

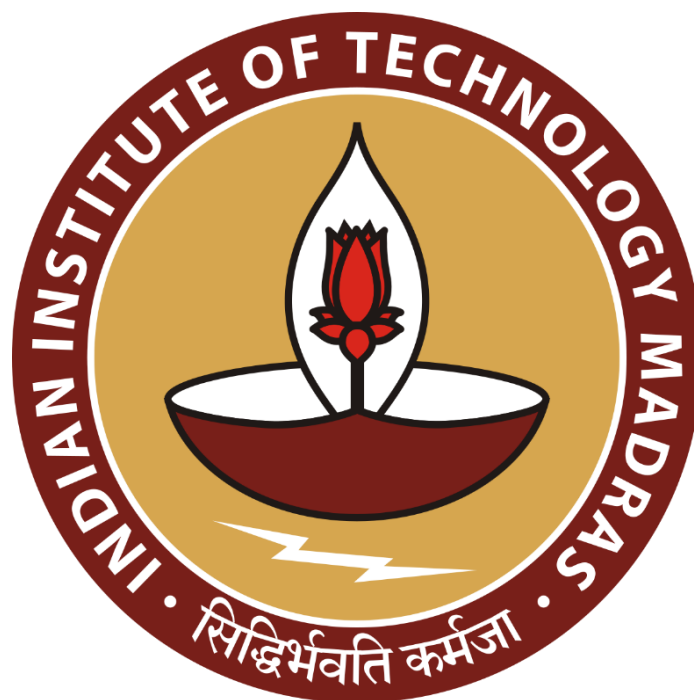
Sales Performance Analysis and Marketing Optimization for a Retail Home Decor Business

Mid-Term Submission for the BDM Capstone Project

Submitted by

Name: **Saransh Saini**

Roll Number: **22F1001123**



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Content

1	Executive Summary and Title	3
2	Proof of Originality	3
2.1	Video Interaction with the Founder	3
2.2	Letter from Business	4
2.3	Images of the Firm	5
3	Metadata	5
3.1	Origins of Data	5
3.2	Data Format	6
3.3	Data Cleaning and Preprocessing	7
4	Descriptive Statistics	9
4.1	Dataset Overview	9
4.2	Central Tendency and Distribution	9
4.3	Sales and Quantity by Month	10
4.4	SKU Categories	10
4.5	Payment Method Breakdown	11
5	Explanation of the Analysis Procedure	11
5.1	Categorical Influence on Sales	11
5.2	SKU Performance Classification	12
5.3	Market Basket Analysis using Apriori Probability	12
5.4	Customer Geo-Mapping	12
5.5	ARIMA Sales Forecasting	12
6	Results and Findings	13

Declaration Statement

I am working on a Project Title “**Sales Performance Analysis and Marketing Optimization for a Retail Home I Business**”. I extend my appreciation to **Noble Homes Decor**, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered through primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the information of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project’s completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I agree that all the recommendations are business-specific and limited to this project exclusively, and cannot be utilized for any other purpose with an IIT Madras tag. I understand that IIT Madras does not endorse this.

Signature of Candidate:



Name: Saransh Saini

Date: 28 September 2024

1. Executive Summary and Title

Noble Homes Décor, a premium retail outlet located in Neb Sarai, Delhi, following a B2C business model, they specialize in handloom and home décor products such as curtains, carpets, mattresses, and more. Since its inception in **February 2024**, the business has faced challenges in terms of low sales, insufficient marketing, and limited brand visibility. This project, titled “**Sales Analysis and Marketing Optimization of Retail Home Decor Business**” seeks to address these challenges by analysing sales data to uncover valuable insights that can guide strategic decisions.

After in-depth discussions with the founder **Mr. Radha Krishnam**, it became clear that the issues are rooted both internally and externally. **Internally**, the business lacks a data-driven approach to tracking sales trends and customer preferences, which impedes decision-making. **Externally**, Noble Homes Décor faces fierce competition from well-established businesses, as well as the presence of larger shopping complexes like Vishal Mega Mart. In such a competitive environment, building a strong brand presence and developing a robust marketing strategy is crucial for growth. This analysis aims to provide actionable insights that can help Noble Homes Décor overcome these obstacles and pave the way for sustained success.


