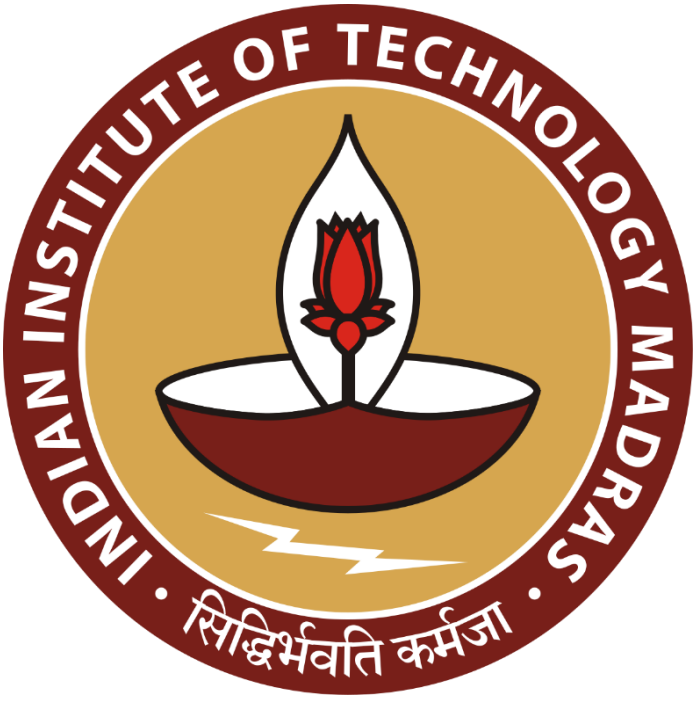
Sales Performance Analysis and Marketing Optimization for a Retail Home Decor Business

Final 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

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 Home Décor is a retail B2C business based in Neb Sarai, Delhi. It specializes in handloom and home décor products such as carpets, curtains, mattresses, and bedsheets. Since its inception in February 2024, the shop has encountered significant challenges in boosting sales and establishing a strong brand identity. These ongoing issues have led the business to contemplate relocating to help lighten the financial burden.

To address these concerns, it was decided that a detailed data analysis of historical sales records would be conducted to uncover insights and recommend strategies for improvement. Despite some available digital data, the business decided not to share it due to concerns about leaking sensitive information related to business partners. Instead, 501 physical customer bills, which spanned from February 2024 to September 2024, were provided for analysis, which required manual data entry and an extensive data cleaning and preprocessing operation to ensure accuracy and consistency.

This report outlines the detailed analytical process conducted to explore solutions to the problems faced by the business and enhance sales growth and marketing optimization. The analysis focuses on five core areas, categorical sales influence, product performance, customer behavior, market opportunities, and future forecasting. Each analysis is aided with visualizations to provide clear and actionable insights.

The objective here is to derive data-driven recommendations aimed at enhancing sales performance and optimizing marketing strategies for Noble Home Décor. By utilizing these insights, the business can make informed decisions to overcome its current challenges and achieve sustainable growth.

Link to Proof of Originality and Data Analysis File: Go to Folder

Noble Homes Décor, Neb Sarai, Delhi

2. 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. Before starting the analysis, it is necessary to prepare the data in order to avoid anomalies and misleading results. A number of data cleaning and preprocessing operations were performed resulting in a data matrix of 13 columns and 821 rows. Now, we can go ahead with the analysis procedure. Below is a detailed explanation of each analysis method used.

i. Categorical Influence on Sales

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

Procedure: In the data preprocessing section all SKUs were classified into 7 distinct categories. This analysis was aimed to understand how each category influences the total sales. The problem that arises here is that simple aggregation of sales by categories would induce a bias towards the categories containing a large number of SKUs. Instead, normalization of the analysis by *Sales Density* was conducted, this would ensure that categories with high number of SKUs like *Home Décor*, won't dominate the analysis, giving a fair and clear picture of the sales split.

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

ii. SKU Performance Classification

AIM: To identify which SKUs are the drivers of business, contributing the most towards the growth of the business.

Procedure: The widely used ABC analysis would help understand which SKUs dictate the sales the most. This would not only help the business in their inventory management but also help develop strong strategies to promote high-performing products. This classification was performed on the entire dataset as well as on a month-by-month basis, also it was applied both on sales data and quantity purchased.

iii. **Market Basket Analysis using Apriori Probability**

AIM: To identify products that are frequently bought together, helping us understand customer behavior.

Procedure: 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: Understand customer distribution and try to identify regions of interest.

