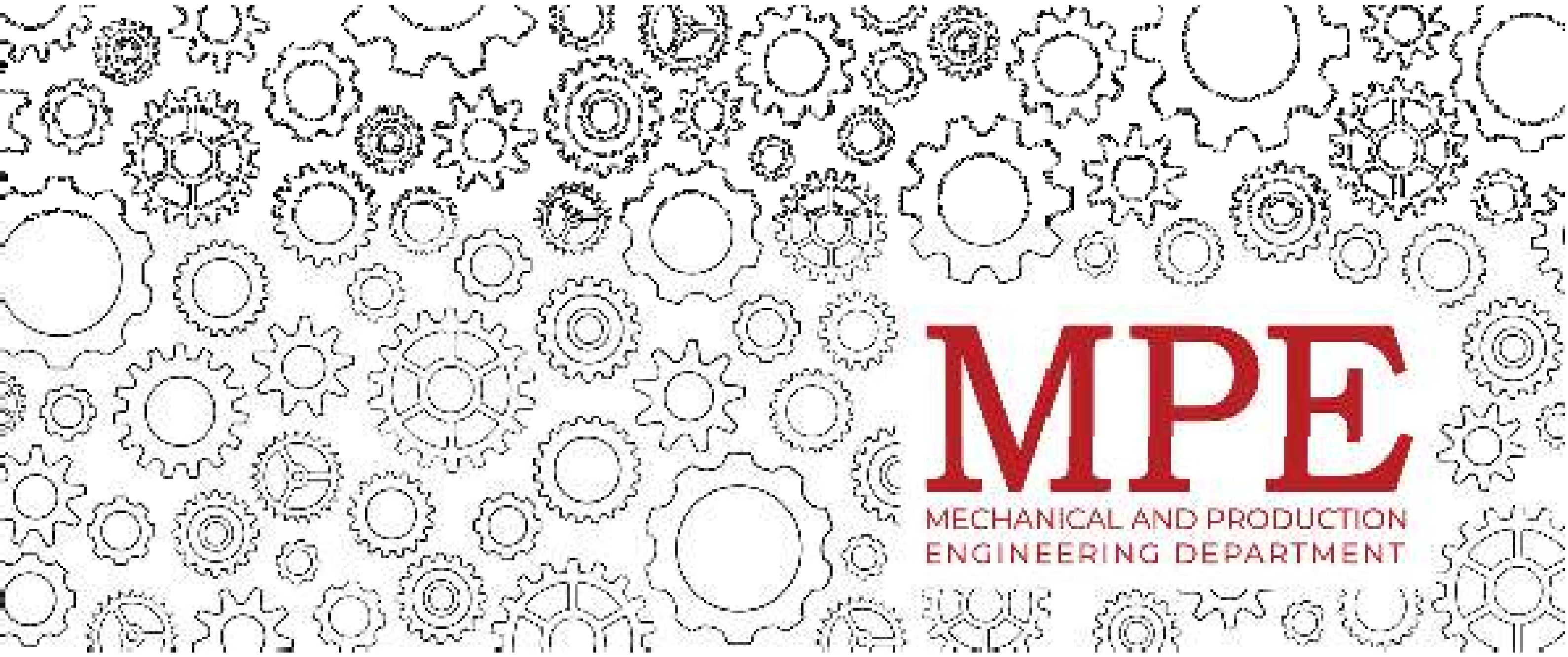
**Islamic University of Technology (IUT)**

The Organization of Islamic Cooperation (OIC)

# MPE FINAL YEAR THESIS



Reducing Food Wastage and Enhancing Restaurants' Operation in Dhaka City: A Data Driven Strategy For Pizza Demand Forecasting Using Machine Learning

2024

# Reducing Food Wastage and Enhancing Restaurants' Operation in Dhaka City: A Data Driven Strategy For Pizza Demand Forecasting Using Machine Learning

**A Thesis by**


**Tawhidul Islam Sazid**

**Abdullah Al Nabil**

**MD. Hasin Al Muhib**

**Department of Mechanical and Production Engineering**

**Islamic University of Technology**


January(2024)

2024

# Reducing Food Wastage and Enhancing Restaurants' Operation in Dhaka City: A Data Driven Strategy For Pizza Demand Forecasting Using Machine Learning

**TAWHIDUL ISLAM SAZID , 190012147**
**ABDULLAH AL NABIL, 190012140**
**MD. HASIN AL MUHIB, 190012138**

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Bachelor of Science in Industrial and Production Engineering

**DEPARTMENT OF MECHANICAL AND PRODUCTION ENGINEERING**

January (2024)

# CERTIFICATE OF RESEARCH

*This thesis titled "**REDUCING FOOD WASTAGE AND ENHANCING RESTAURANT'S OPERATIONS IN BANGLADESH: A CASE STUDY ON MULTIPLE PIZZA RESTAURANTS FOR DEMAND FORECASTING AND STAFF ASSIGNMENT USING MACHINE LEARNING**" submitted by TAWHIDUL ISLAM SAZID (190012147), ABDULLAH AL NABIL (190012140) and MD. HASIN AL MUHIB (190012138) has been accepted as satisfactory in partial fulfillment of the requirement for the Degree of Bachelor of Science in Industrial and Production Engineering.*

***Supervisor***

______________________________

**Prof. Dr. A.R.M Harunur Rashid**

*Professor*

***Head of the Department***

______________________________

**Prof. Dr. Md. Hamidur Rahman**

*Professor*

Department of Mechanical and Production Engineering (MPE)
Islamic University of Technology (IUT)

# DECLARATION

*I hereby declare that this thesis entitled "**REDUCING FOOD WASTAGE AND ENHANCING RESTAURANT'S OPERATIONS IN BANGLADESH: A CASE STUDY ON MULTIPLE PIZZA RESTAURANTS FOR DEMAND FORECASTING AND STAFF ASSIGNMENT USING MACHINE LEARNING**" is an authentic report of study carried out as requirement for the award of degree B.Sc. (Industrial and Production Engineering) at Islamic University of Technology, Gazipur, Dhaka, under the supervision of* **Prof. Dr. A.R.M Harunur Rashid**, *Professor, MPE, IUT in the year 2024*

*The matter embodied in this thesis has not been submitted in part or full to any other institute for award of any degree.*

*Tawhidul Islam Sazid*
*190012147*

*Abdullah Al Nabil*
*190012140*

*Md. Hasin Al Muhib*
*190012138*

**CO-PO Mapping of ME 4800 -Theis and Project**

| COs | Course Outcomes (CO) Statement | (PO) | Addressed by | |
|---|---|---|---|---|
| **CO1** | Discover and Locate research problems and illustrate them via figures/tables or projections/ideas through field visit and literature review and determine/Setting aim and objectives of the project/work/research in specific, measurable, achievable, realistic and timeframe manner. | **PO2** | Thesis Book | |
| | | | Performance by research | |
| | | | | |
| | | | Presentation and soft skill | |
| **CO2** | Design research solutions of the problems towards achieving the objectives and its application. Design systems, components or processes that meets related needs in the field of mechanical engineering | **PO3** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO3** | Review, debate, compare and contrast the relevant literature contents. Relevance of this research/study. Methods, tools, and techniques used by past researchers and justification of use of them in this work. | **PO4** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO4** | Analyse data and exhibit results using tables, diagrams, graphs with their interpretation.<br>Investigate the designed solutions to solve the problems through case study/survey study/experimentation/simulation using modern tools and techniques. | **PO5** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO5** | Apply outcome of the study to assess societal, health, safety, legal and cultural issue and consequent possibilities relevant to mechanical engineering practice. | **PO6** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO6** | Relate the solution/s to objectives of the research/work for improving desired performances including economic, social and environmental benefits. | **PO7** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO7** | Apply moral values and research/professional ethics throughout the work, and justify to genuine referencing on sources, and demonstration of own contribution. | **PO8** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO8** | Perform own self and manage group activities from the beginning to the end of the research/work as a quality work. | **PO9** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO9** | Compile and arrange the work outputs, write the report/thesis, a sample journal paper, and present the work to wider audience using modern communication tools and techniques. | **PO10** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO10** | Organize and control cost and time of the work/project/research and coordinate them until the end of it. | **PO11** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |
| **CO11** | Recognize the necessity of life-long learning in career development in dynamic real-world situations from the experience of completing this project. | **PO12** | Thesis Book | |
| | | | Performance by research | |
| | | | Presentation and soft skill | |

Student Name /ID:

1……………………………………..

2. ……………………………………..

3. ……………………………………..

Signature of the Supervisor: …………………………

Name of the Supervisor: ……………………………..

**K-P-A Mapping of ME 4800 -Theis and Project**

| COs | POs | Related Ks | | | | | | | | Related Ps | | | | | | | Related As | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | K1 | K2 | K3 | K4 | K5 | K6 | K7 | K8 | P1 | P2 | P3 | P4 | P5 | P6 | P7 | A1 | A2 | A3 | A4 | A5 |
| CO1 | PO2 | | | | | | | | | | | | | | | | | | | | |
| CO2 | PO3 | | | | | | | | | | | | | | | | | | | | |
| CO3 | PO4 | | | | | | | | | | | | | | | | | | | | |
| CO4 | PO5 | | | | | | | | | | | | | | | | | | | | |
| CO5 | PO6 | | | | | | | | | | | | | | | | | | | | |
| CO6 | PO6 | | | | | | | | | | | | | | | | | | | | |
| CO7 | PO8 | | | | | | | | | | | | | | | | | | | | |
| CO8 | PO9 | | | | | | | | | | | | | | | | | | | | |
| CO9 | PO10 | | | | | | | | | | | | | | | | | | | | |
| CO10 | PO11 | | | | | | | | | | | | | | | | | | | | |
| CO11 | PO12 | | | | | | | | | | | | | | | | | | | | |

Student Name /ID:

1……………………………………..

2. ……………………………………..

3. ……………………………………..

Signature of the Supervisor: …………………………

Name of the Supervisor: ……………………………..

# ACKNOWLEDGEMENT

In the Name of Allah, the Most Beneficent, the Most Merciful

First of all, I am grateful to ALLAH (SWT), the most benevolent and kind to provide me with the strength and ability to write this dissertation. I want to thank my project supervisor, **Prof. Dr. A.R.M Harunur Rashid**, for his strong and patient support through unpredictable problems during the project and his precious advice when I faced difficulties. His generosity, kindness and strong supervision during work made me feel less stressed in confronting unexpected troubles and being more productive in my personal life.

In the next step, I would like to express my deep acknowledgement to my father and mother for their continued support and dedication towards my higher studies.

## ABSTRACT

In the framework of many pizza businesses in Bangladesh, this thesis examines the crucial problems of food waste. The study adopts a broad strategy to enhance demand forecasting by combining machine learning techniques with data-driven insights. In addition to conducting in-depth interviews with restaurant operators and sales personnel and distributing questionnaires intended to understand customer tastes and behavior, the research comprises the methodical gathering of sales statistics from numerous pizza shops. Through the use of several machine learning models and extensive data preparation, the research seeks to identify the best model for the intricate characteristics of Bangladesh's restaurant business. Using real-world pizza restaurants as a case study, machine learning models are applied in a practical way. Their performance is closely examined in comparison to sales data, and their effects on decreasing food waste and increasing operational efficiency are assessed. The expected results are intended to be both a possible benchmark for comparable situations around the world and a source of useful information for Bangladesh's restaurant industry. This thesis offers a thorough methodology, a thorough analysis, and a thorough discussion in an effort to go beyond simply presenting a data-driven solution. The goal is to provide insightful information that will direct future investigations into the fields of operational optimization and food waste reduction. The research hopes to promote sustainable practices in the restaurant industry by addressing these important factors.

# Table of Contents

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER - ONE
# INTRODUCTION

## 1.1 Background of the Study

Restaurants, particularly those that serve pizza, are becoming more and more popular in Bangladesh. Restaurants face difficulties as a result of this expanding trend, despite its greatness. One major issue is that food waste can occasionally be excessive, which is bad for both the environment and the restaurant's bottom line.

Considering a hectic day at a pizza restaurant. Occasionally, an excess of pizza is produced and remains unsold or uneaten. At other times, they might not be able to meet demand, in which case customers might go elsewhere. Determining the appropriate quantity to prepare and allocating the appropriate workforce to attend to customers are challenging but vital to a restaurant's success.

This study investigates the use of machine learning, to assist in resolving these issues. We're interested in learning how this tool forecasts the quantity and sizes of pizza that customers will order. We want to help pizza shops in Bangladesh save money, cut down on waste, and improve customer satisfaction by doing this.

Bangladesh is a distinct nation with distinct culinary customs. This study takes into account the unique conditions present, such as varying lifestyles and preferences. We think that utilizing ML to comprehend and adjust to these elements can significantly impact the way restaurants run.

In simpler terms, our goal is to find a smart and practical solution for pizza restaurants in Bangladesh to serve customers better, save money, and waste less food. This study might also show how this approach could help other types of restaurants and businesses.

## 1.2 Problem Statement

The restaurant business in Bangladesh experiences notable inefficiencies as a result of a heavy reliance on conventional forecasting methods and, in many situations, no forecasting methodology at all. This causes a large amount of raw material waste, which makes providing smooth client service challenging. We suggest that a machine learning model could estimate demand very accurately and effectively, hence solving these significant challenges, since there is an evident need for a sophisticated solution. This study attempts to determine and put into practice the best machine learning-based approach to improve demand forecasting in restaurants in Bangladesh, with an emphasis on reducing waste of raw materials, boosting operational effectiveness, Offering a comprehensive and workable solution to the issues this region's restaurant business faces is the aim of the extensive investigation.

## 1.3 Goals & Objectives

Goals and Objectives via questionnaires

1. What are the main variables driving the demand for pizza in Bangladeshi restaurants, and how can these variables be included in forecasting models?
2. In order to improve demand forecasting, how might outside variables like Bangladesh's weather, holidays, cultural events, and social media trends be incorporated into machine learning models?
3. In the case of Bangladeshi restaurants selling pizza, how accurate is machine learning-based demand forecasting compared to traditional methods?

## 1.4 Scope and Limitations

The scopes of our study are:

- investigating the application of machine learning algorithms to precisely forecast the demand for pizza while taking into account regional preferences, advertising tactics, and consumer preferences.
- investigating machine learning techniques to optimize workforce levels through the analysis of past data, flexibility in response to changing demand patterns, and consideration of outside variables affecting consumer traffic.
- selecting the best demand forecasting model by putting different machine learning algorithms into practice and optimizing them
- enhancing forecasting accuracy by adding external variables to the machine learning models, such as weather, holidays, cultural events, and economic situations.

Limitations of our study are:

- it may not be possible to generalize the study's conclusions and suggestions to other kinds of eateries or to a wide range of cuisines outside of the pizza environment.
- The quantity and caliber of data greatly influence how effective machine learning models are. The models' efficacy may be impacted by difficulties in getting complete and accurate data from all participating establishments.

