

Food Demand Forecasting

Jyothy Das

21/02/2025

A project report on
Genpact machine learning hackathon project
'Food demand forecasting'.

Contents

1	Abstract	5
2	Problem Definition	6
2.1	Overview	6
2.2	Problem statement	6
3	Introduction	7
4	Literature Survey	7
4.1	Key Areas of Study	7
4.2	Key Findings and Gaps	7
4.3	Relevance to Current Study	7
5	Methodology	7
5.1	Data Collection	7
5.1.1	Data Overview	7
6	Exploratory Data Analysis	7
6.1	Data Transformation	9
6.1.1	Data Synchronization:	9
6.1.2	Column Removal	9
6.1.3	Column Rename	9
6.1.4	9
6.1.5	9
6.1.6	9
6.1.7	9
6.2	Model Development	9
7	Results and Discussion	9
8	Conclusion	10
9	References	10

List of Figures

List of Tables

1 Abstract

Accurate demand forecasting is critical for meal delivery companies to optimize procurement and staffing decisions. This project aims to predict the demand for meals across multiple fulfillment centers for the next 10 weeks using historical order data, meal characteristics, and center-specific attributes.

2 Problem Definition

Meal delivery companies operate in a dynamic and competitive market where efficient inventory and workforce management are crucial for profitability and customer satisfaction. A key challenge is accurately forecasting the demand for various meals at different fulfillment centers to ensure proper stock levels while minimizing waste due to perishable ingredients. Additionally, staffing decisions depend on anticipated order volumes, making demand forecasting a critical aspect of operational planning.

This project focuses on developing a machine learning-based solution to predict the number of orders for different meal-center combinations over the next 10 weeks.

2.1 Overview

Underestimating demand may lead to stock shortages and customer dissatisfaction, whereas overestimating demand can result in excessive inventory, food wastage, and increased operational costs.

To address this challenge, the project involves analyzing historical order data (covering Weeks 1 to 145), meal-specific details (category, sub-category, pricing, discount), and fulfillment center information (location, city, operational factors). By training machine learning models on this data, the goal is to develop an accurate forecasting model that predicts demand for the upcoming 10 weeks (Weeks 146-155). The insights gained from this model will support better decision-making in procurement, inventory management, and staffing, ultimately improving operational efficiency and reducing waste.

2.2 Problem statement

Accurate demand forecasting is essential for meal delivery companies to ensure efficient procurement of raw materials and optimal workforce allocation at fulfillment centers. Since raw materials are perishable and replenished on a weekly basis, precise predictions of future demand can significantly reduce waste, control inventory costs, and improve service reliability.

This project aims to develop a machine learning-based forecasting model to predict the demand for different meal-center combinations for the next 10 weeks (Weeks 146-155). The model will be trained using historical demand data (Weeks 1-145), meal-specific attributes (such as category, sub-category,

price, and discount), and fulfillment center information (such as location and operational characteristics). The predictions generated will assist the company in making data-driven decisions for procurement planning and workforce management, ultimately enhancing operational efficiency and customer satisfaction.

3 Introduction

4 Literature Survey

4.1 Key Areas of Study

4.2 Key Findings and Gaps

4.3 Relevance to Current Study

5 Methodology

5.1 Data Collection

5.1.1 Data Overview

6 Exploratory Data Analysis

- meal id - The distribution is uneven, with certain meal IDs having significantly higher demand than others. - Some meals are ordered frequently, while others have much lower demand.
- center id - The distribution is somewhat uniform but shows some peaks, indicating that certain centers have more orders than others. - Some centers receive a significantly higher number of orders.
- city code - The distribution is not uniform, with a few city codes having notably higher order counts. - Some cities contribute more to the demand, indicating regional preferences.
- region code - The distribution shows clear peaks, indicating that only a few regions dominate food demand. - Some regions may have signifi-

cantly lower demand, possibly due to fewer service centers or population differences.

- op area (operational area of the center) - Most centers operate in a mid-range operational area, with fewer in extreme small or large areas. - Centers with mid-range operational areas tend to have higher demand.
- id - Since the 'id's are unique, the grap appears uniformly distributed. - No specific insights can be drawn from this graph.
- week - Demand slightly increased over time.
- checkout price - There are distinct peaks at certain price points, indicating price clustering. - Some meals are frequently ordered at specific price ranges, showing customer price sensitivity.
- base price - Similar to checkout price, there are peaks at certain values - This suggests price standardization for certain meals.
- emailer for promotion - The majority of the values are 0, meaning most meals are not promoted via email. - A small percentage of meals are promoted, and their effectiveness can be further analyzed.
- homepage featured - Similar to email promotions, most meals are not featured on the homepage. - A very small fraction of meals get featured, additional study is required to assess its influence on demand.
- num orders - The distribution is highly skewed, with most meals having a low number of orders. - A small number of meals receive a very high number of orders, indicating popular items.

Column 1	Column 2
Data 1	Data 2
Data 4	Data 5
Data 6	Data 7

Column 1	Column 2
Data 1	Data 2
Data 4	Data 5
Data 6	Data 7

Column 1	Column 2
Data 1	Data 2
Data 4	Data 5
Data 6	Data 7

Non-null Counts:

Missing Values: There are no missing values, as all 29,999 rows are complete.

6.1 Data Transformation

6.1.1 Data Synchronization:

Efforts have been made to synchronize the state and district names throughout the dataset, ensuring a uniform format.

6.1.2 Column Removal

6.1.3 Column Rename

To enhance clarity and improve readability, several columns were renamed to more intuitive and consistent names.

6.1.4

6.1.5

6.1.6

6.1.7

6.2 Model Development

Insert Model Development details here.

7 Results and Discussion

Insert Results here.

8 Conclusion

Insert Conclusion here.

9 References

Insert References here.