Procedure: The dataset contains only data for 14 addresses (1.7% of total rows). Despite the limitations, useful insights can be drawn from it. The first course of action would be plotting these addresses on the map and trying to group them into the required number of clusters.

v. **ARIMA Sales Forecasting**

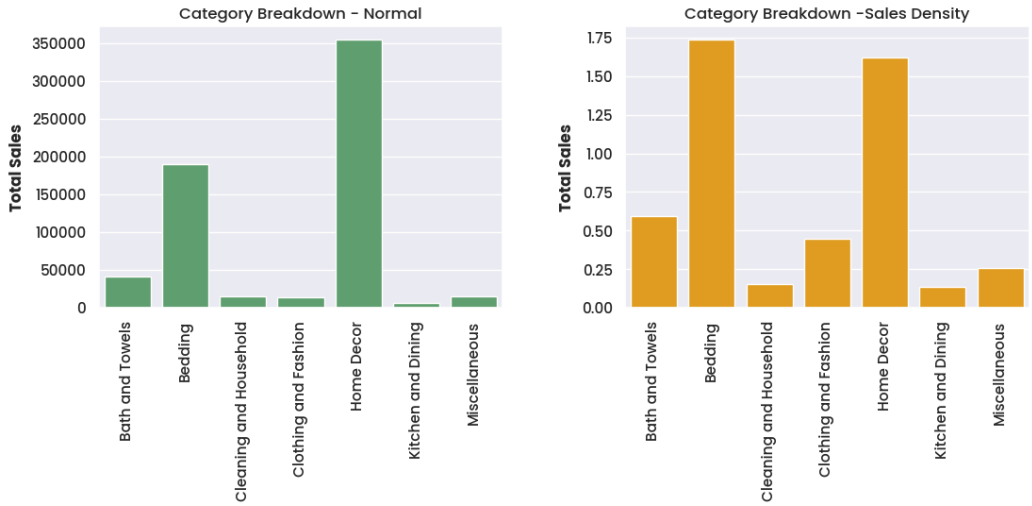
AIM: Create a model, capable of predicting the future sales values the business can experience in the following months.

Procedure: An ARIMA (AutoRegressive Integrated Moving Averages) model was fitted over the monthly sales data. Since we have days without any financial transactions in the dataset, fitting the model over daily sales data won't be easy. To resolve this issue interpolation was used to fill in those missing gaps, allowing us to go ahead with the process. For the ARIMA model parameters, *p=1*, *d=0,* and *q=1* parameters were chosen based on the ACF and PACF graphs.

3. Results and Findings

The above-explained analysis methods were applied on the preprocessed data using *Python* programming language and its libraries such as *Pandas, Numpy, Folium, Mlxtend, Networkx*, and *StatsModels*. The visualizations that are going to be presented here were created using *Matplotlib, Seaborn,* and *Tableau* software. These visualizations will help us understand underlying patterns and trends, helping us to solve these problems.

i. Categorical Influence on Sales



Explanation

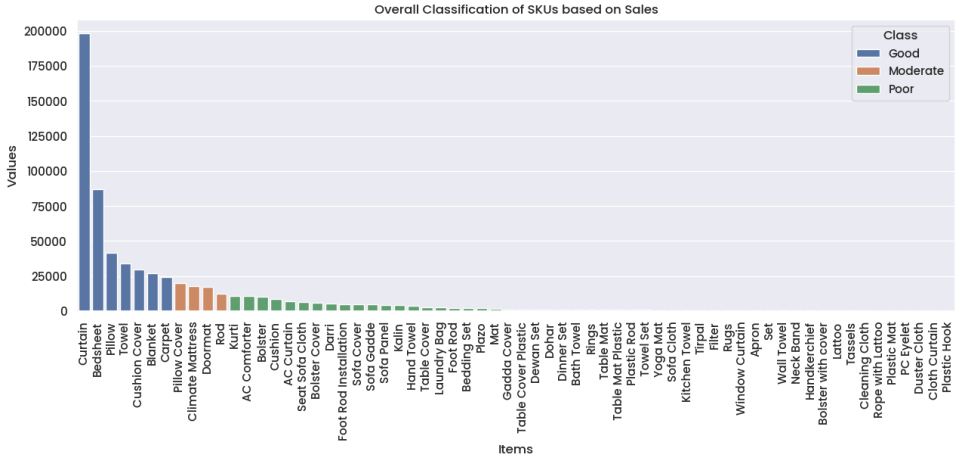
The visualization above showcases two bar charts, the first one visualizing the distribution of sales among different categories using standard methods of aggregation. Whereas, the second visualization explains the distribution of sales based on sales density of each visualisation. Sales density normalizes the sales based on the number of skus in the category.

