

Forecasting in a Product Manufacturing Company

Swati Hadke¹, Rashmi Joshi², Saniya Shende³

¹Assistant Professor, Department of Electronics and Telecommunication Engineering, G H Raison College of Engineering and Management, Nagpur, India

²Co-founder, Zappkode Solutions, Nagpur, India

³Student, Department of Electronics and Telecommunication Engineering, G H Raison College of Engineering and Management, Nagpur, India

Abstract - The growing complexity of modern supply chains and the increasing volume of transactional data have made data-driven decision-making a necessity for manufacturing companies. This paper presents a practical account of data cleaning and interactive dashboard development developed as part of a forecasting and trend analysis system for a product manufacturing company. The raw operational data collected from multiple business units was inconsistent, incomplete, and unstructured, making it unsuitable for direct analysis or forecasting. A systematic data cleaning pipeline was designed and implemented using Microsoft Excel and Power BI to address these issues. Once the data was cleaned and transformed into a reliable format, interactive dashboards were created in Power BI to visualize key performance indicators, sales trends, and inventory insights. These dashboards served as the analytical foundation upon which the senior team built advanced forecasting and trend analysis models. The outcomes demonstrated that clean, well-structured data not only improves forecast accuracy but also enables faster and more informed strategic decisions. This paper focuses exclusively on the data preparation and visualization phases, emphasizing the critical role these stages play in any data analytics pipeline.

Key Words: *Data Cleaning, Power BI, Dashboard Development, Trend Analysis, Sales Forecasting, Data Analytics, Inventory Management.*

1. INTRODUCTION

In today's competitive manufacturing landscape, organizations are increasingly relying on data analytics to gain actionable insights and maintain a strategic advantage. Product manufacturing companies, in particular, generate enormous volumes of transactional data daily - spanning sales records, stock levels, customer demand patterns, and seasonal variations.

However, the true value of this data can only be unlocked when it is clean, consistent, and properly structured for analysis.

This paper documents the work conducted as part of a data analytics project focused on building a forecasting and trend analysis platform for a product manufacturing firm operating in a competitive retail market. The company faced pressing challenges in predicting future demand accurately and managing inventory efficiently - problems rooted in unreliable, fragmented, and poorly maintained data.

The specific contributions of this research within the broader project were twofold: (1) designing and executing a comprehensive data cleaning and preprocessing pipeline, and (2) developing interactive dashboards using Power BI to present the cleaned data in an intuitive and decision-friendly format. The forecasting algorithms and trend analysis modules were developed independently by senior team members, building upon the clean data foundation that was prepared during this phase of the project.

This paper is organized as follows: Section 2 presents an analogy to contextualize the role of data cleaning; Section 3 defines the problem statement; Section 4 outlines the study objectives; Section 5 discusses design and methodology; Section 6 covers tools and technologies used; Section 7 addresses future scope; and Section 8 concludes the paper.

2. ANALOGY

To understand the importance of data cleaning in a data analytics pipeline, consider the process of making a high-quality photograph from raw film. Before the image can be printed and displayed, the film must go through careful processing - removing exposure errors, correcting color distortions, and eliminating noise. If the film is not processed correctly, even the best printing equipment will produce a blurry, inaccurate image.

Data cleaning serves a strikingly similar purpose in analytics. Raw data collected from real-world business

operations is rarely perfect - it contains missing values, inconsistencies, duplicate entries, and formatting errors. Just as unprocessed film cannot produce a clear photograph, unclean data cannot produce accurate forecasts or meaningful insights. The cleaning and transformation of data is therefore not merely a preparatory step but the foundational layer upon which reliable analytics and forecasting depend.

Furthermore, an interactive dashboard can be compared to the frame and lighting in which the photograph is displayed. It does not change the content of the image but determines how clearly and effectively that content is communicated to the viewer. A well-designed Power BI dashboard similarly does not alter the underlying data but presents it in a manner that is immediately comprehensible and actionable for business stakeholders.

3. PROBLEM STATEMENT

The product manufacturing company at the center of this project faced a set of interconnected operational challenges that were hindering its ability to scale and compete effectively:

Inaccurate Sales Forecasting: The company struggled to predict future sales with sufficient accuracy due to the unpredictable nature of seasonal demand, rapidly shifting consumer preferences, and dynamic market conditions. These inaccuracies resulted in persistent inventory imbalances - overstock in some product categories and stockouts in others - ultimately leading to revenue loss and reduced customer satisfaction.

