

**A Project Report on**

**Inventory Forecasting and Analytics for Tata motors company – A case study for one Dealership**

**Submitted in partial fulfilment for award of degree of**

**Master of Science**

**In Business Analytics**

**Submitted by**   
**Thanish Shekar**R22MSB07

**Under the Guidance of**

**Dr. Jay Bharatheesh Simha**

Chief mentor, RACE, REVA University & CTO ABIBA Systems

REVA Academy for Corporate Excellence

**REVA University**

Rukmini Knowledge Park, Kattigenahalli,

Yelahanka, Bangalore – 560064

**August 2025**

# 

# Candidate’s Declaration

I, **Thanish Shekar**, hereby declare that I have completed the project work towards **Master of Science in Business Analytics** at REVA University on the topic entitled **Inventory Forecasting and Analytics for Tata motors company – A case study for one Dealership** under the supervision of **Dr. Jay Bharatheesh Simha, Chief mentor, RACE, REVA University & CTO ABIBA Systems.** This report embodies the original work done by me in partial fulfilment of the requirements for the award of the degree for the academic year 2025.

Place: Bengaluru Thanish Shekar

Date:

A close-up of a signature

AI-generated content may be incorrect.

A black and orange text

Description automatically generated

# Acknowledgment of Project Ownership and Usage Rights

I, Thanish Shekar, a student enrolled in the Master of Science in Business Analytics Program and 2023-2025 Year at RACE, hereby acknowledge that any project, including but not limited to software, hardware, research, or other intellectual property created by me during my academic tenure at RACE, is the property of RACE, REVA University.

I understand and agree that RACE has the exclusive rights to use, reproduce, modify, or distribute the aforementioned projects for academic, research, and further development purposes. This includes the right to monetize, commercialize, or otherwise exploit the projects as deemed fit by RACE.

I acknowledge that I have no objection to RACE, REVA University, using, reproducing, or further developing the projects for the benefit of the institution and its academic community. I further affirm that any commercial or research activities related to the projects conducted by RACE shall not require additional consent or approval from me.

This acknowledgment is made willingly and without any reservations. I am grateful for the education and opportunities provided by RACE, and I recognize the importance of contributing to the academic and research goals of the institution.

Place: Bengaluru Thanish Shekar

Date:

A close-up of a signature

AI-generated content may be incorrect.



# Certificate

This is to certify that the project work entitled **Inventory Forecasting and Analytics for Tata motors company – A case study for one Dealership** has been carried out by **Thanish Shekar** with **R22MSB07**, who is a bonafide student of REVA University, is submitting the **first-year** project report in fulfilment for the award of **Master of Science in Business Analytics** during the academic year **2025**. The Project report has been tested for plagiarism and has passed the plagiarism test with a similarity score of less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said degree.

Signature of the Guide Signature of the Director

Dr. Jay Bharatheesh Simha Shinu Abhi

Guide Director

External Viva

Names of the Examiners

1. <Name> <Designation> <Signature>
2. <Name> <Designation> <Signature>

Place: Bengaluru

Date:



# Acknowledgement

I am profoundly grateful to Dr. Shinu Abhi, Director of Corporate Training, for the guidance and assistance provided during my course and throughout my project.

I owe a great debt of gratitude to our mentor, Dr. Jay Bharatheesh Simha, whose invaluable advice as my project guide significantly aided my understanding and execution of this project. and all other mentors for their indispensable guidance and suggestions in learning Artificial Intelligence. Their invaluable project-related advice was instrumental in its success.

I would like to thank my manager and my office colleagues who were supportive with providing me with the business requirements for this project.

I’m grateful for the platform provided by the founder and Hon’ble Chancellor, **Dr. P Shayma Raju**, Hon’ble Vice Chancellor, **Dr. Sanjay Chitnis**, and Registrar, **Dr. K S Narayanaswamy,**

My heartfelt thanks to Dr. Rashmi Agarwal and to all members of the RACE program office, who were consistently helpful in all aspects related to the program.

Finally, this acknowledgment would be incomplete without expressing my deep appreciation for my family and friends. Their unwavering support, encouragement, and understanding have been pivotal in the successful completion of this project.

Place: Bengaluru

Date:



# Similarity Index Report

This is to certify that this project report titled **Inventory Forecasting and Analytics for Tata motors company – A case study for one Dealership** was scanned for similarity detection. The process and outcome are given below. The plagiarism report is attached in the appendix.

Software Used:

Date of Report Generation:

Similarity Index in %:

Total word count:

Name of the Guide: Dr. Jay Bharatheesh Simha

Place: Bengaluru Thanish Shekar

Date:

A close-up of a signature

AI-generated content may be incorrect.

Verified by:

Signature

Dr. Shinu Abhi,

Director, Corporate Training

# List of Abbreviations

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Abbreviation** | **Long Form** |
| 1 | MAE | Mean Absolute Error |
| 2 | RMSE | Root Mean squared Error |
| 3 | AI | Artificial Intelligence |
| 4 | ML | Machine Learning |
| 5 | SARIMA | Seasonal Autoregressive Integrated Moving Average |
| 6 | ADF | Azure Data Factory |
| 7 | ADLS | Azure Data Lake Storage |
| 8 | CRISP-DM | Cross-Industry Process for Data Mining |
| 9 | KPI | Key Performance Indicator |
| 10 | SQL | Structured Query Language |
| 11 | EDA | Exploratory Data Analysis |
| 12 | SF | Snowflake |
| 13 | SME | Subject Matter Expert |
| 14 | ETL | Extract, Transform, Load |

# List of Figures

|  |  |  |
| --- | --- | --- |
| **No.** | **Name** | **Page No.** |
| Fig.5.1 | Project Methodology (CRISP-DM) | 18 |
| Fig.5.2 | Architecture | 20 |
| Fig.7.1 | Data Dictionary | 25 |
| Fig.8.1 | Entity-Relationship Diagram for the Core Data Model | 29 |
| Fig.8.2 | Data Preparation Roadmap for Inventory Segmentation | 30 |
| Fig.8.3 | ABC Classification | 31 |
| Fig.8.4 | XYZ Classification | 32 |
| Fig.8.5 | Distribution of Coefficient of Variation | 33 |
| Fig.8.6 | XYZ Classification Share | 33 |
| Fig.8.7 | Total Quantity Sold by XYZ Class | 34 |
| Fig.8.8 | Monthly Demand Trend by XYZ Class | 34 |
| Fig.8.9 | FSN Classification by Part Count | 35 |
| Fig.8.10 | Distribution of Recency Days | 36 |
| Fig.9.1 | XGBoost Model | 38 |
| Fig.9.2 | SARIMA Model | 38 |
| Fig.9.3 | Holt-Winters Model | 39 |

# List of Tables

|  |  |  |
| --- | --- | --- |
| **No.** | **Name** | **Page No.** |
| Table 7.1 | Dealer Information | 26 |
| Table 7.2 | Time Based Information | 27 |
| Table 7.3 | Job Card & Service Information | 27 |
| Table 7.4 | Finance & Invoice Information | 28 |
| Table 7.5 | Spare parts Information | 28 |
| Table 10.1 | Model Performance Table | 41 |
| Table 12.1 | Summary Table | 46 |

# Abstract

In today's dynamic market, effective inventory management presents a significant strategic challenge for businesses seeking to maintain a competitive edge. Traditional inventory control systems often struggle to adapt to fluctuating consumer demand, leading to critical inefficiencies such as increased carrying costs from overstocking and lost revenue from stockout events.

At Tata Motors dealerships, the current planning process primarily depends on the expertise of Subject Matter Experts (SMEs) within each product line. With growing interest among these experts to embrace analytics, this research seeks to supplement business intuition with robust time series forecasting models. Specifically, the study introduces three baseline models to segment inventory items: AFX (high-value, fast-moving), AFY (somewhat reliable), AFZ, and BFX (less predictable) categories. This enables the organization to prioritize critical parts, reduce unnecessary holding costs, and optimize operational expenditure.

The objective of this project is to design and propose an inventory forecasting solution powered by modern analytical and machine learning techniques. By employing the CRISP-DM methodology, the project builds a data pipeline with automated ETL processes, statistical forecasting using Python (including pandas for data preparation), and model development with SARIMA, Holt-Winters, and XGBoost algorithms. Data is centralized in a Snowflake warehouse via Azure Cloud to ensure scalability and reliability, supporting both historical reporting and predictive insights through Stream lit dashboards. This end-to-end approach not only streamlines data flows but also delivers actionable, real-time recommendations for inventory planning, focusing on optimizing stock levels over a rolling three-month horizon.

Among the techniques examined, the SARIMA model provided the highest reliability with an R-squared value of 0.9566, followed by Holt-Winters at 0.8873, and XGBoost at 0.8362. These results validate the effectiveness of the data-driven methodology in delivering precise demand forecasts.

