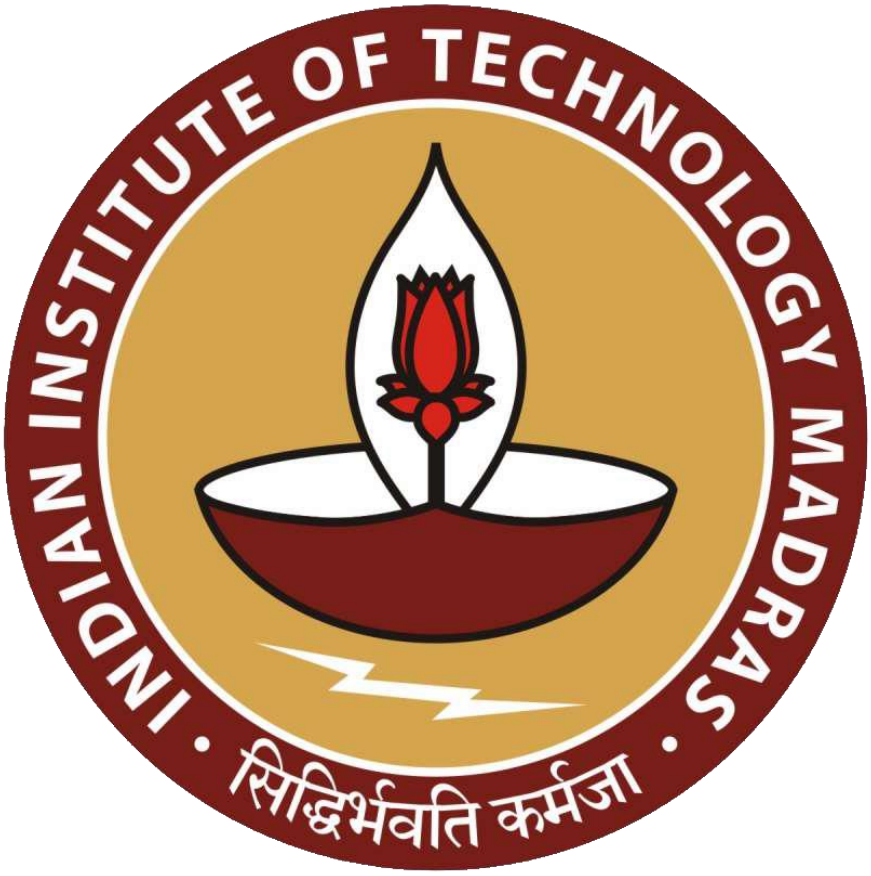
Optimising Operations: Strategies to Overcome Insufficient Workforce, Superfluous Inventory and Inadequate Clientele of a Wholesale Company.

A Midterm Report for BDM Capstone Project

Submitted By:

Ikshwaku Tiwari

21f1003999



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

Executive Summary and Title

Title for the mid-term submission: Optimising Operations: Strategies to Overcome Insufficient Workforce, Superfluous Inventory and Inadequate Clientele of a Wholesale Company.

This midterm report addresses the key challenges faced by Nalanda Dresses, including inventory management, labour allocation, and customer acquisition. The first explanatory section shows and proves the physical existence of the data records and the organisation. Further, the metadata and the descriptive statistics show the data distribution profile, both qualitative and quantitative data collected namely, the sales data, employee distribution and product details. Various data visualisation plots like Box-plot for the sales distribution of the customer companies, Histograms and Bar-plots for the items and cities distribution, etc are created, that show how the data varies, its central tendency, and the maximum and minimum values of each kind of collected data.