2. Proof of Originality

i. Video Interaction with the Founder



Link to Video: [Interaction with Founder - BDM Project](#)

ii. Letter from Business



NOBLE HOMES DECOR- THE HANDLOOM SHOP
C, 8/2, IGNOU Main Rd, Near Bikaner Sweets, Opp. Paryavaran Complex,
Neb Sarai, Sainik Farm, New Delhi, Delhi 110030

Date: 4 October 2024

To,
The Head of the Department,
Indian Institute of Technology Madras,
Chennai, Tamil Nadu - 600036

Subject: Proof of Data Provision for Business Data Management Project

Dear Sir/Madam,


This letter is to certify that Noble Homes Decor, located at C, 8/2, IGNOU Main Road, Neb Sarai, New Delhi, has provided sales data to **Mr. Saransh Saini** for the purpose of his **Business Data Management project at IIT Madras**. The data provided includes sales information from **February 2024 to September 2024**, sourced from our customer bill copies.


The data shared with Mr. Saransh Saini includes transaction details such as items sold, quantities, and total sales amounts. This information has been made available to him solely for academic purposes, as per his request.


We trust this data will assist him in successfully completing his project and contribute to his academic goals. Should you require any further information or verification, please feel free to contact us.


Thank you.



Sincerely,
Radha Krishnam
Owner, Noble Homes Decor
+91 98105 98432



 www.noblehomesdecor.com

 sales@noblehomesdecor.com

 **+91 9873 915477, +91 89203 00454**

 C-8/2, Saiyad ul Ajaib, IGNOU Main Rd, Neb Sarai, Sainik Farm,
New Delhi, Delhi 110030 

Proof of Originality Letter from the Business

iii. Images of the Firm

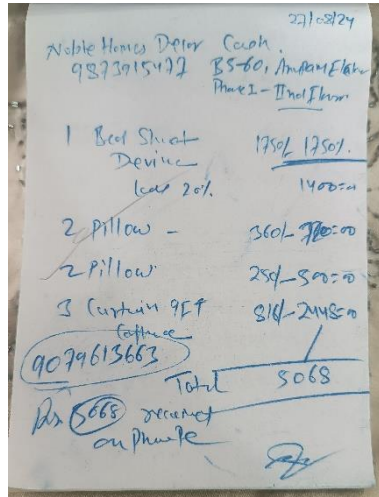


Link to Image Folder: [Images of the Firm](#)

3. Metadata

i. Origins of the Dataset

The dataset for this analysis was derived from Noble Home Décor's customer bill copies. These bills represent the original transactions handed to customers after a successful purchase. Unfortunately, the business doesn't store this data digitally, and access to their limited digital records was restricted due to sensitive information related to other businesses.



The data entry process for these bill copies was entirely manual. In total, I worked with 501 bills, covering transactions between 24th February 2024 and 22nd September 2024.

ii. Data Format

Column	Data Type	Description	Example	Missing Value (%)
Date	DateTime	Date of the transaction (dd-mmm-yy)	02-Mar-24	0%
Bill No	Integer	Numbers assigned during data entry	406	0%
Items	String	Name of the SKU	Table Cover	0%
Unit Price	Integer	Unit Price of the SKU	750	0%
Quantity	Integer	Quantity of SKU sold	3	0%
Total Price	Integer	Unit Price X Quantity	2250	0%
Payment	String	Cash or UPI	Cash	0%
Address	String	Address of the customer	B-31A Raju Complex	98.29%
Extra Info	String	Any extra information about the customer	Ms. Vinita [9934***]	99.51%

Table describing the columns of the dataset

iii. Data Cleaning and Preprocessing

Since the data entry was done manually, it was prone to human errors and required several steps of transformation and preprocessing:

- a. **Data Cleaning:** During the cleaning process, duplicate bills that had been mistakenly entered multiple times were removed. Importantly, the columns Address and Extra Info, despite being sparsely populated, were **NOT** removed due to their significance in further analysis.
- b. **SKU Clustering:** One of the key challenges was that the dataset included various spelling inconsistencies and alternate descriptions for the same SKUs. For instance, curtains appeared under labels like *curtains gajjari*, *reliance curtain*, *curtain long*, *window curtain*, etc. If each one of these labels would have been considered as separate SKUs, the data would have diluted into numerous sparsely purchased SKUs, which would have been worthless to analyze.