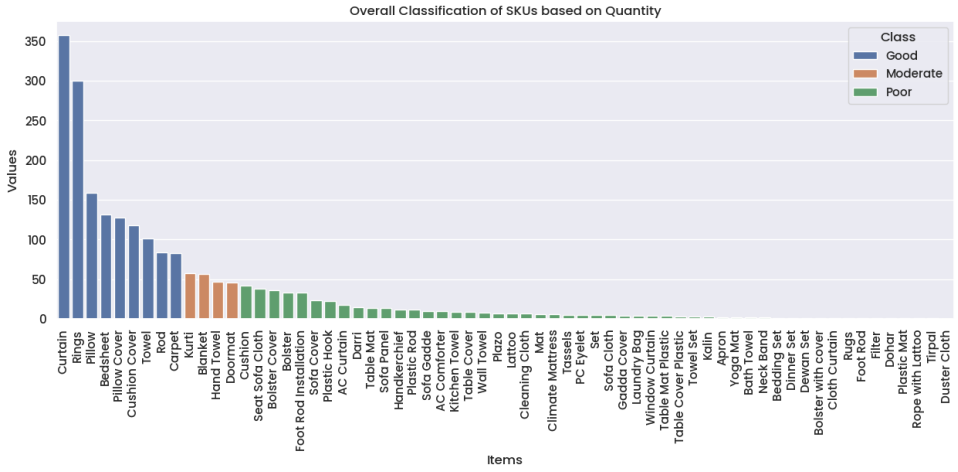
Findings

- *Bedding* category despite having fewer SKUs, holds more influence on Total Sales than the supposedly dominant *Home Décor* category, making it the most profitable category to invest in.
- The *Clothing and Fashion* category appears to be ranked among the least influential categories in the normal breakdown, but actually, it is the fourth most influential category on the chart. This means the category performs extra-ordinarily given the number of SKUs.
- *Bath and Towels* also rise as an influential category on both charts, indicating that given the number of SKUs, its performance is a bit above average.
- *Cleaning and Household*, *Kitchen and Dining*, and *Miscellaneous* categories show little improvement in the sales density graph, indicating they not only have fewer SKUs but also minuscule sales influence, meaning under-average performance.

❖ *Bedding*, *Home Décor*, *Bath and Towels,* and *Clothing and Fashion* are the driving factors of the sales, emerging as the four pillars of the business.

ii. SKU Performance Classification





Explanation

❖ The ABC analysis draws out a classification of SKUs into 3 classes: ***Good Performers*** [70% of the measure], ***Medium Performers*** [10% of the remaining], and ***Poor Performers*** [rest]. The classification is conducted based on total sales and quantity purchased respectively. Alongside overall classifications, monthly classifications are also conducted. These classifications will help the business in inventory planning and creating targeted promotions for Good and Moderate performers, and make strong and intelligent decisions for Poor Performers. Also, we need to understand that individually these classifications won't be of much use to us, rather we have to find a way to combine both these classifications to get the best results out of it. The

classifications are combined into 3 categories: *Good Sales and Good Quantity* **(GS-GQ)[Core Products]**, *Good Sales but Moderate Quantity* **(GS-MQ)[Luxury Products]**, and *Moderate Sales but Good Sales* **(MS-GQ)[Bulk Selling Products]**.

Findings

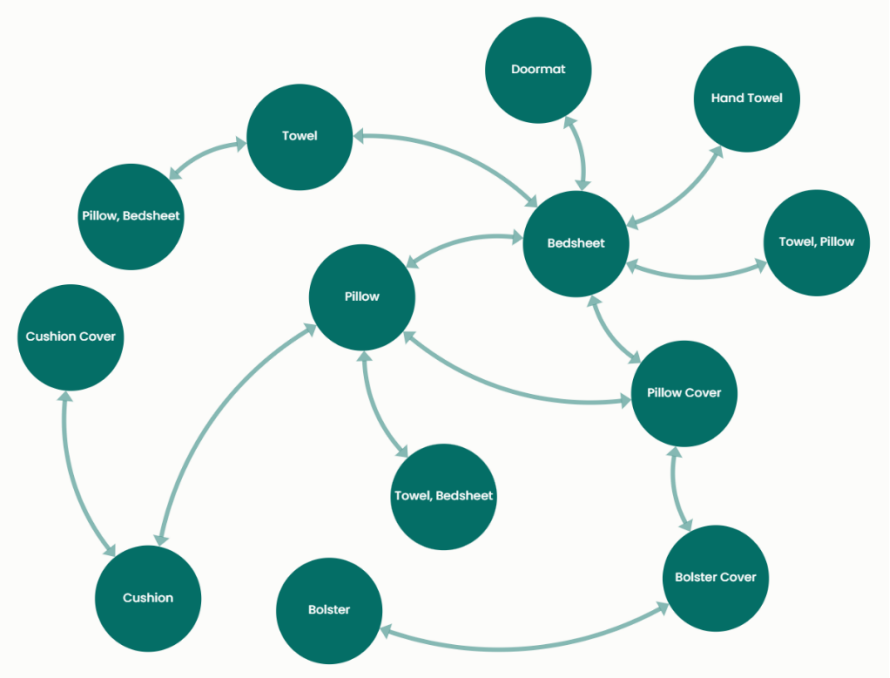
- Firstly, an overall classification of the entire dataset was conducted. This would help us understand evergreen products.

Overall Classification [March-September 2024]	
GS-GQ	Curtain, Bedsheet, Pillow, Towel, Cushion, Cover, Carpet
GS-MQ	Pillow Cover, Rod
MS-GQ	Blanket

- These classifications were then computed on a monthly basis as well, this would allow us to deeply understand the SKU performance. Also, the same combination of classifications would be used here as well.