Keywords: SARIMA, Holt-Winters, XGBoost, Inventory Forecasting, Historical, Predictive Dashboard, ETL, Snowflake, R-squared value, Segmentation

# Table of Contents

[Candidate’s Declaration 2](#_Toc165713943)

[Acknowledgment of Project Ownership and Usage Rights 3](#_Toc165713944)

[Certificate 4](#_Toc165713945)

[Acknowledgement 5](#_Toc165713946)

[Similarity Index Report 6](#_Toc165713947)

[List of Abbreviations 7](#_Toc165713948)

[List of Figures 7](#_Toc165713949)

[List of Tables 8](#_Toc165713950)

[Abstract 9](#_Toc165713951)

[Table of Contents 10](#_Toc165713952)

[Chapter 1: Introduction 11](#_Toc165713953)

[Chapter 2: Literature Review 14](#_Toc165713954)

[Chapter 3: Problem Statement 17](#_Toc165713955)

[Chapter 4: Objectives of the Study 18](#_Toc165713956)

[Chapter 5: Project Methodology 19](#_Toc165713957)

[Chapter 6: Business Understanding 23](#_Toc165713958)

[Chapter 7: Data Understanding 25](#_Toc165713959)

[Chapter 8: Data Preparation 29](#_Toc165713960)

[Chapter 9: Modeling 38](#_Toc165713961)

[Chapter 10: Model Evaluation 41](#_Toc165713962)

[Chapter 11: Analysis and Results 43](#_Toc165713963)

[Chapter 12: Deployment 45](#_Toc165713964)

[Chapter 13: Conclusions and Future Scope 47](#_Toc165713965)

[Bibliography 48](#_Toc165713966)

[Appendix 50](#_Toc165713967)

[Plagiarism Report 51](#_Toc165713968)

[Github Link 53](#_Toc165713971)

# Chapter 1: Introduction

Tata Motors, a leading player in the automotive industry, operates a dense supply chain network that includes a wide array of suppliers, dealerships, and service centers. Within this ecosystem, efficient inventory management is essential to meet fluctuating customer demands and maintain uninterrupted service. In the dynamic landscape of today’s automotive market, efficient inventory management remains a persistent challenge for organizations striving to outperform their competitors. The availability of critical spare parts and components directly influences customer satisfaction, while inventory-related inefficiencies can increase operational costs and tie up valuable capital. As the automotive landscape continues to evolve, the industry peers face mounting pressure to streamline its inventory practices to keep pace with new market expectations and technological advancements.

Traditional, largely manual inventory management methods struggle to respond to this complexity, often resulting in stockouts, excess inventory, long lead times, and inefficient order fulfillment. Most of the current planning process primarily depends on the expertise of Subject Matter Experts (SMEs) within each product line. These SMEs generate forecasts based on historical knowledge and there is a growing need to harness data analytics and predictive modeling for more accurate demand forecasting and smarter inventory decisions. By leveraging advanced analytical tools, The Study aims to minimize carrying costs, improve service quality, and ensure operational efficiency across its dealership, establishing a model for modern inventory management in the sector.

* 1. **Background**

The automotive industry is undergoing rapid transformation, with increasing demands for efficiency and agility in managing complex supply chains. For a leading manufacturer like Tata Motors, which coordinates a vast network of suppliers, dealerships, and service centres, effective inventory management is a mission-critical aspect. The ability to ensure the right spare parts and components are available at the right time not only supports seamless service delivery but also has a direct impact on customer satisfaction and operational costs.

However, the aftermarket segment faces acute challenges due to highly variable and often unpredictable demand patterns. These fluctuations arise from numerous factors, including changing vehicle usage, seasonal trends, and macroeconomic shifts. Traditional inventory management approaches-primarily manual processes and basic rule-of-thumb forecasting-are ill-equipped to handle this degree of variability. The consequences manifest as stockouts, excess inventory, prolonged lead times, and inefficient order fulfilment, all of which erode profitability and undermine customer loyalty. Recognizing these shortcomings, this project is positioned to leverage advanced data analytics and predictive modelling to build an inventory forecasting and end to end analytics system. By enhancing demand forecasting, optimizing inventory levels, and minimizing carrying costs, Tata Motors aims to set a new benchmark for operational excellence across its dealership, ultimately resulting in stronger business outcomes and greater customer value.

* 1. **Need for the Study**

The primary aim of this project is to empower Tata Motors with a data-driven inventory management framework. By leveraging historical transactional data and deploying robust statistical models, the solution aspires to enhance forecast accuracy that help dealerships keep just the right amount of inventory. This approach will enable the dealership to optimize operational costs, improve service efficiency, and ensure higher customer satisfaction.

* 1. **Scope of the Study**

This study focuses on delivering a practical and scalable solution for Tata Motors. Its scope is centred on four key areas:

1. Inventory levels more closely aligned with actual demand; by classifying parts based on their value, sales frequency, and demand patterns, we can create highly focused and accurate forecasts for each specific category.
2. Analytics platform: A unified analytics platform, that handles the entire process-from automatically extracting data to preparing it for analysis, running the forecast models, and finally, displaying the insights on an easy-to-understand dashboard. This gives the dealership a single, reliable environment for making data-driven decisions.
3. Cross Comparison: deploying this solution in parallel with existing inventory management practices, the project enables a rigorous, side-by-side evaluation of business outcomes. We can rigorously measure improvements in forecast accuracy, operational efficiency, and cost savings without disrupting the dealerships current operations.
4. The insights and measurable impact from this initial implementation will serve as a blueprint. This will provide a clear and tested template for scaling the analytics framework across the wider Tata Motors network in the future.
   1. **Contribution to Organization**

This project helps as a strategic advantage to one of the Tata Motors Dealership by embedding data-driven practices into the dealership’s inventory management process. By combining part classification techniques (such as ABC, XYZ, and FSN) with time series forecasting, the solution provides accurate insights for stocking the right parts at the right time.

Furthermore, the implementation of an end-to-end analytics platform and data warehouse creates a unified and scalable foundation for continuous improvement. The pilot initiative not only enables thorough, parallel evaluation of new and existing processes at one dealership but also establishes a best-practices blueprint for expansion across other locations within the organization. Ultimately, this empowers Tata Motors to enhance service levels, improve customer satisfaction, and realize measurable cost savings throughout its dealer network.

# Chapter 2: Literature Review

To ensure a business runs smoothly and profitably, two things are essential: accurately predicting what customers will want and having the right amount of stock on hand to meet that demand. For this study, we looked back over the last decade of research to see how the companies are achieving this. Our review focused on modern, data-driven strategies, including segmentation analysis, which groups similar products together; cloud-based analytics platforms that can handle huge amounts of data; and advanced time series analysis and predictive modelling to forecast future trends. By examining these successful methods, we built a strong foundation for our own approach, ensuring it was grounded in proven, effective techniques.

The research by [1] provides a critical framework for this project, highlighting a persistent source of inefficiency in modern supply chains: the traditional, isolated evolution of demand forecasting and inventory control. Our study directly addresses this identified gap by implementing a tightly integrated system where the outputs from predictive models like SARIMA are used to directly inform inventory policy and optimize stock levels for distinct part segments. Consequently, this work represents a practical application of the "Level 3 integration" that Goltsos et al. describe as both rare and vital. In alignment with their recommendations, the deployed models are evaluated not just on their predictive accuracy, but on their direct and measurable impact on inventory performance.

A study by [2] presents a hybrid approach for demand forecasting in the Indian automotive aftermarket sector by combining machine learning and time-series models. The paper applies a bag of 23 algorithms to over 29,000 SKUs and selects the best model using RMSE-based comparison, followed by an ensemble of top-performing models to improve forecast accuracy. The study emphasizes granular forecasting using real-world sales data and demonstrates the practical utility of the proposed model for supply chain risk mitigation and decision-making.

[3] address the challenge of forecasting pharmaceutical inventory for South Australia’s public health sector using time series models. The study evaluates linear regression, exponential smoothing, and Holt-Winters methods, concluding that the Holt-Winters Seasonal Additive Damped model achieved the best accuracy with an RMSE of 408. A Power BI dashboard was also developed to visualize forecast results, offering a practical tool for inventory decision-making across hospitals and healthcare services.

A working paper by [4] from the MIT Centre for Transportation & Logistics discusses the integration of forecasting and inventory management in multi-echelon supply chains. The study models a distribution network with base-stock policies and highlights how different forecast update frequencies and safety stock allocations impact inventory efficiency. By modelling various levels of forecast accuracy and update frequency, the paper shows how integrated planning improves service levels and reduces inventory holding costs.

The paper by [5] addresses the challenge of predicting spare parts demand in the automotive sector to minimize excess inventory and avoid stockouts. The authors use a hybrid forecasting methodology combining historical trend analysis with ARIMA modelling, focusing on monthly demand data from a tier-1 auto component supplier. Their model was trained and validated using real-world sales data, and results showed that the ARIMA (1,1,1) model produced the lowest MAPE (8.3%), significantly improving forecast reliability. This study demonstrates that simple time series models, when calibrated well, can optimize inventory levels and reduce procurement costs.

