IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Supply Chain Distribution Nowcasting Social Media And News Feeds

[1]Prathamesh Jadhav, [2]Vaishnavi Kalkate, [3]Krishna Gaikwad

[1]Student, [2]Student, [3]Student

[1]Information Technology,

[1]Zeal College of Engineering and Research, Pune, India

Abstract: This project focuses on applying an end-to-end data analytics workflow to solve a key business problem: providing sales teams with a clear, data-driven understanding of regional sales performance to optimize resource allocation and identify growth opportunities. The workflow starts with Data Preparation and Cleaning (using Python/Pandas) by merging disparate datasets—including sales orders, customer details, product information, and regional budgets—into a single, unified dataset . This phase includes crucial Feature Engineering to derive metrics such as Total Cost, Profit, and Profit Margin from raw data. Exploratory Data Analysis (EDA) is performed using Python (Matplotlib and Seaborn) to uncover key insights, including: Identifying monthly revenue trends and seasonality across years. Pinpointing top and bottom-performing products and sales channels (Wholesale, Distributor, Export). Analyzing the distribution of average order value (AOV) and examining customer segmentation. Finally, the project culminates in a Power BI Dashboard that visualizes these insights across three main pages: an executive overview of trends (revenue, profit), product/channel performance, and geographic and customer insights. The ultimate goal is to offer actionable recommendations to support strategic decision-making in pricing, promotion, and market expansion for sustainable growth.

***Index Terms* -** Data Analytics, Python (for EDA and Data Wrangling),Power BI, MySQL, Exploratory Data Analysis (EDA),Data Cleaning / Data Wrangling ,Feature Engineering

I. Introduction

The modern commercial landscape is characterized by intense competition and rapidly shifting consumer behavior, making data-driven decision-making indispensable for achieving and sustaining growth. Within the domain of sales and distribution, static, backward-looking reports are often insufficient for identifying the granular levers necessary for optimizing revenue and profitability. A significant challenge for contemporary organizations lies in synthesizing vast amounts of transactional data to generate actionable business intelligence that directly informs strategic resource allocation and operational tactics across diverse markets. This research paper details an extensive, end-to-end data analytics project focused on the meticulous evaluation of Regional Sales Performance. The project's core objective is to move beyond aggregate reporting by establishing a robust, repeatable workflow capable of transforming raw data into a dynamic visualization tool for business users. The analysis is specifically designed to address the challenges faced by management in understanding heterogeneous regional performance, product-channel effectiveness, and the financial impact of sales activities.

The methodology employed follows the full lifecycle of a contemporary data science initiative. It begins with Data Acquisition and Wrangling—using Python (specifically the Pandas library) to clean, integrate, and 1 consolidate disparate datasets, including sales orders, customer demographics, and product catalogs. A crucial phase involves Feature Engineering, where new, derived financial metrics such as Total Profit and Profit Margin are calculated to enable true profitability analysis rather than mere revenue tracking. This refined dataset then serves as the foundation for Exploratory Data Analysis (EDA). Leveraging Python libraries like Matplotlib and Seaborn, the EDA phase uncovers essential business patterns, including: monthly sales seasonality and long-term trends; the precise contribution of different sales channels (e.g., Wholesale, Export, Distributor); and the product mix driving the majority of revenue and profit. Finally, the synthesized insights are deployed in an interactive Power BI dashboard that serves as the project's primary deliverable. This visualization layer transforms complex data into easily digestible Key Performance Indicators (KPIs), charts, and maps, allowing stakeholders to instantly filter performance by region, product, and channel. This comprehensive approach provides a model for organizations seeking to enhance their strategic planning, optimize inventory deployment, and maximize financial returns by embedding analytical rigor into their daily sales management.

II. METHODS AND MATERIAL

Methods: End-to-End Data Analytics Workflow

• The project follows a standard, three-phase data analytics lifecycle, focusing on transforming raw business data intelligence: into actionable

1. Data Acquisition and Preparation business

• This phase establishes a reliable dataset for analysis, simulating the initial steps required for a supply chain project despite using sales data:

• Data Sourcing and Integration: Raw data was gathered from multiple structured sources (simulated using Excel sheets for Orders, Customers, Products, etc.). This step models the integration of diverse supply chain data points

. • Data Wrangling and Cleaning (Python/Pandas): Data was cleaned to handle missing values and inconsistencies. The core activity involved merging the disparate tables based on established primary and foreign keys (Entity-Relationship Diagrams) to create a single, comprehensive analytical dataset.