Preliminary data analysis is conducted on the sales data, that existed as physical records and were later converted to digital, which include the sales trend over 10 months of data. This plot shows the variance, peaks and drops in sales. The top high-value company customers are shown by creating a tree map, and lastly how the items and sales are related by creating a scatter plot and regression line. A brief description of the ML model that gives positive results is provided. A baseline Linear Regression model was developed to explore this relationship, achieving a mean squared error (MSE) of 0.2014. Advanced machine learning techniques, such as Gradient Boosting, significantly reduced the MSE to 0.1251, indicating improved predictive performance for sales forecasting.

Proof of Originality of Data:

Name of the company: Nalanda Dresses.

Owners: Mr Subodh Shah and Mr Mehul Shah.

Address: 441/4, Nalanda Dresses, Near Avtar Meher Baba Poona Centre, Before KEM Hospital Pune, 411011

Below are a few images clicked that show the company workspace.

Image 2.1: Interior of the shop.

Image 2.2: Display section in the office of the owners.

Image 2.3 Front view of the shop.

The video interaction, letterhead from the organisation, data collected and a few more pictures of the company are in the following drive:
https://drive.google.com/drive/folders/1BSJNzBLUrf5kBM0Mxj5mikSXSQQdMjlf?usp=drive_link

The collected and cleaned data is stored in the Excel sheets:

https://docs.google.com/spreadsheets/d/1Se30g22Tyavg6IwAQe6i3YboaaY7I3y9/edit?usp=sharing&ouid=104995185334591017326&rtpof=true&sd=true

Metadata and Descriptive Statistics:

Column	Data Type	Description	Example
Date	Date	The date of the transaction in 'DD-MM-YYYY' format.	'18-07-2024'
Company Name	String	Name of the company that placed the order.	'Seemati'
City/Town	String	Location of the company's warehouse.	'Chennai'
Sales	String (Currency)	Total revenue generated from the transaction.	'52,485.00'
Items	Float	Number of the items sold in the transaction.	'114.00'

Table 3.1: Sales Data of the Company

Description:

- The first data sheet contains the sales figures provided by the company owners. It comprises 6 columns (including the serial number) and 143 rows.
- The first column is the dates at which the transactions took place. The data is collected from August 2023 to December 2023, and then from April 2024 to August 2024, making it a total of 10 months of sales data.
- The next columns are the company names (their warehouses) and the names of the cities and towns that they are located in. There are a total of 27 warehouses and a total of 17 locations.
- The last 2 columns indicate the sales of each transaction and the dress items sold during the transactions.

Column	Data Type	Description	Example
Jobs	String	Specific roles assigned to workers in production.	'Stitching'
Number of Employees	Integer	Total number of employees assigned to the particular job.	2

Table 3.2: Employee Data

Description:

- The second data sheet consists of the qualitative data collected by observance and interactions.

- There are 3 columns (including the serial number), a total of 8 employees, some with more than one job assignment.

Column	Data Type	Description	Example
Material/Dress	String	The type of fabric used in production/The different kinds of dresses the firm manufactures.	‘Satin’
Price-Range (INR)	String (Range)	The selling price range of the dresses.	‘430-450’