The study by [6] focuses on the problem of inaccurate inventory demand forecasting in the retail sector, which often leads to either overstocking or stockouts. The authors apply a quantitative forecasting methodology using historical sales data, implementing exponential smoothing and multiple regression models to predict future demand. The modeling approach includes testing different smoothing constants and regression variables to capture seasonal trends and external demand drivers. Results indicate that the Holt-Winters exponential smoothing model provided the most accurate forecasts with a MAPE improvement of over 12% compared to the retailer’s existing manual method.

This paper [7] outlines the architectural principles of combining data lakes and data warehouses. This approach integrates ETL pipelines, business intelligence (BI), and machine learning (ML) into a unified platform. By curating data in layered stages—from raw ingestion through curated analytics tables and business-centric products—the Lakehouse facilitates trust, governance, ACID transaction support, and high performance for diverse workloads, including real-time streaming and advanced ML. These architectural advantages directly support the needs of modern inventory analytics projects. This Paper is essential to explain the rationale behind my choice of a modern cloud architecture.

The Medallion Architecture is an industry best practice for structuring data in a Lakehouse or modern data warehouse. It advocates for a multi-layered approach- Bronze (raw data), Silver (validated), and Gold (business-aggregated)-to ensure data reliability and quality for downstream analytics and ML models. Adopting this architecture ensures your analytics platform is robust, auditable, and scalable, which is critical for an automated forecasting system.

This foundational academic paper [9] on Snowflake's architecture, a cloud-native, multi-cluster shared-data architecture that separates compute and storage for independent scaling. This paper provides the technical justification for choosing Snowflake as the data warehouse for my project, especially for handling large-scale historical data and serving both the historical and predictive dashboards efficiently.

This paper [10] presents the architecture of a complete, cloud-based demand forecasting platform designed for e-commerce. It details the end-to-end workflow, from data ingestion and processing in a data lake to distributed model training and serving predictions via APIs. This source directly supports my project's design by providing a case study of an integrated analytics platform that automates the entire forecasting pipeline in a high-volume retail context.

Through these literature reviews and blogs on Demand forecasting, time series analysis and various approaches and techniques, A key takeaway from these reviews is that true value is unlocked not just by improving model accuracy with methods like ML or Time series Forecast, but by building an integrated end-to-end system, performing Exploratory Data Analysis (EDA), and understanding how to effectively split data. The system should leverage a scalable cloud platform, structured with a quality-driven framework like the Medallion Architecture. This study focuses on forecasting inventory for parts in these high-value, fast-moving, and varying demand reliability segments and building a integrated end-to-end platform with dashboard.

# Chapter 3: Problem Statement

*ADISHAKTI CARS PVT LTD, a key dealership in the Tata Motors network, faces a critical inventory management challenge driven by its reliance on traditional, intuition-based forecasting. This manual approach fails to accurately predict the highly volatile and intermittent demand for automotive spare parts, leading to a costly and disruptive cycle of overstocking and stockouts. The inability to anticipate shifts in customer behavior, seasonality, and market trends puts the dealership in a constant reactive state.*

The Impact:

This forecasting inefficiency creates significant downstream consequences that ripple through the entire operation, impacting finances, operations, and customer relationships.

1. Financial Strain: Excess inventory leads to high carrying costs and the risk of parts obsolescence, directly hurting profitability.
2. Operational Friction: Stockouts cause service disruptions and extend repair times, creating bottlenecks for the service team and frustrating customers.
3. Eroding Customer Loyalty: Failure to have the right parts on hand results in long wait times and broken promises, severely damaging customer satisfaction and trust in the Tata Motors brand.

This project addresses these issues by designing and deploying a scalable, analytics-driven framework to replace traditional way with data-driven decision-making. By leveraging machine learning models for demand forecasting and creating intuitive dashboards for visualization, the framework will transform raw sales data into actionable inventory strategies.

The Case Study not only solves Adishakti's immediate problems but also provides a blueprint for optimizing inventory, increasing profitability, and enhancing customer loyalty across the other dealership network.

# Chapter 4: Objectives of the Study

The primary objective of this project is to design and implement an end-to-end analytics framework with a data-driven forecasting model that can help the dealership to maintain optimal inventory levels and improve overall customer service.

To achieve this, the project will deliver five key objectives:

1. **Automate Data Movement:** Design and implement an automated data pipeline using an ETL tool, leveraging a cloud-based architecture with Snowflake for robust data warehousing.
2. **Historical Data Visualization:** Develop an interactive dashboard using Stream lit to visualize historical sales data.
3. **Inventory Segmentation:** Formulate Inventory parts into four segment groups based on parts value, consumption, and demand reliability. The three segments will be categorized as follows,
   1. AFX Segment: High-value, fast-moving parts with reliable demand variability.
   2. AFY Segment: High-value, fast-moving parts with less reliable demand variability.
   3. AFZ Segment: High-value, fast-moving parts with unpredictable demand patterns.
   4. BFX Segment: Medium-value, fast-moving parts with reliable demand variability.
4. **Forecasting Models for Segments:** Build a demand forecasting model for each segment created in the third objective to accurately predict demand over the next three months.
5. **Predictive Insights and Dashboard:** Develop a Predictive dashboard to provide actionable inventory recommendations to business stakeholders based on forecasted demand.

By implementing these objectives, the project aims to address the dealership’s inventory management challenges- reducing order delays, lowering costs, and strengthening customer loyalty through data-driven decision-making.

# Chapter 5: Project Methodology

This project centres on developing a robust, data-driven mathematical forecasting model for predicting parts inventory at the dealership. To ensure a systematic and repeatable approach, the project adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, coupled with a modern cloud architecture for its technical implementation.

The CRISP-DM methodology provides a clear blueprint for managing the project lifecycle from business understanding through data exploration, preparation, modelling, evaluation, and deployment.

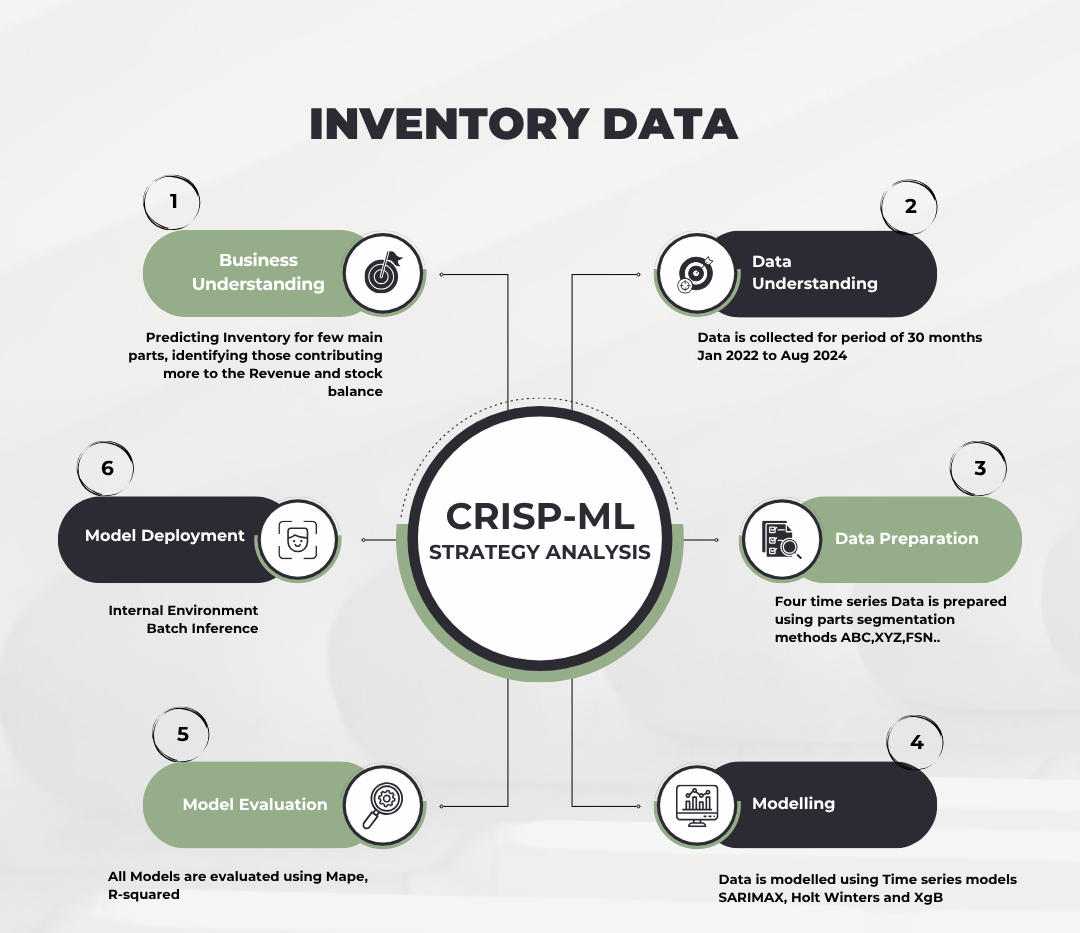


Fig. 5.1: Project Methodology

As illustrated in Fig. 5.1, the CRISP-DM framework ensures a structured approach to manage and execute the projects effectively and efficiently.