Month	**GS-GQ**	**GS-MQ**	**MS-GQ**
March	Curtain, Towel, Bedsheet, Cushion Cover	Carpet, Pillow	Blanket
April	Towel, Kurti		
May	Curtain, Bedsheet, Cushion Cover, Towel, Kurti	Pillow	Blanket
June	Curtain, Rod, Blanket	Kurti	AC Curtain
July	Curtain, Bedsheet, Carpet, Pillow	Cushion Cover, Seat Sofa Cloth	Towel, Sofa Gadde
August	Curtain, Bedsheet, Pillow	Pillow Cover, Cushion Cover	
September	Curtain, Rod		Bedsheet, Pillow

iii. Market Basket Analysis using Apriori Probability



Network Graph of the Market Basket Analysis

Explanation

This graph visualizes the connections between the SKUs that are most likely to be bought together. A two-way array between two SKUs means that they have a connection and are probable to be bought together. The graph has been created using *Kumu* software, a great tool to create graphs and connection charts. The pairs that have been displayed here have a *confidence* value of >0.2 and a *lift* of >1, *confidence* and *lift* being two measures of the likelihood of being connected. To compute these connections the *"mlxtend"* library of Python was used. Association Rules were computed by the Apriori Probabilities using the library.

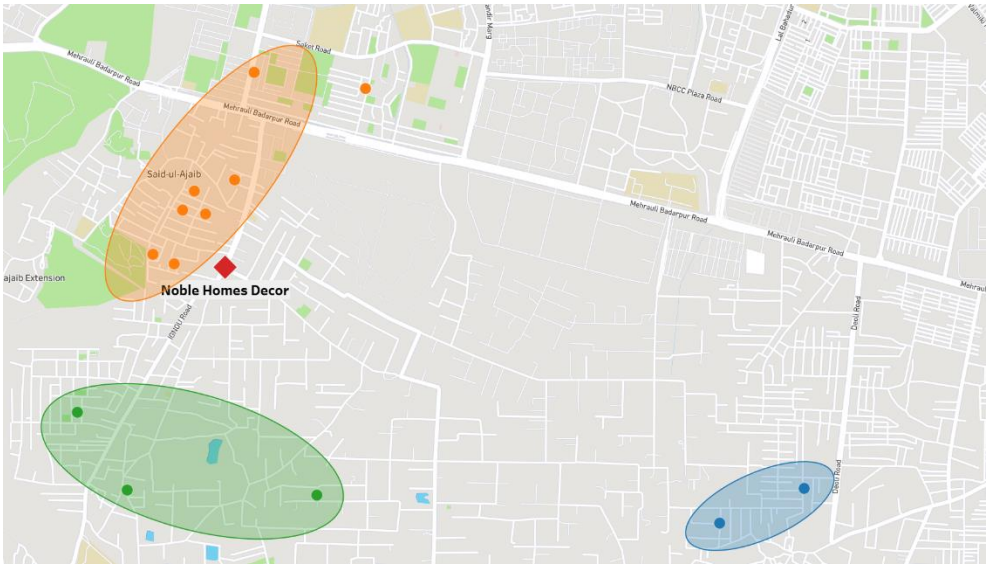
	antecedents	consequents	support	confidence	lift
0	(Doormat)	(Bedsheet)	0.016393	0.235294	1.290152
2	(Hand Towel)	(Bedsheet)	0.012295	0.285714	1.566613
4	(Pillow)	(Bedsheet)	0.047131	0.242105	1.327499
5	(Bedsheet)	(Pillow)	0.047131	0.258427	1.327499
6	(Pillow Cover)	(Bedsheet)	0.022541	0.211538	1.159896
8	(Towel)	(Bedsheet)	0.030738	0.211268	1.158411
10	(Bolster)	(Bolster Cover)	0.024590	0.600000	15.410526
11	(Bolster Cover)	(Bolster)	0.024590	0.631579	15.410526
13	(Bolster Cover)	(Pillow Cover)	0.014344	0.368421	3.457490
17	(Cushion)	(Cushion Cover)	0.016393	0.400000	2.870588
19	(Cushion)	(Pillow)	0.010246	0.250000	1.284211
21	(Pillow Cover)	(Pillow)	0.026639	0.250000	1.284211
24	(Pillow, Towel)	(Bedsheet)	0.010246	0.714286	3.916533
25	(Pillow, Bedsheet)	(Towel)	0.010246	0.217391	1.494182
26	(Towel, Bedsheet)	(Pillow)	0.010246	0.333333	1.712281

As a result, 15 such connections with confidence and lift values satisfying the required amount were discovered. These pairs were then imported into the *Kumu* software and required connections were made.

