**DEPARTMENT OF INFORMATION TECHNOLOGY**

**B.K BIRLA COLLEGE OF ARTS, SCIENCE & COMMERCE (AUTONOMOUS)** **KALYAN, 421301**

***(Affiliated to University of Mumbai)***



# Internship Report

**Organization Name :**  TCS iON

**Intern Name :** Tejas Chandrakant Chaudhari

**Roll No :**  07

**Student ID :** 3836372

**Year :** 2023-2024

## Submission Date : 2 January 2024

**University Name :** Mumbai University

**College Name:**: B.K.Birla College of Arts, Science and Commerce

# TABLE OF CONTENTS

**Chapter 1: Letter of Undertaking**

**Chapter 2: Experience Letter**

**Chapter 3: Acknowledgement**

**Chapter 4: Objective**

**Chapter 5: Introduction**

**Chapter 6: Internship Activities**

**Chapter 7: Steps Involved**

**Chapter 8: Assumptions**

**Chapter 9: Approach / Methodology**

**Chapter 10: Challenges & Opportunities**

**Chapter 11: Reflection on Internship**

**Chapter 12: Conclusion**

# CHAPTER 1: Letter of Undertaking

I, Tejas Chandrakant Chaudhari, a student enrolled in the MSc. DSBDA program with Student ID 3836372, hereby confirm that the internship report I have submitted is solely the result of my own efforts. I did not copy any part of my report, either in whole or in part, from any other student or source, whether for payment or free. Additionally, I affirm that I have not included any plagiarized material in any section of my report.

Furthermore, I confirm the authenticity of the documents provided, asserting that they have been legitimately issued by authorized personnel within the organization. In the event that I am found to be misrepresenting, misleading, or concealing any facts regarding my activities, whether academic or nonacademic but relevant to this course, I acknowledge that the university is authorized to take disciplinary action against me in accordance with university policies and regulations.

I also confirm that I have thoroughly read and understood all the guidelines, rules, and regulations provided by the Course Instructor. I accept full responsibility in the case of any negligence on my part.