* 1. **Methodological Framework: CRISP-DM**

**Business Understanding:** Understand the business goals and objectives for inventory forecasting. Currently the Dealership is facing stockouts, customer dissatisfaction leading to decrease in revenue. This is exercised by SMEs of each product line based on domain expertise. Identify stakeholders and gather requirements. The scope of this project is to segment parts based on parts cost, movement, and demand reliability and streamlining the forecasting workflow with Azure Cloud and Snowflake. The optimal forecasting approach is selected based on MAPE error metrics, ensuring the dealership benefits from the most accurate and actionable forecast results.

**Data Understanding:** Acquire the relevant transactional data needed for the analysis and modeling. Understand the data, its key variables, and its structure. Perform initial data exploration to identify patterns, trends, and anomalies.

**Data Preparation:** Prepare the data for by checking for missing values, normalising and applying feature engineering. Perform EDA to understand the distribution, relationships, seasonality and key characteristics of the transaction data.

**Modeling:** Develop an inventory forecast model. Prepare the dataset and consider the past two months values for testing. Develop and compare multiple models (XGBOOST, SARIMAX, Holt Winters, Linear Regression) for the 4 segments. Identify the top 3 best performing model for forecasting the next 3 months of Part units.

**Evaluation:** The model's predictive performance is evaluated using MAPE and R-squared Metric and the units forecasted are cross-checked with the previous quarterly sales.

**Deployment:** Deploy the best-performing model into the decision-making process to provide actionable insights, where the model is accessed for batch inferencing on monthly basis.

* 1. **Technical Architecture and Implementation**

To execute the CRISP-DM framework, a scalable technical architecture has been designed on the Azure cloud platform, leveraging Snowflake for the data warehouse. This architecture, illustrated in Figure 5.2, provides an end-to-end pipeline that automates data flow, processing, analysis, and deployment. It is designed around the Medallion Architecture principle to progressively refine data through bronze, silver, and gold layers, ensuring quality and governance.

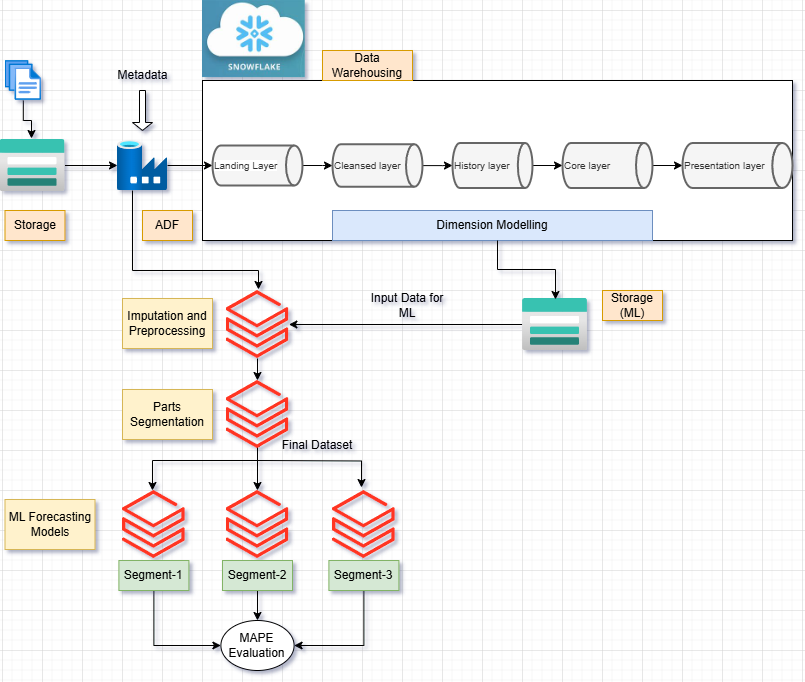


Fig. 5.2: Architecture

The architecture is composed of four distinct stages that align with the project's objectives:

1. **Data Ingestion:** This initial phase focuses on acquiring raw data from its sources. The process is orchestrated using Azure Data Factory (ADF) and all the raw source data are stored in Azure Data Lake Storage.
2. **Data Processing:** The Data is then ingested into Snowflake data warehouse. Here, data is cleansed, validated, de-duplicated, and transformed into a structured, format.
3. **Enterprise layer & Modeling:** The Gold layer in Snowflake contains data that is aggregated and transformed for specific business purposes. The core data science work, including the multi-criteria inventory segmentation and the training of forecasting models (SARIMA, Holt-Winters, XGBoost), is performed within the Azure Databricks environment. Models are trained, evaluated using MAPE and R-squared metrics.
4. **Consumption:** The final stage involves delivering insights to end-users. A batch inferencing process uses the validated model to generate demand forecasts for the next three months. These predictions are written back to the Snowflake. The insights are then made accessible through two distinct, user-friendly dashboards built with Streamlit.
   1. Historical Sales Dashboard: allowing stakeholders to interactively explore and analyze past sales data and inventory performance.
   2. Predictive Dashboard: This dashboard visualizes the forecasted demand and provides actionable recommendations, enabling data-driven inventory planning and decision-making.

**Tools, Technologies and ML Models**

1. Data Acquisition and Understanding: Python, Pandas, SQL, pySpark
2. Cloud: Azure (ADLS, Databricks, ADF), Snowflake (warehouse)
3. Modeling: Scikit-learn, pmdarima

# Chapter 6: Business Understanding

Tata Motors, a leading global automotive manufacturer, operates through a vast and complex supply chain network that includes numerous suppliers, service centers, and dealerships. This project focuses on a key dealership within this network, ADISHAKTI CARS PVT LTD. As the primary interface for sales and after-sales service, dealerships like Adishakti operate in the challenging automotive aftermarket, which is characterized by highly variable and often unpredictable demand for a wide array of spare parts.

The efficiency of a dealership's inventory management is paramount, directly impacting its profitability, operational workflow, and its ability to uphold the Tata Motors brand promise of quality and reliability.

The primary operational challenge stems from a disconnect between the centralized forecasting process and the dealership's local market dynamics. Currently, Adishakti operates based on an inventory plan generated by Subject Matter Experts (SMEs) at the Head Regional Office, who utilize advanced forecasting models and business knowledge. While this top-down approach provides a strategic baseline for many dealerships, it often lacks the granularity to cope with the high volatility of local aftermarket demand experienced specifically at Adishakti.

This gap between the central plan and on-the-ground reality creates significant inefficiencies, leading to a cascade of negative business impacts on finances, operations, and customer relationships.

This project will therefore introduce a data-driven inventory management framework at the dealership level, using local data to optimize forecasting and classification, thereby enhancing efficiency, reducing costs, and improving customer satisfaction.

The inventory includes a vast portfolio of small, medium, and heavy parts, with numerous variants and versions for each vehicle model, making procurement and management inherently complex. To ensure a focused and impactful analysis, the scope of this project is strictly defined:

1. It will concentrate exclusively on aftermarket parts.
2. The data used will be limited to that of the “Adishakti Cars Pvt limited- Hebbal” service center.

The aim of this project is to develop a supplementary, data-driven forecasting layer that empowers the dealership with localized insights. The goal is to accurately predict the demand for inventory parts over a rolling three-month period, with a specific focus on four high-priority segments (AFX, AFY, AFZ, and BFX) that are classified by their value, consumption rate, and demand predictability. By enhancing forecast accuracy at this granular level, the project seeks to avoid stockouts, reduce carrying costs, and improve overall operational efficiency. This system is designed not to replace the expertise of SMEs, but to complement their strategic input with reliable, objective, and data-driven insights to foster more effective inventory decisions.

# Chapter 7: Data Understanding

This chapter outlines the initial phase of the CRISP-DM methodology, focusing on the acquisition and exploration of the dataset used for this project. A thorough understanding of the data's characteristics, structure, and quality is fundamental to developing an effective and reliable forecasting model.

* 1. **Data Source and Acquisition**

The dataset for this project consists of transactional sales data sourced directly from the ADISHAKTI CARS PVT LTD dealership. This historical data captures the complete journey of a customer's vehicle service, from the creation of a job card to the final invoicing of parts used. The dataset encompasses a continuous period of 32 months, from January 2022 to August 2024.