Findings

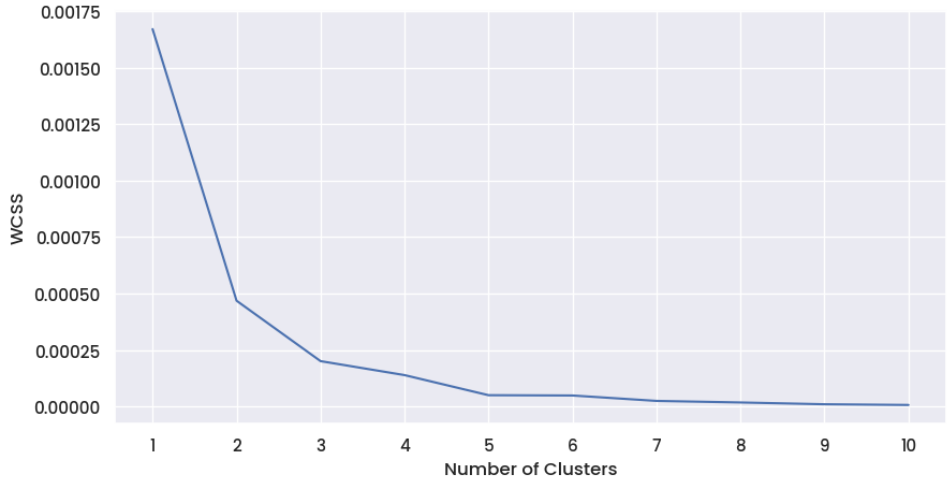
- *Bedsheet* is the central connection of the network. It is frequently bought alongside *pillows, pillow covers, towels, hand towels,* and *doormats*.
- There are some obvious pairs likes *Pillow* and *Pillow Covers*, *Cushion* and *Cushion Covers*, *Bolster* and *Bolster Covers*. These pairs are self-explanatory, now having the backing of analysis.
- There are few interesting pairs such as *Bedsheets with Doormats*, *Pillow Covers with Bolster Cover* and *Pillow with Cushion*, they can be useful for cross-promotion.

iv. Customer Geo-Mapping

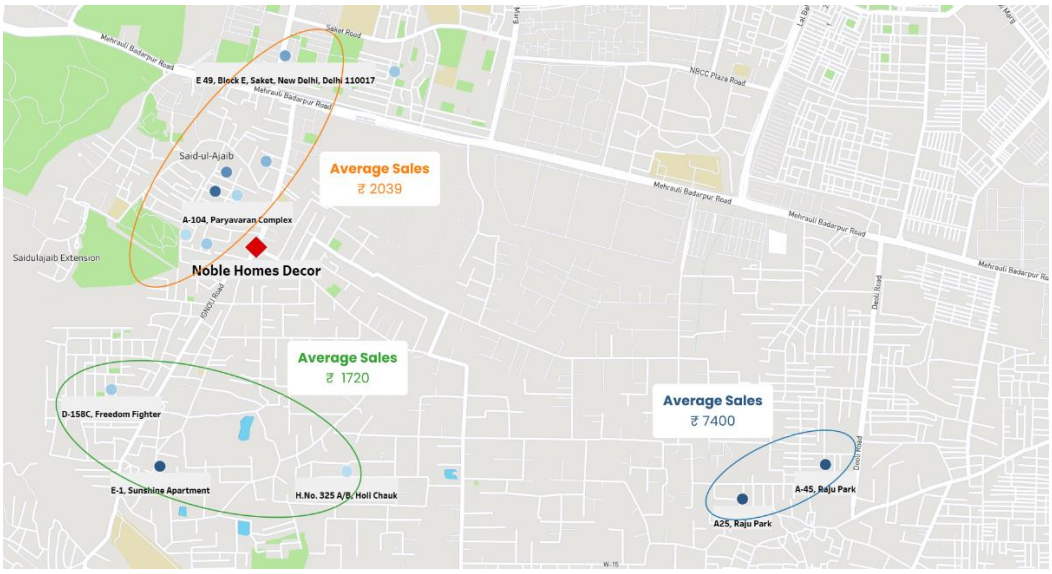


Explanation

The dataset I am using contains information about addresses for just 14 transactions making for just 1.71% of the total row count. Using this sparse data the above map was created by plotting the longitudes and latitudes of those addresses on the map. Then these data points were put under the KMeans clustering algorithm to check if they form any clusters, and surely, they do. To decide what number of clusters to form, the Elbow Curve method was used.



As it is clearly visible on this graph, there is a sudden dip in the WCSS value at 3 clusters, after that it more or less smooths out. This is called the "**Elbow**" of the graph and the point at which it is evident is the appropriate number of clusters to use. This allows us to partition our address locations into 3 regions of interest. Now let's look at each of their values to our business, this is measured using the average total sales of that region.

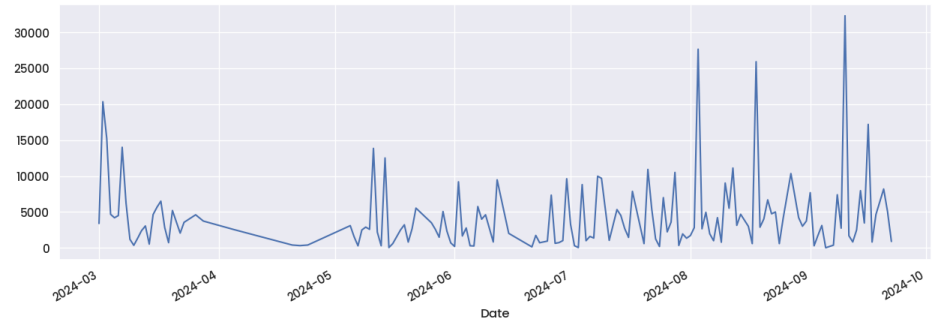


Findings

- The *Orange Cluster* without a doubt has the highest number of customers present. Being much closer to the shop than other clusters is also a potent factor behind this finding.
- The *Orange Cluster* though having the largest customer base, pales in comparison to the *Blue Cluster* when total sales are at question. This indicates that the *Blue Cluster* is yet another important region of interest.