# CHAPTER - TWO
# LITERATURE REVIEW

Table 01: Key Findings and Limitations of Core Papers

| **Title of the Papers** | **Key Findings** | **Limitations** |
|---|---|---|
| “Demand forecasting in restaurants using machine learning and statistical analysis” - *Takenaka, T. (2019)* | i. The study provides valuable insights into the practical applicability of machine learning and statistical analysis in demand forecasting for restaurants.<br><br>ii. Evaluation of demand forecasting methods: Bayesian Linear Regression, Boosted Decision Tree Regression, Decision Forest Regression, and Stepwise method.<br><br>iii. Practical applicability with forecasting rates exceeding 85%.<br><br>iv. Utilization of internal and external data for enhanced accuracy. | i. Limited generalizability due to the small number of restaurants studied.<br><br>ii.Limited exploration of the impact of demand forecasting on other aspects of restaurant operations. |
| “A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants” - *Pilz, J. (2022)* | i. The paper introduces two Bayesian generalized additive models for forecasting demand at the level of individual menu items using POS data without additional data inputs.<br><br>ii. The first model assumes normally distributed sales, while the second assumes a negative | i. The paper focuses on near-term forecasting, primarily addressing daily sales rather than long-term trends or strategic planning. |

| | | |
|---|---|---|
| | binomial distribution, which is suited to the count nature of sales data.<br><br>iii. The Bayesian approach is computationally efficient because it focuses on maximum a posteriori estimates, which allows for frequent model retraining.<br><br>iv. The models are evaluated using POS data from two different restaurants, showing that they provide a reasonable fit for the observed time series data | |
| "Daily retail demand forecasting using machine learning with emphasis on calendric special days" - *Stuckenschmidt, H. (2020)* | i. Machine learning methods show promise as a viable alternative for retail demand forecasting, especially when accounting for variations due to special calendric days.<br><br>ii. ML methods significantly outperform traditional forecasting approaches, particularly on special days, potentially reducing error by 10% to 20%. | i. The study did not explore automated model-building processes or hyperparameter optimization, which could lead to further improvements in accuracy.<br><br>ii. This study did not include a detailed feature importance analysis to gain domain-specific insights. |
| "Data analytics in the supply chain management: Review of machine learning applications in demand forecasting." - *Alan Priyatna, I. (2020)* | i. The paper discusses the impact of machine learning on supply chain management, specifically in demand forecasting.<br><br>ii. Highlights the disruption caused by technology in business | |

| | | |
|---|---|---|
| | models, emphasizing the role of big data and ML in SCM.<br><br>iii. Neural networks, artificial neural networks, support vector regression, and support vector machine are the most used ML algorithms in demand forecasting.<br><br>iv. These four algorithms account for 77% of the ML applications in demand forecasting.<br><br>v. ML algorithms provide more accurate forecasts and reduce computational costs compared to traditional models. | |
| "Machine Learning and statistics: A Study for assessing innovative demand forecasting models." - *Kamphues, J. (2021)* | i. The study compares six prevalent forecasting models, including ETS, SARIMAX, XGBoost, Random Forest, LSTM, and MLP, and evaluates their performance in forecasting demand patterns.<br><br>ii. The results show that the MLP (Multilayer Perceptron) model delivered the highest overall performance score, indicating its suitability for a wider range of products.<br><br>iii. ML-based models RF and XGBoost proved to be low in effort, while DL methods (LSTM and | i. The study does not consider the integration of other parameters that correlate with demand, such as weather or economic data, which could impact the forecast quality.<br><br>ii. Deep Learning methods require high computational effort, which may limit their practical applicability in certain environments. |

| | | |
|---|---|---|
| | MLP) required higher computational effort. | |
| " Time series analysis for supply chain planning in restaurants" - *Renisha, B. (2020, October)* | i. Development of a machine learning system for forecasting raw material procurement based on dish sales prediction.<br><br>ii. Utilization of time series data and machine learning models (Holt Winters and STL algorithms) to accurately forecast product procurement needs. | i. The study focuses on a specific geographical area (Chennai, India), limiting the generalizability of the findings to a global scale.<br><br>ii. The effectiveness of the proposed machine learning system may be influenced by the availability and accuracy of sales data and raw material information.<br><br>iii. The paper needs to address the potential challenges or barriers to implementing the proposed system in real-world restaurant settings |

# CHAPTER - THREE
# METHODOLOGY

## 3.1 Introduction

There were a number of crucial stages in the methodology that comprised our pizza demand forecasting model. We initiated the process by amassing exhaustive sales data from ten distinct establishments, in addition to weather data obtained from NASA and government holiday data. Subsequently, we employed Python packages like numpy and pandas to perform meticulous data cleaning and preprocessing, ensuring the dataset's integrity and preparedness for analysis. Subsequently, we conducted exploratory data analysis (EDA) to detect noteworthy patterns and connections, which then guided our feature engineering procedure. This stage was crucial in the process of constructing and selecting pertinent attributes that would augment the model's ability to make accurate predictions. We utilized K-fold cross-validation for model training and employed RandomizedSearchCV to optimize hyperparameters, hence enhancing the performance of our chosen models: Random Forest Regressor, Gradient Boosting Regressor, XGBoost Regressor, and AdaBoost Regressor. In order to enhance precision, we merged these models into ensemble techniques, namely Voting and Stacking Regressors. By employing a comprehensive and systematic methodology, we were able to ensure that our models were precise and applicable to a wide range of scenarios. This allowed us to generate dependable predictions that effectively reduced food waste and optimized the operations of restaurants.

## 3.2 Research Design

Initially, we gathered data from multiple restaurants and performed data cleaning and preprocessing using Python's Numpy and Pandas modules. Upon completing the data cleaning process, we analyzed the patterns and fluctuations and investigated any external variables that may influence pizza sales. In the Feature Engineering area, we generated significant features that have an impact on our sales volume. Subsequently, we employed Pearson's correlation coefficient to identify significant features. Subsequently, we partitioned our dataset into two segments: the training data, which encompassed 80% of the original data, and the testing data, which comprised 20% of the original data. Within the training data, we further partitioned the data into separate sets for training and validation purposes. The training data consists of 80% of the original training data, while the validation data consists of 20% of the original training data. Next, we chose six widely used and popular machine learning models based on trees and proceeded to train each of them. After completing the training process, we narrowed our selection to four models based on the evaluation metrics: R2 Score, MSE, MAE, and MAPE. Subsequently, we fine-tuned the parameters of these four models using Randomised SearchCV. Once we obtained the optimal settings, we proceeded to retrain the models and assess their performance and error metrics. After carefully choosing three models out of four, we combined them to create two further models: a Voting model and a Stacking model. Subsequently, we evaluated our chosen models by comparing them to voting and stacking models using performance metrics and other error metrics. After conducting a thorough comparison of many models, we identified the most optimal model. We then proceeded to evaluate the performance of this model using our actual testing data.

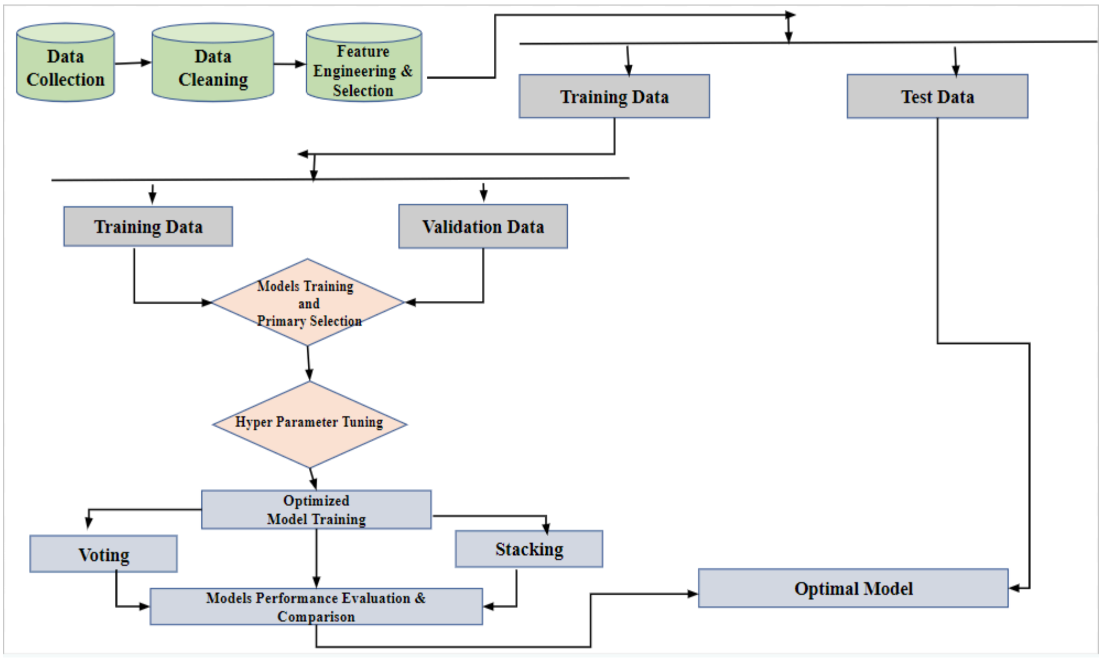



Figure 01: Flow Chart of Methodology

## 3.3 Data Collection

The data is the foundation of the machine learning model used for demand forecasting, and ensuring the data is of high quality is one of the most challenging tasks in data collection. Access to comprehensive and precise data will result in the development of strong prediction models that can accurately forecast future sales. In order to address the issue of food loss and improve restaurant operations, reliable demand projections are essential. These forecasts enable restaurants to optimize their inventory, minimize food spoilage, and enhance customer satisfaction by ensuring that desired items are readily available.

### *3.3.1 Sales Data:*

This study uses sales data of pizzas from 10 different restaurants in Dhaka City as the core dataset. In order to keep the examination consisting of a variety of different types of restaurants, typical chains as well as local eateries were chosen to be included in it. This diversity ensures that the data captures a wide spectrum of sales patterns and consumer behaviors.

**Data Collection Process:**

1. Restaurant Selection: Ten restaurants were selected based on their willingness to participate and their relevance to the study (i.e., they mostly specialize in selling pizza).
2. Types of data: The sales data include daily records spanning the previous two years, providing a comprehensive account of:
   Ordered_Date: The exact day on which each sales record was made.
   Ordered_Quantity: The aggregate quantity of pizzas sold on a daily basis.
   Customer_ID: Each customer has a unique ID for each day.
   Pizza_ID: Each pizza has a unique ID.
   Pizza_Name: Each Pizza has a unique name.
   Pizza_Size: Each pizza has different sizes.

**Difficulties Faced:**

Ensuring data privacy and confidentiality: It was crucial in establishing trust with restaurant owners and securing access to precise sales data.
Data Consistency: Ensuring that data from various sources maintained uniform formatting and included same time periods.

### *3.3.2 National Holiday Data:*

Public holidays can have a substantial impact on restaurant revenues since they lead to alterations in customer behavior and dining habits. We acquired national holiday data from the official government website of Bangladesh.

**Data Collection Process:**

1. **Data Source:** The authoritative government website that provides a comprehensive list of all national holidays.
2. **Types of data**: The holiday dataset contains data for the last two years, providing comprehensive information on:
   - Holiday Name: The designated title for each official holiday.
   - Date: The precise day on which the holiday occurs.
   - Holiday Type: Categorization of the holiday based on its nature (e.g., national, religious, cultural)

**Difficulties Faced:**

1. **Accuracy:** Ensuring the holiday data was complete and accurate.
2. **Data Format:** Standardizing the format of holiday data to ensure compatibility with sales and weather datasets.

#### *3.3.3 Weather Data:*

Weather conditions have a substantial impact on consumer behavior, which in turn affects the demand for food. In order to integrate this component into our forecasting model, we gathered past weather data for Dhaka City from the NASA website.

Data Collection Process:

1. Data Source: NASA's website, which offers dependable and comprehensive meteorological information.
2. Types of data: The weather dataset contains daily records from the previous two years, providing detailed information on:
    - Temperature: The daily maximum and minimum temperatures measured in degrees Celsius.
    - Precipitation: It refers to the quantity of rainfall measured in millimeters.
    - Humidity: The average proportion of moisture in the air on a daily basis.
    - Wind Speed: The average speed of the wind on a daily basis, measured in kilometers per hour.
    - Weather Conditions: Descriptive terms such as sunny, overcast, rainy, etc.

**Difficulties Faced**:

1. Data Granularity: Ensuring that the weather data matched the granularity of the sales data (daily records).
2. Data Integration: Aligning weather data with sales data by date to ensure accurate merging of datasets.

## 3.4 Data Cleaning and Preprocessing

#### 3.4.1 What is Data Cleaning and Preprocessing

Data cleaning and preprocessing are crucial stages in the machine learning pipeline, exerting a direct impact on the effectiveness and dependability of prediction models. Their value extends across various dimensions, encompassing data quality assurance, as well as improved model performance and robustness.

Data cleaning is the process of detecting and correcting flaws, inconsistencies, and missing information in a dataset. Accurate and comprehensive data is essential for machine learning models to effectively learn significant patterns. Erroneous or noisy data can result in models that produce false predictions, as they may absorb inaccurate information throughout their learning process. Data cleaning is essential for ensuring that the models are trained on precise and pertinent information, hence enhancing their validity.

Preprocessing is the process of converting raw data into a format that is appropriate for analysis. This encompasses the process of standardizing numerical features to a uniform scale, converting categorical variables into numerical values, and generating new features that more accurately depict the underlying patterns. Thorough preprocessing guarantees that all features have a suitable impact on the model, avoiding any individual feature from having an excessive influence on the model's predictions. This stage is vital for optimizing the overall performance of the machine learning model.

Preprocessing, which involves managing outliers and maintaining uniform data formatting, is effective in mitigating both model bias and variance. Bias is the discrepancy that arises when a complex real-world situation is simplified using a model, while variance is the model's susceptibility to changes in the training data. Preprocessing aids in mitigating bias and variation by rectifying data inconsistencies and eliminating outliers, resulting in models that exhibit strong generalization capabilities towards novel, unknown data. Striking the right balance is crucial in order to create resilient models that do not suffer from overfitting or underfitting of the training data.

Utilizing clean and preprocessed data results in more streamlined model training. Preprocessing decreases the complexity of the dataset by eliminating irrelevant or redundant information and emphasizing key elements. The reduced data facilitates expedited and more effective training procedures, enabling prompt iterations and enhancements of the model. Effective training is crucial in iterative development contexts, where models are continually enhanced using new data and feedback.

Optimally preprocessed data can enhance the comprehensibility of machine learning models. Clear and standardized features facilitate comprehension of the interactions between variables and the decision-making process of the model. Transparency is essential for stakeholders that require trust and understanding of the model's predictions, particularly in vital fields like healthcare, finance, or operations management.

Data cleaning and preprocessing are essential for guaranteeing the quality, uniformity, and appropriateness of data for machine learning. These procedures improve the performance of the model, decrease bias and variance, enable effective feature engineering, and guarantee the robustness and efficiency of the training process. In the end, these advancements result in forecasts that are more precise, dependable, and easily understood. These predictions are crucial for effectively utilizing machine learning to address practical issues like reducing food waste through demand forecasting and improving restaurant operations.

### 3.4.2 Data Cleaning and Preprocessing Tools

Data cleaning and preprocessing are essential and crucial stages in the preparation of datasets for machine learning. These stages guarantee the accuracy, consistency, and suitable structure of the data, hence improving the performance and dependability of predictive models. Our thesis focused on utilizing Python and its libraries, numpy and pandas, to effectively address the issue of food wastage and improve the operational efficiency of restaurants in Dhaka City. We employed a data-driven strategy that involved machine learning techniques to forecast pizza demand. Python's adaptability, along with the robust features offered by numpy and pandas, made them the most suitable options for managing our varied and intricate datasets.

#### 3.4.2.1 Python Programming Language

Python is a versatile programming language that is frequently used in the fields of data science and machine learning because of its simplicity, clarity, and vast library ecosystem. It enables effortless incorporation of diverse data processing and machine learning libraries, which is crucial for creating advanced models. Python's user-friendly interface and extensive range of resources make it highly ideal for data scientists of all levels, from beginners to experts. Python was the main programming language used in our work. It provided the framework for all data cleaning and preprocessing tasks as well as model training and evaluation. The comprehensive assistance it provided for data manipulation and analysis jobs enabled us to efficiently and successfully manage substantial amounts of data.

#### 3.4.2.2 Numpy

Numpy, also known as Numerical Python, is an essential component of numerical computation in the Python programming language. It offers extensive assistance for handling arrays, matrices, and many mathematical operations. The architecture of this product prioritizes both performance and user-friendliness, making it well-suited for managing extensive datasets and executing complex numerical calculations.

An important benefit of numpy is its effective array operations. Numpy arrays differ from Python lists in that they are stored in contiguous memory blocks, enabling faster access and manipulation. Efficiency is essential when dealing with extensive datasets, as it greatly decreases calculation time. In our project, we utilized numpy arrays to manage and standardize numerical features. Normalisation is a crucial preprocessing technique that rescales numerical features to a standardized range, such as [0, 1] or [-1, 1], to ensure that all features have an equal impact on the model's learning process. This phase ensures that no individual characteristic has a disproportionate impact on the model's outcome as a result of its magnitude.

Numpy is very proficient in managing missing values, which is a prevalent problem in datasets found in real-world scenarios. Missing values can occur due to different factors, such as poor data entry or data corruption. For our research, we employed numpy methods to detect and fill in missing values, thereby preserving the integrity of our datasets. Methods such as mean or median imputation, which involve replacing missing values with the average or middle value of the corresponding attribute, aid in preserving the coherence of the data. Numpy offers extensive capabilities for executing intricate mathematical computations, including the calculation of temperature ranges, means, and standard deviations. These calculations are crucial for comprehending the fundamental patterns within the data.