* 1. **Dataset Characteristics**

The dataset provides a granular view of the dealership's day-to-day service and sales operations. Its primary characteristics are summarized below:

1. Total Records: The complete dataset contains 237,566 individual transaction records.
2. Total Columns: The dataset is comprised of 28 distinct columns, covering various dimensions of the sales and service process.
3. Unique Parts: The data includes transactions for 3,785 unique part numbers.
4. Granularity: Each record in the dataset represents a single spare part line item on a customer invoice, linked to a specific service job card.
   1. **Data Dictionary and Key Data Types**

To effectively prepare the data for modeling, it is essential to understand the type and purpose of each column. The Fig 7.1 below describes the key features that are most relevant to this forecasting project.

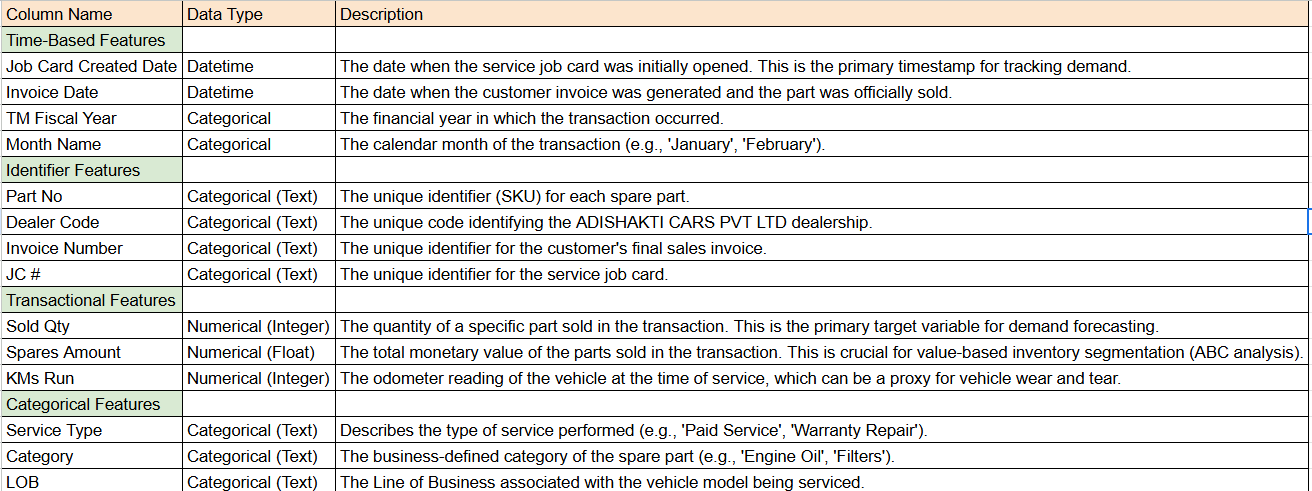


Fig. 7.1: Data Dictionary

The dataset’s 28 columns can be logically grouped into five key areas, each providing a different dimension of information crucial for forecasting and analysis.

* + 1. **Dealer Information**

This group of attributes provides the organizational context, identifying the specific dealership and its position within the broader Tata Motors network.

Table 7.1: Dealer Information

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Partner Type** | **Dealer** | **Dealer Code** | **Division** | **Region** |
| Dealer | ADISHAKTI CARS PVT LTD-3007720 | 3007720 | 3007720-Sv&Pa-Bengaluru-AtiCar | South1 |
| Dealer | ADISHAKTI CARS PVT LTD-3007720 | 3007720 | 3007720-Sv&Pa-Bengaluru-AtiCar | South1 |
| Dealer | ADISHAKTI CARS PVT LTD-3007720 | 3007720 | 3007720-Sv&Pa-Bengaluru-AtiCar | South1 |
| Dealer | ADISHAKTI CARS PVT LTD-3007720 | 3007720 | 3007720-Sv&Pa-Bengaluru-AtiCar | South1 |

* + 1. **Time-Based Attributes**

These columns form the backbone of the time-series analysis, allowing for the mapping of sales data over time.

Table 7.2: Time Based Information

|  |  |  |
| --- | --- | --- |
| **TM Fiscal Year** | **Month Name** | **Month** |
| 2022-23 | April | 2022 / 04 |
| 2022-23 | April | 2022 / 04 |
| 2022-23 | April | 2022 / 04 |
| 2022-23 | April | 2022 / 04 |

* + 1. **Job Card & Service Details**

This set of features provides context about the specific service event that triggered the demand for spare parts. Understanding the nature of the service can help explain variations in demand for certain parts. For example, parts demand might differ significantly between a routine paid service and an accidental repair.

Table 7.3: Job Card & Service Information

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **JC #** | **Job Card Created Date** | **JC Open Days** | **Job Card Closed Date** | **Service Type** | **KMs Run** |
| JC-AtiCar-BP1-2122-006244 | 28-09-2021 11:38 | 209 | 25-04-2022 19:32 | Accident | 39501 |
| JC-AtiCar-BP1-2122-006244 | 28-09-2021 11:38 | 209 | 25-04-2022 19:32 | Accident | 39501 |
| JC-AtiCar-BP1-2122-006244 | 28-09-2021 11:38 | 209 | 25-04-2022 19:32 | Accident | 39501 |
| JC-AtiCar-BP1-2122-006244 | 28-09-2021 11:38 | 209 | 25-04-2022 19:32 | Accident | 39501 |

* + 1. **Financial and Invoice Details**

These columns are crucial for understanding the business impact of each transaction, they provide the monetary value of sales and confirm the status of transactions.

Table 7.4: Finance & Invoice Information

|  |  |  |
| --- | --- | --- |
| **TM Spares Amount** | **Invoice Status** | **Invoice Number** |
| 179.66102 | New | IATIEN2223001268 |
| 10432.2034 | New | IATIEN2223001268 |
| 256.25 | New | IATIEN2223001268 |
| 9475 | New | IATIEN2223001268 |
| 561.71875 | New | IATIEN2223001268 |

* + 1. **Inventory & Spare Parts Data**

This is the most critical category for this project, as it contains the specific details about the items being sold. The Part No acts as the unique stock-keeping unit (SKU), while the Sold Qty is the primary target variable that the forecasting models will be trained to predict.

Table 7.5: Spare parts Information

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part No** | **Part Desc** | **Category** | **TM Part Indicator** | **Sold Qty** | **PPL** | **PL** |
| 543888506378B8 | BRIGHT\_CHROM-FRONT BUMPER WITH HUMANITY | Spare Part | Y | 1 | Nano | Nano CX |
| 8855AXLT0001 | 1010R FAST ACTIVATOR(1L) | Nano Spares | Y | 30 | Nano | Nano CX |
| 8855AXLT0002 | 256S FAST ACTIVATOR(1L) | Nano Spares | Y | 300 | Nano | Nano CX |
| 8855AXLT0032 | 679S ULTRA PRODDUCTIVE CLEAR(5L) | Nano Spares | Y | 1000 | Nano | Nano CX |

# Chapter 8: Data Preparation

* 1. **Data Dictionary and Key Data Types**

The source dataset, containing 237,566 records for 3,785 unique parts captured over 32 months from Jan 2022 to Aug 2024, includes a diverse mix of liquid parts and spare parts of varying sizes. To manage this complexity and ensure data quality, a structured, multi-layered data architecture is implemented within Snowflake.

1. **Landing Layer (SDL):** The ETL process begins by ingesting the raw, daily sales data into a landing zone within Snowflake. This layer, often referred to as the Bronze layer, holds the data in its original, unaltered format. This serves as a source of record and allows for reprocessing if needed without having to re-extract data from the source systems.
2. **Cleansed Layer:** From the landing layer, the data is processed and moved to a cleansed layer. Using a series of SQL scripts, several key cleaning operations are performed:
   1. **Missing Value Handling:** Identifying and imputing or removing records with critical missing information.
   2. **Duplicate Removal:** Ensuring each transaction is represented only once.
   3. **Data Type Conversion:** Correcting data types for columns like dates and numerical values to ensure consistency (e.g., converting text-based dates to a proper DATETIME format).
3. **History Layer (SCD Type 2):** The cleansed data is then loaded into a history layer. This layer is designed using a Slowly Changing Dimension (SCD) Type 2 methodology. This approach is crucial as it preserves the complete history of all transactions and changes over time, rather than overwriting old data. By maintaining historical records, we ensure that any analysis or model training reflects the true state of the business at any given point in the past.
4. **Core Layer (Normalized Data Model):** In the final transformation step, the data moves to the Core layer. Here, the large, flat transactional table is normalized by breaking it down into smaller, logically distinct tables. This process reduces data redundancy, improves data integrity, and optimizes query performance. This normalized structure forms the foundation of our analytical data model.
   1. **Data Model: Entity-Relationship (ER) Diagram**

The structure of the Core layer is represented by the Entity-Relationship (ER) diagram shown in Figure 8.1. This diagram illustrates how the different business entities relate to one another, forming a star schema. The central fact table, invoice\_line\_details, captures the core transactional event—the sale of a part. This table is linked via foreign keys to multiple dimension tables, each of which describes a different aspect of the transaction.