Inefficient Inventory Management: Without a reliable forecasting system, inventory decisions were largely reactive rather than proactive. This reactive approach led to overstocking of slow-moving products and understocking of high-demand items, increasing holding costs and causing missed sales opportunities across multiple product lines.

Poor Data Quality: A deeper investigation revealed that the root cause of both forecasting failures and inventory mismanagement was the poor quality of the underlying data. Data collected from multiple sources - including sales systems, warehouse records, and product catalogs - was inconsistent in format, riddled with missing values, and lacked standardization. This made it impossible to apply any meaningful analytical or forecasting method directly.

Absence of Visual Reporting Tools: The company lacked real-time, interactive dashboards to monitor sales performance, track key performance indicators (KPIs), or identify emerging trends. Decision-makers relied on static, manually prepared reports that were time-consuming to create and often outdated by the time they were reviewed.

4. OBJECTIVE OF THE STUDY

The primary objectives of this study are as follows:

Data Cleaning: To identify, diagnose, and resolve data quality issues - including missing values, duplicate records, formatting inconsistencies, and outliers - in the raw datasets provided by the product manufacturing company.

Data Transformation: To restructure and standardize the cleaned data into a consistent, analysis-ready format suitable for downstream processing, forecasting model development, and trend analysis.

Dashboard Development: To design and develop interactive, visually intuitive dashboards in Power BI that present key business metrics - including sales performance, inventory levels, and product-wise trends - in a manner accessible to both technical and non-technical stakeholders.

Support for Forecasting: To provide a clean and reliable data foundation upon which senior analysts can build advanced machine learning-based forecasting models for demand prediction and trend analysis.

Operational Insights: To enable faster, evidence-based decision-making by transforming fragmented raw data into structured visual intelligence.

5. DESIGN AND METHODOLOGY

The project followed a structured and iterative data analytics pipeline, moving progressively from raw data collection to a polished, analytics-ready state. The methodology involved the following phases:

Phase 1 - Data Collection: Raw operational data was collected from the product manufacturing company's internal systems. This included sales transaction records, inventory snapshots, product master data, and historical demand figures. Data was sourced from multiple departments and consolidated into a central repository for analysis.

Phase 2 - Data Exploration and Profiling: Before initiating cleaning, an exploratory analysis was conducted to understand the structure, volume, and

quality of the collected datasets. This involved identifying column types, value distributions, completeness rates, and potential anomalies. Tools used at this stage included Microsoft Excel for initial inspection and Power BI for data profiling.

Phase 3 - Data Cleaning: This was the most critical phase of this study's contribution. The cleaning process addressed the following issues: (a) Missing Values - null and blank entries were handled through imputation or removal depending on the context and column significance; (b) Duplicate Records - exact and near-duplicate rows were detected and eliminated; (c) Inconsistent Formatting - date formats, product codes, category labels, and numerical representations were standardized across all datasets; (d) Outlier Detection - statistical methods were applied to identify and appropriately handle extreme values that could distort analysis results; (e) Data Type Correction - columns were assigned their correct data types to enable accurate calculations and aggregations.



Fig. 1: Data Cleaning Pipeline

Phase 4 - Data Transformation: Once cleaned, the data underwent transformation to improve its analytical utility. New derived columns were created where needed, product hierarchies were established, and multiple datasets were merged or joined based on common keys.

Phase 5 - Dashboard Development: The dashboards were designed with the end user in mind - providing clear visual representations of sales trends, product-wise performance, inventory levels, and regional distribution. Filters, slicers, and drill-through options were incorporated to enable dynamic exploration of data across different time periods and product categories.



Fig. 4: Power BI Dashboard Components

Phase 6 - Validation and Handoff: The cleaned datasets and dashboards were reviewed and validated by senior team members before being used as input for the forecasting and trend analysis modules. This phase ensured that the data met the quality standards required for machine learning model training.

Figure 1 presents the overall design and methodology pipeline adopted in this project.



Fig. 1: Design and Methodology Pipeline

6. TOOLS AND TECHNOLOGY USED

The following tools and technologies were used during the data cleaning and dashboard development phases of the project:

Microsoft Excel: Used for the initial inspection of raw datasets, basic data profiling, and preliminary cleaning tasks such as removing duplicates, correcting cell formats, and handling simple missing value cases.