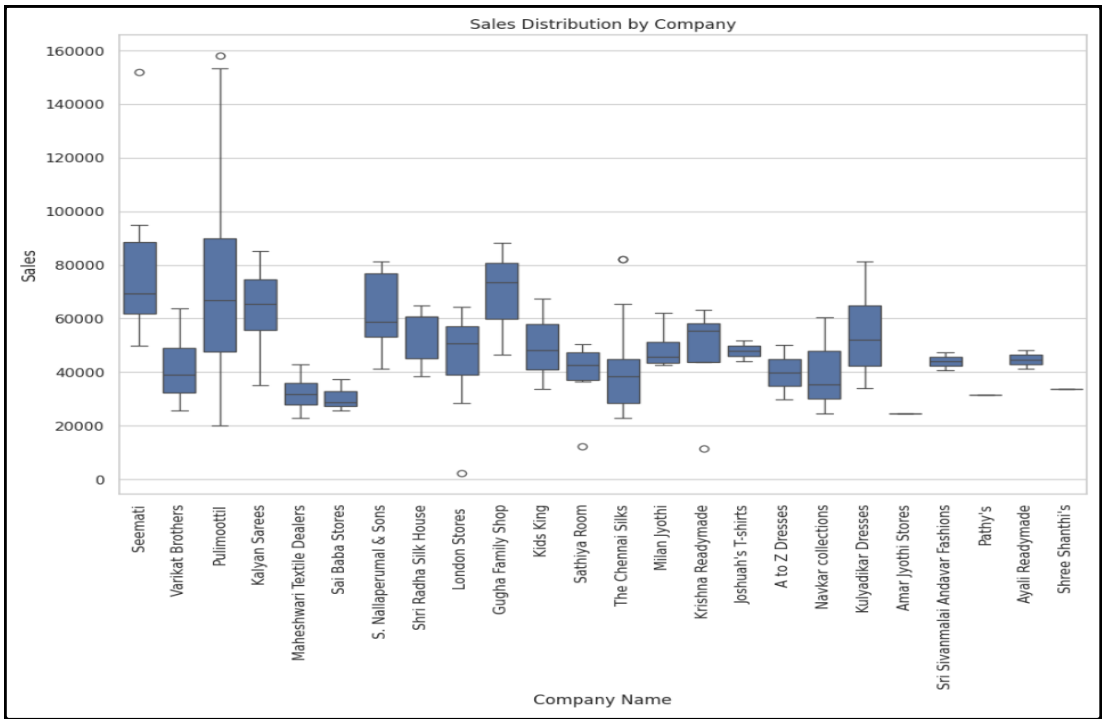
Table 3.3: Dress Material Data

Description:

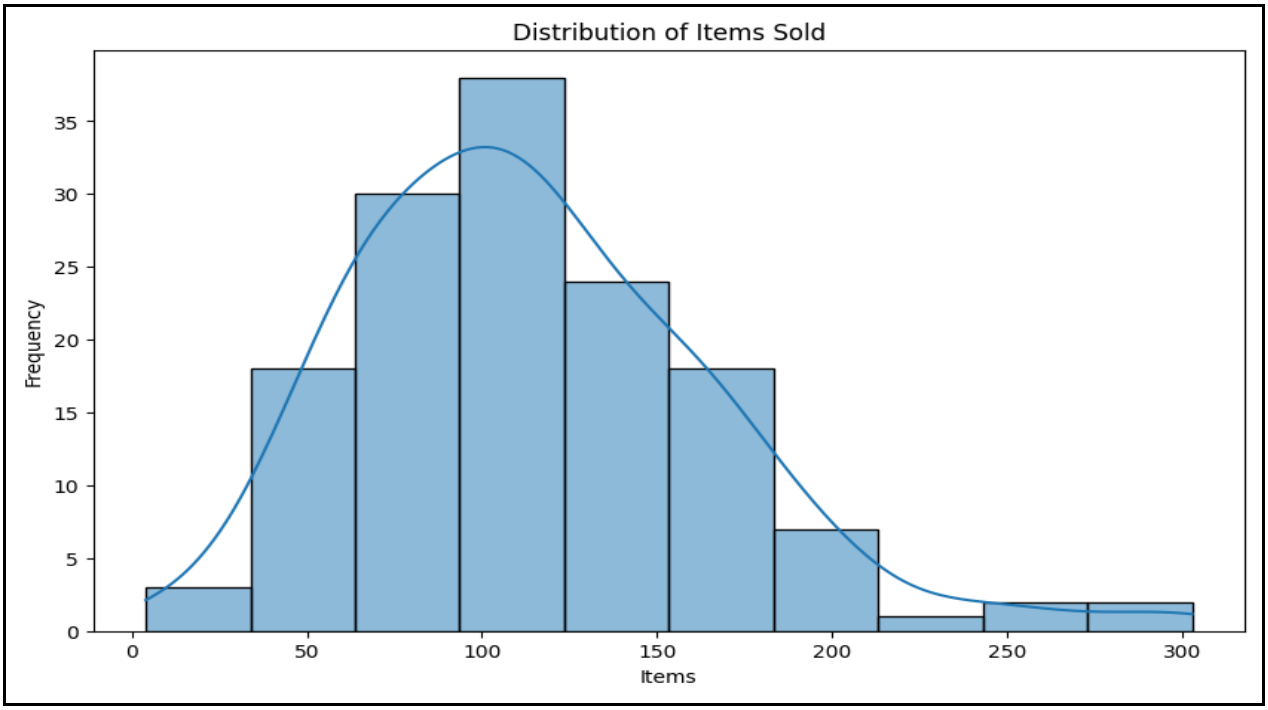
- The final data sheet describes the types of dresses sold, which depend on the material used to produce the dress.
- The 2 columns in this dataset indicate the types of materials used and the price range at which the final product is sold.

