0807 |
| Afghanistan | Middle East and | Baby Food | 12 | 354509.5575 | 944082.5329 |
| Afghanistan | Middle East and | Beverages | 1 | 99000.8508 | 299906.3965 |
| Afghanistan | Middle East and | Beverages | 2 | 102651.6181 | 310973.1517 |
| Afghanistan | Middle East and | Beverages | 3 | 62391.81445 | 188996.0421 |
| Afghanistan | Middle East and | Beverages | 4 | 67885.73128 | 205666.4671 |
| Afghanistan | Middle East and | Beverages | 5 | 66517.97064 | 201544.9339 |
| Afghanistan | Middle East and | Beverages | 6 | 68414.43053 | 207298.8898 |
| Afghanistan | Middle East and | Beverages | 7 | 103522.1607 | 313661.8052 |
| Afghanistan | Middle East and | Beverages | 8 | 108758.9115 | 329554.3875 |
| Afghanistan | Middle East and | Beverages | 9 | 71734.37573 | 217381.9845 |
| Afghanistan | Middle East and | Beverages | 10 | 101237.3381 | 306749.9045 |
| Afghanistan | Middle East and | Beverages | 11 | 98847.15494 | 299532.714 |
| Afghanistan | Middle East and | Beverages | 12 | 103359.529 | 313210.8459 |
| Afghanistan | Middle East and | Cereal | 1 | 661684.6907 | 1536399.52 |
| Afghanistan | Middle East and | Cereal | 2 | 588844.8435 | 1367274.634 |
| Afghanistan | Middle East and | Cereal | 3 | 481568.5753 | 1118186.697 |
| Afghanistan | Middle East and | Cereal | 4 | 734965.7002 | 1706552.098 |
| Afghanistan | Middle East and | Cereal | 5 | 619003.2566 | 1437292.155 |
| Afghanistan | Middle East and | Cereal | 6 | 888135.4948 | 2062197.247 |
| Afghanistan | Middle East and | Cereal | 7 | 849054.1246 | 1971453.613 |

Data Studio provides a robust platform for continuous improvement and monitoring of predictions and insights.

IV. RESULT

In this study, the project used SARIMA (Seasonal Autoregressive Integrated Moving Average) models to forecast sales. The sales data was pre-processed, which made it possible to create accurate prediction models. The resulting forecasts of total profit and total revenue were carefully recorded in a structure arranged in a Google Sheets spreadsheet as shown in fig. 4.

Figure 4 : Results stored in sheets

The automated workflow was created through the seamless integration of Google Colab, Google Sheets and Google Data Studio. This made it easy to instantly update forecasts in a spreadsheet, which were then reflected in interactive dashboards created by Google Data Studio as shown in fig 5. This dynamic aggregation ensured that visual representations of sales forecasts were constantly updated and accurately represented prevailing trends and patterns.

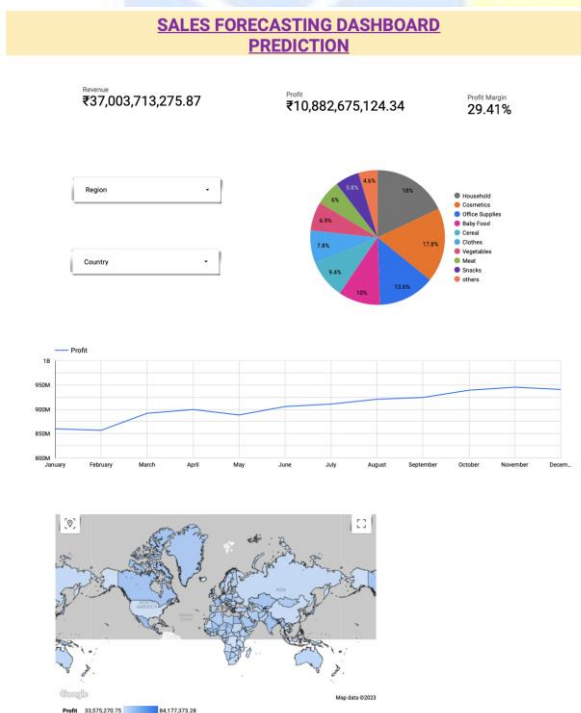


Figure 5 : Dashboard for predicted data

Through an iterative process of analysis and visualization, the project provided comprehensive results that enabled companies to make sound decisions, optimize inventory and improve sales strategies. This effort underscores the most important role of data methods in the field of sales forecasting.

V. CONCLUSION

In summary, it can be stated that this study made a comprehensive study of sales forecasting and its direct effects on companies, focusing mainly on the SARIMA methodology. Through the practical application of SARIMA, the main objectives of the research were to improve sales forecasts, facilitate decision-making processes, optimize inventory management strategies and design effective sales methods. The results strongly highlight the significant potential of SARIMA to effectively address these critical aspects. By creating accurate and effective sales forecasts, SARIMA enables organizations to make informed decisions and anticipate market fluctuations.

In addition, the research agenda highlighted in the articles emphasizes the impact of changes in business analysis, especially SARIMA, on the decision landscape of business units. This emphasis emphasizes the significant impact that analytics can have on the overall performance of organizations. Importantly, it underscores the need for continued in-depth research to fully understand the game's impact and establish best practices to improve SARIMA's capabilities.

Overall the implemented project is a concrete example of the growing importance of sales forecasting in business. This places SARIMA and business analytics as uniquely necessary tools to guide informed decision-making processes related to sales-oriented activities. Because of the insights provided, this study not only adds value, but also provides a cornerstone for future research projects, thus playing a key role in the progressive development of sales forecasting methods and techniques.

