International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

SALES DATA ANALYSIS USING PANDAS

[1]J.priyanka,[2]A.Sai lasya,[3]T.Deepak,[4]Sk.Jahangeer Basha

[1]Assistant Professor, [234]Students

Department of CSE(Data Science)

Siddhartha institute of technology & sciences,narapally

jalagapriyanka.cse@siddhartha.co.in, 23TQ1A6705@siddhartha.co.in, 23TQ1A6752@siddhartha.co.in, 23TQ1A6740@siddhartha.co.in

ABSTRACT

Sales data analysis plays an important role in helping organizations understand their business performance and make better decisions. With the increasing amount of data generated by businesses, analyzing sales data efficiently has become essential. This project focuses on Sales Data Analysis using Pandas The main objective of this project is to analyze sales data and extract meaningful insights such as total sales, product performance, region-wise sales distribution, and monthly sales trends. The dataset used in this project contains information such as date, product name, region, quantity sold, and sales revenue. Using the Pandas library, the dataset is cleaned, processed, and analyzed to identify important patterns and trends. Various data analysis techniques such as grouping, aggregation, and filtering are used to evaluate the sales performance of different products and regions. Data visualization techniques using Matplotlib are applied to generate charts such as bar charts, pie charts, line graphs, and scatter plots to present the analysis results in a clear and understandable manner. The results obtained from this analysis help in identifying best-selling products, understanding regional sales distribution, and observing monthly sales growth patterns. This project demonstrates how data analysis tools can be effectively used to transform raw sales data into valuable business insights that support better decision-making.

I INTRODUCTION

Sales data analysis is an important process used by businesses to understand their sales performance and make better decisions. By analyzing sales data, organizations can identify trends, best-selling products, and regional sales patterns. This helps companies improve their strategies and increase profitability. In this project, Sales Data Analysis is performed using Pandas. The dataset contains information such as date, product, region,

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

quantity sold, and sales revenue. Using Pandas, the data is processed and analyzed to extract useful insights. This project mainly focuses on analyzing sales data to understand product performance and sales distribution across different regions and time periods. The data is processed using the Pandas library, which helps in organizing, cleaning, and analyzing the dataset efficiently. Visualization tools such as Matplotlib are used to represent the results in graphical form. These visualizations make it easier to identify important patterns and trends in the sales data, helping businesses make informed decisions.

II LITERATURE SURVEY

Sales data analysis is an important process used by organizations to understand their business performance and market trends. Businesses collect large volumes of sales data from different sources such as retail stores, online platforms, and customer transactions. By analyzing this data, companies can identify best-selling products, customer preferences, and seasonal sales patterns. Sales analysis helps organizations improve marketing strategies, increase revenue, and make better business decisions.

Tools Used for Data Analysis Various tools and technologies are used for performing data analysis. Some of the commonly used tools include Microsoft Excel, Python, R, Power BI, and Tableau. Excel is useful for small-scale data analysis, while Power BI and Tableau are mainly used for creating dashboards and visual reports. However, for large datasets and advanced analysis, programming languages such as Python are widely used because they provide powerful data processing and analytical capabilities.

Role of Python and Pandas Python has become one of the most popular programming languages in the field of data science and data analysis. It is widely used due to its simple syntax, flexibility, and powerful libraries. One of the most important libraries for data analysis is Pandas. Pandas provides data structures such as DataFrame and Series that help users organize, clean, and analyze data efficiently. It supports operations like filtering, grouping, aggregation, and transformation, which makes data analysis easier and faster.

III SYSTEM ANALYSIS

system analysis is the process of examining a system to understand how it works, identify problems, and find ways to improve its efficiency and performance. it involves studying the system's components, processes, and interactions, gathering requirements, and evaluating possible solutions. the main goal is to ensure the system meets user needs effectively and operates smoothly, helping organizations make better decisions and optimize their operations.system analysis is the process of examining a system to understand how it works,

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

identify problems, and find ways to improve its efficiency and performance. it involves studying the system's components, processes, and interactions, gathering requirements, and evaluating possible solutions. the main goal is to ensure the system meets user needs effectively and operates smoothly, helping organizations make better decisions and optimize their operations. it also includes analyzing data flow, identifying bottlenecks, and suggesting improvements. by carefully studying the system, analysts can design better solutions, reduce errors, save time, and increase productivity. overall, system analysis is essential for creating reliable and efficient systems that support organizational goals.