- ❖ The *Green Cluster* on the other hand has fewer customers but contains a reasonable average sales amount value to validate it. Meaning this cluster can be considered as a moderate region of interest.
- ❖ On further, investigation with the founder, it was discovered that the region enclosed by the *Blue Cluster* [Raju Park] is one of the rich sectors of the district.
- ❖ Also, on investigating the *Orange Cluster*, I was informed that the region belongs to a large complex of flats and apartments called the *Paryavaran Complex*. Though not on the rich side, several customers from this region have visited the shop after being referred to by their know people in the neighborhood.

v. ARIMA Sales Forecasting



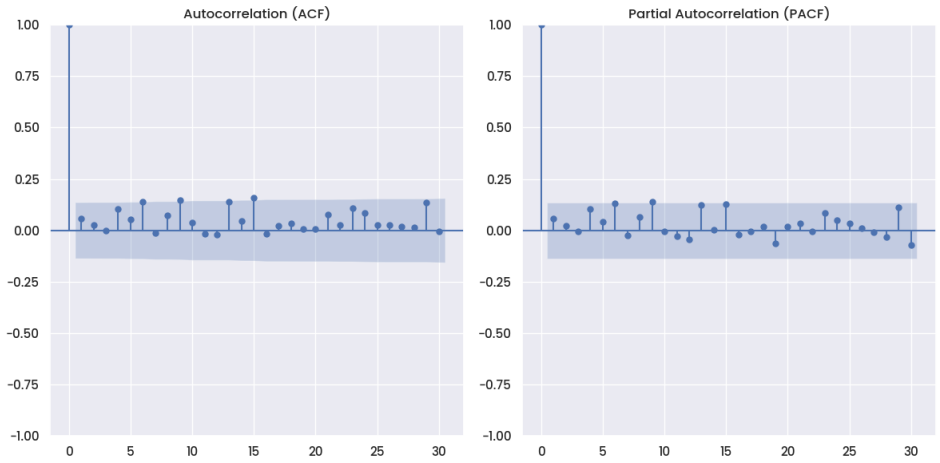
Explanation

ARIMA Time Series Forecasting is a well-recognized branch of Forecasting Models. Having the data for about 7 months, it should be possible for us to prepare a well-functioning ARIMA model using Daily Sales data. But the problem here is that our data is not continuous, we have several days with no sales data. As a result, we have this bizarre chart of daily sales displaying massive fluctuations. To solve this issue, I decided to **interpolate** the missing values. Here rather than filling the missing data values with 0, use interpolation to estimate sales on those dates.

ADF Statistics and p-value were computed in order to determine if the data is compatible enough to create an ARIMA model out of it. A negative ADF Statistic and p-value < 0.05 are necessary for a good fit.

Test Results	
ADF Statistic	-13.426574
p-Value	0.000000

Next, we have to determine the parameters for our ARIMA model. For that PACF and ACF graphs would be useful.



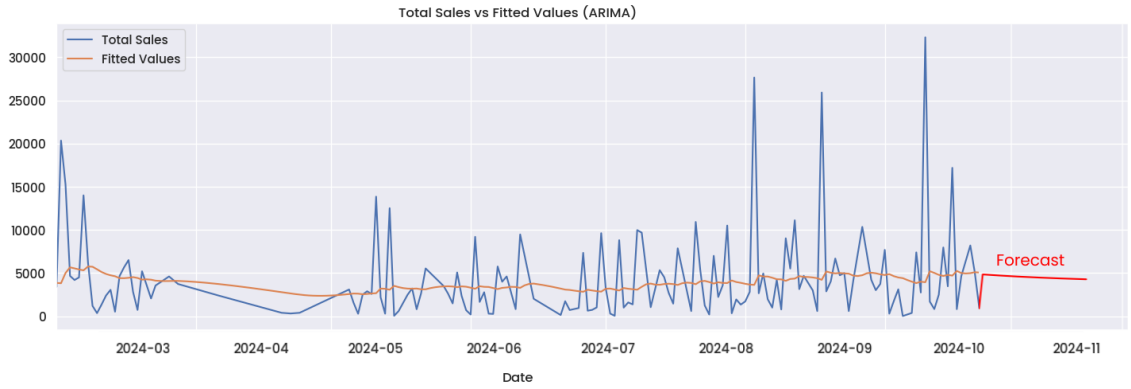
The parameters at stake here are AR(p), d, and MA(q).

- The AR(p) parameter would be 1, because of just one significant spike in the PACF graph.
- The MA(q) parameter would be 1, because of just one significant spike in the ACF graph.
- The d parameter or the Degree of Differencing can be assumed to be 0 as the dataset looks pretty stationary.