In addition, numpy facilitates vectorized operations, enabling the concurrent processing of numerous data points. This feature is especially advantageous in the field of machine learning, where there is a requirement to carry out operations on extensive datasets quickly and effectively. In our project, we employed vectorized operations to efficiently apply transformations and calculations to full arrays, resulting in a notable acceleration of the preprocessing procedure. Numpy's optimized speed guarantees that data processing activities are executed significantly faster than typical Python loops.

#### 3.4.2.3 Pandas

Pandas is a sophisticated data manipulation library that is based on numpy. It offers important data structures such as Series and DataFrame, which are crucial for managing structured data. The architecture of the software prioritizes user-friendly features and robust data manipulation capabilities, making it highly popular among data scientists and analysts.

A notable attribute of pandas is its adeptness in efficiently managing missing data. The library provides a variety of functions, including fillna() and dropna(), for handling missing values. For our project, we employed these functions to preprocess the dataset by either filling in missing values or eliminating entries with substantial gaps. Ensuring the completeness and trustworthiness of the dataset is crucial since it directly affects the accuracy of the predictive models.

Pandas offers a wide range of tools for manipulating data and creating new features. The task of combining datasets from many sources, including sales data, weather data, and holiday data, was efficiently managed utilizing the pandas.merge() function in our project. This function facilitated the integration of these datasets by utilizing a shared identifier, such as the date, to ensure that all pertinent information was properly matched and prepared for analysis. Pandas methods such as apply(), groupby(), and pd.get_dummies() aided the process of feature engineering, which is a vital step in preparing data for machine learning. As an illustration, we developed additional functionalities to measure the impact of vacations on pizza sales. This involved creating binary indicators for holidays and categorizing different weather conditions into qualitative descriptors.

Furthermore, pandas has exceptional proficiency in data filtering and aggregation, which are crucial for upholding data consistency and precision. For our project, we employed the pandas library to selectively extract data for certain time intervals, consolidate sales data, and convert date formats. These responsibilities guaranteed that the data was both free from errors and organised in a manner that optimized the model's capacity to acquire knowledge from it.

Pandas' connection with other libraries, like as numpy and matplotlib, amplifies its functionality. This integration enables a smooth workflow, where data manipulation, numerical computation, and visualization may all be carried out within a unified environment. We utilized this integration in our project to visually represent trends and patterns in the data, which offered useful insights that informed our feature engineering and model building procedures.

Ultimately, numpy and pandas are essential tools for data cleansing and preprocessing in our project. The efficient array operations and robust mathematical functions of numpy, along with the extensive data manipulation capabilities of pandas, offered a comprehensive foundation for preparing our datasets. These technologies guaranteed the accuracy, consistency, and suitability of our data for machine learning. As a result, we were able to create a dependable demand forecasting model that effectively reduces food wastage and improves restaurant operations in Dhaka City.

**3.4.3 Data Cleaning and Preprocessing in Sales Data**

The data cleaning and preparation procedure is essential to ensure the integrity and uniformity of the information utilized in our machine learning model for predicting pizza demand. For our research, we gathered sales data from ten distinct eateries in Dhaka City, each exhibiting variations in both size and popularity. To maintain the integrity and trustworthiness of the dataset, a meticulous data cleaning and preparation technique was required due to the diverse nature of the data. We employed Python's numpy and pandas modules to execute these operations owing to their robust data manipulation capabilities.

At first, we consolidated the ten separate datasets into one unified and extensive dataset. The process of unifying the dataset was crucial in order to establish a consistent and integrated dataset that could be examined and utilised for the purpose of training machine learning models. During the data cleansing phase, we detected other issues that required attention. An important problem arose when it was discovered that around 10% of the data in the pizza_sizes column was absent. The absence of data can have a significant negative effect on the effectiveness of machine learning models, resulting in imprecise forecasts. In addition, the pizza_sales column comprised a combination of numerical and qualitative values, necessitating standardisation to ensure uniform analysis.

In order to handle the presence of several data types in the pizza_sales column, we initially transformed the category values into numerical values. This change was essential to ensure the accurate processing of the data by the machine learning algorithms. Converting categorical variables into numerical values is a common preprocessing step that simplifies the mathematical calculations needed by machine learning models.

Next, we decided to use the column's mode value in place of the missing values in the pizza_sizes column. The mode, which refers to the value that occurs most frequently, was chosen as it is a reliable measure for replacing missing categorical data. This strategy aids in preserving the distribution of the data and mitigating the potential for creating bias. Upon substituting the null values with the mode value, we generated a Kernel Density Estimate (KDE) graph to visually represent the distribution of the pizza_sizes data both before and after the imputation process. The KDE graph indicated that the imputation was effective, as the overall distribution of the data remained unchanged.

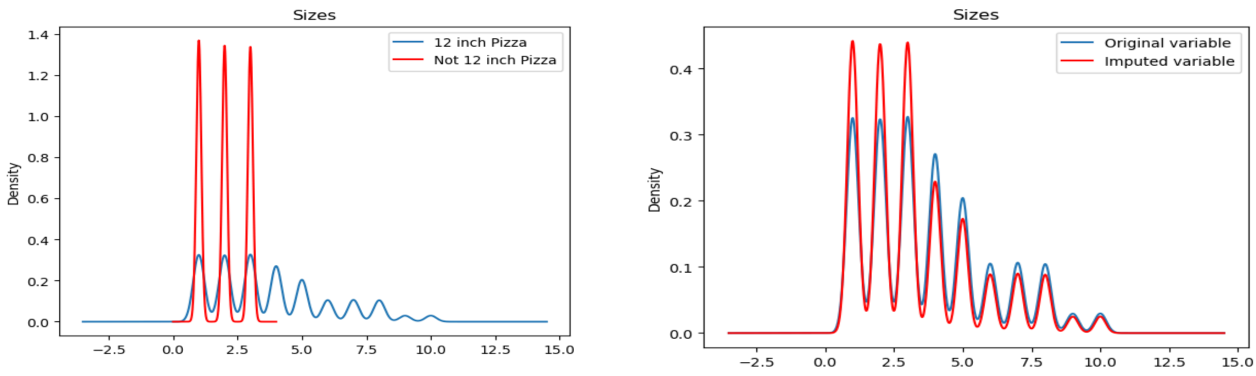



fig: Before and After Handling Null Values

During the data preprocessing phase, we applied standardisation techniques to the dataset in order to achieve consistency and uniformity among all establishments. It was necessary to establish uniform range values for all the restaurant data, as this is essential for removing discrepancies caused by variations in restaurant size and sales volume. Standardisation guarantees that the model treats the data from each restaurant identically, avoiding any individual restaurant from having an excessive impact on the model's predictions. In addition, we established a coherent time period for the dataset, starting on January 3, 2022, and ending on January 19, 2024. The selected time period was intended to guarantee that the data encompasses a wide-ranging chronology, enabling the model to acquire knowledge from a varied range of circumstances and patterns.

We made sure that our dataset was reliable, consistent, and appropriate for training machine learning models by carefully cleaning and preparing it. Python's numpy and pandas packages facilitated rapid handling of intricate data manipulation tasks. These procedures were crucial in the development of a dependable demand forecasting model with the goal of minimising food wastage and improving restaurant operations in Dhaka City. By meticulously cleaning and preparing the data, we established a strong basis for the following stages of our analysis, guaranteeing that the insights obtained from the model would be both reliable and practical.

## 3.5 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an essential stage in the machine learning (ML) process, which aims to comprehend the attributes and organisation of the dataset prior to model creation. Exploratory Data Analysis (EDA) encompasses the process of condensing the primary attributes of the data, graphically representing its distributions, recognising patterns, and discerning connections between variables. EDA is essential for multiple reasons. Firstly, it aids in identifying and managing missing values, outliers, and inconsistencies, hence maintaining the quality of the data. Furthermore, it offers valuable understanding of the fundamental data patterns, enabling informed decision-making on feature selection and engineering. Furthermore, Exploratory Data Analysis (EDA) assists in the selection of suitable Machine Learning (ML) models and evaluation metrics by uncovering the characteristics of the data and any obstacles. EDA ultimately improves the interpretability of results and enhances the performance and reliability of ML models by providing guidance for data preprocessing and modelling decisions through a comprehensive understanding of the data.

### 3.5.1 Visualization Tools

#### 3.5.1.1 Matplotlib:

Matplotlib is a versatile toolkit that allows for the creation of static visualisations of high quality, suitable for publication purposes. The software provides a diverse range of charting features that allow for intricate customisation of plots, such as line plots, bar charts, histograms, and scatter plots. Matplotlib played a crucial role in visualising the weekly sales data for each restaurant in our project. By utilising this method, we were able to visually represent the fluctuations in sales over a period of time and discern recurring patterns, such as surges coinciding with national holidays and festivals. The library's wide range of customisation features allowed us to modify the plots in order to emphasise specific trends and patterns, hence enhancing the comprehensibility and practicality of the data. Moreover, the quick handling of big datasets by matplotlib was essential for our time series study, guaranteeing that the visualizations remained lucid and informative despite the abundance of data points.

#### 3.5.1.2 Seaborn:

Seaborn is a statistical data visualisation library that is constructed on the foundation of matplotlib. The software offers a sophisticated interface for creating visually appealing and instructive statistical visuals. Seaborn is renowned for its capacity to generate intricate visualisations using minimal code, owing to its seamless interaction with pandas data structures and its emphasis on the aesthetics of data visualisation. For our research, we utilised seaborn to produce Kernel Density Estimate (KDE) plots and heatmaps. We utilised KDE plots to visually represent the distribution of pizza sizes both before and after addressing missing values, hence validating the efficacy of our data cleansing techniques. Heatmaps, however, were employed to investigate the relationships between various attributes, such as temperature and sales volumes. The utilisation of seaborn's pre-existing themes and colour palettes elevated the visual attractiveness of our plots, resulting in not only instructive but also visually pleasant representations.

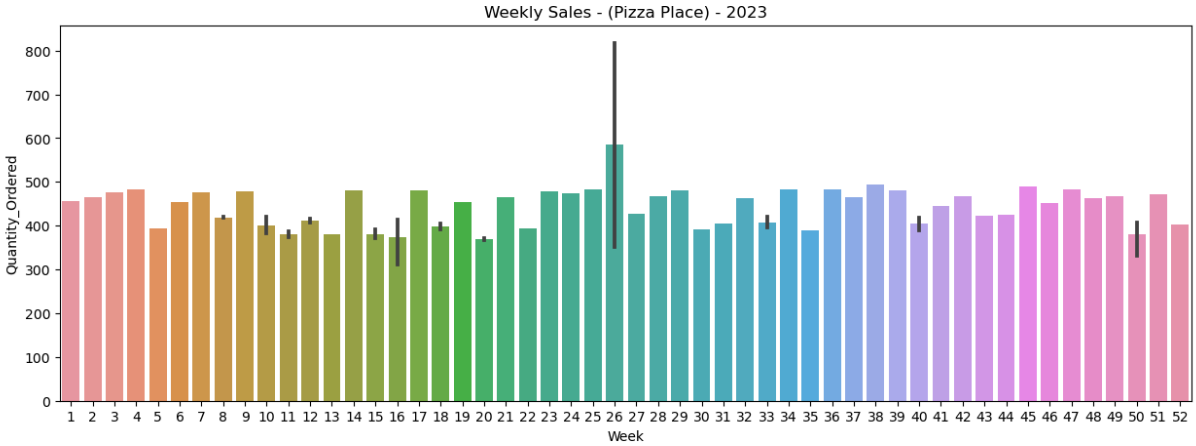
### 3.5.2 Exploratory Data Analysis on Sales Data

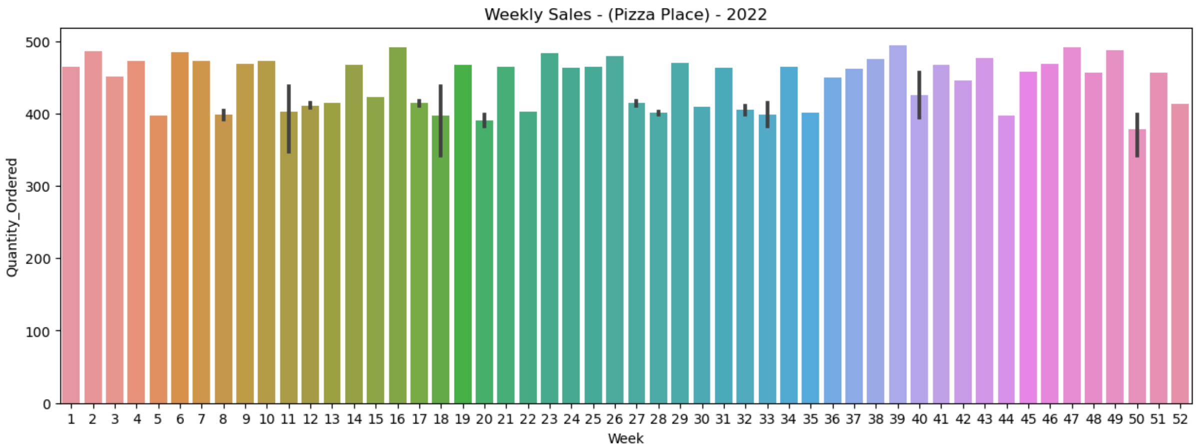
Initially, our main emphasis was on analyzing time series data to reveal patterns such as seasonality, trends, and other temporal variations in the sales data. We initiated the process by graphing the weekly sales data for each restaurant using bar charts. This visualization facilitated the observation of temporal variations in sales, unveiling significant crests and troughs. An important finding from this investigation was the notable surge in sales around national holidays and festivals. These time intervals regularly exhibited elevated sales volumes, indicating a robust link between public holidays and heightened pizza demand. This discovery emphasized the significance of integrating holiday data into our forecasting model in order to accurately reflect the anticipated surges in demand. We also looked at how daily weather patterns affected pizza sales. we discovered a connection between them.

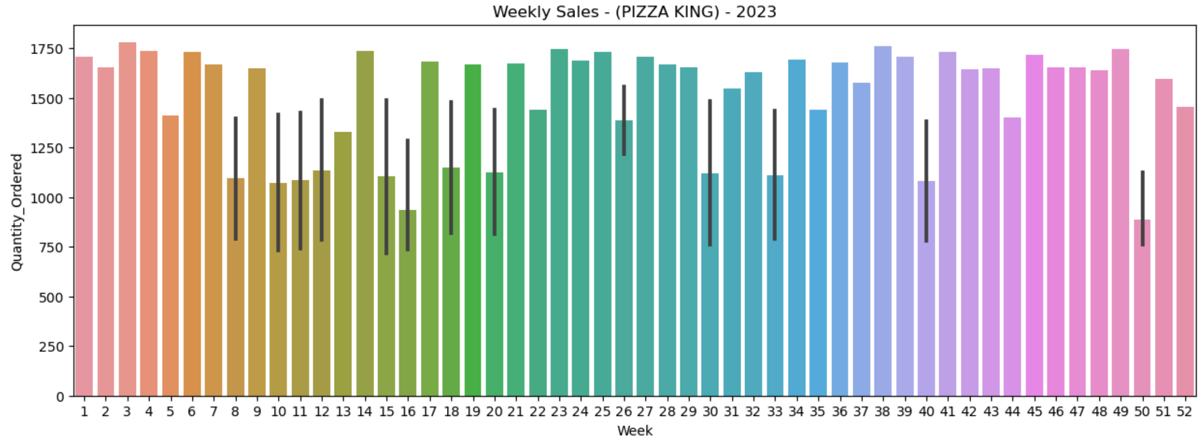
Furthermore, a seasonal trend showing lower sales volumes in the summer than in other seasons was revealed by our time series study. We created a number of bar charts that compared sales quantities during various seasons in order to investigate this further. This seasonal analysis verified that summer sales were, in fact, lower, most likely as a result of the hot weather discouraging customers from ordering pizza. The seasonal trend was a crucial component that our model took into account since it showed a consistent pattern that could be used to improve forecasting.