Sincerely,

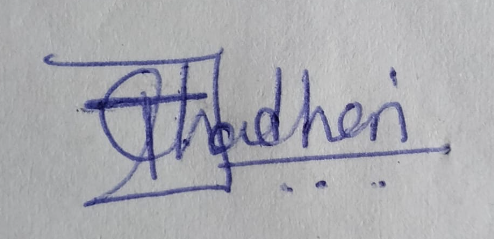
Tejas Chandrakant Chaudhari

MSc. DSBDA Student

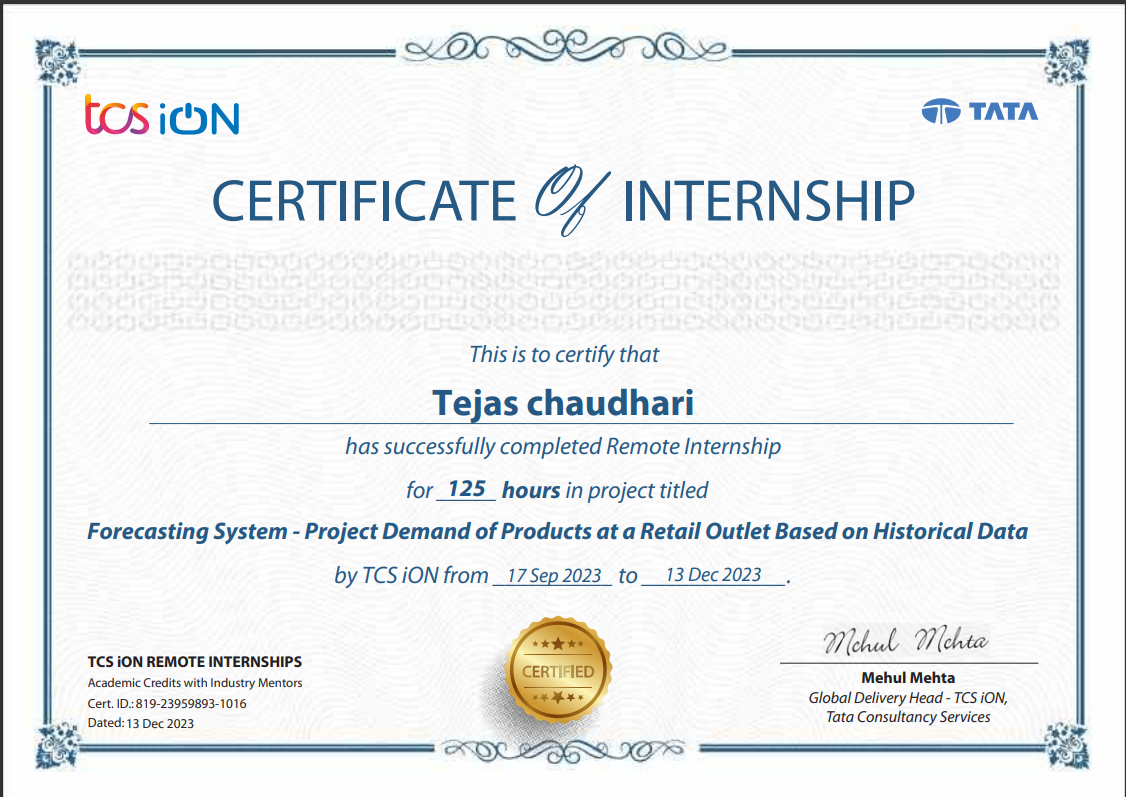
Student ID: 3836372

Date: 02-01-2024

Sign:



# CHAPTER 2: Experience Letter



# CHAPTER 3: Acknowledgement

I express my profound gratitude to my industry mentor for their unwavering support and guidance throughout this project. Their provision of an excellent platform has been instrumental in completing this endeavor. I'm sincerely thankful for their consistent assistance in addressing my queries at every phase. I also want to thank all my friends who helped me with valuable suggestions during this project.

I would like to convey my sincere thanks to Prof. Esmita Gupta, Head of the Department, and the esteemed faculty of the Department of Information Technology at B K Birla College of Arts, Science & Commerce (Autonomous), Kalyan. Their generous opportunity to delve into the corporate world and gain insights into IT business has been truly invaluable.

A special note of gratitude goes to the Department of Information Technology for entrusting me with the 'Forecasting System - Project Demand of Products at a Retail Outlet Based on Historical Data.' Their unwavering support, continual inspiration, and insightful guidance were indispensable in navigating the complexities of this project. Their wealth of knowledge and experience served as a guiding beacon, assisting me in overcoming numerous challenges.

Additionally, I am deeply thankful to my friends whose perspectives and contributions during extensive project discussions played a pivotal role in refining the smart glasses, particularly focusing on user-friendliness

# Chapter 4: Objective

The core aim of this project is to architect and deploy a resilient Forecasting System that demonstrates a high capability in accurately anticipating product demands based on historical data analysis. The system is designed to harness advanced forecasting models and analytical techniques, empowering it to offer dependable insights into forthcoming demand patterns. Utilizing the wealth of historical data, the objective is to elevate strategic decision-making processes, streamline inventory management, and significantly augment overall operational efficiency. Positioned as a pivotal tool, this Forecasting System is poised to efficiently forecast and address customer demands within the dynamic landscape of modern business

# Chapter 5: Introduction

In the ever-evolving landscape of retail and inventory management, accurately predicting and understanding product demand stands as a pivotal factor for businesses striving to not just survive but to excel. This project is centered on the domain of demand forecasting, aiming to leverage the vast potential inherent in extensive datasets to provide precise insights into future sales trends.

The dataset utilized for this forecasting endeavor comprises a substantial 913,000 rows structured around four pivotal columns: Date, Store, Item, and Sales. Each row within this dataset represents a unique transaction, capturing intricate details regarding the specific location and product involved in the sale. This dataset functions as the fundamental cornerstone upon which our forecasting models are constructed, offering a comprehensive panorama of past consumer behaviors and purchasing patterns.