To solve this problem **OpenRefine**'s clustering techniques were used. OpenRefine provides 2 broad categories of clustering methods:

1. Token-based (n-gram, key collision, phonetic fingerprint, etc)
2. Character-based, (Levenshtein distance, PPM, etc)

All of these methods were used with particular parameters to find the best clusters available. After this process, we were left with just **64 SKUs**.

c. Category Classification

Once the SKUs were clustered, I classified them into broader, yet more manageable categories. This would make our understanding of trends and patterns much easier.

After carefully examining the SKUs all of them were classified into these 7 categories: *Home Decor*, *Bedding*, *Bath and Towels*, *Cleaning and Household*, *Clothing and Fashion*, *Kitchen and Dining*, and *Miscellaneous*.

d. Quantity Standardisation

The quantity field included values belonging to multiple units such as *kg*, *g* & *ft*. Having data values in these units of measurement can cause severe problems in further analysis and this needs to be dealt

with. It was discovered that unit prices are set according to *IKg* and *lft*, in other words, *IKg* and *lft* can be interpreted as *1 unit*.

So, going ahead with this I converted all *Kg* and *ft* to float values, and all *g* values were scaled to *kg*.

e. Latitude & Longitude Mapping

The Address field holds the addresses of customers, containing only 14 non-null rows. Although few in number these addresses will come in handy during Geo-Spatial analysis. The problem is that these addresses in string format are of little use to us, so mapping them to their respective Latitudes and Longitudes was required.

After all the data preprocessing the dataset looks like this

Column	Data Type	Description	Example	Missing Value (%)
Date	DateTime	Date of the transaction (dd-mmm-yy)	02-Mar-24	0%
Bill No	Integer	Numbers assigned during data entry	406	0%
Items	String	Name of the SKU	Table Cover	0%
Category	String	The category of the SKU	Bedding	0%
Unit Price	Integer	Unit Price of the SKU	750	0%
Quantity	Integer	Quantity of SKU sold	3	0%
Std. Qnt	Float	Standardised Quantities of SKUs	0.85	0%
Total Price	Integer	Unit Price X Quantity	2250	0%
Payment	String	Cash or UPI	Cash	0%

Address	String	Address of the customer	B-31A Raju Complex	98.29%
Latitude	Float	Latitudes of the Address	28.5167	98.29%
Longitude	Float	Longitude of the Address	77.2034	98.29%
Extra Info	String	Any extra information about the customer	Ms. Vinita [9934***]	99.51%

Link to Dataset [[Go to Dataset](#)]

4. Descriptive Statistics

i. Dataset Overview

- **Number of SKUs:** 64
- **Time Frame:** February 2024 to September 2024
- **Number of Transactions:** 501
- **Total Revenue:** ₹ 6,42,256
- **Average Transaction Value:** ₹ 782.28

ii. Central Tendency and Distribution

	Unit Price	Std Qnt.	Total Price
Mean	₹ 405.49	2.601	₹ 782.28
Minimum	₹ 3	0.03	₹ 25
25%	₹ 175	1	₹ 240
50%	₹ 300	1	₹ 450
75%	₹ 550	2	₹ 800
Maximum	₹ 3400	150	₹ 11200
Std. Dev	384.83	7.83	1107.38