Finally, the ARIMA model is trained and here we can see the results.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:            Total Sales   No. Observations:                  206
Model:                 ARIMA(1, 0, 1)   Log Likelihood               -2020.687
Date:                Tue, 15 Oct 2024   AIC                           4049.375
Time:                        02:21:33   BIC                           4062.686
Sample:                    03-01-2024   HQIC                          4054.758
                         - 09-22-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       3861.6815    714.015      5.408      0.000    2462.237    5261.126
ar.L1          0.9710      0.045     21.713      0.000       0.883       1.059
ma.L1         -0.9259      0.071    -13.126      0.000      -1.064      -0.788
sigma2     1.942e+07      0.514   3.78e+07      0.000    1.94e+07    1.94e+07
===================================================================================
Ljung-Box (L1) (Q):                   0.06   Jarque-Bera (JB):              2206.15
Prob(Q):                              0.81   Prob(JB):                         0.00
Heteroskedasticity (H):               3.96   Skew:                             3.30
Prob(H) (two-sided):                  0.00   Kurtosis:                        17.61
===================================================================================
```

The p-values for AR, MA, and sigma values are all < 0.05 meaning our ARIMA model is pretty solid and is capable of understanding the underlying patterns in the data.



The graph above shows that the model had a good fit on the data. The red line is the forecast line.

4. Interpretation of Results and Recommendations

All the above-conducted analysis allows us to take a dive into this ocean of data and bring out pearls of insights that not only would give us a much better understanding of the business but also result in effective data-driven decision-making.

Below are the insights and the recommendations derived from them:

i. Categorical Influence on Sales

- The *Bedding, Home Décor, Bath and Towel,* and *Clothing and Fashion* are the chief sales driving categories of the business.

  Recommendation: Invest more resources and promotional activities on these four categories. It would be beneficial to invest more on our string points rather than equally distribution.

- Despite having fewer SKUs, *Bedding* and *Clothing and Fashion* have shown exceptional performance.

  Recommendation: Focus on increasing the number of SKUs in these categories, especially in clothing. It was observed that a lot of times customers made standalone *Clothing and Fashion* purchases indicating strong demand for them.

- *Cleaning and Household*, *Kitchen and Dining*, and *Miscellaneous* categories show little improvement in the sales density graph, indicating they not only have fewer SKUs but also minuscule sales.

Recommendation: Try running promotional offers where heavy discounts or free items are given along with every purchase of items belonging to a high-valued category. This will not only allow us to leverage the demand for those high-valued products to promote the under-performers but also boost the sales of the high-valued products as well.

ii. SKU Performance Classification

- On the overall scale of 7 months *Curtain, Bedsheet, Pillow, Towel, Cushion Cover,* and *Carpet* are those products that have good sales and good quantities purchased. I will label them as **Core Products.**

  Recommendation: Ample stocks of these core products should be kept in the inventory at all times. The business should decide upon the effective re-ordering points to ensure a constant availability of these items.

- Out of all these core products *Curtains* seem to dominate both the sales and quantities purchased metrics, making it the highest-valued product in our inventory.

  Recommendation: The business should focus on creating its overall brand image as the **"Best Shop for the Best Curtains"**. Rather than focusing their resources on several underperforming SKUs, focus entirely on curtains. The inventory should have an incredibly wide range of options for curtains, ranging from mid-range to premium quality products. As it is evident that the business is already being recognised by their curtains, the shop's interior should also be adjusted in such a way that it communicated that wide range of options for curtains the moment the customer walks in. Also, they can even start selling customized curtains on special customer requests for a suitable price. The upcoming festival season is a critical opportunity for the business to monopolize the curtain market and establish themselves as a big player on this niche market. They should run massive marketing campaigns to communicate this with the customer base and attract them.

- *Pillow Covers* have good sales but lacks in purchased quantity, this means that pillow covers may be purchased along with *Pillows,* which is a GS-GQ product, sometimes.

Recommendation: The possibility that Pillow Covers are being purchased along with Pillows can give us the opportunity to bundle them together as a combo offer. This may boost the sales of both products.

- *Blankets* have good purchased quantity but moderate sales, this may indicate that most of the blankets are being sold in large bundles to another organization.

  Recommendation: Blankets being sold at wholesale price to other firms can be considered good business, but we can leverage these business connections to create further business opportunities. We can provide them with product samples of bedding products like *Pillows, Bedsheets*, etc, this would allow us to market these products to our business partners.

- *Bedsheet* is a GS-GQ product in the month of March, May, July & August. It seems that there is no seasonality that dictates the sales of Bedsheets, this means it is an evergreen product and we would be able to sell it no matter the season.

  Recommendation: *Bedsheet* is a high-value product just like *Curtains* that will be in demand no matter the season. To boost up sales even more we should have a wide variety of bedsheets of all price ranges in our inventory.

- In the month of June, we can see that *AC Curtains* were in demand being purchased in large quantities but the sales remained on the lower end. This would again indicate that these products were sold in mass to another business at wholesale price. The month of July is the hottest month of the year, which would cause extensive usage of Air Conditioners, this would then raise the demand for *AC Curtains*.