To navigate the intricacies embedded within our dataset, we employed three robust forecasting models: SARIMAX, Holt-Winters Exponential Smoothing (HWES), and Prophet. Each model contributes its distinct strengths and methodologies to the analysis. Through a methodical and exhaustive evaluation process, we rigorously examined the capabilities of these models. The objective was to pinpoint the most effective approach for accurately forecasting demand within our specific contextual framework

# Chapter 6: Internship Activities

1. Viewing all Welcome kit
2. Internship Pre-assessment
3. Understanding the project and there requirements
4. Daily activity reports and Intermediate report
5. Project Test Assessment
6. Project Report
7. Internship Feedback form

# Chapter 7: Steps Involved in Project

1. Understanding the project & its requirements
2. Searching for the appropriate dataset
3. Data Exploration & Data Preparation
4. Modelling Prerequisites: Before applying forecasting models, assessing data seasonality and trends for sales is crucial. To examine seasonality, we employed two statistical tests - the Augmented Dickey-Fuller (ADF) test & the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.
5. Modelling
6. Fine-tuning

# Chapter 8: Assumptions

1. Availability of Accurate Data
2. Modern Technology Infrastructure
3. Precise Outcome Prediction
4. Efficient Risk Management
5. Flexibility to Accommodate Varied Scenarios
6. Dedication to Continual Improvement

# Chapter 9: Approach / Methodology

**9.1 Creating the dataset**

For this project, the dataset was sourced from the Kaggle Store Item Demand Forecasting Challenge, comprising 913,000 rows organized into four crucial columns—Date, Store, Item, and Sales. This dataset represents a comprehensive repository of historical sales data. The dataset is described as follows:

* Date: Sales date without holiday effects or store closures.
* Store: Store ID, providing a unique identifier for each store.
* Item: Product ID, offering a distinct identification number for each product.
* Sales: Quantity of products sold, indicating the number of products sold from a specific store on a specific date.

**9.2** **Cleaning and preparing the dataset**

The subsequent phase of the project involved a meticulous examination of the raw sales data, focusing on addressing missing values, outliers, and inconsistencies. This comprehensive cleaning process was crucial to uphold the dataset's quality and reliability, establishing a strong foundation for subsequent analysis and modeling.

Furthermore, anticipating the unique data prerequisites of each modeling approach, the dataset underwent tailored transformations to align with the specific requirements of individual models before the modeling phase commenced

**9.3 Data Analysis:-**

Following the completion of the data cleaning phase, an in-depth analysis of the historical sales data was conducted. Employing exploratory data analysis (EDA) techniques, the goal was to uncover trends, patterns, and seasonal variations within the dataset. These insights were pivotal in guiding subsequent modeling decisions.

The data analysis unfolded as follows:

1.Exploratory Data Analysis (EDA):

1. Descriptive Statistics: The initial exploration involved calculating descriptive statistics like mean, median, and standard deviation. This allowed for a broad understanding of key variables' central tendencies and dispersions
2. Correlation Analysis: Conducting correlation analysis helped examine relationships between variables, including sales and external factors like promotions or holidays. Understanding these correlations aided in identifying potential demand drivers.
3. Time Series Decomposition: Time series decomposition was executed to unveil underlying components such as trend, seasonality, and residual variations. This approach facilitated a nuanced comprehension of the sales data evolution.