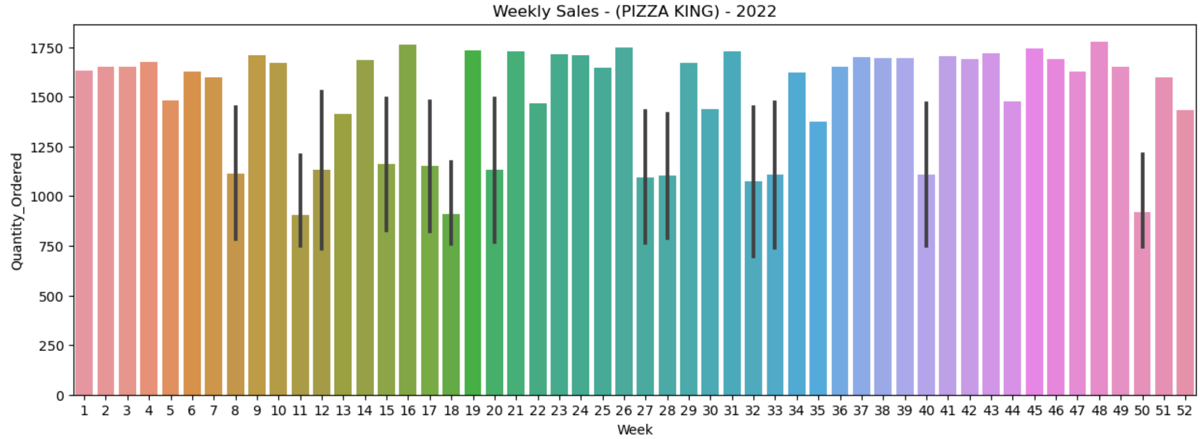
These analysis and visualizations provided us with important insights that we used in our feature engineering process. We were able to develop features that successfully captured these trends by determining the considerable influence that seasons, holidays, and weather patterns had on sales. As features in our machine learning models, we added binary indicators for seasonal flags, holidays, and temperature ranges. These characteristics improved the prediction accuracy of our models by allowing them to take into consideration the changes in sales caused by these outside influences.

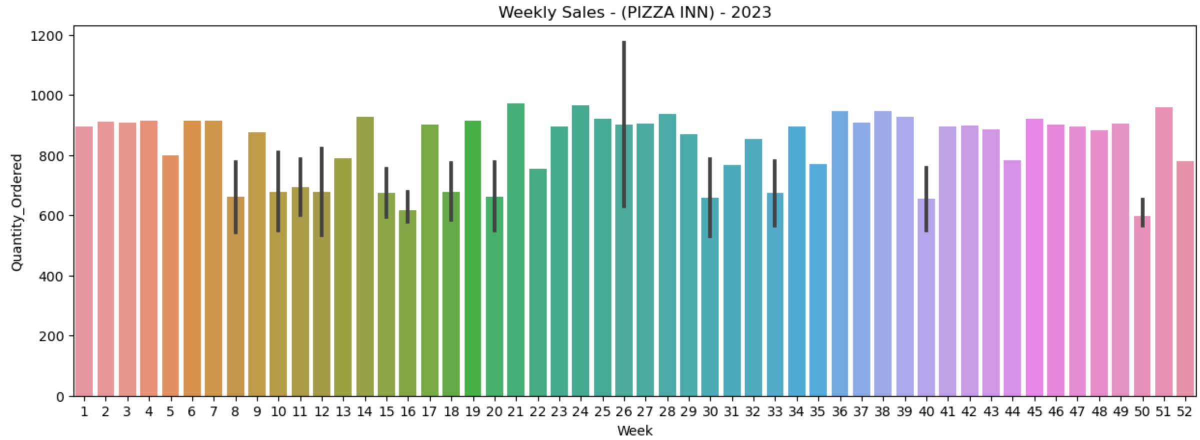
The visualization Graphs are shown below:

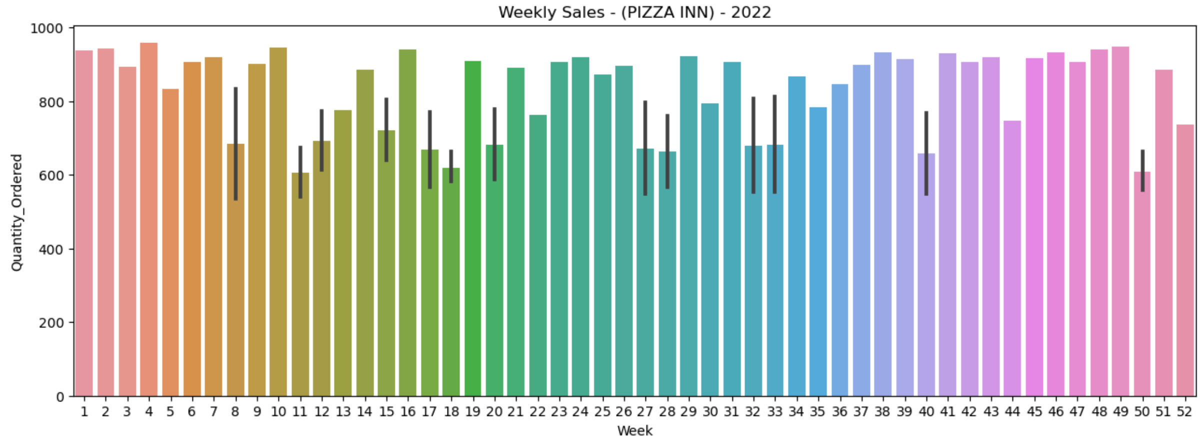


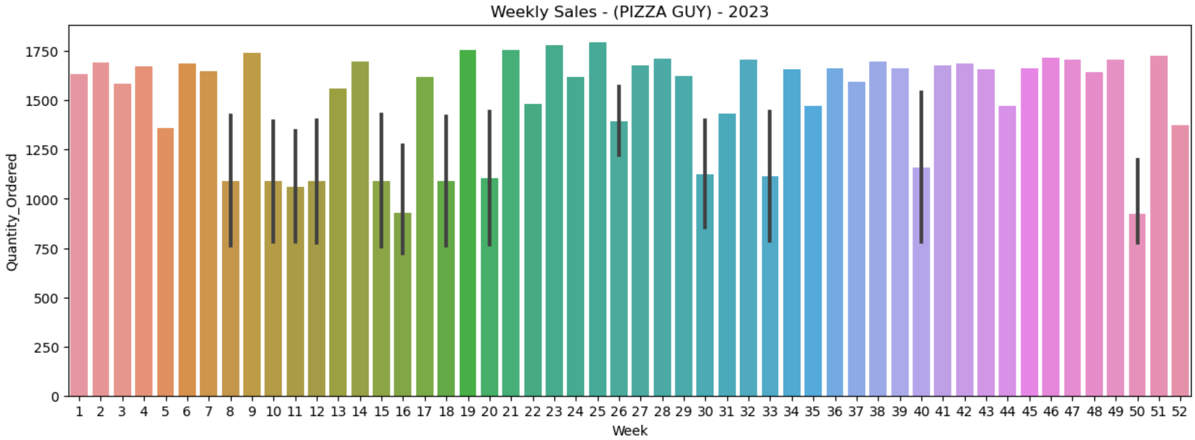



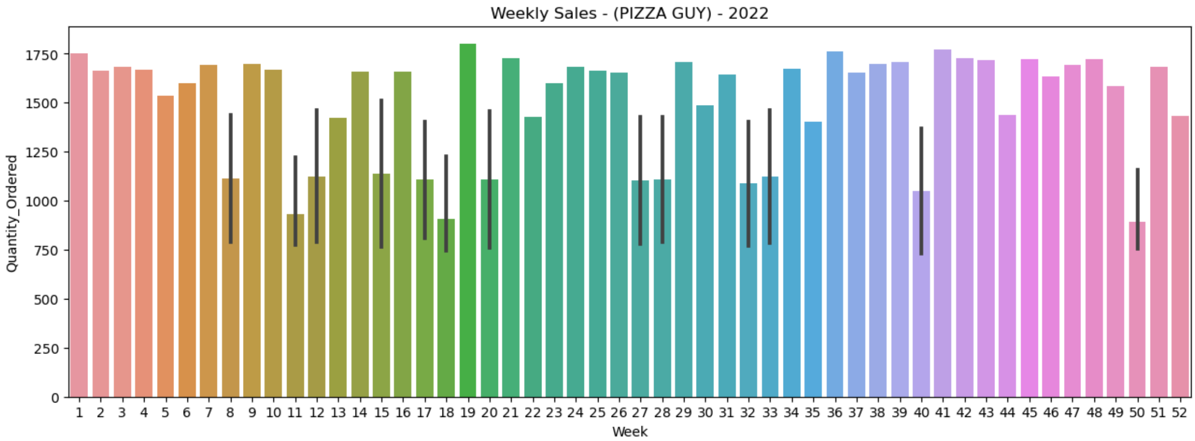

Weekly Sales - (PIZZA KING) - 2023
Quantity_Ordered
Week

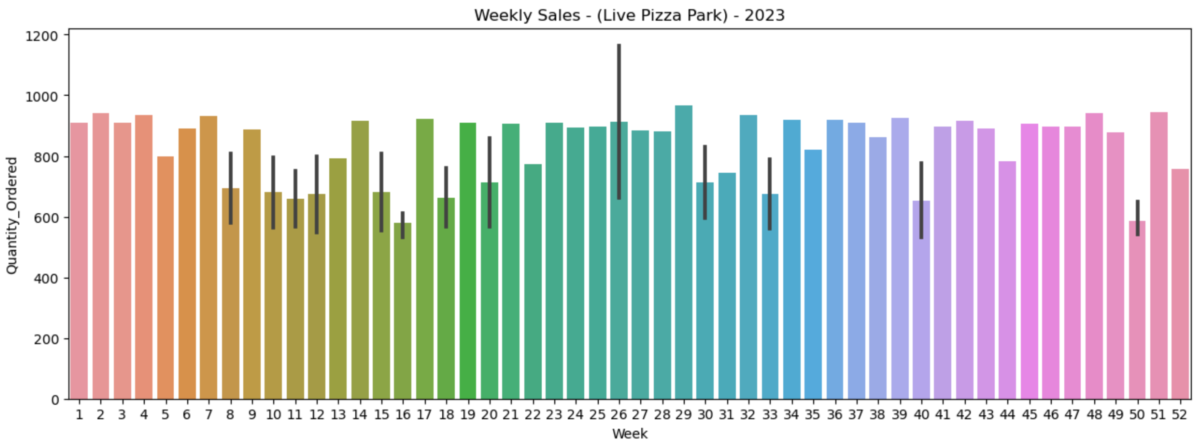

Weekly Sales - (PIZZA KING) - 2022
Quantity_Ordered
Week

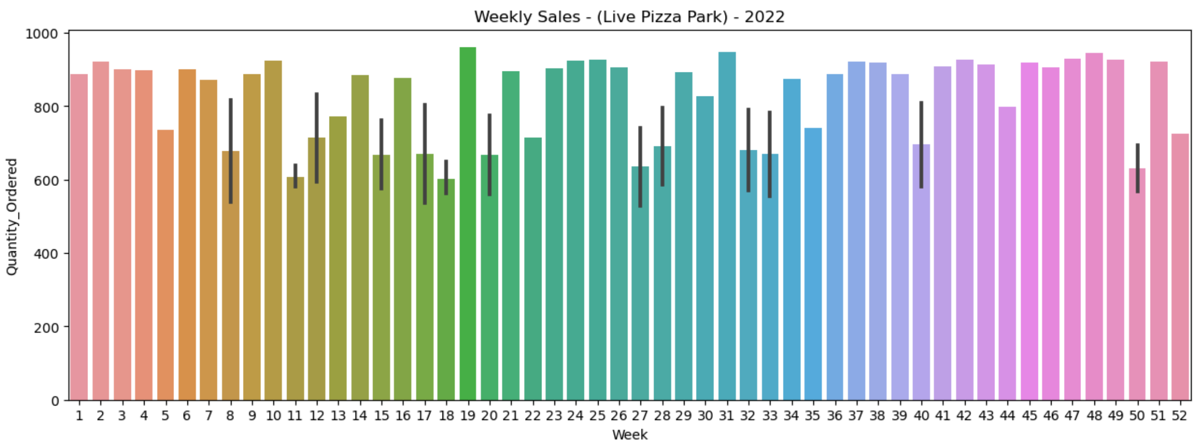

Weekly Sales - (PIZZA INN) - 2023
Quantity_Ordered
Week

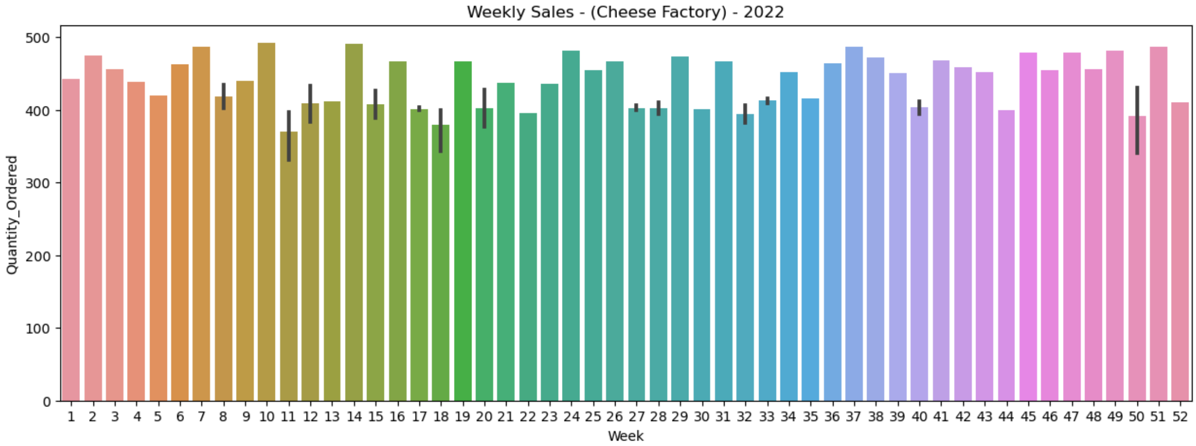

Weekly Sales - (PIZZA INN) - 2022
Quantity_Ordered
Week

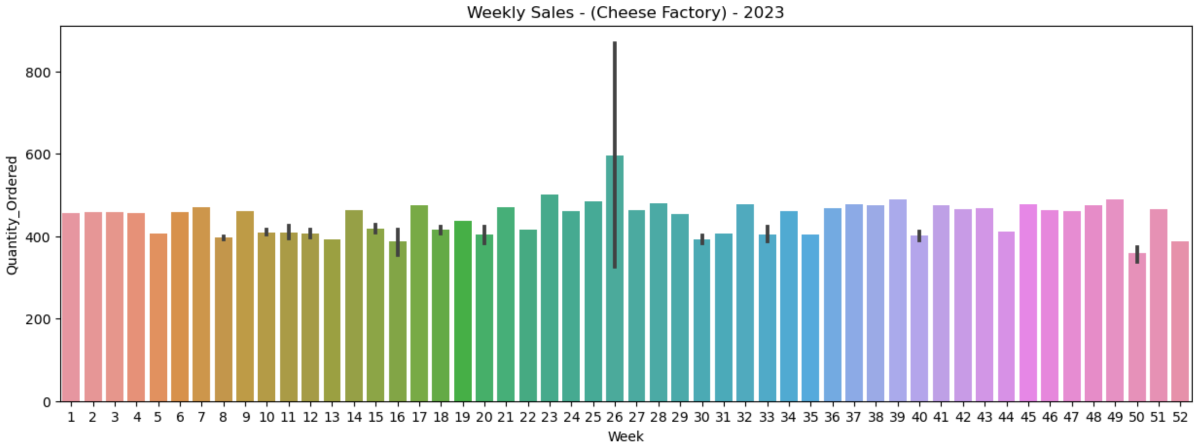

Weekly Sales - (PIZZA GUY) - 2023
Quantity_Ordered
Week

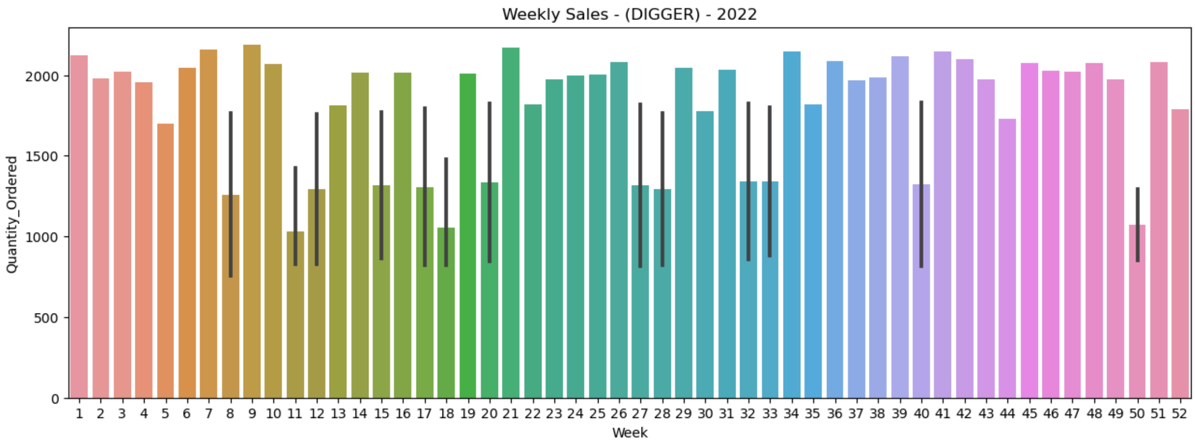

Weekly Sales - (PIZZA GUY) - 2022
Quantity_Ordered
Week

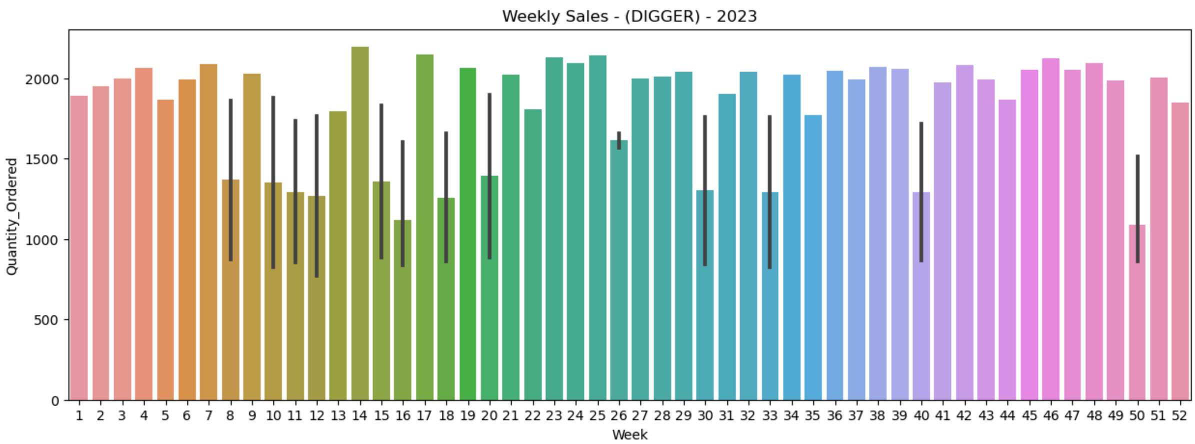

Weekly Sales - (Live Pizza Park) - 2023
Quantity_Ordered
Week

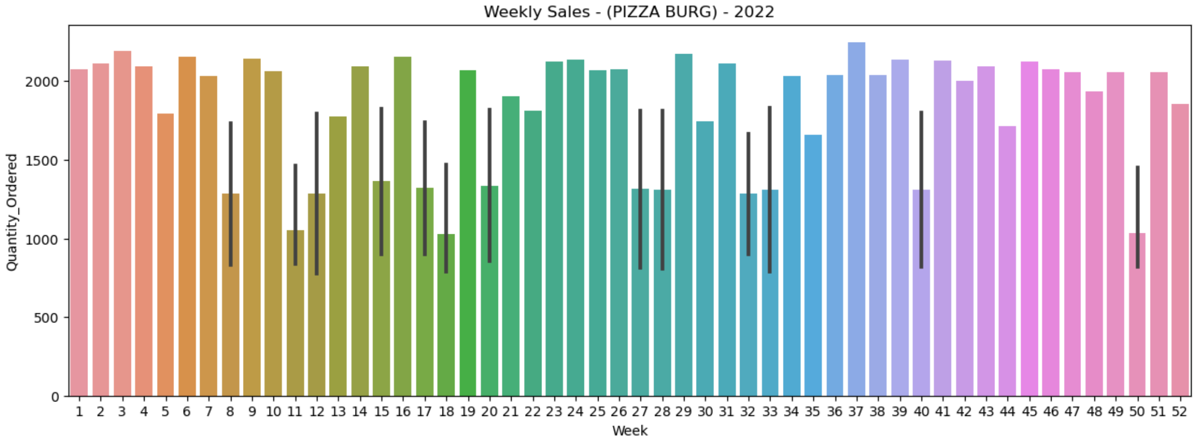

Weekly Sales - (Live Pizza Park) - 2022
Quantity_Ordered
Week

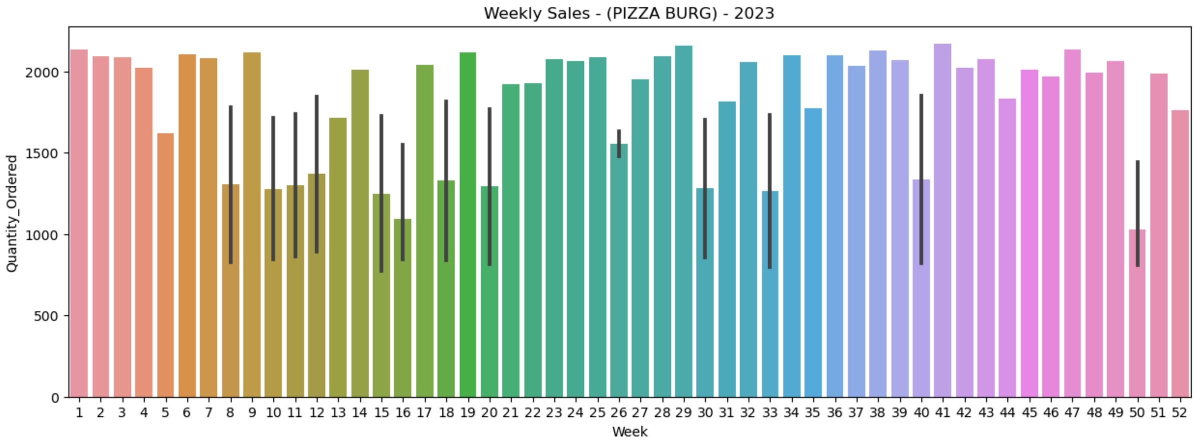

Weekly Sales - (Cheese Factory) - 2022
Quantity_Ordered
Week

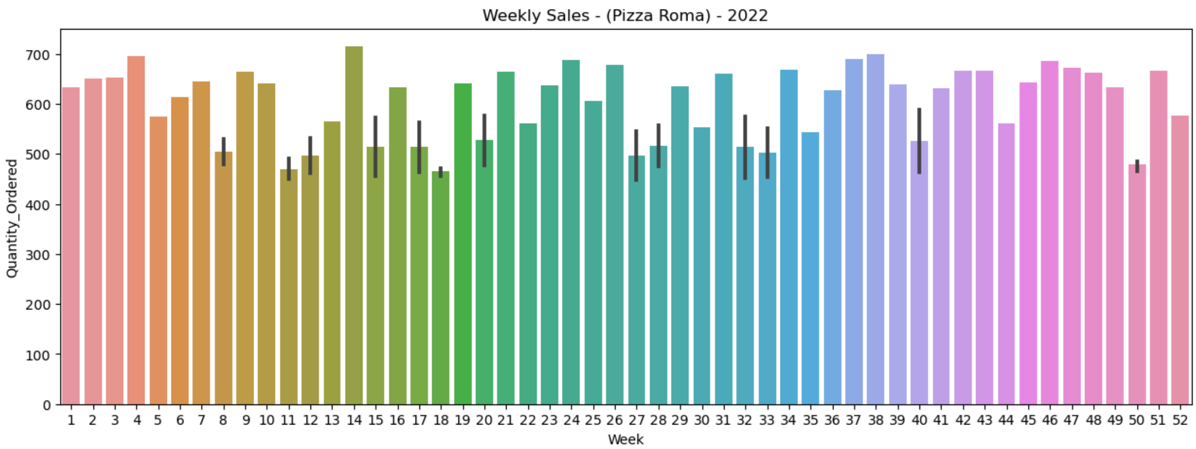

Weekly Sales - (Cheese Factory) - 2023
Quantity_Ordered
Week

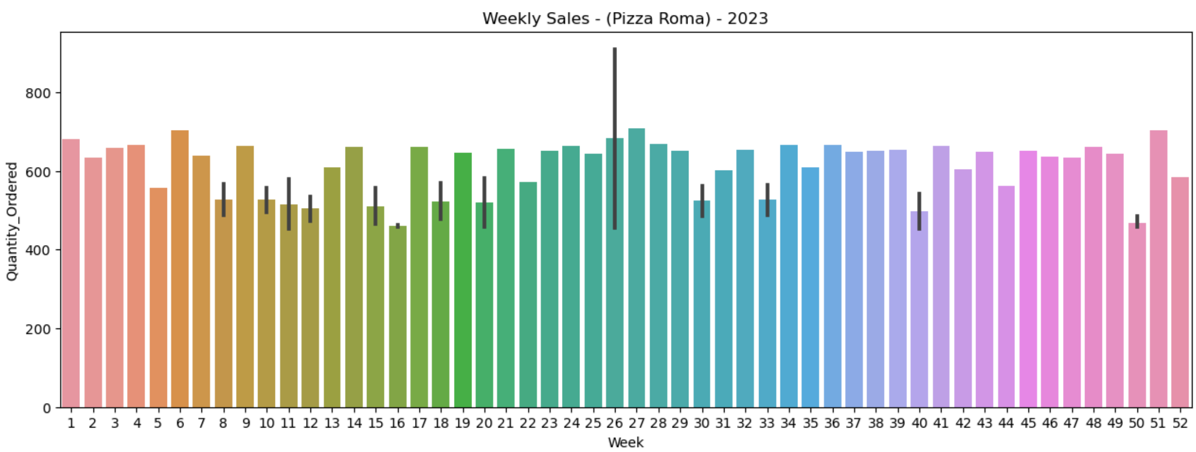

Weekly Sales - (DIGGER) - 2022
Quantity_Ordered
Week


Weekly Sales - (DIGGER) - 2023
Quantity_Ordered
Week


Weekly Sales - (PIZZA BURG) - 2022
Quantity_Ordered
Week


Weekly Sales - (PIZZA BURG) - 2023
Quantity_Ordered
Week


Weekly Sales - (Pizza Roma) - 2022
Quantity_Ordered
Week


Weekly Sales - (Pizza Roma) - 2023
Quantity_Ordered
Week

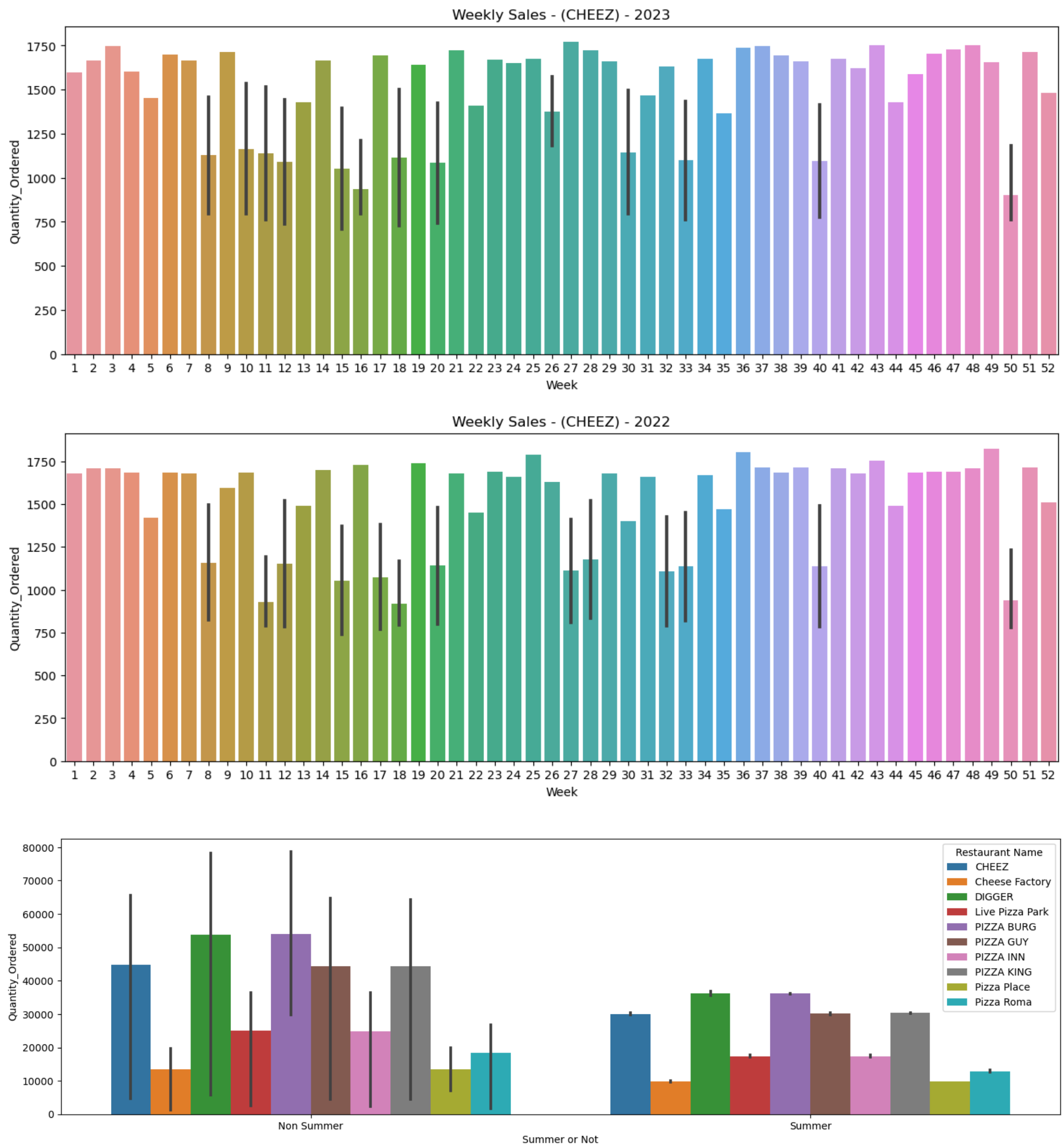



Fig: Sales vs Season

Ultimately, our exploratory data analysis (EDA) yielded a thorough comprehension of the temporal and contextual variables that impact pizza sales. Through a thorough analysis of time series data and the use of bar charts to visualize sales patterns, we were able to identify important trends and linkages. This information was then used to construct a strong demand forecasting model. The knowledge gained from these insights played a crucial role in successfully accomplishing our objective of minimizing food waste and improving restaurant efficiency. This ensured that the model was adequately prepared to accurately forecast future sales by analyzing past data and external factors.

### 3.6 Feature Engineering

Feature engineering involves the conversion of raw data into significant features that might enhance the effectiveness of machine learning models. The process entails the careful selection, construction, and adjustment of characteristics in order to capture relevant information and trends within the data, hence improving the model's capacity to generate precise forecasts. Feature engineering plays a vital role in machine learning as it directly influences the performance, interpretability, and generalization of models. Robust features can enhance models' ability to derive significant insights from intricate data, manage non-linear relationships and interactions, decrease dimensionality, and enhance computational efficiency. In the end, proficient feature engineering enables models to acquire knowledge more effectively from data, resulting in improved predictions and insights.

We developed a number of crucial features during the feature engineering process that greatly increased our machine learning model's capacity to estimate pizza demand. In the feature construction phase of this procedure, we created additional features like "Festivals," "Summer or Not," and "Daily Temperature." Our exploratory data analysis revealed high relationships between these characteristics and daily pizza sales, which led us to identify these qualities as critical. Sales, for example, had a tendency to rise during festivals and fall throughout the summer. The daily temperature also had a discernible effect on sales volumes. By building these features, we were able to identify these significant trends in the data.

Finding the most pertinent properties for our model was the goal of the following stage, feature selection. In our situation, the output feature was the daily sales of pizza, and this required calculating the correlation between each input characteristic and the output feature. We made sure that our model would be based on the most informative parts of the data by concentrating on features with strong correlation, which improved the model's accuracy and resilience. By removing less useful features and keeping the ones that had the highest predictive value, this phase helped us to improve the quality of our feature set.