  Recommendation: To increase sales, we can introduce discount offers and marketing campaigns to attract customers. We should have AC Curtains of varying sizes and designs in our inventory, with ample stock to satisfy the demands.

iii. Market Basket Analysis using Apriori Probability

- There are some obvious connections between some pairs of products like *Pillow* and *Pillow Cover*, *Bolster* and *Bolster Covers* (extremely strong connection), and *Cushion* and *Cushion Covers*.

Recommendation: We can bundle these products together and market them as a combo offer, or we can even put a small amount of discount on these combos to boost up the sales even more.

- *Pillows* and *Bedsheets* are often bought together and have a strong connection. This connection is understandable as people would be keener to buy pillows along with bedsheets.

  Recommendation: Here bundling items as a combo won't be a wise decision because both Bedsheets and Pillows are GS-GQ (Good Sales and Good Quantity Purchased) products, bundling together at a lower price would result in loss. Instead, we can give some discount when a customer wishes to purchase both bedsheets and pillows. This would also create a positive brand image with the customers.

iv. Customer Geo-Mapping

- The *Orange Cluster* has the largest number of customers without a doubt, but on the other hand, the average sales are quite mediocre. This indicates that perhaps middle-class families reside in that region.

  Recommendation: We should provide free home delivery services to this region. The reason being first, the area is right next door and free delivery won't cost anything. Secondly, providing these services would create a positive image for the brand, also it would help in the promotion of the shop by word-of-mouth.

- The *Blue Cluster*, is quite interesting, though fewer in number the average sales are the highest. This indicates that the region may have rich households.

  Recommendation: Due to the presence of rich households, it is necessary to form a good brand image in that region. This can be accomplished by aggressive marketing in that region, providing delivery services at a cheap cost, and introducing a loyalty program, where customers would be given discounts or free items if they collect enough loyalty points.

- The *Green Cluster* is quite normal, it contains fewer addresses and decent average sales. This would mean this area is a bit like the *Orange Cluster*, with middle-class families.

Recommendation: We should apply the same measures I mentioned above for the *Orange Cluster*, the reason being they both may have customers with the same level of economic status. But providing free delivery services may be a topic to have a proper discussion about.

- The fact that we had only 14 rows of data to conduct this analysis testifies that the business does not follow a data-driven approach to management.

  Recommendation: The business should start collecting more information about its customers to better understand their behavior patterns. This is crucial to ensure a strong data-driven approach to business management.

v. ARIMA Sales Forecasting

- The forecast values indicate a potential decline in total sales in the coming months on (October and November).

  Recommendation: The coming months will be the season of festivals and the business should experience significant growth in that time frame. To ensure superior sales in this period proper marketing strategies need to be employed with efficient inventory management. All the above-mentioned marketing strategies and inventory management could help the business maximize sales.

- The ARIMA model indicates significant fluctuations in sales, suggesting that some external factors (seasonality, holidays, or marketing efforts) may be impacting customer behavior.

  Recommendation: To tackle this, we need to implement a flexible stock management system that adjusts inventory levels based on forecasted demand. In periods of low sales, consider introducing flash sales or limited-time offers to drive customer interest and smooth out the fluctuations.

5. Conclusion

The analysis of Noble Homes Décor has provided valuable insights into the sales patterns, product performance, categorical influence on sales, customer behavior, and marketing strategies, contributing to a deeper understanding of the challenges and opportunities the business faces. By leveraging the data, we were able to identify several key trends and recommend actionable strategies to improve performance.

First, the sales data analysis confirmed that sales fluctuated significantly across different months, with certain peak periods, especially around seasonal and festive times.

The ABC analysis effectively classified SKUs into high, moderate, and low contributors to overall revenue. The insight gained from this analysis should inform better stocking decisions and identify regions of focus. By ensuring that the top-selling SKUs are always well-stocked and highlighted, Noble Homes Décor can maximize revenue from its best-performing items.

Next, the Market Basket Analysis allows us to understand the customer purchasing behavior. We identified the SKUs that are frequently bought together, forming 15 pairs using Python's *mlxend* library.

In terms of customer clusters, the geo-mapping analysis revealed distinct behavior patterns in different regions. Targeted marketing strategies such as offering free delivery to clusters with a larger customer base and relatively low average sales, or loyalty programs in regions with fewer but high-value customers, can help increase repeat purchases and improve customer retention.

The ARIMA forecasting model indicates a potential decline in sales in the coming months if no intervention is made. To mitigate this, the shop should capitalize on the insights from the analyses and implement more aggressive marketing campaigns in both offline and online channels.

In conclusion, Noble Homes Décor has significant potential for growth and profitability through better inventory management, targeted marketing strategies, and leveraging customer insights. While challenges like fluctuating sales and increased competition exist, adopting a data-driven approach will provide the business with a competitive edge in optimizing both sales and customer satisfaction.