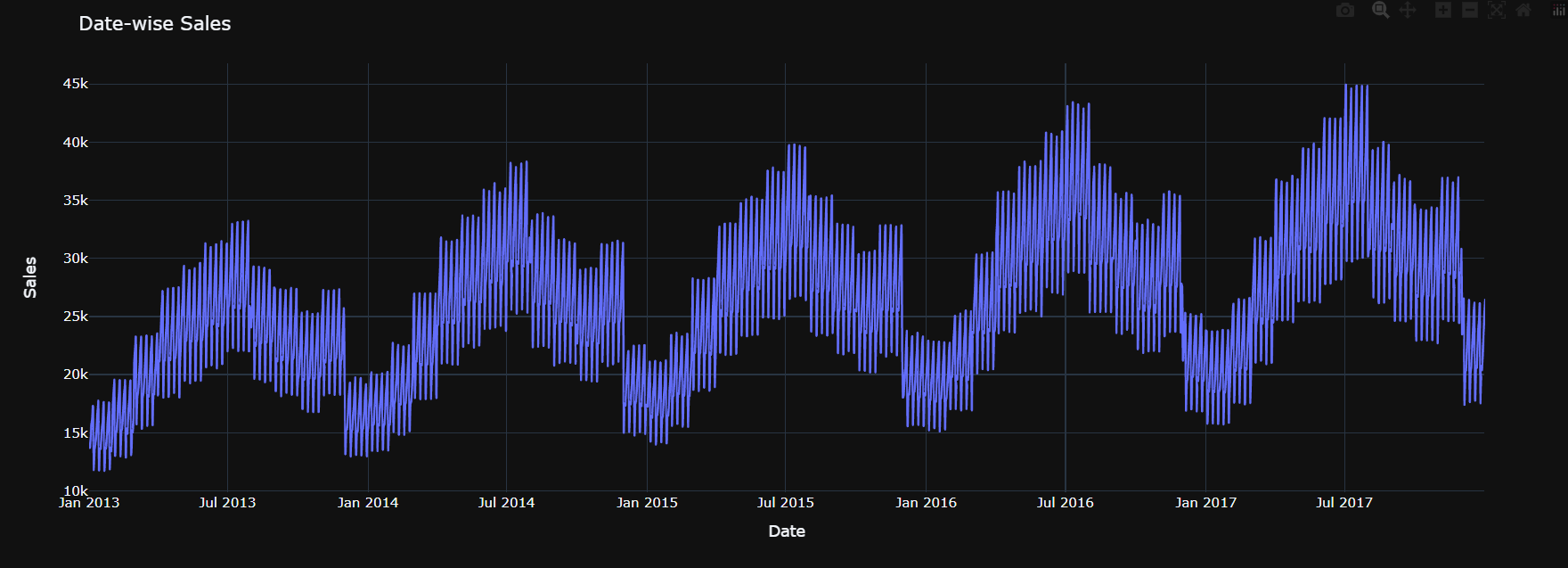
2.Data Visualization:

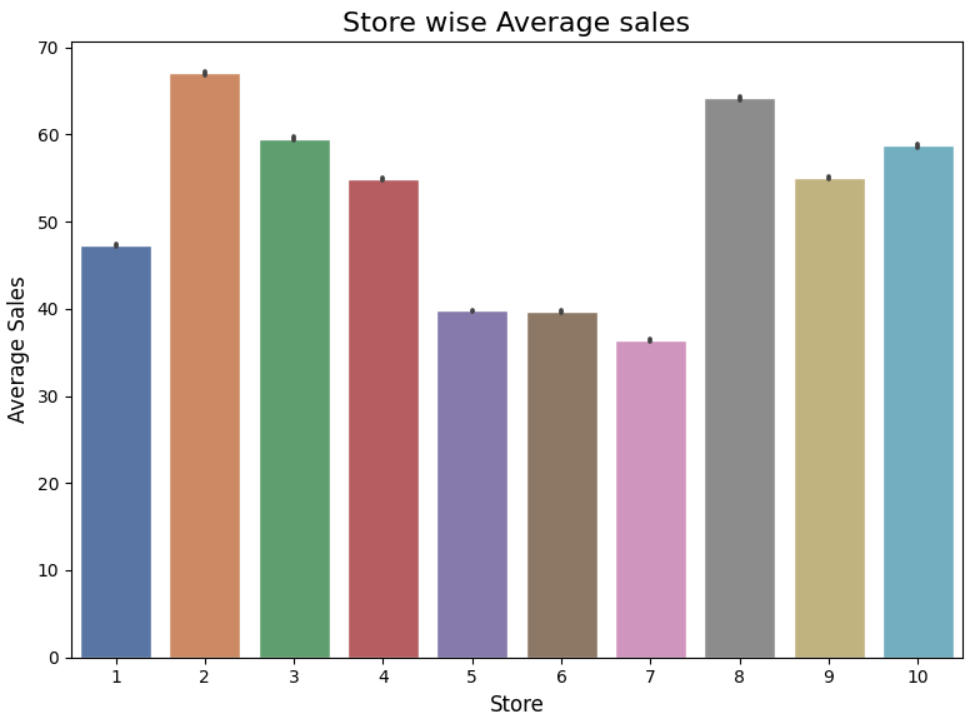
1. Time Series Plots: Visualization of sales trends over time was achieved through time series plots, enabling the identification of long-term patterns, anomalies, and potential outliers.
2. Seasonal Plots: Seasonal plots were generated to visualize recurring patterns, aiding in the identification of seasonality within the sales data. These insights provided potential features for the forecasting models.
3. Distribution Plots: Distribution plots were utilized to analyze the distribution of sales data, identifying skewness and kurtosis. This understanding guided the selection of suitable forecasting models.