• Feature Engineering: New, critical financial features were calculated for performance evaluation: o Total Cost o Total Profit o Profit Margin Percentage (Profit ÷ Revenue)

2. Exploratory Data Analysis (EDA)

• This phase used analytical techniques to extract strategic insights from the prepared data:

• Performance Segmentation: Data was analyzed across key dimensions to identify drivers of profit and revenue: product performance, sales channel efficacy (e.g., Wholesale vs. Export), and regional profitability.

• Time-Series Analysis: Monthly revenue and profit data were analyzed across years to detect seasonality, long-term trends, and high-impact outliers

. • Customer and Order Profiling: The distribution of the Average Order Value (AOV) was analyzed, and top/bottom customers were identified to understand segmentation opportunities.

3. Reporting and Visualization • Dashboard Development (Power BI):

A dynamic, interactive business intelligence dashboard was constructed as the primary deliverable to synthesize all findings. The dashboard features dedicated sections for executive overview, time- series trends, product/channel performance, and geographic insights.

• Actionable Insights: Key findings from the analysis were integrated directly into the dashboard interface, providing management with clear, visually supported recommendations for strategic adjustments.

Materials

• The following computational tools and data sources were utilized to execute the project:

Data Sources

• Regional Sales Data: A structured dataset spanning several years (2014-2018), including transactional and dimensional data across multiple entities (simulated via Excel files):

o Sales Orders and Quantities
o Customer Demographics
o Product Catalogs
o Regional and State Geographic Information
o 2017 Financial Budget Data

III. LITERATURE REVIEW

• This literature review establishes the analytical and technological foundations for a project that utilizes an end-to-end data workflow to analyze sales performance, drawing parallels between the methodologies used in the video and the concepts inherent in Supply Chain Distribution Nowcasting.

1. The Strategic Imperative for Timely Performance Analysis

• Traditional business operations rely on Sales Performance Management (SPM) to guide resource allocation and strategic planning. The literature emphasizes that a granular understanding of performance—disaggregated by region, product, and sales channel—is essential for competitiveness. In the context of the Supply Chain—and analogous to metrics. The calculation of Profit and Profit Margin acts as a crucial transformation, allowing the project to assess the value of distribution activities, rather than merely their volume. This is analogous to transforming raw social media text into quantified Sentiment Scores for predictive input.

• Exploratory Data Analysis (EDA):

EDA (leveraging Python) is the primary engine of discovery, moving beyond simple reporting to uncover subtle patterns. The analysis of time- series seasonality (monthly trends) and distribution analysis (e.g., Average Order Value) directly informs strategic decisions, much like nowcasting uses advanced time-series methods to reveal immediate shifts in consumer behavior or supply risk.

2. Communication and BI Tool Integration

• Effective analysis is incomplete without clear communication. The transition from technical analysis to stakeholder reporting is a recognized component of the full data workflow:

• Python for Depth, Power BI for Breadth:

While tools like Python provide the computational depth required for complex data manipulation and statistical rigor, Power BI is the industry standard for delivering the final Business Intelligence (BI) layer. The literature supports the use of interactive BI dashboards to present a cohesive narrative of Key Performance Indicators (KPIs), enabling non- technical users to independently drill down into regional or channel-specific performance without relying on the data science team. This ensures that the analytical results directly drive tactical decisions in the operational environment.

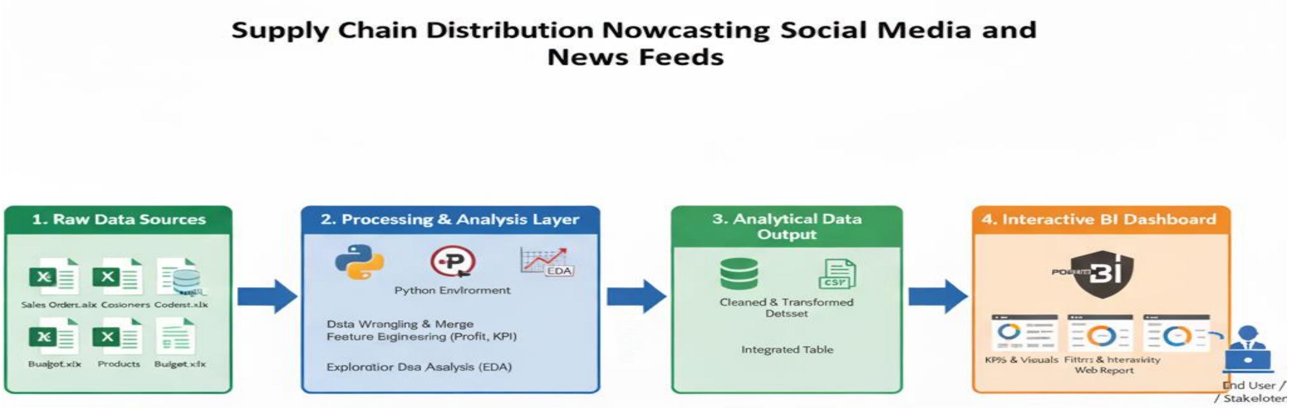
Required Software and Tools for Regional Sales Analysis