Existing system

The existing system refers to the current way a process or operation is carried out in an organization before any changes or improvements are made. it includes all the methods, procedures, tools, and technologies that are presently being used to perform tasks. studying the existing system helps identify its strengths, weaknesses, and limitations, such as inefficiencies, errors, or delays. understanding the current system is essential for analyzing problems and designing a new or improved system that meets user requirements more effectively and increases overall productivity.

.**Disadvantages of existing system**

- inefficient and time-consuming processes
- higher chances of errors and mistakes
- lack of accuracy and consistency
- difficulty in storing and retrieving data
- limited scalability for growing needs

Proposed system

The proposed system is the new or improved way of performing a process that aims to overcome the limitations of the existing system. it uses modern techniques, tools, or technologies to make operations more efficient, accurate, and reliable. the proposed system focuses on automating tasks, reducing errors, improving data management, and providing faster access to information. it is designed to meet user requirements effectively, enhance productivity, and support better decision-making, ultimately making the overall system more organized and efficient.

Advantages of proposed system

- improves efficiency and saves time

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

- reduces errors and increases accuracy
- better data management and easy retrieval
- faster decision-making with timely information
- enhances productivity and performance

IV METHODOLOGY

Research Design The methodology of the Sales Data Analysis using Pandas project describes the step-by-step process used to analyze the sales dataset and generate useful insights. The project follows a structured approach that includes data collection, preprocessing, analysis, visualization, and result generation. Each phase plays an important role in converting raw data into meaningful information.

Phase 1: Data Collection In this phase, the sales dataset is collected for analysis. The dataset contains information such as product name, sales amount, quantity sold, date of sale, and region. The data is stored in a CSV file format which can be easily accessed using Python. This dataset serves as the input for the entire analysis process.

Phase 2: Data Loading The collected dataset is loaded into the system using the Pandas library in Python. Pandas provides powerful functions to read and manage structured data. The CSV file is imported into a DataFrame, which allows efficient manipulation and analysis of the data.

Phase 3: Data Preprocessing Before performing analysis, the data needs to be cleaned and prepared. In this phase, missing values are handled, duplicate records are removed, and incorrect data types are corrected. Data preprocessing ensures that the dataset is accurate, consistent, and ready for analysis.

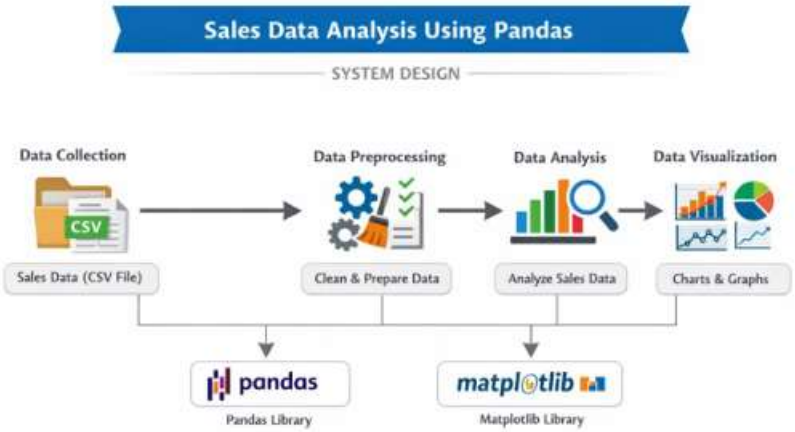
System Architecture\

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

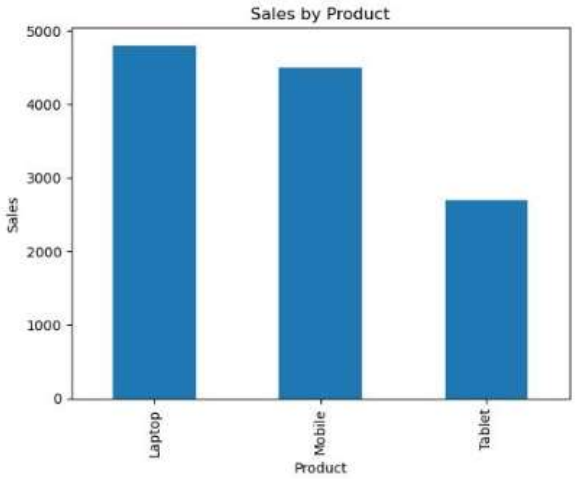
ISSN: 3068-272X ww.ijdim.com Original Research Paper