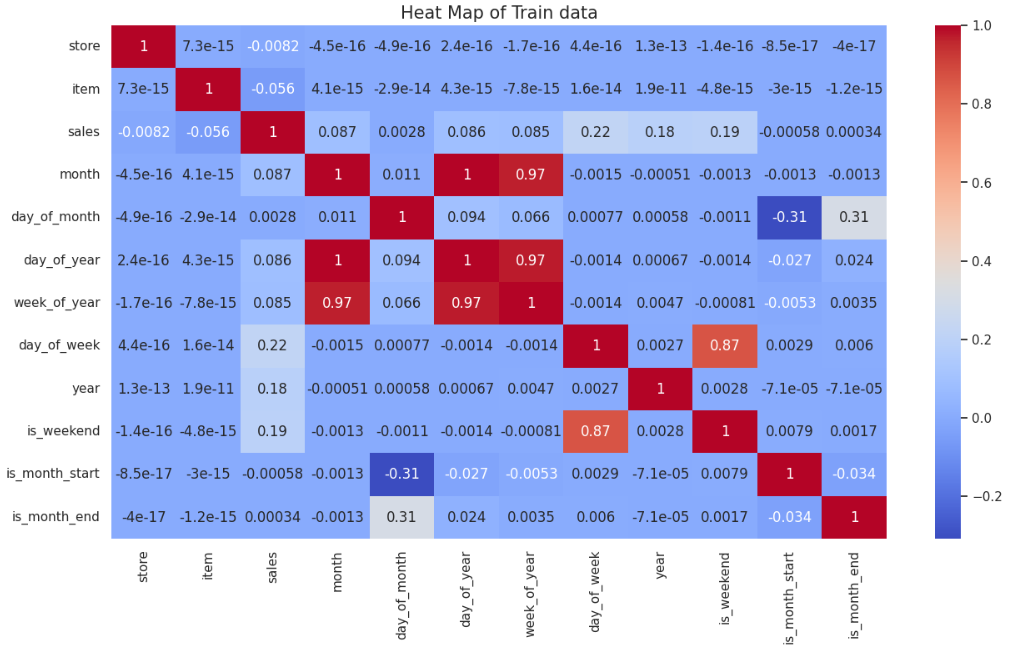
3. Insights and Observations:-

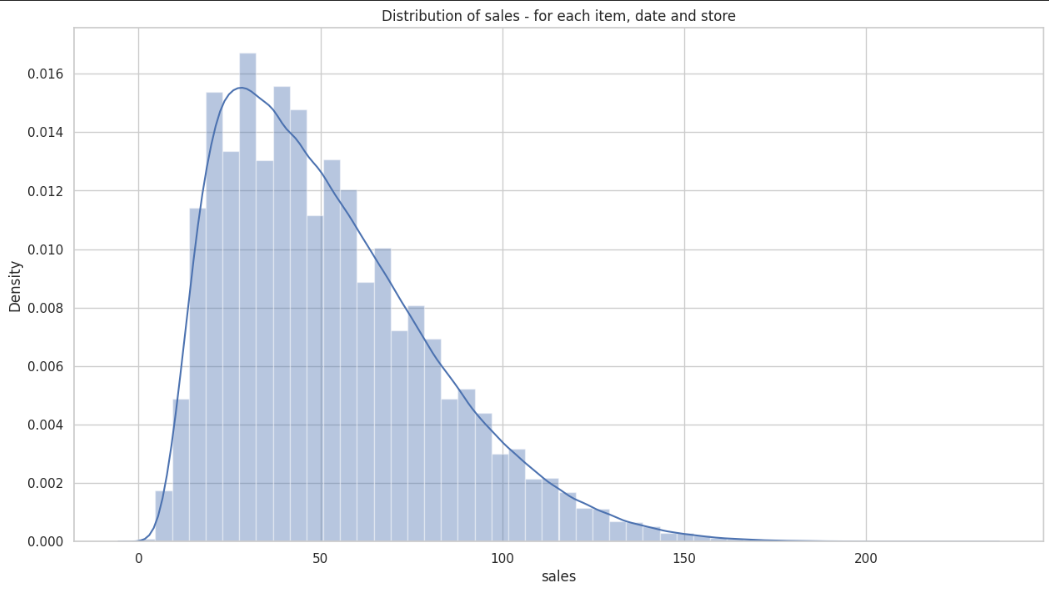
The EDA process yielded significant insights, including:

1. Clear seasonal trends in sales, with noticeable peaks during specific periods.
2. Identifiable trends and patterns suitable for modeling purposes.
3. Consistent sales patterns across all ten stores on a monthly, yearly, and quarterly basis.
4. Similar sales patterns observed across stores, implying a consistent trend in sales performance."
5. Some plots from EDA as follows :-









**9.4 Prerequisite of modeling :-**

Before applying forecasting models, it was essential to assess data seasonality and trends in sales. We employed two statistical tests, namely the

* Augmented Dickey-Fuller (ADF) test
* Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.

Both tests indicated the presence of seasonality within the data, confirming its suitability for applying forecasting models

**9.5 Splitting the data**

The dataset was split in an 80:20 ratio for each model, adhering to the fundamental requirements. In this approach, 80% of the data was allocated for training the model, while the remaining 20% was set aside for testing purposes. This division facilitated the assessment of the model's efficiency and accuracy.

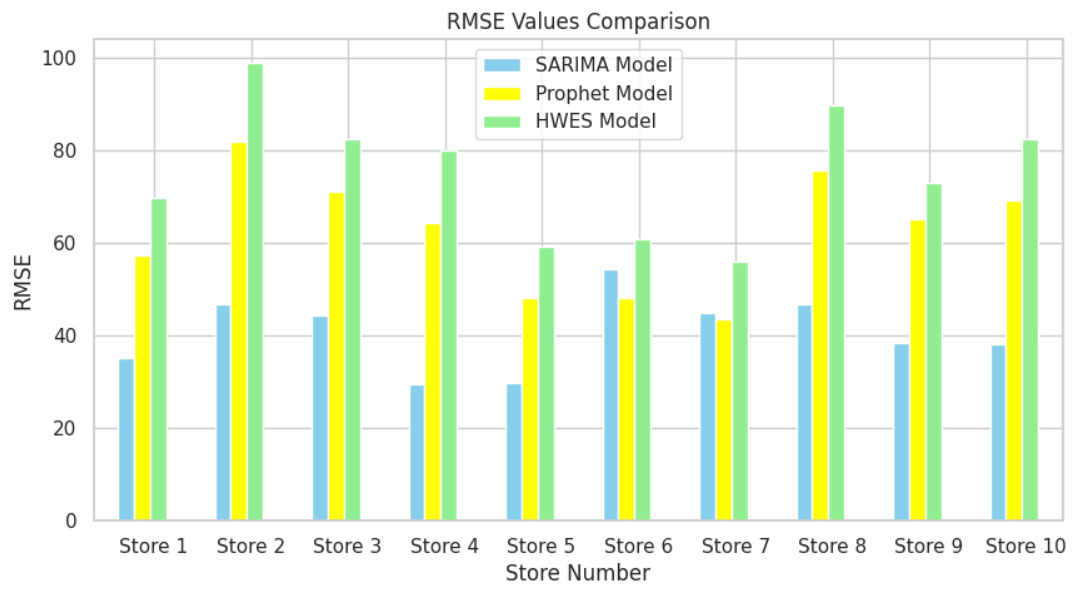
**9.6 Modelling**

The project's focal point involved the development of forecasting models to predict future demand. Three distinct models—SARIMAX, Holt-Winters Exponential Smoothing (HWES), and Prophet—were selected for evaluation and implementation. Each model underwent parameter fine-tuning to enhance predictive accuracy.