Descriptive Statistics of the Dataset

iii. Sales and Quantity by Month

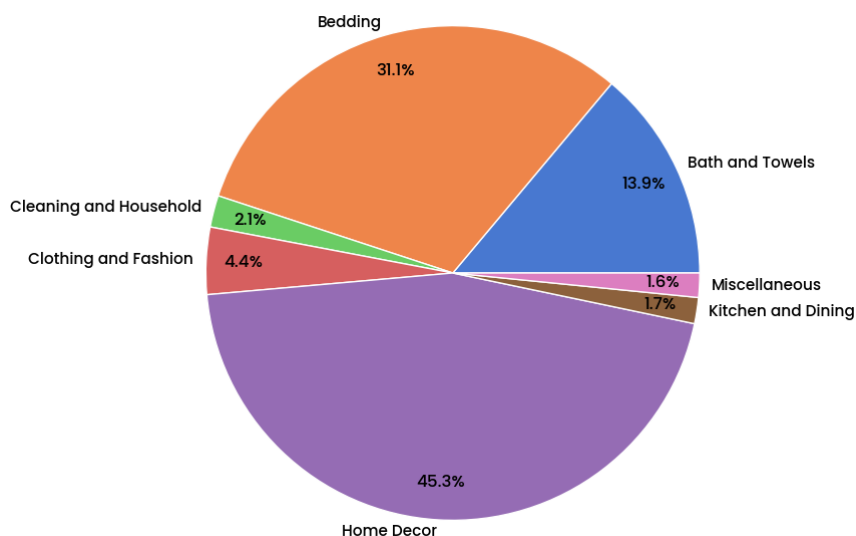


Bar Charts of Months vs Sales and Quantity

As these graphs claim the amount of sales and number of goods sold in the months of February and April, are significantly low as compared to other months. In the case, of February the heavy fluctuation can be credited to the fact that only two rows of data is present for the month. Moreover, the shop started operations in February of 2024, which further explains the low figures. But, in the case of April, a massive fluctuation can be observed. After discussing this issue with the founder, it was clarified that due to some issues with their business partners, the operations were put on a temporary pause for a few days leading to this fluctuation.

iv. SKU Categories

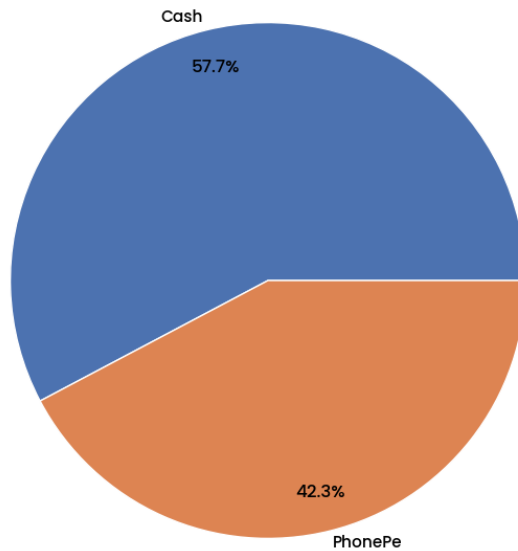
The 64 SKUs are categorized into 7 categories



Category-wise SKU Distribution

The pie chart clearly conveys that the **Home Décor** and **Bedding** categories account for more than 75% of the SKUs. This suggests that the shop majorly deals not only in *Home Décor* products but also in *Bedding* products. Out of the remaining 24% *Bath and Towels* category looks like a big player.

v. Payment Method Breakdown



The pie chart shows that the majority of transactions are being completed using *Cash* payments.

Payment Method Breakdown

5. Explanation of Analysis Procedure

In this analysis, a combination of advanced methods has been deployed, each of which contributes to addressing the challenges faced by Noble Homes Décor. This analysis would try to uncover sales trends, product performance, and customer behaviour. Below is a detailed explanation of the analysis methods used with proper justification.

i. Categorical Influence on Sales

AIM: To understand the influence of each category on total sales.

Justification: This breaking down of sales by category was necessary to capture the distinct behaviour of different product categories. The problem that arises here is that simple aggregation of sales by categories would induce a bias towards categories containing large numbers of SKUs. To avoid this, normalization of the analysis by sales density was conducted, this would ensure that categories like *Home Décor* won't automatically dominate the analysis.

$$\text{Sales Density} = \frac{\text{Total Sales for Category}}{\text{Total Sales for All Categories}} \div \frac{\text{Number of SKUs in Category}}{\text{Total Number of SKUs}}$$

ii. SKU Performance Classification

AIM: Use ABC analysis to classify each SKU as Good, Moderate, or Poor Performance based on their sales and quantity purchased.

Justification: ABC analysis is a widely recognized method to classify inventory based on sales and quantity. This analysis was applied both overall and on a month-by-month basis, offering us clarity on which products should be prioritized for restocking, promotions, or even discontinuation.

iii. Market Basket Analysis using Apriori Probability

AIM: To identify products that are frequently bought together and help understand customer behaviour.

Justification: Market Basket Analysis is one of the fundamental techniques under **Association Rule Mining**. This analysis will help identify products that are frequently bought together, thus aiding the business in creating bundled promotions and cross-selling opportunities. For this, **the Apriori Probability** algorithm was used with the *confidence* and *lift* filter being *0.2* and *1* respectively.

iv. Customer Geo-Mapping

AIM: Map out the addresses in the dataset and try to understand the customer distribution.