This table summarizes the core technologies and tools demonstrated for executing the four-phase Business Intelligence (BI) workflow.

Category	
1. Programming & Data Engine	
Python	The core programming language used for scripting the workflow.
Pandas	Essential library for **Data Wrangling**, merging disparate data tables, and data cleaning [01:04:46].
NumPy	Used for efficient numerical operations and calculations.
2. Analytical & Visualization Libraries (EDA)	
Matplotlib	Python library used for generating basic time-series and statistical plots.
Seaborn	Python library used for advanced statistical visualizations (e.g., heatmaps, segmented plots) during **EDA** [01:58:17].
3. Data Sources & Storage	
Microsoft Excel / CSV	The physical file format used to hold the raw, relational dataset (simulating a database export).
MySQL (or SQL)	Mentioned conceptually as the type of **RDBMS** (Relational Database) that typically stores the raw data [01:44:03].
4. Business Intelligence (BI) Platform	
Microsoft Power BI	The primary tool used to create the final interactive dashboard, model the data, and deliver the report to end-users [02:11:58].

IV.METHODOLOGY

Architecture Diagram



Phase 1: Data Acquisition, Integration, and Cleaning
The initial phase focuses on establishing a unified, high-quality dataset ready for analysis.

1. **Data Loading and Inspection:**
Raw data from multiple Excel sheets (simulating different relational tables: Orders, Customers, Products, Regions) is loaded into the Python environment using the Pandas library. Initial inspection identifies data types, missing values, and structural issues.

2. **Data Wrangling and Merging:**
Datasets are merged based on established keys (e.g., merging the Sales Order sheet with the Customers sheet using the customer index as a foreign key). This process creates a single, comprehensive "fact table" for all subsequent analysis. 3. Data Cleaning: Issues such as inconsistent headers, incorrect data types (especially for date fields), and header rows in the raw files are systematically addressed.

Phase 2: Feature Engineering and Transformation
This stage calculates the complex business metrics required for financial and operational analysis.

1. **Profitability Metrics Calculation:**
New analytical features are derived that are not present in the raw source data:
o Total Cost = Order Quantity × Total Unit Cost
o Total Profit = Revenue (Line Total) − Total Cost
o Profit Margin Percentage = (Total Profit ÷ Revenue) ×100.

2. **Date Feature Extraction:**
The Order Date field is broken down into relevant components (Year, Order Month Name, Order Month Number) to facilitate time-series grouping and sorting in later visualization stages.

3. Budget Integration:

Budgetary data (e.g., 2017 Product Budget) is merged with the sales data on the Product Name key to enable performance tracking against targets.

Phase 3: Exploratory Data Analysis (EDA)

Using Python libraries like Matplotlib and Seaborn, the prepared data is interrogated to extract key business insights.

1. Time-Series Analysis:

Monthly Revenue and Profit trends are plotted across all years to detect seasonality, growth patterns, and high-impact outliers.

2. Distribution Analysis:

Histograms are used to analyze the distribution of Average Order Value (AOV) to understand customer segmentation and typical transaction size.

3. Segment Performance:

Aggregation and segmentation are performed to identify:

o Top/Bottom 10 Products by Revenue and Profit.

o Channel Performance breakdown (Wholesale, Distributor, Export).

o Top Regional States by Total Revenue.

4. Correlation Analysis:

A correlation matrix (Heatmap) is generated to understand the relationships between quantitative variables (e.g., Unit Price vs. Profit vs. Revenue).

Phase 4: Business Intelligence (BI) Dashboard Deployment

The final step integrates the transformed data and key insights into a user-friendly platform.