1. SARIMAX Model: The Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model was chosen due to its proficiency in capturing both seasonality and trend within time series data. This decision stemmed from the identified seasonality observed in the sales data during the exploratory data analysis (EDA) phase.
2. Holt-Winters Exponential Smoothing (HWES) Model: The HWES method, specifically the triple exponential smoothing technique, was considered for its capability in managing seasonality, trend, and assigning weighted significance to recent observations. This model was selected to compare and evaluate its performance against SARIMAX.
3. Prophet Model: Prophet, an adaptable forecasting model developed by Facebook, was selected for its versatility in handling seasonality and holidays. It excels in scenarios with datasets exhibiting pronounced seasonal patterns and containing missing values. Additionally, the model's optimal parameters are automatically determined within the backend of the code

**9.7 Fine Tuning:-**

Fine-tuning proved to be instrumental in elevating the performance of every forecasting model. The iterative refinements carried out during this phase significantly enhanced prediction accuracy while customizing each model to suit the distinct characteristics of the sales data. The configurations selected through this process formed the bedrock for the final models designated for deployment in the demand forecasting project. This fine-tuning process underscores the critical role of meticulous parameter adjustments in attaining optimal forecasting outcomes



# Chapter 10: Challenges & Opportunities

1. Challenge: The initial challenge encountered in this project revolved around managing the vast volume of data. To effectively utilize it, the data was segmented into 10 store-specific datasets for analysis and modeling.
2. Opportunity: Identifying areas for enhancement has unveiled promising opportunities to expand the capabilities of the demand forecasting project. By integrating advanced deep learning (DL) modeling techniques, exploring additional features, and continuously refining model parameters, the forecasting system can evolve into a more robust and adaptable tool. These enhancements directly align with the project's overarching objective of furnishing precise and actionable insights to bolster effective inventory management strategies

# Chapter 11: Reflection on Internship

The reflection specifically highlights the practical application of theoretical knowledge in the domain of time series forecasting. This encompasses the utilization of a blend of statistical and machine learning techniques. The phrase 'working on actual data' underscores the internship's hands-on nature, emphasizing interns' engagement with real-world datasets. This practical involvement grants them a deeper understanding of the challenges and implications inherent in time series forecasting.

# Chapter 12: Conclusion

The selection of a suitable forecasting model is a critical aspect of any demand forecasting project. After extensive modeling and evaluation, the SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) model emerged as the top-performing model among SARIMAX, HWES (Holt-Winters Exponential Smoothing), and Prophet. This selection was based on its consistent demonstration of superior prediction accuracy across the diverse stores involved in the analysis.

The performance of the SARIMAX model was gauged by several evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics are crucial in assessing how well the model's predictions align with the actual demand data. Lower values of MAE, MSE, and RMSE indicate a higher accuracy in forecasting, and the SARIMAX model consistently showcased lower error values compared to HWES and Prophet across all stores.

In summary, the demand forecasting project successfully fulfilled its primary objective of developing an effective forecasting model. The SARIMAX model emerged as a robust choice, demonstrating superior performance in predicting product demand across diverse stores. The methodology employed in the project, encompassing data cleaning, exploratory analysis, modeling, and fine-tuning, serves as a comprehensive roadmap for future endeavors in demand forecasting.

Moreover, the project's exploration of Deep Learning (DL) techniques as part of the future scope introduces an exciting avenue for further research and enhancement of forecasting accuracy. DL methods, with their capacity to handle complex patterns and relationships within data, offer promising prospects for improving the precision of demand forecasting models.

Overall, the project contributes valuable insights, methodologies, and a successful model deployment that could significantly enhance decision-making processes in inventory management within retail environments. The incorporation of advanced techniques and the project's findings serve as a valuable foundation for future improvements in forecasting and inventory optimization strategies.