Justification: Despite the limited number of addresses available in the data (only 14), the geo-mapping analysis provided some valuable insights into customer distribution. The Folium library in Python was used to visualize these addresses on a Satellite Map. For instance, certain clusters of addresses may indicate neighborhoods that could be targeted for local marketing efforts, while small clusters of addresses may indicate possible new regions to focus marketing.

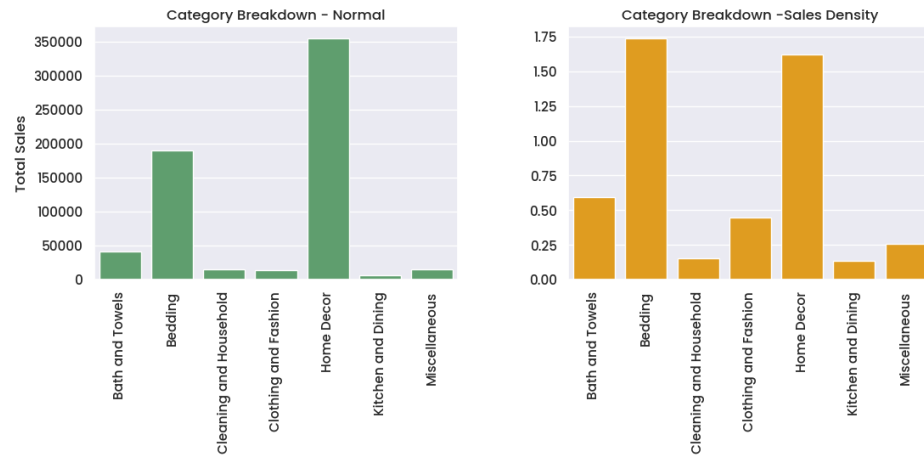
v. ARIMA Sales Forecasting

AIM: Predict future sales that the business can experience in the next few months.

Justification: Future sales prediction is necessary for both inventory and financial planning. An ARIMA (AutoRegressive Integrated Moving Average) model was employed on the monthly sales data for future prediction using time series analysis. With parameters $p=2$, $d=1$, and $q=1$, the model is capable of capturing the patterns in past sales data.

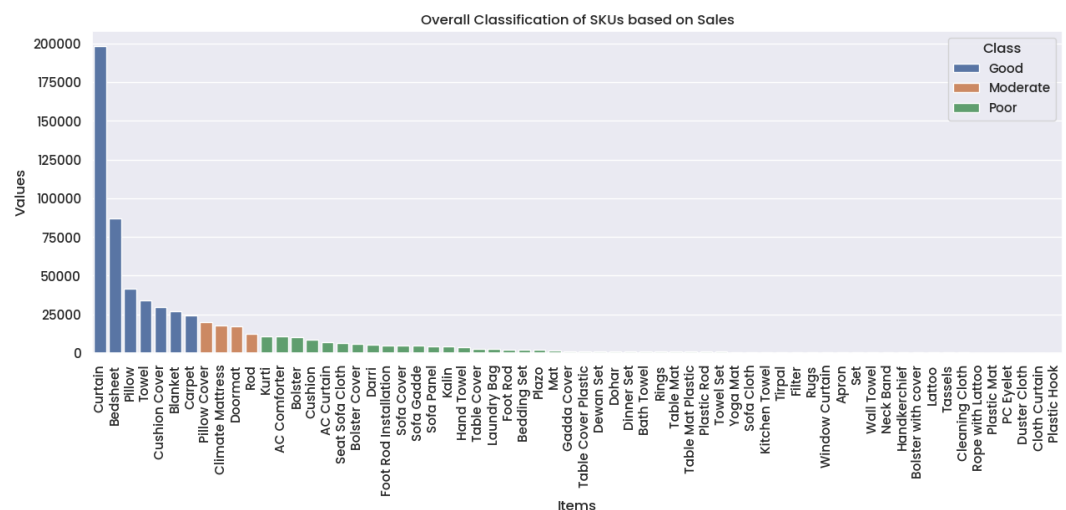
6. Results and Findings

i. Categorical Influence on Sales



As we can witness there is a clear and stark contrast between the plots. The sales density graph uncovers that the *Beddings* category, despite having fewer SKUs, holds more influence on total sales compared to the dominant *Home Décor* category. Moreover, the *Clothing and Fashion* category which appears to have minuscule influence in the standard breakdown, is actually ranked the fourth most influential category on the chart. Using this information, we can safely claim that the categories *Bedding*, *Home Decor*, *Bath and Towels*, and *Clothing and Fashion* emerge as the four key pillars of the business.