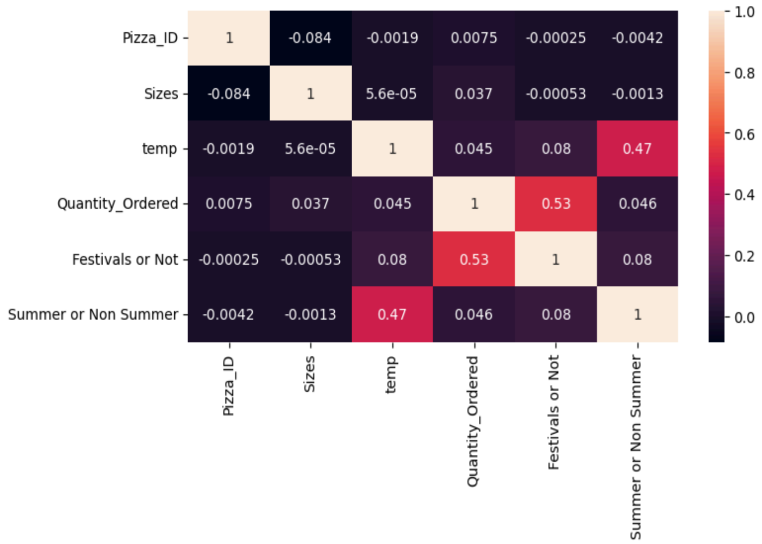



Fig: Correlation Matrix

From this figure we can see that Pizza_ID has a very low correlation coefficient with respect to Quantity_Ordered. So we removed this feature from our dataset.

The last step of our feature engineering approach involved feature encoding. This phase was essential for transforming category information into a numerical format that could be analyzed by our machine learning algorithms. We utilized binary encoding for binary categorical features such as "Festivals or Not" and "Summer or Non-Summer." This straightforward yet efficient technique assigns a binary value of either 0 or 1 to each category, hence simplifying the interpretation and utilization of the data in the model. In addition, we utilized OneHotEncoding for the "Restaurant Name" feature, which is also a categorical variable. This technique generates a novel binary column for every category, guaranteeing that the model does not make any assumptions about the hierarchical relationship between the names of the restaurants. By meticulously encoding these characteristics, we maintained the categorical information while adapting it for numerical analysis, thus improving the model's performance.

In summary, our feature engineering approach, which included the creation, choice, and transformation of features, played a crucial role in preparing our dataset for successful machine learning. By adopting a thorough strategy, we were able to effectively capture the crucial elements of the data. This ensured that our model could make precise predictions about pizza demand, leading to a reduction in food wastage and the optimization of restaurant operations.

## 3.7 Machine Learning Models

In our work, we used six machine learning models. Those are Decision Tree Regressor, Random Forest Regressor, KNN Regressor, Gradient Boosting Regressor, XgBoost Regressor and AdaBoost Regressor. The algorithms of these models are mentioned below.

### 3.7.1 Decision Tree Regressor

A non-parametric method for handling regression issues in supervised learning is the Decision Tree Regressor. A decision model that resembles a tree is created by first dividing the data into smaller groups based on the input feature values. Each leaf node in the tree represents a forecast of a continuous value, each internal node in the tree represents a decision made using an attribute, and each branch in the tree shows the outcome of that decision. Decision trees are very intuitive and simple to understand, which makes them an excellent tool for understanding the underlying structure of the data. Feature scaling is not required for these models to process numerical and categorical input. However, they have a propensity to overfit, which means that as the tree becomes overly complex, they may inadvertently record irrelevant information rather than the actual pattern in the data. Many techniques, including pruning, setting minimum sample split sizes, and limiting the tree's depth, are frequently used to alleviate the overfitting problem.

### 3.7.2 Random Forest Regressor

By combining many decision trees, the Random Forest Regressor is an ensemble learning strategy that improves prediction accuracy. In the training phase, a "forest" of decision trees is created by the system. In regression tasks, it determines the mean of the tree's predictions. Using a method known as bootstrap aggregation, also referred to as bagging, each tree in the forest is trained using a random subset of the data and characteristics. By preventing individual trees from being unduly dependent on any one portion of the training data, this technique reduces overfitting and improves generalizability. Because of their great accuracy, noise resistance, and capacity to handle large, highly dimensional datasets, Random Forests are incredibly effective for both classification and regression applications.

### 3.7.3 KNN Regressor

The K-Nearest Neighbors (KNN) Regressor is a straightforward technique for instance-based learning. It predicts by taking the average target values of the k-nearest neighbors in the feature space. The value of k, which represents the number of neighbors, is a hyperparameter that governs the level of complexity of the model. KNN is a versatile and non-parametric method because it does not rely on any assumptions about the underlying data distribution. The method is characterized by its high level of intuitiveness and ease of implementation. However, it might be computationally demanding, particularly for extensive datasets, due to the need for calculating point distances during prediction. Moreover, the performance of KNN can be greatly influenced by the selection of the distance metric and the existence of irrelevant or redundant features, typically requiring feature scaling and meticulous feature selection.

### 3.7.4 Gradient Boosting Regressor

The Gradient Boosting Regressor is an ensemble method that constructs models in a sequential manner, with each subsequent model rectifying the mistakes produced by the preceding models. The process involves iteratively improving a loss function using gradient descent. At each iteration, a fresh tree is incorporated into the ensemble with the aim of minimizing the residual error of the combined model. Gradient Boosting is capable of generating highly precise models, particularly for intricate datasets. Nevertheless, it is more susceptible to overfitting compared to other ensemble methods and can be computationally demanding, necessitating meticulous adjustment of hyperparameters such as the learning rate, number of trees, and tree depth. Notwithstanding these difficulties, Gradient Boosting is extensively utilized in machine learning contests and practical applications because of its exceptional prediction capability.

### 3.7.5 XgBoost Regressor

The XGBoost Regressor is a sophisticated implementation of gradient boosting that is optimized for maximum efficiency, flexibility, and portability. XGBoost distinguishes itself by its rapidity and effectiveness, which can be attributed to its optimization approaches, including tree pruning, parallelized tree creation, and regularization. These characteristics aid in mitigating overfitting and enhancing the accuracy of the model. XGBoost exhibits robust handling of missing information and offers extensive support for diverse goal functions, rendering it very adaptable for various regression workloads. Its scalability and capacity to handle extensive datasets have established it as a favored option in machine learning contests and business applications. Nevertheless, the intricacy of the model can pose difficulties in terms of fine-tuning and comprehension when compared to more straightforward models.

### 3.7.6 AdaBoostRegressor

The AdaBoostRegressor, also known as Adaptive Boosting, is an ensemble technique that combines several weak learners, usually decision trees, to create a powerful prediction model. The algorithm operates by iteratively modifying the weights of training cases, with a particular emphasis on the errors made by prior learners. During each iteration, a fresh weak learner is incorporated into the ensemble, and its predictions are merged with the existing ensemble using weights. The main advantage of AdaBoost is its capacity to enhance the accuracy of weak models and its resistance to overfitting, especially when used in conjunction with decision trees. Nevertheless, this model is susceptible to noisy data and outliers, as it assigns greater importance to cases that are difficult to forecast. Nevertheless, AdaBoost continues to be a potent technique for enhancing the accuracy of models.

### 3.7.7 Stacking Regressor

The Stacking Regressor is a method of ensemble learning that integrates numerous regression models by training a meta-model to aggregate their predictions. In a conventional stacking configuration, many foundational models (such as decision trees, linear models, or neural networks) are trained on the identical dataset, and their forecasts are utilized as inputs for the meta-model, which is typically a simple linear regression or another robust model. This strategy exploits the advantages of many algorithms, which has the ability to encompass a broader spectrum of patterns in the data. Stacking can greatly enhance forecast accuracy by minimizing both the variance and bias of the combined model. Nevertheless, the process necessitates meticulous selection and fine-tuning of both the foundational models and the meta-model. Additionally, it can be computationally demanding as it involves training several models.

### 3.7.8 Voting Regressor

The Voting Regressor is a technique that aggregates the predictions of numerous regression models to get a unified output. Contrary to stacking, which involves a meta-model learning how to merge predictions, the voting regressor simply calculates the average of its constituent models' predictions. There are two methods to accomplish this: averaging (hard voting) or weighting each model's prediction according to its performance (soft voting). The implementation of the Voting Regressor is simple and can enhance the resilience and precision of forecasts by utilizing the variety of distinct models. Ensemble learning mitigates the danger of overfitting commonly observed in individual models and is capable of effectively addressing various regression challenges. However, it may not exhibit the same level of performance as more advanced ensemble approaches such as stacking, which has the ability to intelligently integrate predictions.

### 3.8 Model Training and Primary Selection

The primary selection of machine learning models is critical for ensuring that only the most promising models are developed and deployed. Using K-fold cross-validation, we divided our training data into training and validation sets for our study. By ensuring that the model is tested across several data subsets, this approach improves the model's robustness and generalizability.

Each model was first trained and validated with the default parameters that the corresponding algorithms supplied. A number of models were used, including the following: XgBoost, AdaBoost, Gradient Boosting, K-Nearest Neighbours (KNN) Regressor, Decision Tree, Random Forest Regression. The dataset was split into K subsets for the K-fold cross-validation process. The model was then trained K times, utilizing the remaining K-1 subsets as the training set and a different subset as the validation set each time. This method offers a more accurate evaluation of the model's performance while reducing the possibility of overfitting.

Table 02: MODEL TRAINING & PRIMARY SELECTION

| MODEL NAME | R2 SCORE | MSE | MAE | MAPE |
|---|---|---|---|---|
| **Decision Tree** | **0.819564** | **409.939534** | **13.611507** | **0.246625** |
| **Random Forest** | **0.751296** | **583.284424** | **13.425933** | **0.245355** |
| **KNN** | **0.570619** | **975.525975** | **16.985130** | **0.268526** |
| **XgBoost** | **0.836504** | **371.452998** | 13.246496 | **0.242591** |
| **AdaBoost** | **0.679716** | **727.666040** | **19.280404** | **0.355446** |
| **Gradient Boosting** | **0.797103** | **460.970263** | **14.283177** | **0.266659** |

We chose the four models that showed the best accuracy and resilience based on assessment measures including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). After being shortlisted, these models underwent additional testing and analysis to make sure they performed as well as possible for our demand forecasting objective. The development of a dependable and efficient forecasting model that may drastically cut down on food waste and improve restaurant operations requires this stringent selection procedure.

## 3.9 Hyper Parameter Tuning and Optimized Model Training

### 3.9.1 Hyper Parameter Tuning

Hyperparameter tuning is an essential and crucial stage in the process of developing a machine learning model. It has a substantial impact on the performance and efficacy of the final model. Model parameters are learned from the input during training, whereas hyperparameters are predetermined before the learning process starts and dictate the behaviour of the training algorithm. Hyperparameters encompass several parameters that influence the behavior and performance of machine learning algorithms. Some examples of hyperparameters include the learning rate in gradient-based algorithms, the number of trees in ensemble methods such as Random Forest and Gradient Boosting, the depth of trees, and the number of nearest neighbors in K-Nearest Neighbours (KNN).

The main objective of hyperparameter tuning is to identify the ideal combination of hyperparameters that maximizes the model's performance on the validation set. This procedure frequently entails a compromise between bias and variance, wherein a finely adjusted model achieves equilibrium, diminishing both underfitting and overfitting. Grid search and random search are two frequently used techniques for hyperparameter tuning.

Grid search entails defining a range of potential values for each hyperparameter and methodically testing every conceivable combination to get the optimal set that yields the highest performance. Although grid search is comprehensive, it can be computationally demanding and time-consuming, particularly for models with numerous hyperparameters or when the search space is extensive. Nevertheless, grid search continues to be a favored option due to its simplicity and comprehensive methodology.

Random search, however, employs a random selection process to choose a combination of hyperparameters for evaluation within the given ranges. This approach is frequently more effective than grid search, particularly when there is a large number of hyperparameters. Studies have demonstrated that random search is capable of discovering optimal hyperparameter values in a faster and sometimes more efficient manner compared to grid search. This is achieved by investigating a wider range of possibilities inside the search space.

Hyperparameter tuning involves the utilization of advanced methods such as Bayesian optimisation and evolutionary algorithms. Bayesian optimization constructs a statistical model of the goal function and utilizes this model to choose the most favorable hyperparameters for assessment. It exhibits greater sample efficiency compared to grid and random search, rendering it appropriate for applications that involve costly evaluations. Genetic algorithms imitate the mechanism of natural selection by progressively refining a group of potential hyperparameter sets through actions such as mutation, crossover, and selection.

Hyperparameter optimization is crucial for intricate models such as deep neural networks and ensemble methods, as the selection of hyperparameters can significantly impact the model's performance. In deep learning, hyperparameters such as the learning rate, batch size, and the number of layers and units in each layer have a substantial impact on the model's convergence and accuracy. In ensemble approaches such as Random Forest and Gradient Boosting, the optimal values for the number of trees, maximum tree depth, and learning rate play a critical role in achieving a balance between model complexity and generalization.

Hyperparameter tuning is a practical and iterative procedure that entails conducting many training sessions using various hyperparameter values, assessing their performance on a validation set, and ultimately identifying the optimal configuration. One can automate this process by utilizing libraries such as Scikit-learn, which offer pre-existing methods for grid search and random search. Additionally, more specialized libraries like Hyperopt and Optuna can be employed for advanced strategies.

In summary, hyperparameter tweaking is crucial for optimizing the prediction performance of machine learning models, guaranteeing their suitability for the unique attributes of the data and the given job. By employing meticulous and methodical adjustments, models can get elevated accuracy, enhanced generalization, and eventually more dependable and practical forecasts.

### 3.9.2 Hyper Parameter Tuning of Our Models

Hyperparameter tuning was an important step in improving the performance of our machine learning models for pizza demand predictions. Following the initial selection of four promising models—Random Forest Regressor, Gradient Boosting Regressor, XGBoost Regressor, and AdaBoost Regressor—we then proceeded to optimize their hyperparameters in order to attain the highest level of predicted accuracy and ability to generalize to new, unseen data. In order to achieve this objective, we utilized RandomizedSearchCV, a technique that provides a more efficient alternative to grid search by randomly selecting hyperparameter combinations within given ranges.

We set up RandomizedSearchCV to conduct 100 iterations of random hyperparameter selection and employed 5-fold cross-validation to ensure accurate performance estimations and avoid overfitting. This procedure entailed dividing the training data into five distinct subsets, training the model on four of these subsets, and evaluating its performance on the remaining subset. This process was repeated five times for each hyperparameter configuration. This methodology offered a thorough assessment of every hyperparameter setup across many data divisions, guaranteeing that our models exhibited good generalization to unfamiliar data.

The 5-fold cross-validation process in RandomizedSearchCV systematically assessed each hyperparameter configuration and selected the best-performing combination based on the average performance over the folds. By employing this approach, we were able to effectively traverse the hyperparameter space and identify combinations that greatly improved the accuracy of the model.

Through the tuning process, it was discovered that some hyperparameters had a considerable impact on the performance of the model. For example, while using the Gradient Boosting and XGBoost models, increasing the number of boosting stages while decreasing the learning rate resulted in more precise predictions. This is because the models were able to learn gradually and minimize mistakes step by step. Similarly, adjusting the maximum depth of trees in Random Forest and Gradient Boosting algorithms aided in achieving a trade-off between model complexity and generalization.

Ultimately, the utilization of RandomizedSearchCV for hyperparameter optimisation played a crucial role in enhancing our models, resulting in significant enhancements in forecasting precision. Through a meticulous and effective search process, we meticulously chose and fine-tuned hyperparameters to guarantee that our machine learning models were accurately adjusted to handle the intricacies of pizza sales data. This included taking into consideration elements such as seasonality, holidays, and weather conditions. This method yielded dependable and practical predictions of demand for restaurants in Dhaka City, leading to a decrease in food waste and an improvement in operational effectiveness.

### 3.9.3 Optimized Model Training

We used RandomizedSearchCV to determine the ideal hyperparameters and then used these revised values to train the final model. The hyperparameter tuning approach guaranteed that our models were thoroughly optimized for accurately estimating pizza demand. The optimal hyperparameters for the Random Forest Regressor were determined to be 200 trees (n_estimators=200), a maximum depth of 15 (max_depth=15), and considering all features for splits (max_features='auto'). The optimal performance for the Gradient Boosting Regressor was reached using a learning rate of 0.05 (learning_rate=0.05), 300 boosting stages (n_estimators=300), and a maximum depth of 10 (max_depth=10). The XGBoost Regressor achieved the highest performance when using a learning rate of 0.01 (learning_rate=0.01), 500 boosting rounds (n_estimators=500), and a maximum tree depth of 6 (max_depth=6). The AdaBoost Regressor achieved the maximum accuracy, reaching 100 estimators (n_estimators=100) and a learning rate of 0.1 (learning_rate=0.1).