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 8.1: Entity-Relationship Diagram for the Core Data Model

This normalized model provides a clear and efficient structure for analysis. For example, to analyze sales for a specific part, one can join the invoice\_line\_details table with the Parts dimension table. Similarly, to understand sales trends related to a particular service type, a join can be made with the Service dimension table.

* 1. **Inventory Segmentation Strategy**

A critical step in data preparation is the strategic segmentation of inventory to apply tailored forecasting strategies rather than a one-size-fits-all approach. Following the project objectives, the data is categorized using a combination of established inventory classification techniques:

ABC analysis (based on item value), FSN analysis (based on consumption rate), and XYZ analysis (based on demand variability). This multi-criteria approach ensures that parts are grouped by their financial impact, sales velocity, and forecastability.

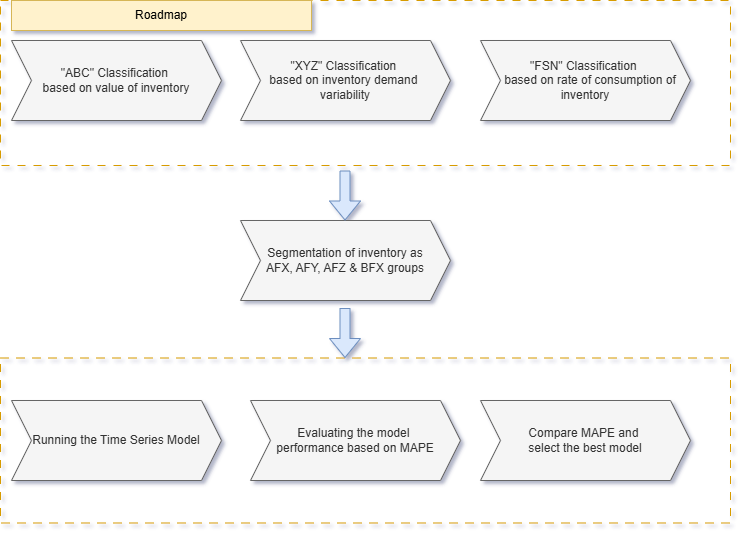


Fig. 8.2: Data Preparation Roadmap for Inventory Segmentation

The sequential roadmap for this process is illustrated in Figure 8.2. The ultimate goal is to combine these individual classifications to create four distinct, high-priority segments—AFX, AFY, AFZ, and BFX - each with unique characteristics. These carefully curated segments form the final datasets that will be fed into the time-series modeling phase of the project.

1. **ABC Classification (Value-Based)**

The initial step in the segmentation process is to classify all 3,785 unique parts based on their value of inventory cost to the dealership. The methodology involves calculating the total sales value for each unique part number over the 32-month analysis period, using the Spares Amount from the transactional data.

Based on this cumulative value, parts are categorized into three distinct classes:

* 1. **Class A:** The most valuable products. These are the parts that collectively account for the top 80% of the dealership's total spares revenue.
  2. **Class B:** The moderately valuable products. This group represents the next 10% of the total spares revenue (from 80% to 90% of the cumulative total).
  3. **Class C:** The least valuable products. This is the long tail of parts that collectively account for the final 10% of the total spares revenue.

This classification confirms that a relatively small number of Class A parts drive the majority of the business's revenue. To ensure the project delivers the highest business impact, the forecasting efforts will focus primarily on Class A and Class B items.

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 8.3: ABC Classification

It is clearly seen that the 80-20 rule of Pareto principle is valid with this data. From the Fig. 8.3, 13% of parts belonging to A cat contributes 80% of the total sales -15Cr, parts belonging to B cat contributes 10% to the total sales- 2Cr and parts belonging to C cat contributes 75% to the total sales – 2Cr

1. **XYZ Analysis on Demand Variability**

XYZ analysis is performed on the demand variability and classified based on the Coefficient of Variation.

The classification is based on the following rules:

1. If CoV is <= 0.5 classified as X-class
2. If CoV is between 0.5 and 1.0 as Y-class
3. And rest of the parts as Z-Class.

A statistical summary of this classification is presented in Fig 8.4, which shows the distribution of CoV values and the number of parts within each class.

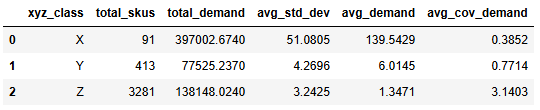


Fig. 8.4: XYZ Classification

The overall distribution of the Coefficient of Variation across all parts is shown in Figure 8.5. The histogram is right-skewed, indicating that a majority of parts exhibit low-to-moderate demand variability, while a smaller number of parts have very high, unpredictable demand patterns.

A graph with a line graph

AI-generated content may be incorrect.

Fig. 8.5: Distribution of Coefficient of Variation

The resulting share of parts in each category is visualized in Figure 8.6.

A pie chart with a number of percentages

AI-generated content may be incorrect.

Fig. 8.6: XYZ Classification Share

While Z-class parts are the most numerous and unpredictable, it is essential to understand the sales volume contributed by each class. As shown in Figure 8.7, the X-Class parts, despite being fewer, account for the highest total quantity sold. This highlights their importance to the business and the need for highly accurate forecasting to maintain service levels.

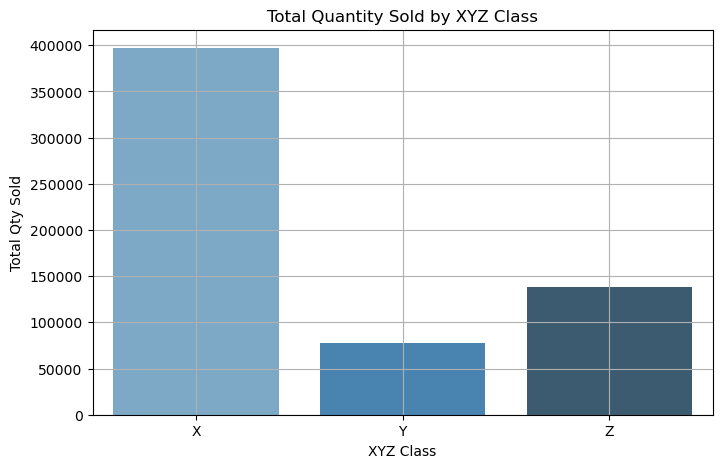


Fig. 8.7: Total Quantity Sold by XYZ Class

Finally, Figure 8.8 provides a visual confirmation of the classification by plotting the monthly sales trends for each class. The X-Class trend is relatively stable, the Y-Class shows moderate fluctuations, and the Z-Class exhibits the highly erratic and spiky demand pattern characteristic of unpredictable items.

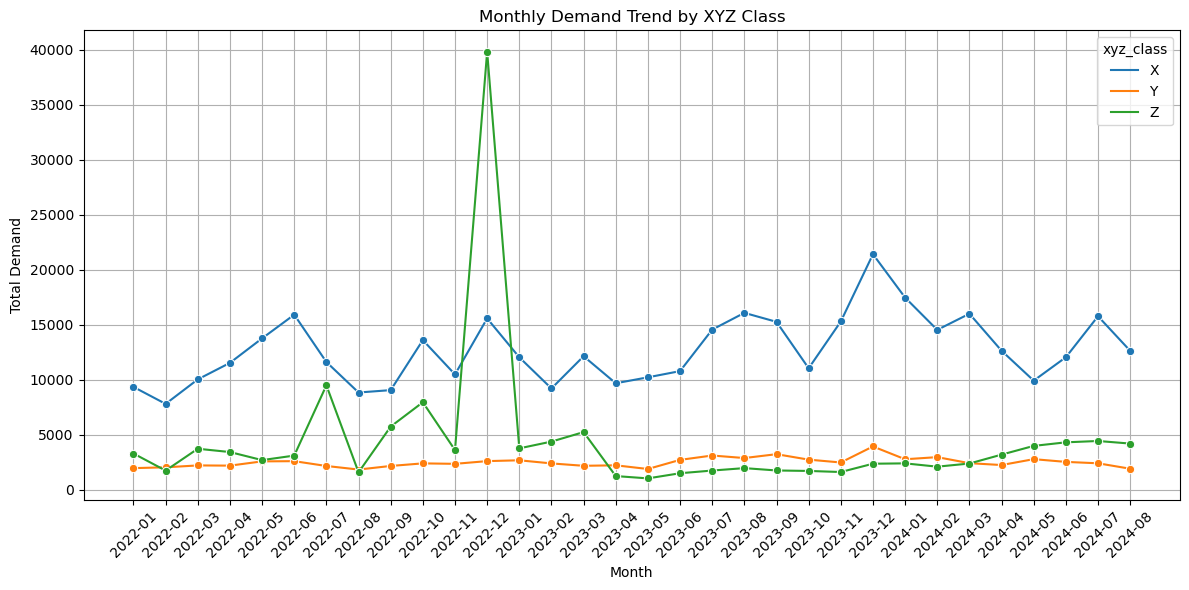


Fig. 8.8: Monthly Demand Trend by XYZ Class

1. **FSN Classification**

FSN (Fast-moving, Slow-moving, Non-moving) classification is a critical inventory management technique used to categorize spare parts based on their consumption pattern over time.