The system design of the Sales Data Analysis using Pandas project explains how the sales dataset is processed and analyzed. The data is first loaded from a CSV file using the Pandas library in Python. After cleaning and preparing the data, analysis is performed to find useful sales insights. Finally, the results are displayed using charts and graphs created with Matplotlib to better understand sales trends**.**

This module is responsible for collecting the sales dataset used for analysis. The data is stored in a CSV file that contains information such as product name, sales amount, quantity, date of sale, and region. The dataset acts as the primary input for the entire analysis process

V RESULTS & OUTPUT

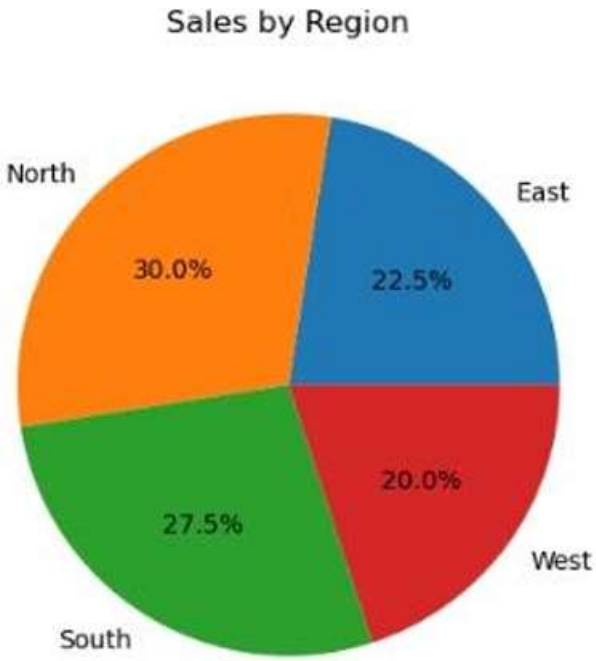


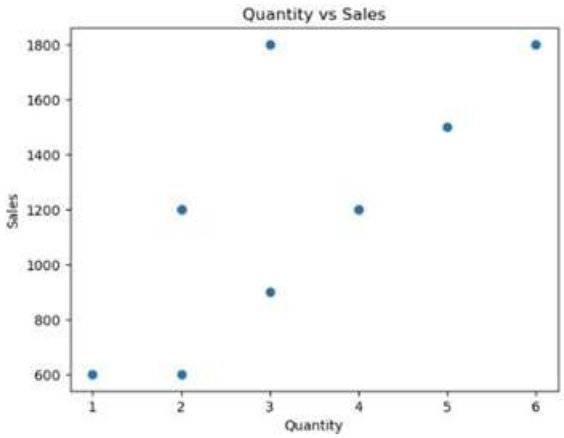
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper





VI CONCLUSION

The Sales Data Analysis using Pandas project effectively transforms raw sales data into actionable insights that help businesses improve decision-making and operational efficiency. By analyzing historical sales data, the project identifies top-performing products, underperforming categories, and regional sales trends. It provides a clear understanding of customer purchasing behavior, seasonal demand fluctuations, and overall business performance. The use of Pandas for data cleaning, aggregation, and manipulation ensures accurate results, while visualizations such as bar charts, pie charts, and scatter plots make

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

complex data easy to interpret. This project also lays the groundwork for future enhancements, including predictive modeling, real-time analytics, interactive dashboards, automated reporting, and integration of customer behavior or multi-source datasets. Ultimately, it reduces manual effort, improves analysis accuracy, supports data-driven strategies, and enables businesses to optimize sales, marketing, and inventory management for sustained growth**.**

REFERENCE

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International

Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.