The following graphs show the descriptive statistics for the corresponding data tables created. The image numbers can be linked with the above metadata provided. The graphs were created using the Seaborn and the Matplotlib libraries in Python.

Graph 3.1.1 shows the sales figures for each consumer company. The box plot shows how the sales figures are concentrated showing companies with maximum sales like Pulimoottil and companies with few transactions and minimum sales like Pathy’s. The graph's visible outliers indicate poor sales or promotional sales.

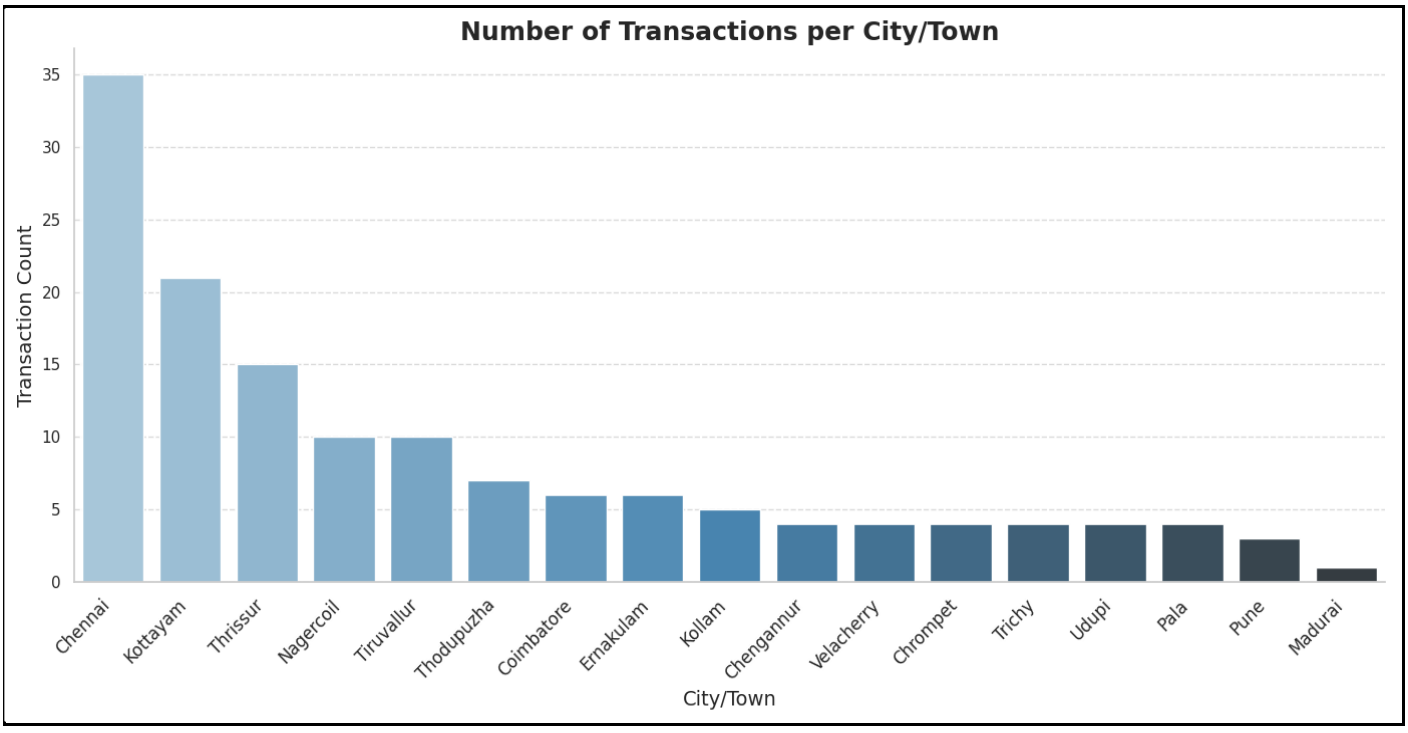


Graph 3.1.1: Box Plot for the sales distribution across the companies.

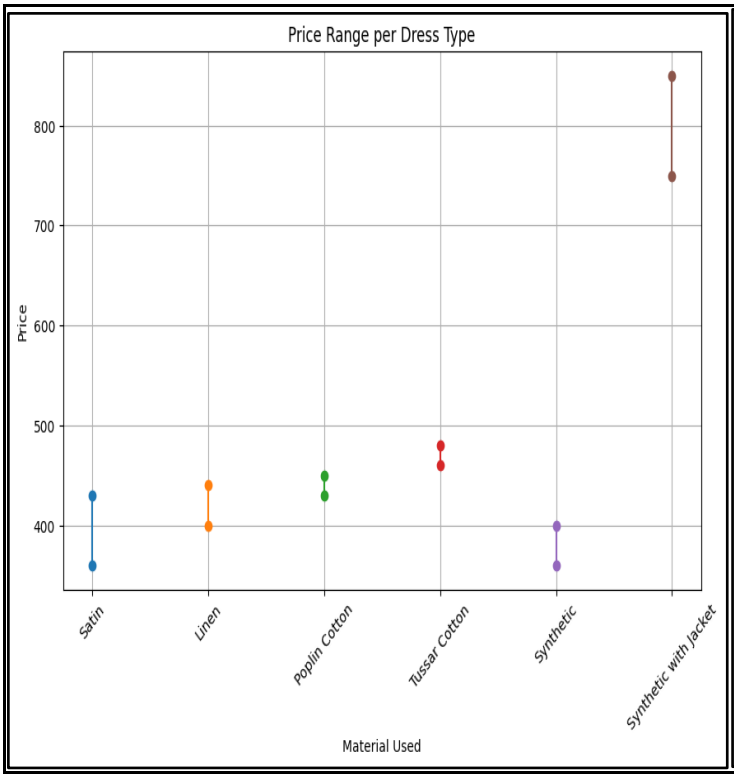


Graph 3.1.2: Histogram of the items sold.