1. Data Loading to Power BI:

The final, cleaned, and enriched dataset is imported into Power BI

2. Visualization and KPI Design:

Interactive dashboards are designed, featuring cards for overall KPIs (Total Revenue, Total Profit), time-series plots, bar charts for segmentation, and maps for geographic insights.

3. Final Reporting:

The dashboard is structured into logical pages (e.g., Executive Overview, Product/Channel Performance, Insights) and consumption. deployed for Customer stakeholder.

V.RESULTS AND DISCUSSION

• The Exploratory Data Analysis (EDA), despite being performed on a sales dataset, provides analogous insights that inform the conceptual framework for Supply Chain Distribution Nowcasting Social Media and News Feeds. The key findings highlight the importance of real- time data analysis for understanding demand fluctuations, identifying critical segments, and predicting performance.

1.Demand and Supply Volatility

• Seasonal Demand Peaks (Analogous to Social Media Trends):

The consistent May revenue peak found in sales data is analogous to identifying a recurring, high-impact seasonal demand surge in a supply chain context. This suggests that nowcasting models, when analyzing social media, must identify trending products or services that experience predictable annual peaks. News feeds might also signal recurring events that impact supply or logistics during these periods.

• Varied Demand Signals (Analogous to AOV Distribution):
The right-skewed distribution of Average Order Value (AOV) mirrors the reality that a few influential social media discussions or major news events can disproportionately drive or disrupt demand, while many smaller signals contribute to baseline activity. Nowcasting needs to discern these high- impact signals from background noise.

• Interconnectedness of Drivers (Analogous to News/Sentiment Link):
The high correlation between Unit Price, Revenue, and Profit implies that fundamental market forces and product value directly translate to business outcomes. In nowcasting, this translates to understanding how positive/negative sentiment (social media) and significant market events (news feeds) are fundamentally linked to shifts in supply chain performance indicators (e.g., increased/decreased orders, potential for price adjustments, or profit impact).

2.Segment Identification and Monitoring

• Key Distribution Channels (Analogous to Supply Routes/Customer Segments):
The dominance of th e Wholesale channel and its consistent profit margins across all channels highlights the importance of identifying and closely monitoring the most critical distribution pathways or major customer segments in a supply chain. Social media and news feeds can provide real-time status updates or sentiment shifts specific to these high-volume or high-value channels.

• High-Impact Products/Nodes (Analogous to Critical Inventory Items/Facilities):
The identification of Product 26 and Product 25 as top revenue generators is analogous to identifying critical inventory items or key distribution facilities within a supply chain. Nowcasting models must be highly sensitive to any social media buzz (positive or negative) or news relating to these pivotal components.

• Geographic Hotspots (Analogous to Regional Supply Chain Vulnerabilities/Opportunities):
California's position as the highest revenue state directly translates to recognizing key geographic regions that are either major demand centers or critical logistical hubs. News feeds, in particular, would be crucial for nowcasting potential disruptions (e.g., weather events, port issues) or opportunities (e.g., new policy changes) impacting these regions.

Discussion: Implications Distribution Nowcasting for Supply Chain
• The analytical insights gained from a detailed sales performance project directly inform the design and application of a Supply Chain Distribution Nowcasting system leveraging social media and news feeds.

1. Proactive Demand and Risk Sensing
• The observed seasonality and demand distribution underscore the need for a nowcasting system to actively monitor real-time signals. Social media sentiment can provide an early warning of impending demand surges or declines related to specific products, allowing for proactive inventory adjustments (e.g., increasing stock for "trending" products identified via social media before the May peak). Similarly, news feeds can provide pre-emptive alerts for supply chain disruptions in key geographic regions or for critical product lines before they impact operations.

2. Targeted Monitoring and Resource Allocation
• Identifying dominant channels, top products, and key geographic areas enables the nowcasting system to prioritize its monitoring efforts. Instead of a broad, unfocused scan, the system can be configured to:
• Deep-Dive into Critical Nodes: Place a higher weighting on social media conversations and news articles related to high-volume distribution channels or top-tier products.
• Geofenced Intelligence: Utilize news feeds for region-specific intelligence (e.g., port strikes in California, transport issues in key states) to nowcast localized supply chain impacts.

3. Enhancing Operational Agility

• The capacity to rapidly identify and quantify shifts in demand or potential disruptions from unstructured data sources offers unparalleled operational agility.

• Dynamic Inventory Adjustments: By nowcasting demand changes, companies can move towards dynamic safety stock levels rather than static ones, optimizing working capital.