ii. SKU Performance Classification



This is the result of the ABC analysis conducted over the entire dataset with monthly breakdowns also conducted. Some of the key insights that can be drawn are: *Curtains* dominate the sales with a total revenue nearing ₹ 2,00,000. Following them are bedding products like *Bedsheets* and *Pillows*. *Mattresses*

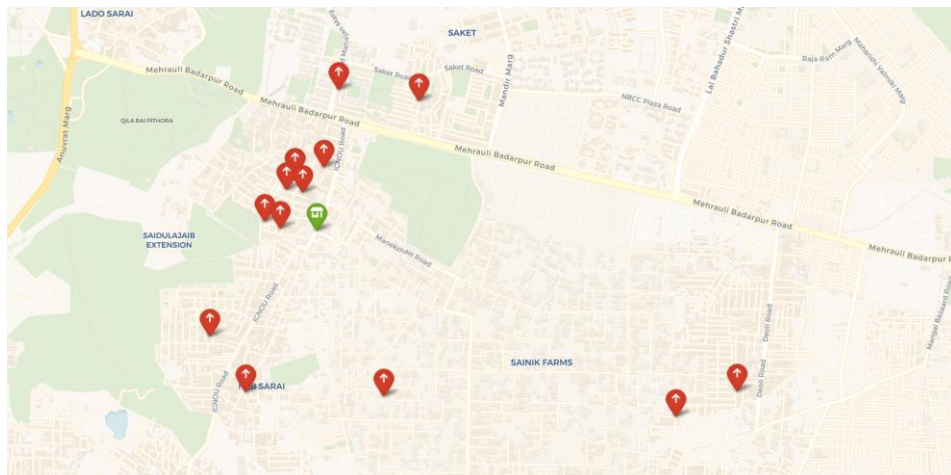
and *Doormats* fall under moderate categories indicating that these items must be purchased alongside the Good Performing products. The green-colored SKUs are Poor performers, contributing just 20% to the total sales. The business ought to make decisions regarding where to continue offering these products aided by some promotions or discontinue them.

iii. Market Basket Analysis using Apriori Probability



This is a network graph created using *Kumu* software. As we can witness, *Bedsheets* seems to be the central product that is frequently bought alongside *pillows*, *doormats*, *hand towels*, *pillow covers*, etc. *Bolsters* and *Bolster Covers* project a strong relationship, just like *Cushions* and *Cushion Covers*. An interesting detail that's observed here is the connection between *Towels* and *Bedsheets*. Though unusual, this can indicate towards presence of a hotel or other hospitality businesses as a customer.

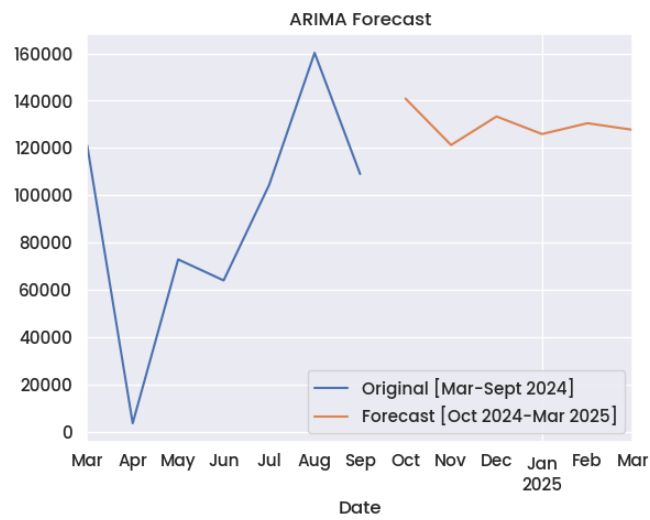
iv. Customer Geo-Mapping



The map shows clusters of customer addresses in the neighborhood. The majority of these customers are located within a few kilometres of the business. Atleast 3 clusters are visible, indicating potential points of targeted promotions.

Moreover, the business can expand its range of deliveries as it gains prominent brand value.

v. ARIMA Sales Forecasting



The ARIMA model predicts a gradual decline in business performance over the next 6 months. Strong and efficient decisions are required to boost sales and improve business performance.