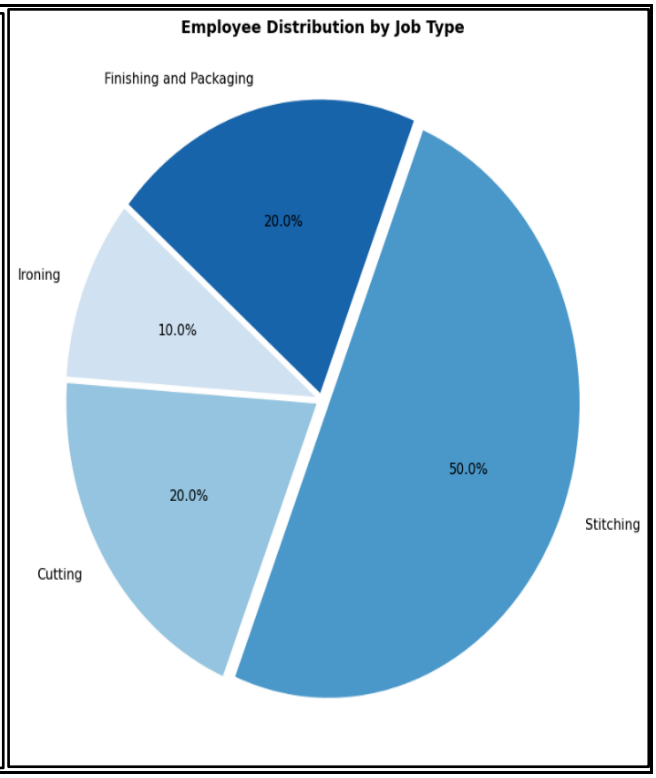
Graph 3.1.2 shows the sales volume and distribution. The graph is roughly right-skewed, peaking at around 100 items. Graph 3.1.3 shows the transactions per city, with Chennai having the maximum transactions.



Graph 3.1.3: Bar Chart for the cumulative transactions for each city.



Graph 3.2: The price range of each material.



Graph 3.3: Employee Job Distribution

Graph 3.2 is a dot plot that signifies the different ranges of dresses manufactured by the company. Most dresses have a similar range apart from the premium synthetic dresses with jackets.

Graph 3.3 shows the labour distribution in the company, with ‘stitching’ having the maximum workforce contribution. The pie chart visualises the distribution accurately due to limited categories and makes it easier to interpret the relative share of each category.

Data Provenance and Analysis Methods:

The organisation's problems included inventory management, labour and customer acquisition. The primary data available from the organisation was the sales figures. There were no records of inventory replenishment. Halting the production of dresses would result in labour loss.

The sales figures were recorded in physical ledgers. Databases or digital records weren’t available. The photographs of the data were collected, transcribed manually and cleaned in Excel sheets. The data collected ranges from August 2023 to December 2023, and then from April 2024 to August 2024, making it a total of 10 months of sales data. There is a break in the range due to the unavailability of data for that timeframe.

Considering the excess inventory challenge, the viable method explained in the proposal was creating a demand forecasting model using sales prediction. The sales data has viable features that contribute towards building the model. To check whether a linear relationship exists between the ‘Items Sold’ and the ‘Sales’, a scatter plot with a regression line fit is created.

Further, to see the sales trend throughout the timeframe of available data, a line plot with separate plots for August 2023-December 2023 and April 2024-August 2024 is created. This plot highlights various factors that may influence peaks and drops in sales, like seasonal trends, promotional sales, etc.

Using this information a Machine Learning model is created that considers sales as the target variable. Describing the model creation process, extracting the days and months from the date feature contributes to

adding new features to the model and encoding the categorical variables (Companies and Cities) into a numerical format. Since the target variable consists of greater values, a log transformation on the sales is done to evaluate the model in transformed space.

Linear regression was used as a baseline model to establish whether a linear relationship exists between the log-transformed sales data and the rest of the features. The equation for Linear Regression after log transformation becomes:

$$log(1+Sales) = \beta 0+\beta 1X1+\beta 2X2+\cdots+\beta pXp+\epsilon,$$

Where log(1+Sales) is the dependent (target variable). MSE (mean square error) is used as a performance metric for this baseline model.