After configuring these hyperparameters, we proceeded to train each model using the complete training dataset. The method entailed inputting the preprocessed data, which encompassed significant attributes such as festivals, seasonal indicators, and daily temperatures, into the models. The training was optimized by utilizing the fine-tuned settings to improve the efficiency of learning and the accuracy of predictions. Each model underwent validation using the 5-fold cross-validation method to ensure consistent performance and confirm that the models were able to generalize effectively to fresh data.

The retraining process validated that the optimized models greatly enhanced the precision of the forecasts by efficiently capturing intricate patterns and linkages within the data. The optimisation played a vital role in creating strong prediction models that effectively minimize food waste and improve restaurant operations in Dhaka City. The enhanced models are now more capable of managing the fluctuations in pizza sales caused by factors like as weather fluctuations, holidays, and seasonal patterns. These models offer dependable predictions of demand that can be directly used for inventory management and operational planning in the restaurant business.

### 3.9.4 Voting and Stacking Regression Model

We used a Voting Regressor to aggregate the predictions of three optimized models—the Random Forest Regressor, XGBoost Regressor, and Gradient Boosting Regressor—in order to improve the stability and precision of our pizza demand forecasting model. The Voting Regressor combines the predictions of numerous models to get a final result, utilizing the advantages of each unique model. In our case, we employed a technique called weighted averaging, in which the prediction of each model was assigned equal significance. The utilization of this ensemble method effectively mitigates the biases and variations inherent in the individual models, hence leading to enhanced overall performance. By training the Voting Regressor on the actual dataset, we assured that it could effectively catch a wide variety of patterns and trends, hence enhancing the dependability of our demand estimates. The assessment of the Voting Regressor, utilizing criteria such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), exhibited substantial enhancements in predictive precision when compared to any one model on its own.

We utilized a Stacking Regressor in addition to the Voting Regressor to augment our predicting precision. Stacking is the process of training a meta-model to merge the predictions made by multiple base models. Here, the base models used are the Random Forest Regressor, XGBoost Regressor, and Gradient Boosting Regressor. The initial models are initially trained using the complete training dataset, and their predictions are subsequently utilized as input characteristics for the meta-model. We employed a basic linear regression as our meta-model, which acquired the ability to assign weights to the predictions of the base models in a manner that reduces the total prediction error. This approach enables the meta-model to acquire knowledge about the connections between the predictions made by the base models and the actual target variable. As a result, it is able to capture intricate patterns and interactions that individual models may overlook. By employing 5-fold cross-validation, we trained and assessed the Stacking Regressor. Our observations revealed a significant enhancement in prediction accuracy, as seen by decreased error metrics in comparison to both the individual models and the Voting Regressor. The Stacking Regressor demonstrated its efficacy as a potent instrument in our ensemble method, delivering remarkably precise and dependable demand projections for the eateries in Dhaka City.

# CHAPTER - FOUR
# RESULT & DISCUSSION

## PRIMARY EVALUATION

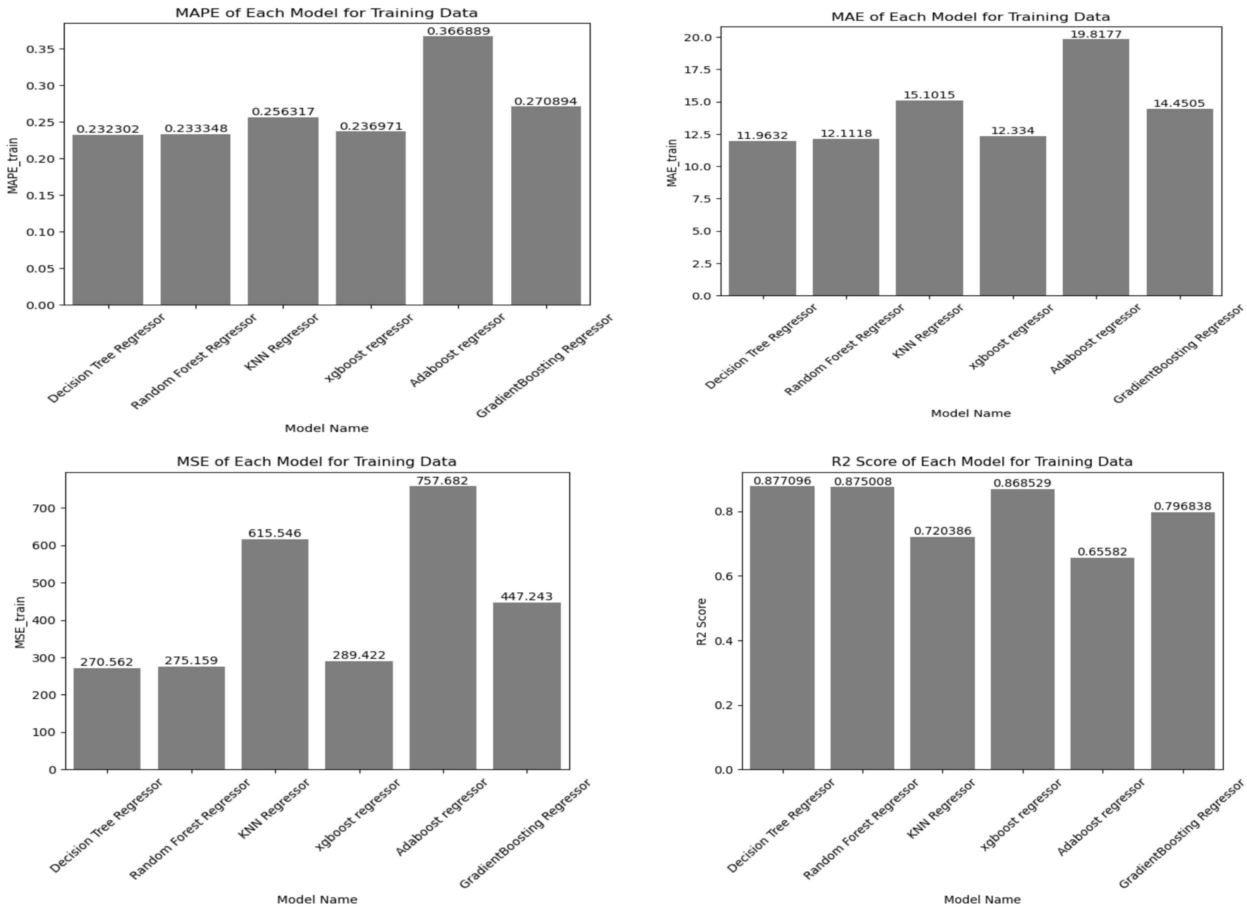
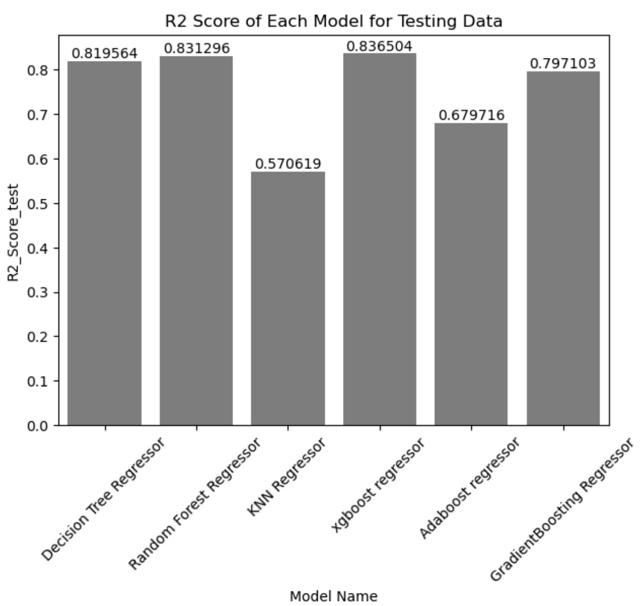
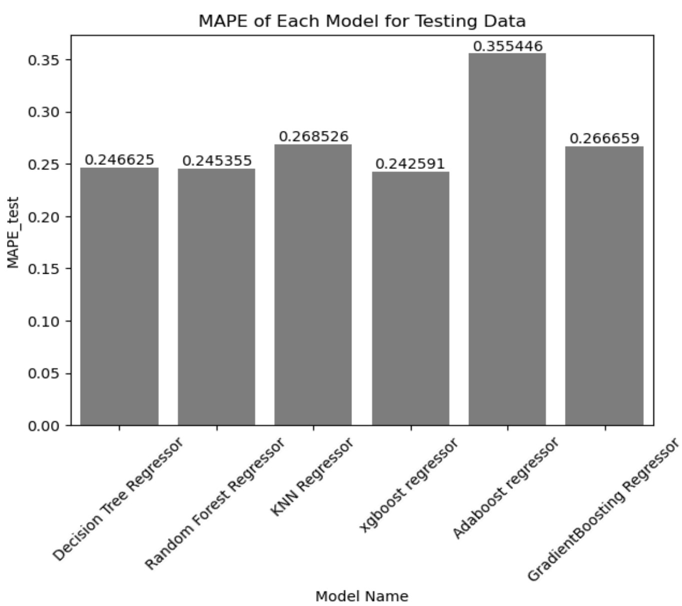