The recency score for each unique part is determined by calculating the number of days between the most recent invoice date in the entire dataset (which serves as a fixed reference date) and the last invoice date recorded for that specific part. A lower recency score indicates a more frequently consumed item.

Based on this recency score, each part is classified according to the following criteria:

* 1. Fast-moving (F): Parts sold within the last 30 days.
  2. Slow-moving (S): Parts sold between 31 and 180 days ago.
  3. Non-moving (N): Parts that have not been sold for more than 180 days.

The result of FSN analysis is captured in Fig. 8.9 FSN Classification

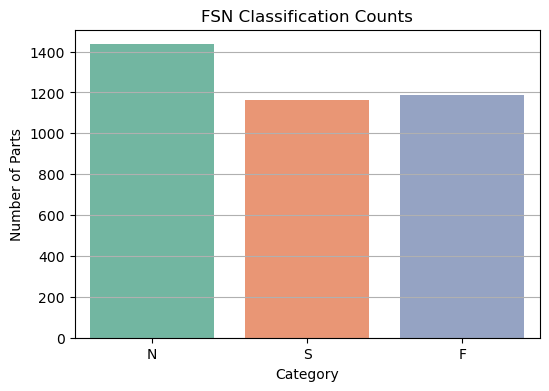


Fig. 8.9: FSN Classification by Part Count

The distribution of these recency scores across all parts is visualized in the boxplot below (Figure 8.9), showing the spread of consumption patterns.

A graph of a number of days

AI-generated content may be incorrect.

Fig. 8.10: Distribution of Recency Days

* 1. **Final Segmentation and Dataset Creation**

The final step in the data preparation phase is to merge the individual ABC, FSN, and XYZ classifications. Combining these three dimensions reveals a total of 20 distinct inventory segments within the dataset.

While all 20 segments provide useful information, to align with the project objectives of focusing on high-impact items, this study narrows its scope to four specific priority segments: AFX, AFY, AFZ, and BFX.

These segments represent the most critical parts for the dealership, combining high-to-medium value, high consumption, and varying levels of demand predictability.

# Chapter 9: Modeling

In this project, the modeling process focuses on forecasting the short-term (3-month) demand for those selected segments i.e. AFX, AFY, AFZ & BFX segments. Each model output is compared using MAE & R-squared metric.

The Total count of parts for which the forecasting will be applied is 347 parts out of the 3785 parts.

* 1. **Data Preparation and Train-Test Split**
* For modeling, the data was reshaped into long format, where each row represented the demand for a particular part in a specific month.
* Feature engineering included lagged variables (up to 3 months), rolling averages, month, year, and quarter indicators.
* For evaluation, a standard train-test split was applied: the last three months' data served as the test set for out-of-sample validation.

The below listed time series models are fitted on the training data set. Mean absolute error & R-squared metric is used to evaluate the model performance.

1. XGBoost
2. SARIMA
3. Holt-Winters
   1. **Model Training & Forecasting**
4. **XGBoost Model**

* Feature set included lagged demand values, rolling means, and calendar features.
* The model used an ensemble of decision trees (1,000 estimators, learning rate=0.01) and early stopping on validation error.
* Performance metrics: RMSE, MAE, and R-squared were calculated from actual vs. predicted values on the test set.
* Forecasting for the next 3 months used the last available lags and rolling means for each part.

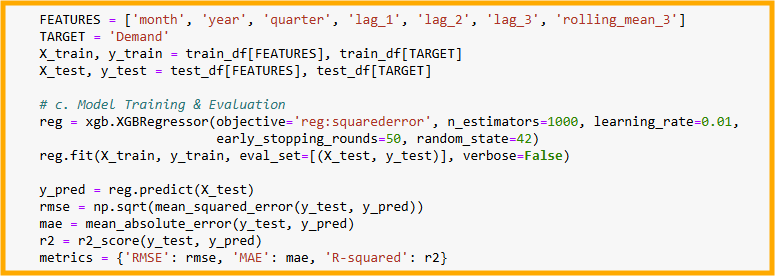


Fig. 9.1: XGBoost Model

1. **SARIMA Model**

* For each part, SARIMA hyperparameters were chosen via auto-arima, accounting for seasonality (12-month period).
* Applied only to series with sufficient length and variance; a fallback moving average was used when SARIMA was not feasible.
* Forecasts were generated for each part for the next 3 months.
* Model assessment included RMSE, MAE, and R-squared.

A screen shot of a computer code

AI-generated content may be incorrect.

Fig. 9.2: SARIMA Model

1. **Holt-Winters Model**

* Required a minimum of 25 monthly points per part.
* Fit separate additive trend/seasonal models for each part, with 12-period seasonality.
* Similar fallback (to moving average) applied where insufficient data existed.
* Forecasts produced for all qualifying parts, evaluated using RMSE, MAE, R-squared.

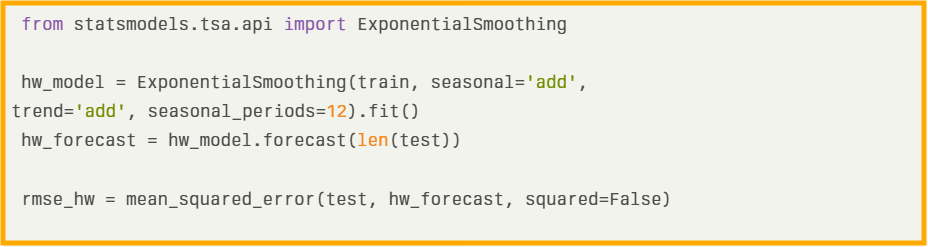


Fig. 9.3: Holt-Winters Model

# Chapter 10: Model Evaluation

This chapter presents a comprehensive evaluation of the time series forecasting models developed for Tata Motors dealership inventory management.

The evaluation framework is based on the following principles:

* Separation of Training and Test Data: Models are trained on historical data (Jan 2022 – May 2024) and evaluated on the most recent 3-month hold-out test set (June–August 2024).
* Comparative Metrics: Performance is benchmarked via Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R² for out-of-sample predictions.
* Business Alignment: Evaluation emphasizes not only numerical accuracy but support for actionable business insight specifically, cost reduction and service reliability.

Three core models were developed and benchmarked for demand prediction on segmented (AFX/AFY/AFZ/BFX) time series –

* XGBoost Regressor (Gradient Boosted Trees, with lag features),
* SARIMA (Seasonal Autoregressive Integrated Moving Average),
* Holt-Winters Exponential Smoothing (Additive trend and seasonality)

The model is evaluated based on the RMSE, MAE and R² metric given below in the Table 10.1.

Table 10.1: Model Performance Table

| **Model** | **RMSE** | **MAE** | **R²** |
| --- | --- | --- | --- |
| XGBoost | 156.49 | 14.45 | 0.836 |
| SARIMA | 87.57 | 8.32 | 0.957 |
| Holt-Winters | 129.82 | 14.71 | 0.887 |

* SARIMA outperformed both XGBoost and Holt-Winters across all key metrics, achieving the lowest RMSE and MAE, and the highest R², indicating a strong fit and precise predictive power.
* XGBoost: while robust and flexible did not match the classical SARIMA’s ability to capture temporal patterns across disaggregated inventory time series.
* Holt-Winters performed strongly for items with seasonality but were less adaptable to sudden structural changes or high volatility present in some segments.

# Chapter 11: Analysis and Results

This case study details an integrated data-driven approach for optimal inventory management at a Tata Motors dealership, leveraging advanced classification, segmentation, and time series forecasting.

The analysis followed a structured data preparation roadmap:

* **ABC Analysis:** Parts were categorized based on their revenue contribution identifying high-value (A), moderate-value (B), and low-value (C) parts.
* **XYZ Analysis:** Segmentation based on demand variability, capturing stable (X), moderate (Y), and highly variable (Z) demand patterns.
* **FSN Classification:** Parts were additionally classified as Fast, Slow, or Non-moving, based on their rate of consumption.

**Final Segmentation:** Combining the above, four composite groups were created for modeling,

AFX: 61 parts - high value, stable demand, fast-moving

AFY: 144 parts - high/moderate value, moderate demand variability

AFZ: 132 parts - high/moderate value, high demand variability

BFX: 10 parts - moderate value, stable and fast-moving

From the full list of 3,785 parts, 347 were selected for focused forecasting based on these classifications.

After segmentation, three forecasting algorithms were benchmarked for each group using a three-month rolling forecast, assessing performance with RMSE, MAE, and R-squared.

**Key Takeaways**

The segmentation logic (AFX, AFY, AFZ, BFX) ensures that part-level forecasting is tailored to specific inventory and demand patterns, enabling smarter reordering and stock optimization.

SARIMA is the preferred model for this use case, recommended for automated deployment across the classified segments due to its accuracy and stability.