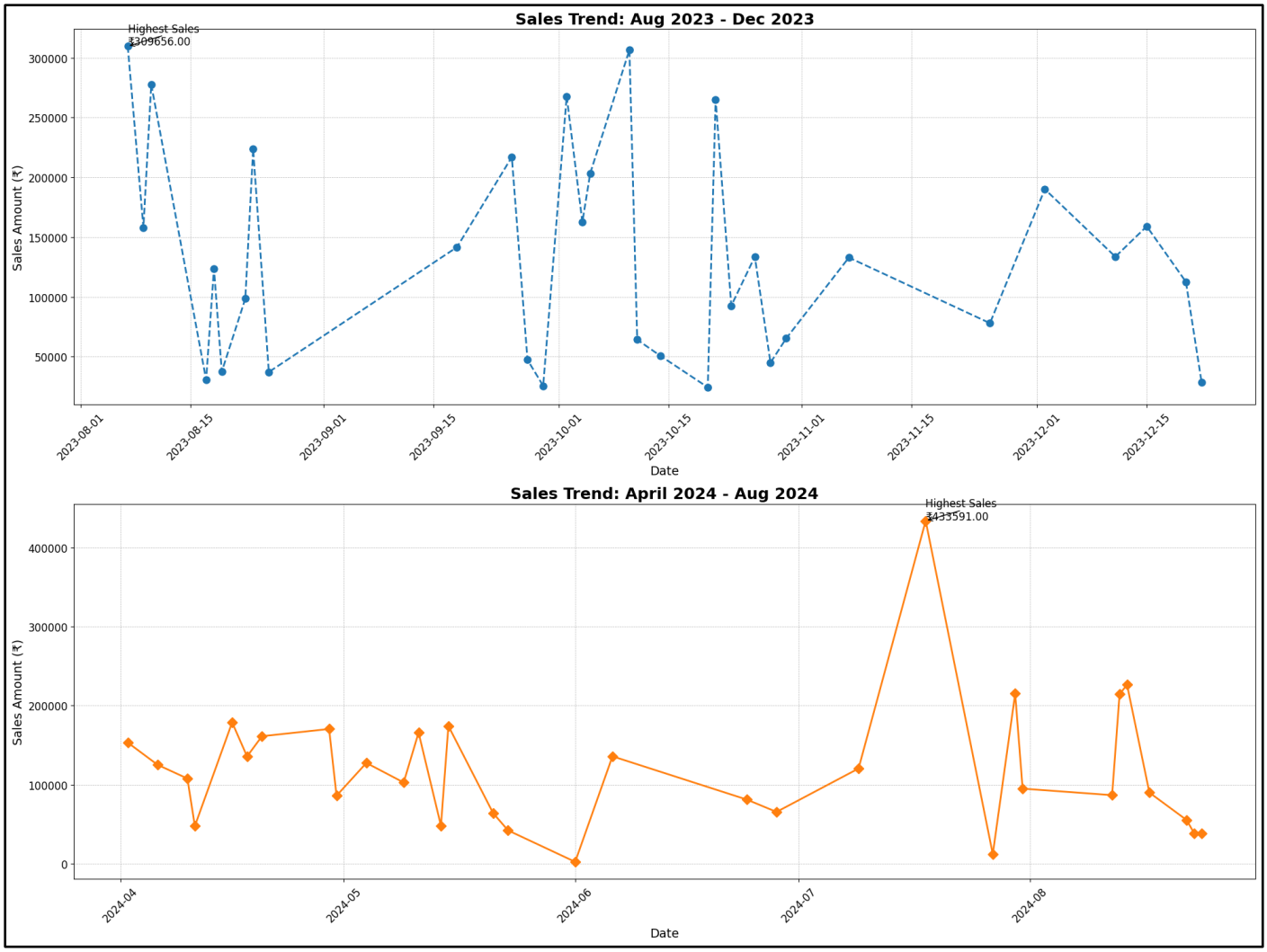
Proceeding further, to understand the company's clientele a treemap has been plotted to understand the sales of the best 5 companies. The treemap is used for easier interpretation (since the tree map uses 'space' as a measure) of the sales volume of each company. This will tell which companies have the highest sales and the geographical location corresponding to each company.