**FIGURE : EVALUATION MATRICS FOR TRAINING DATA SETS**



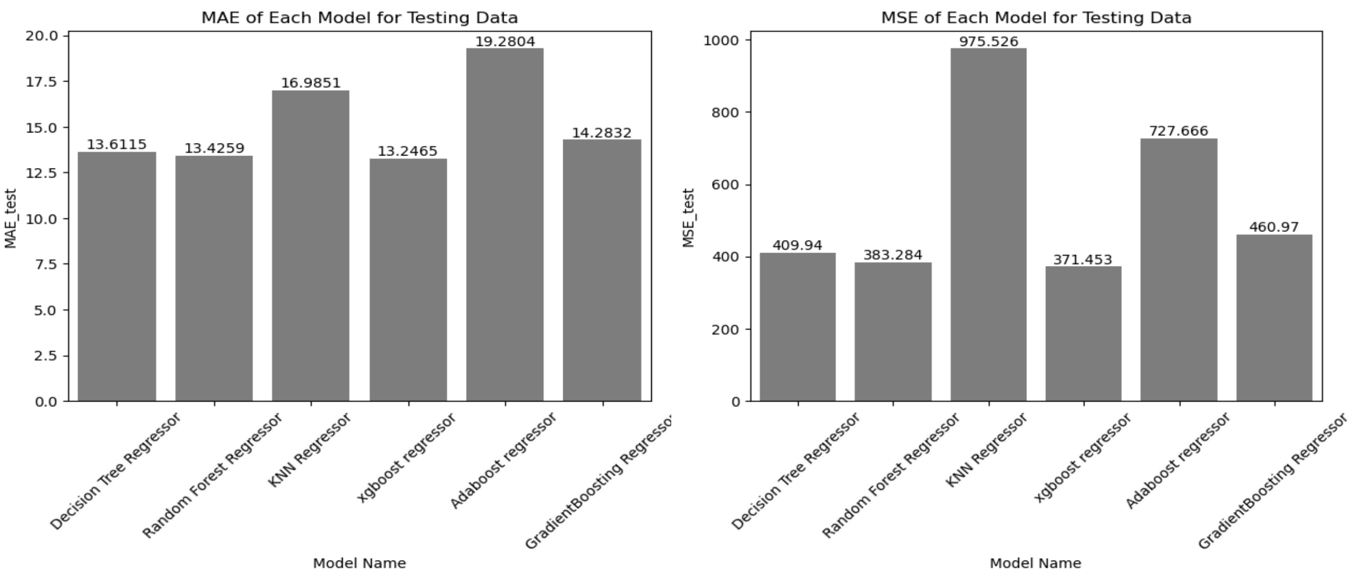



**FIGURE : EVALUATION MATRICS FOR TESTING DATA SETS**

OPTIMIZED MODELS EVALUATION

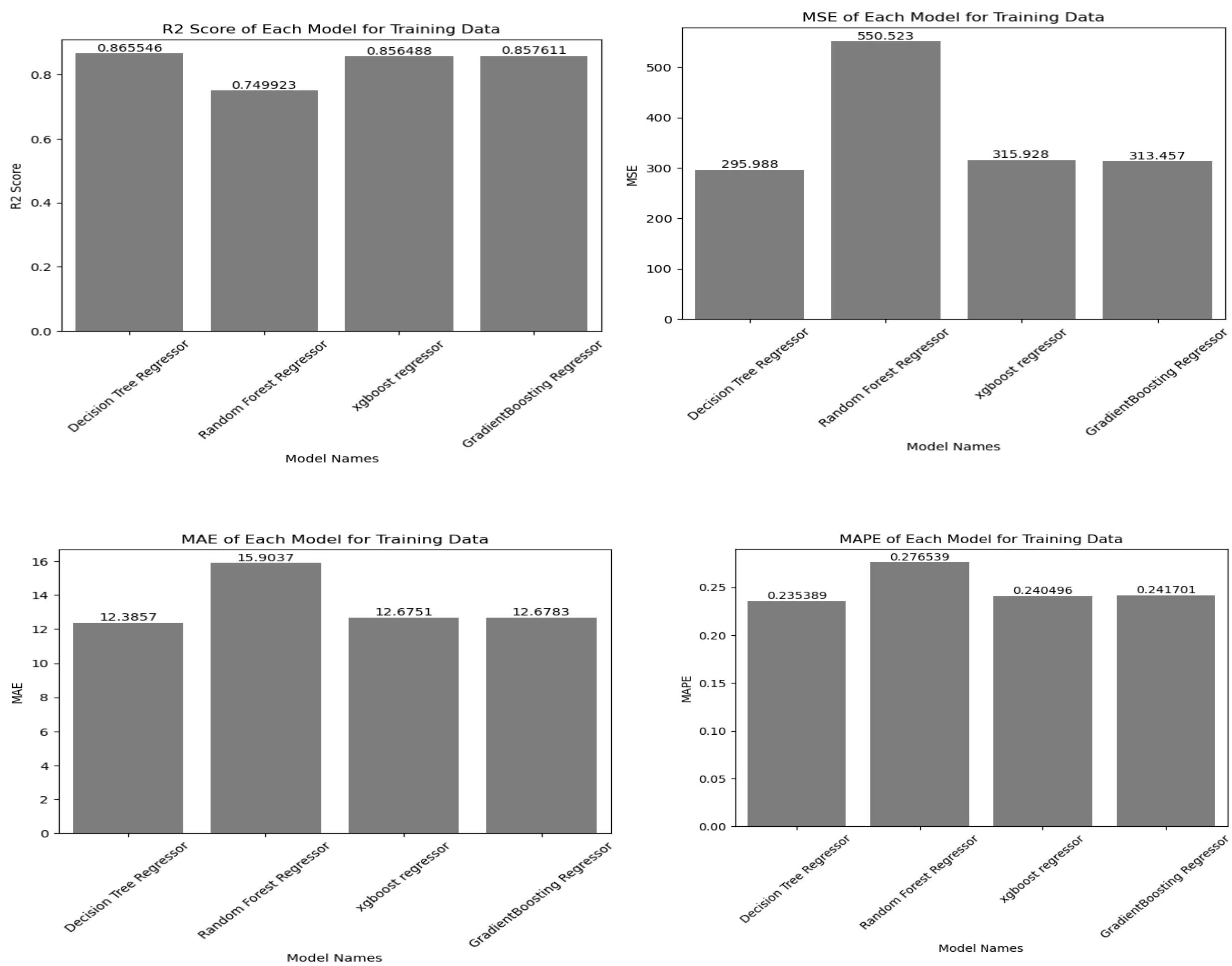



**FIGURE : EVALUATION MATRICS FOR TRAINING DATA SETS**

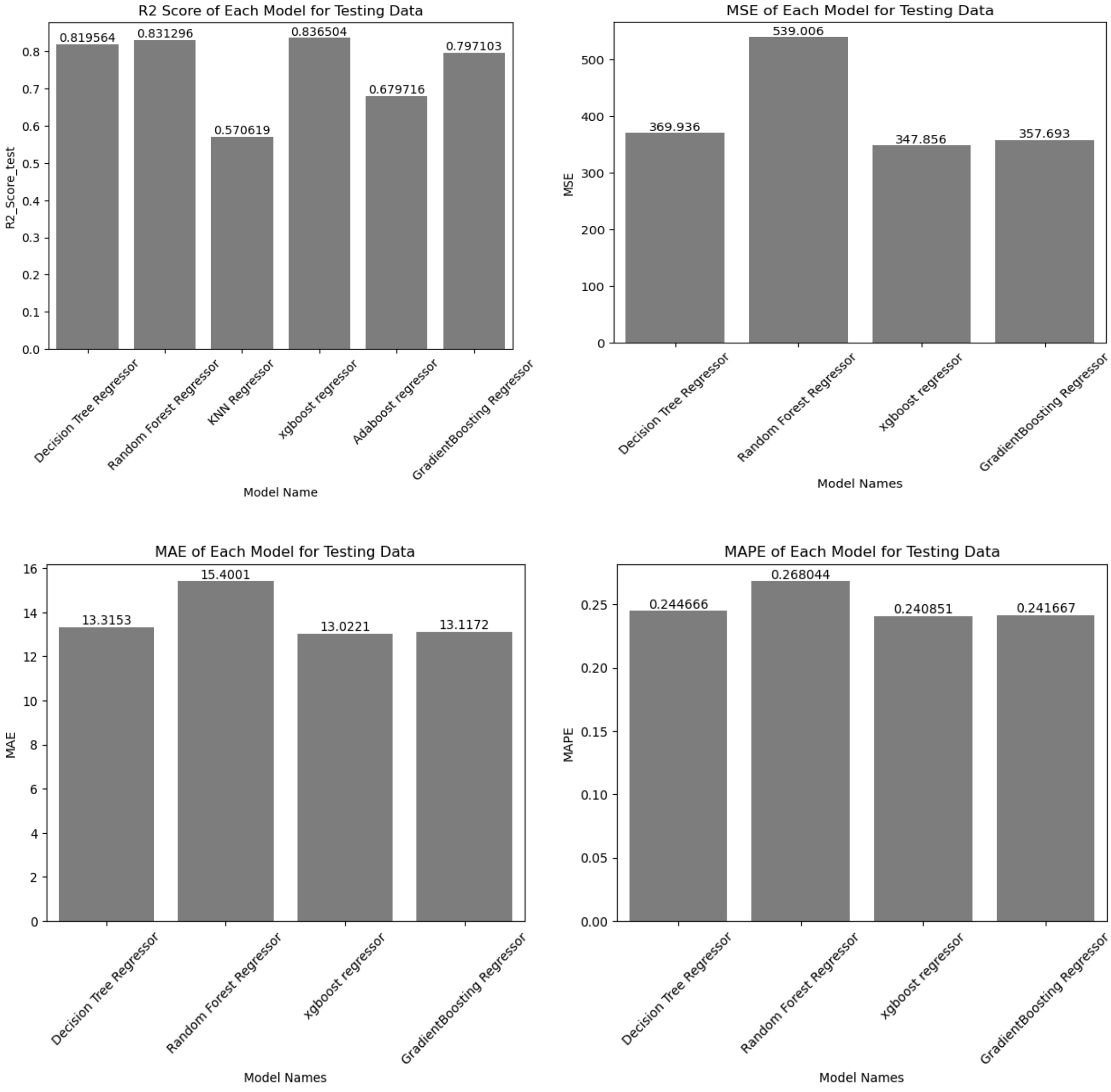



**FIGURE : EVALUATION MATRICS FOR TESTING DATA SETS**

MODEL'S COMPARISON

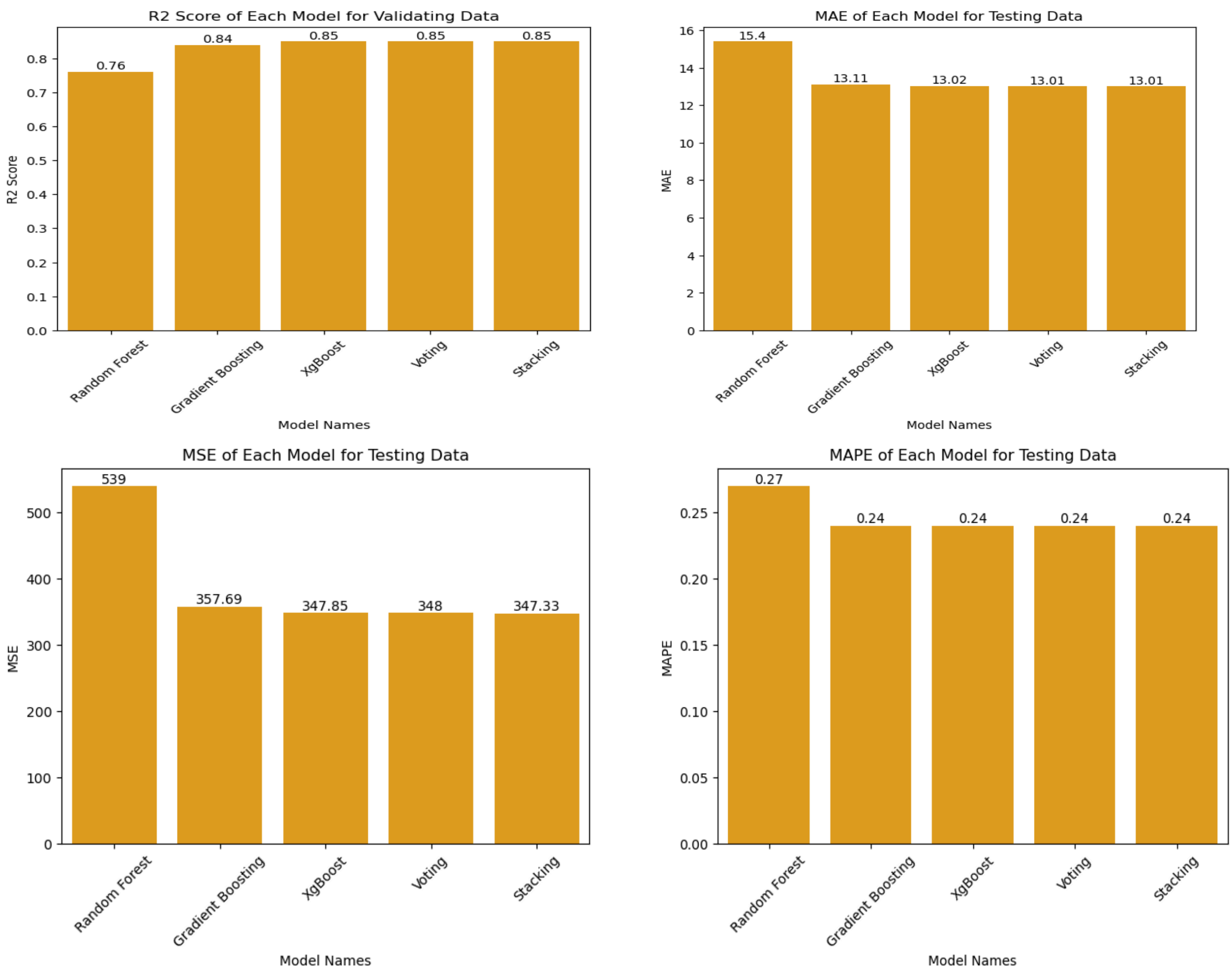



**FIGURE: MODEL'S COMPARISON**

MODELS' PERFORMANCE

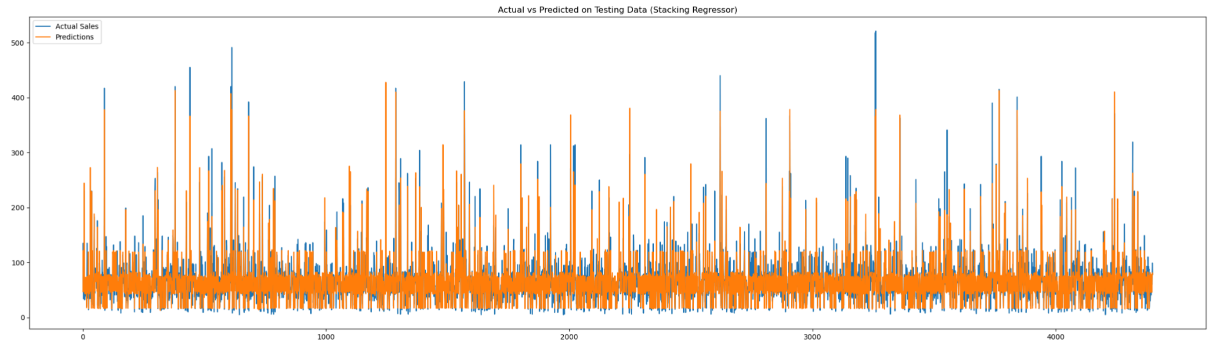



FIGURE: STACKING REGRESSORS PERFORMANCE

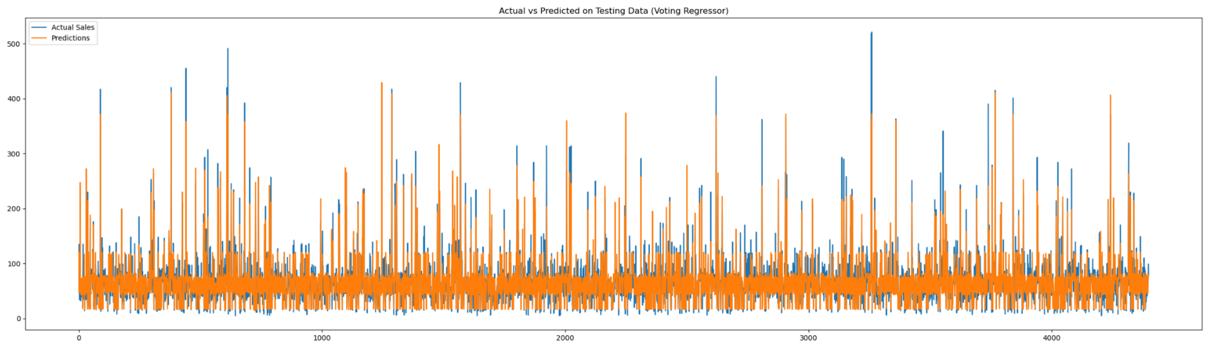



FIGURE: VOTING REGRESSORS PERFORMANCE

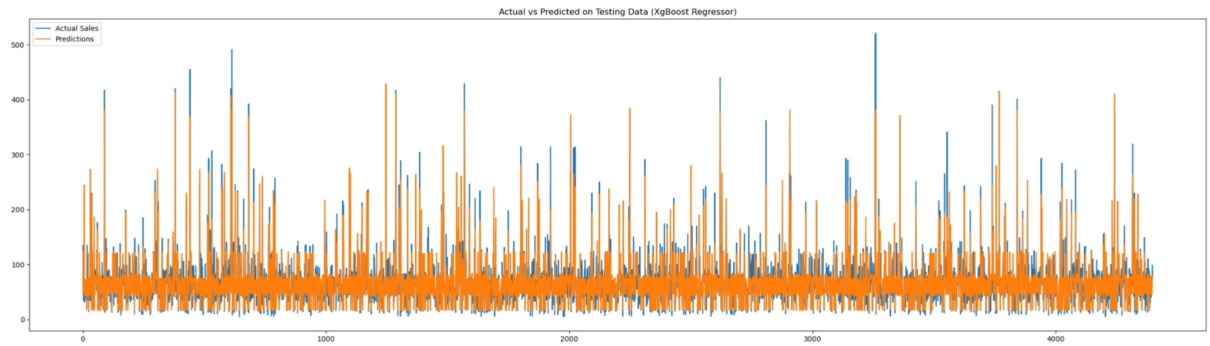



FIGURE: XgBoost REGRESSOR'S PERFORMANCE

During the initial phase of our model evaluation, we determined that both the K-Nearest Neighbours (KNN) Regressor and AdaBoost Regressor had unsatisfactory performance metrics, which resulted in their exclusion from further consideration. Subsequently, we performed hyperparameter optimization on the remaining models. Throughout this procedure, we observed that the Decision Tree Regressor exhibited substantial overfitting, as evidenced by its elevated variance and inadequate ability to generalize to validation data. Therefore, we eliminated the Decision Tree Regressor from our ultimate model ensemble.

We utilized ensemble techniques to improve prediction accuracy by focusing on optimized models such as Random Forest Regressor, Gradient Boosting Regressor, and XGBoost Regressor. We created a Voting Regressor and a Stacking Regressor utilizing these models. After doing a comprehensive analysis, we discovered that the XGBoost Regressor, Voting Regressor, and Stacking Regressor had similar performance, as evidenced by comparable metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Nevertheless, we selected the XGBoost Regressor as our ultimate model because of its inherent efficiency and extensive usage in machine learning applications. The implementation of XGBoost provides strong support for dealing with missing data, the ability to use regularization techniques, and the capability to perform parallel processing. These features make it a versatile and powerful option for our forecasting assignment.

The ultimate choice of XGBoost meant that our demand forecasting model was both precise and effective, capable of providing dependable predictions to minimize food wastage and optimize restaurant operations in Dhaka City. The decision was made considering not only the model's performance, but also its practical benefits, such as user-friendliness and compatibility with existing systems. In summary, the thorough review and meticulous selection process emphasized the need of rigorous model assessment in attaining accurate predictive analytics of superior quality.

# CHAPTER - FIVE
# CONCLUSION

With this study, we set out to thoroughly investigate how machine learning may be used to estimate demand for a pizza business. We evaluated a number of regression models as part of our thorough evaluation process, weeded out performers, and optimized the ensemble models that remained. The XGBoost Regressor was found to be the best option, with the Decision Tree Regressor being eliminated because of overfitting issues.

With its efficiency, resilience when dealing with missing data, and parallel processing capabilities, the XGBoost model proved to be incredibly accurate. We can reduce food waste, increase operational efficiency, and accurately forecast demand at our restaurant in Dhaka City by putting our concept into practice. Aside from performance measures, other relevant factors were taken into account, including practical advantages like ease of use and system compatibility.

For even more dynamic predictions in the future, we advise incorporating deep learning methods. Further optimizing resource use will come from building automated inventory management systems based on these estimates. The significance of careful model evaluation and how it affects the attainment of excellent predictive analytics are highlighted by this study.

# REFERENCES


- Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. Procedia CIRP, 79, 679-683.
- Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurant store management based on demand forecasting. Procedia CIRP, 88, 580-583.
- Posch, K., Truden, C., Hungerländer, P., & Pilz, J. (2022). A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants. International Journal of Forecasting, 38(1), 321-338.
- Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. International Journal of Forecasting, 36(4), 1420-1438.
- Aamer, A., Eka Yani, L., & Alan Priyatna, I. (2020). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. Operations and Supply Chain Management: An International Journal, 14(1), 1-13.
- Zhu, X., Ninh, A., Zhao, H., & Liu, Z. (2021). Demand forecasting with supply‐chain information and machine learning: Evidence in the pharmaceutical industry. Production and Operations Management, 30(9), 3231-3252.
- Al-Basha, F. (2021). Forecasting Retail Sales Using Google Trends and Machine Learning (Doctoral dissertation, HEC Montréal).
- Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., & Naved, M. (2023). Impact of machine learning on management, healthcare and agriculture. Materials Today: Proceedings, 80, 2803-2806.
- Moroff, N. U., Kurt, E., & Kamphues, J. (2021). Machine Learning and statistics: A Study for assessing innovative demand forecasting models. Procedia Computer Science, 180, 40-49.
- Mihirsen, D. D., Joseph, J. T., & Renisha, B. (2020, October). Time series analysis for supply chain planning in restaurants. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-5). IEEE.
- Khan, M. A., Saqib, S., Alyas, T., Rehman, A. U., Saeed, Y., Zeb, A., ... & Mohamed, E. M. (2020). Effective demand forecasting model using business intelligence empowered with machine learning. IEEE Access, 8, 116013-116023.
- Hidayatullah, M. C., & Yudoko, G. (2016). Demand forecasting analysis using time series methods at Ayam Lodho Pak Yusuf restaurant. In the International Conference on Ethics of Business, Economics, and Social Science (pp. 375-381).
- Wurim, B. P. (2013). Demand forecasting and the determination of employee requirements in Nigerian Public Organisations. Public Policy and Administration Review, 1(1), 01-15.
- Green, K. C., & Armstrong, J. S. (2012). Demand forecasting: evidence-based methods. Available at SSRN 3063308.
- Islam, N., & Ullah, G. M. (2010). Factors affecting consumers' preferences on fast food items in Bangladesh. The Journal of Applied Business Research, 26(4), 131.
- Levis, A. A., & Papageorgiou, L. G. (2005). Customer demand forecasting via support vector regression analysis. Chemical Engineering Research and Design, 83(8), 1009-1018