XGBoost should be considered when new parts or market conditions introduce more complex or irregular sales behaviors.

Holt-Winters remains a strong option for classically seasonal parts, most notably those in the AFX group—where simple patterns dominate.

This approach enables actionable, data-driven decisions for inventory management and operational efficiency at dealership level, reducing stockouts and overstock, and optimizing customer satisfaction.

# Chapter 12: Deployment

The end-to-end deployment, as depicted in the architecture, follows a robust batch inference process for monthly inventory forecasting and segmentation at the dealership level.

1. Data Ingestion and Warehousing
   1. Storage: Raw inventory transactional files (like CSV/XLSX) are uploaded by business users into a cloud storage location.
   2. ADF (Azure Data Factory): Orchestrates the ETL process. It picks up these files on a scheduled basis, typically once per month.
   3. Snowflake: Data passes through multiple layers - Landing Layer, Cleansed Layer, History Layer for historical tracking and core layer for dimensional modeling and reporting.
2. Segmentation and Forecasting
   1. Parts Segmentation: Using predefined business logic (ABC-by-revenue, XYZ-by-demand variability, FSN-by-consumption), parts are classified and grouped into AFX, AFY, AFZ, and BFX segments. The segmented, processed data is sent to the Machine Learning storage layer for monthly inference.
   2. ML Forecasting Models: For each segment group (AFX, AFY, AFZ, BFX), batch time series models (SARIMA, XGBoost, Holt-Winters, etc.) are triggered to train (if necessary) and generate forecasts for the next 2-3 months.
   3. Batch Process: This entire pipeline is orchestrated to run once a month, ensuring all new transactional data since the last run is included along with the rolling data of past 32 months.

Results are written back to presentation or ML storage layers for business-facing dashboards, enabling decision-makers to review stock recommendations.

Summary Table: Monthly Batch Inference Flow

Table 12.1: Summary Table

| **Step** | **Tool/Layer** |  | **Function** |
| --- | --- | --- | --- |
| Data Upload | Cloud Storage |  | Raw file ingestion |
| ETL | Azure Data Factory |  | Cleansing, loading to Snowflake DW |
| Segmentation | Processing Scripts in Databricks |  | ABC/XYZ/FSN, creation of AFX/AFY/AFZ/BFX segments |
| Batch ML Inference | SARIMA/XGBoost/etc. |  | Forecasting demand 2-3 months ahead |
| Evaluation & Output | DW + Dashboards |  | Model ranking, report generation |

This architecture enables repeatable, scalable, and auditable deployment of inventory forecasting optimizing inventory for each segment monthly, with results pushed to business dashboards for actionable insights.

# Chapter 13: Conclusions and Future Scope

* 1. **Conclusion**

The project aimed at understanding the inventory forecasting and analytics challenge for a Tata Motors dealership by leveraging advanced data segmentation, robust machine learning, and time series modeling techniques. The study involved grouping parts into ABC (value-based), XYZ (demand variability), and FSN (consumption rate) segmentations, enabling targeted forecasting and stock strategies. The deployment of XGBoost, SARIMA, and Holt-Winters models provided a comprehensive comparative analysis. Among these, SARIMA delivered the highest accuracy (RMSE ≈87.6, MAE ≈8.3, R² ≈0.96), outperforming both XGBoost and Holt-Winters in real dealership demand prediction tasks.

By automating data flows using Azure-based ETL pipelines and designing dashboards for actionable insights, the solution not only optimized stock levels minimizing both excess and stockouts—but also reduced carrying costs and improved the overall operational efficiency of the dealership. This approach ensures timely availability of parts, directly enhancing customer satisfaction and service turnaround.

* 1. **Future Scope**

Explore the potential of training and testing machine learning (ML) and artificial intelligence (AI) models to enhance overall performance, leading to greater cost savings.

Integrate real-time data feeds from multiple dealerships to support proactive, system-wide inventory optimization and replenishment.

Incorporate additional predictive variables (promotions, macroeconomic factors, vehicle sales data) to further refine demand forecasts.

Enhance the ETL process for better streamlining and Auditing.

# Bibliography

[1] Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. (2022). Inventory–forecasting: Mind the gap. European Journal of Operational Research, 299(2), 397–419. <https://doi.org/10.1016/j.ejor.2021.07.040>

[2] Sharma, A. K., ManjuKiran, P. S. J., Maheshwari, P., & Divakar, V. (2021). Demand forecasting using coupling of machine learning and time series models for the automotive aftermarket sector. 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT). IEEE. DOI: 10.1109/ICEECCOT52851.2021.9708010

[3] Rushton, R., Lorraine, O., Tiong, J., Karim, M., Dixon, R., Greenshields, W., Marotti, R., & Bretaña, N. A. (2023). Forecasting inventory for the state-wide pharmaceutical service of South Australia. Procedia Computer Science, 219, 1257–1264. <https://doi.org/10.1016/j.procs.2023.01.409>

[4] Graves, S. C. (2022). Inventory Planning with Forecast Updates in a Multi-Echelon Supply Chain. MIT Center for Transportation & Logistics Working Paper. https://ctl.mit.edu (Note: official DOI not available as it's a working paper, not a peer-reviewed journal article.)

[5] Manavalan, E., & Jayakrishna, K. (2020). Development of inventory forecasting model for spare parts in the automotive industry. Procedia Manufacturing, 43, 223–230. <https://doi.org/10.1016/j.promfg.2020.02.150>

[6] Ramaa, A., Subramanya, K. N., & Rangaswamy, T. M. (2014). Inventory Management Models: A Case Study of Retail Sector in India. Procedia Engineering, 97, 1784–1791. <https://doi.org/10.1016/j.proeng.2014.12.333>

[7] Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021). Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. 11th Conference on Innovative Data Systems Research (CIDR ’21). <https://www.cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf>

[8] Databricks. (2023). What Is the Medallion Architecture? (Industry Whitepaper). While not a peer-reviewed conference paper, this is the canonical source for the architecture.

[9] Dageville, B., Cruanes, T., Zukowski, M., Antonov, V., Avanes, A., Bock, A., … & Sundaram, R. (2016). The Snowflake Elastic Data Warehouse. Proceedings of the 2016 International Conference on Management of Data (SIGMOD ’16), 215–226. ACM. <https://doi.org/10.1145/2882903.2903741>

[10] Chen, L., Li, J., & Wang, S. (2020). A Cloud-Based Platform for Large-Scale Demand Forecasting in E-commerce. In Proceedings of the IEEE International Conference on Big Data (Big Data), pp. 4513-4522.

# Appendix

## Plagiarism Report[[1]](#footnote-1)

**Plagiarism Report** with below 15% Similarly index to be attached in the annexure. The title page and last pages with the similarity index report are attached.

## Github Link

**Additional Instructions (To be removed while submitting)**

**Formatting for Printing:**

1. The printing of your capstone project must be done with a RACE-approved printer only. Details will be shared.
2. Cover page shall be in White hardbound (for both 1st and 2nd year projects) as per the template shared.
3. The report shall be printed on an A4 size Executive Bond sheet.
4. The font used shall be Times New Roman, and the font size shall be 12. For the Heading, use **Times New Roman 14 in Bold** and for the subheading, use **Times New Roman 12 in Bold**.
5. The top, bottom, left and right margins shall be 1” each.
6. The line spacing shall be fixed at 1.5 lines.
7. Table line spacing shall be single line spacing.
8. Page numbers shall be placed at the bottom right position.
9. Chapters shall be numbered 1, 2, 3, etc. The tables and charts shall be in the format of 1.1, 1.2, etc. i.e., 1.1 indicate that it is the first table in Chapter 1; 2.1 Indicates the first table in Chapter 2. Similarly, chart no. 1.1 indicates the first chart in Chapter 1.
10. The project report shall be a minimum of 40 pages and shall not exceed 75 pages for second-year projects and a minimum of 30 pages, and a maximum of 50 pages for first-year projects.
11. You must submit three hard copies duly signed by the mentor and guide (scanned signature will be sufficient) and the Director along with a soft copy in pdf format. (Two copies to submit to the university and one is your copy).
12. Title of the study, objectives, analysis, findings, and suggestions must tally.
13. Each chapter must start on a fresh page.

**Mandatory Inclusions**

1. **Plagiarism Report** with below 15% Similarly index to be attached in the annexure. The title page and last pages with the similarity index report are attached.
2. **Publication** in a Journal/Conference Presented/White Paper – **Full paper** extracted from the journal / full conference paper and **the certificate** must be attached.
   1. Those who have **published**: (Attach full paper)

[Authors Name], “[Article Title].” [Journal Name/Conference Name], [Volume Number], [Issue Number], [Year], [Pages], DOI.

For those who have **not published** yet may add submission information: (Attach full paper)

[Authors Name], “[Article Title].” [Journal Name/Conference Name], [Date of submission]

1. Github Link
2. Any other annexures (optional).

1. Turnitn report to be attached from the University. [↑](#footnote-ref-1)