Power BI: The primary tool for data cleaning at scale and for building interactive dashboards. Power Query within Power BI was used extensively for data transformation, type conversion, column

splitting, and merging of multiple data tables. DAX (Data Analysis Expressions) was used to create calculated columns and measures for KPI computation.

MySQL: Used for structured querying of relational datasets - including filtering records by criteria, aggregating sales data by time period or product category, and joining tables across different business domains.

Jupyter Notebook (Python): Used by the senior team for exploratory data analysis, statistical modeling, and building machine learning-based forecasting models using libraries such as Pandas, NumPy, Matplotlib, and Scikit-learn.

VS Code: Used for web scraping tasks performed by team members to collect competitor and market data from e-commerce platforms. Python-based libraries such as BeautifulSoup and Requests were used for this purpose.



Fig. 3: Tools and Technology Stack

7. FUTURE SCOPE

The current work focused on data cleaning and dashboard creation forms a strong foundation for several future enhancements that can significantly increase the value delivered by this analytics platform:

Automated Data Pipeline: The manual data collection and cleaning process can be replaced by an automated ETL (Extract, Transform, Load) pipeline that continuously ingests, cleans, and refreshes data in real time, eliminating human effort and reducing the risk of errors.

Advanced Forecasting Integration: The cleaned data can be fed into more sophisticated machine learning models - such as LSTM (Long Short-Term Memory) networks or Prophet-based time series models - to generate more accurate and longer-horizon demand forecasts.

Real-Time Dashboard Updates: The Power BI dashboards can be connected to live data sources through APIs or database streams to provide

stakeholders with up-to-the-minute insights rather than relying on periodic manual updates.

Sentiment and Competitor Analysis: The web-scraped data from competitor platforms can be further analyzed using natural language processing (NLP) techniques to extract consumer sentiment, trending product attributes, and pricing strategies - enriching the forecasting model with external market signals.

Mobile and Cloud Deployment: Deploying the dashboards on cloud platforms and making them accessible via mobile devices would enable on-the-go decision-making for managers and supply chain stakeholders.

8. CONCLUSION

This paper documented the data cleaning and interactive dashboard development work carried out as part of a larger forecasting and trend analysis project for a product manufacturing company. The experience highlighted a fundamental truth in data analytics: no analysis, however sophisticated, can produce trustworthy results without a clean and well-structured data foundation.

Through the systematic application of data profiling, cleaning, and transformation techniques using Excel, Power BI, and MySQL, the raw and fragmented data collected from multiple business units was converted into a consistent, analysis-ready dataset. The interactive dashboards built on top of this data empowered stakeholders to visualize sales trends, monitor inventory levels, and track business performance in an intuitive and dynamic manner.

While the forecasting and trend analysis models were developed by senior team members, this preparatory work was indispensable to their success. It demonstrated that data cleaning and visualization are not auxiliary tasks in an analytics project - they are its most critical and time-intensive components. The insights drawn from this research underscore the importance of investing in data quality and visual reporting as preconditions for any meaningful application of artificial intelligence or machine learning in business operations.

REFERENCES

- [1] Montijn van de Ven, P. L. Machado, A. Athanasopoulou, B. Aysolmaz, and O. Turetken, "Key performance indicators for business models,"

Information Systems and e-Business Management,
2023. DOI: <https://doi.org/10.1007/s10257-023-00650-2>

[2] C. Lotfi, S. Srinivasan, M. Ertz, and I. Latrous, "Web Scraping and Its Applications in Data Analysis," SCRS Conference Proceedings on Intelligent Systems, pp. 381-394, January 2021. DOI: 10.52458/978-93-91842-08-6-38

[3] G. Shmueli and O. R. Koppius, "Predictive Analytics in Information Systems Research," MIS Quarterly, vol. 35, no. 3, pp. 553-572, 2011.

[4] T. Dasu and T. Johnson, Exploratory Data Mining and Data Cleaning. Wiley-Interscience, 2003.

[5] Microsoft Corporation, "Power BI Documentation," Microsoft Docs, 2024. [Online]. Available: <https://docs.microsoft.com/en-us/power-bi/>

[6] R. Kimball and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd ed. Wiley, 2013.

[7] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 Competition: 100,000 time series and 61 forecasting methods," International Journal of Forecasting, vol. 36, no. 1, pp. 54-74, 2020.

[8] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly, vol. 36, no. 4, pp. 1165-1188, 2012.