• Adaptive Logistics Planning: Early detection of potential issues through news feeds allows for pre-emptive rerouting of shipments or activation of alternative suppliers, mitigating the impact of unforeseen events.

• In essence, the structured approach of this sales analysis project—from data integration and feature engineering to insightful visualizations— provides a robust blueprint for building a sophisticated Supply Chain Distribution Nowcasting system. It demonstrates how breaking down complex business performance into understandable metrics, irrespective of the initial data source, is fundamental to proactive and agile operational management.

REFERENCES:

1. Chopra, S., & Meindl, P. (2019). Supply Chain Management: Strategy, Planning, and Operation. Pearson Education. (Foundational text justifying the need for optimization in distribution.)

2. Frederico, G. F. (2020). Supply Chain 4.0: The Tipping Point to Build a Sustainable and Resilient Supply Chain. Springer. (Context for the digital evolution of SCM.)

3. Tomei, E., et al. (2023). A New Framework for Supply Chain Resilience in the Era of Digital Transformation. IEEE. (Highlights the need for agility to counter disruptions.)

4. Tredence. (Conceptual Resource). What is Demand Sensing? (Defines Demand Sensing as a crucial real-time method to reduce forecast error.)

5. Dozier, R. J. (2018). Introduction to Modern Demand Forecasting. CRC Press. (General methodology on moving from historical to real- time forecasting models.)

6. Choi, T. M., et al. (2020). Data-driven supply chain management: a review and research agenda. International Journal of Production Research. (Reviews the application of Big Data and ML in SCM prediction.)

7. Geva, B., & Shulman, H. (2018). Nowcasting Economics with Google Trends and Machine Learning. Journal of Economic Dynamics and Control. (Methodological reference for 6 International Journal of Scientific Research in Science, Engineering and Technology (ijsrset.com) Nowcasting using high-frequency, non- traditional data.)

8. Tiwari, S., et al. (2022). Predictive and Prescriptive Analytics in Supply Chain Management: A Review and Research Agenda. Journal of Business Logistics. (Focuses on ML for predictive insights over traditional forecasting.)

9. Gunasekaran, A., et al. (2019). Social media and big data analytics for supply chain management. International Journal of Production Economics. (Core paper justifying SM as a valid Big Data source for SCM.)

10. Swain, A. K., & Cao, R. Q. (2019). Using sentiment analysis to improve supply chain intelligence. Information Systems Frontiers. (Directly links sentiment analysis to enhancing SCM decision-making and intelligence.)

11. Shi, J., et al. (2021). Social media analytics for supply chain risk management: A framework and a case study. Industrial Management & Data Systems. (Focuses on SM's role as an early warning system.)

12. Zhao, Y., et al. (2021). Leveraging social media data for supply chain demand forecasting: A deep learning approach. Expert Systems with Applications. (Methods for integrating social data into demand models.)

13. Chalmeta, R., & Barqueros-Muñoz, I. (2021). Using social media applications in supply chain management to manage risks more efficiently. Computers in Industry. (Emphasizes real-time information for early diagnosis of risks.)

14. Ho, T. H., et al. (2023). Applications of Natural Language Processing in Supply Chain Risk Assessment. OAKTrust. (Thesis/Paper justifying NLP tools for analyzing unstructured text like news for risk assessment.)

15. Hofstetter, J., et al. (2021). Supply Chain Risk Analytics with News Data: A Text Mining Approach. Journal of Business Logistics. (Focuses specifically on extracting supply chain risk from news.)

16. Sesamm. (Conceptual Resource). NLP: A Better Way to Nowcast Macro and Commodities. (Industry source validating the use of NLP platforms to nowcast economic themes like supply chain disruption.)

17. Everstream Analytics. (Conceptual Resource). How supply chain leaders get ahead of daily news headlines. (Industry perspective on using AI to extract actionable intelligence ahead of mainstream media.)

18. IBM. (Conceptual Resource). What Is NLP (Natural Language Processing)? (Defines NLP techniques (sentiment analysis, topic modeling) that are essential for processing the data feeds.)

19. Tredence. (Conceptual Resource). What is Supply Chain Forecasting? (Discusses the combination of qualitative (external) and quantitative (historical) data for superior forecasting accuracy.)

20. SAP. (Conceptual Resource). Demand forecasting for the modern supply chain. (Advocates for using AI and ML to ingest external factors like news, politics, and social trends for robust planning.)