In addition, a key aspect of this study is the inclusion of panel visualization. Integrating data visualization with dashboards improves the availability and interpretability of sales forecast results. Visual representation of complex data models and forecasts gives stakeholders a clear view, helping them make effective decisions. These dashboards act as a bridge between advanced analytical results and practical decision-making, increasing the value of research findings and their practical applicability.

VI. REFERENCES

- [1] Uppala, M. S., Belavagi, M. C., & Attigeri, G. (2022). Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison. *IEEE Access*, 10, 3224938. doi:10.1109/ACCESS.2022.3224938
- [2] Hindle, G., Kunc, M., Mortensen, M., Oztekin, A., & Vidgen, R. (2019). Business analytics: Defining the field and identifying a research agenda. *European Journal of Operational Research*, 275(2), 519-532. doi:10.1016/j.ejor.2019.10.001
- [3] Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organizations. *Electronic Journal of Information Systems*, 17(1), 1-22. doi:10.1057/ejis.2014.17
- [4] Bichler, M., Heinzl, A., & van der Aalst, W. M. P. (2016). Business analytics and data science: Once again? *Business & Information Systems Engineering*, 58(3), 185-194. doi:10.1007/s12599-016-0461-1
- [5] Potančok, M., Pour, J., & Wui. (2020). Factors influencing business analytics solutions and views on business problems. *Data*, 6(8), 82. doi:10.3390/data6080082
- [6] Thiesing, F. M., & Vornberger, O. (1997). Sales forecasting using neural networks. *Proceedings of the International Joint Conference on Neural Networks*, 1, 877-882. doi:10.1109/ICNN.1997.614234
- [7] Krishna, A., V. A., Aich, A., & Hegde, C. (2018). Sales-forecasting of retail stores using machine learning techniques. *Proceedings of the 2018 2nd International Conference on Smart Systems and Signal Processing (ICSSSP)*, 1-5. doi:10.1109/ICSSSP.2018.8768765
- [8] Pavlyshenko, B. M. (2018). Machine-learning models for sales time series forecasting. *Data*, 4(1), 15. doi:10.3390/data4010015
- [9] Cheriyan, S., Ibrahim, S., Mohanan, S., & Susan Treesa. (2018). Intelligent sales prediction using machine learning techniques. *Proceedings of the 2018 3rd International Conference on Computing, Communication, and Automation (ICCCA)*, 1-6. doi:10.1109/ICCCA.2018.8659115
- [10] Cadavid, J. P. U., Lamouri, S., & Grabot, B. (2020). Trends in machine learning applied to demand & sales forecasting: A review. *HAL*, hal-01881362.
- [11] Rohaan, D., Topan, E., & Groothuis-Oudshoorn, C. G. M. (2021). Using supervised machine learning for B2B sales forecasting: A case study of spare parts sales forecasting at an after-sales service provider. *Expert Systems with Applications*, 174, 115925. doi:10.1016/j.eswa.2021.115925
- [12] Yu, Q., Wang, K., Strandhagen, J. O., & Wang, Y. (2019). Application of long short-term memory neural network to sales forecasting in retail—A case study. *Advances in Intelligent Systems and Computing*, 846, 2-10. doi:10.1007/978-981-10-5768-7_2

- [13] Martinovic, J., & Damjanovic, V. (2006). The sales forecasting techniques. Proceedings of the 2006 International Conference on Management of Manufacturing Systems, MOMS 2006, 262-267. doi:10.1109/MOMSYS.2006.1656765
- [14] Apriani, D., Aan, M., & Saputra, W. E. (2021). Data visualization using Google Data Studio. International Journal of Computer Science and Information Technology Management, 2(1), 68-75. doi:10.34306/ijcitsm.v2i1.68
- [15] Azis, N., Wahidin, A. J., Cakranegara, P. A., & Muditomo, A. (2018). Visualization of tourist visit time series data using Google Data Studio. Mantik: Jurnal Ilmiah Matematika dan Teknologi Informasi, 13(1), 1-6. Available: <https://iocscience.org/ejournal/index.php/mantik/index>
- [16] Arunraj, N. S., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX model to forecast daily sales in food retail industry. International Journal of Operations Research and Information Systems, 6(4), 1-11. doi:10.4018/IJORIS.2016040101
- [17] Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. Journal of the Japan Institute of Metals, 86(1), 100058. doi:10.1016/j.jjimei.2022.100058
- [18] Singh, B., Kumar, P., Sharma, K. P., & Sharma, N. (2020). Sales forecast for Amazon sales with time series modeling. Proceedings of the 2020 11th International Conference on Computing, Communication and Automation (ICCCA), 1-6. doi:10.1109/ICCCA.2020.9071463
- [19] Vaswani, A., Prasad, P. C., & Padhi, P. K. (2008). Time series analysis: An application of SARIMA model in general trade to forecast sales. Available: <https://jusst.org/time-series-analysis-an-application-of-sarima-model-in-general-trade-to-forecast-sales/>
- [20] Singh, K., & Wajgi, R. (2016). Data analysis and visualization of sales data. Proceedings of the 2016 1st International Conference on Pattern Recognition and Machine Learning (ICPRML), 1-6. doi:10.1109/STARTUP.2016.7583967
- [21] Data Sets for Testing (Till 5 Million Records)—Sales. Accessed: Feb. 5, 2021. [Online]. Available: <https://excelbianalytics.com/wp/downloads-18-sample-csv-files-data-sets-for-testing-sales/>