The final data collection method was purely done qualitatively. To understand the labour distribution in the company, the total number of employees and the roles assigned to them were collected. Since the employees were unable to answer web surveys, audio interactions with different employees were done to understand their sentiments and a survey of questions (in the local language) was provided to each of them. The following is the survey link that was printed and provided to them:
https://docs.google.com/forms/d/e/1FAIpQLSf7hmHPWcRnGtUaGa5LbEkhn_4cYpgnQCPaqE4br9TWQTPAkQ/viewform?usp=sf_link

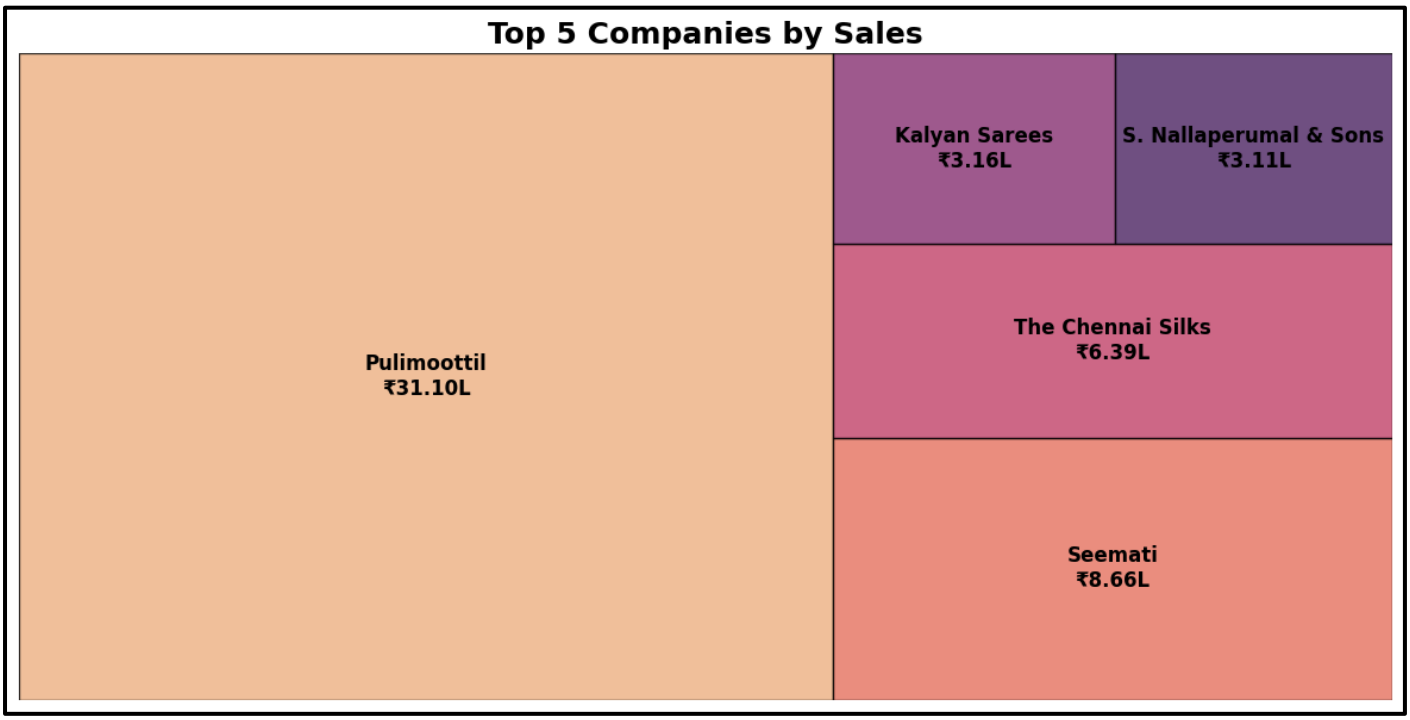
Conducting sentiment analysis using the responses will help understand factors such as employee turnover rate, retention factors, job satisfaction levels, and skills required for hiring new employees.

Results and Interpretation:

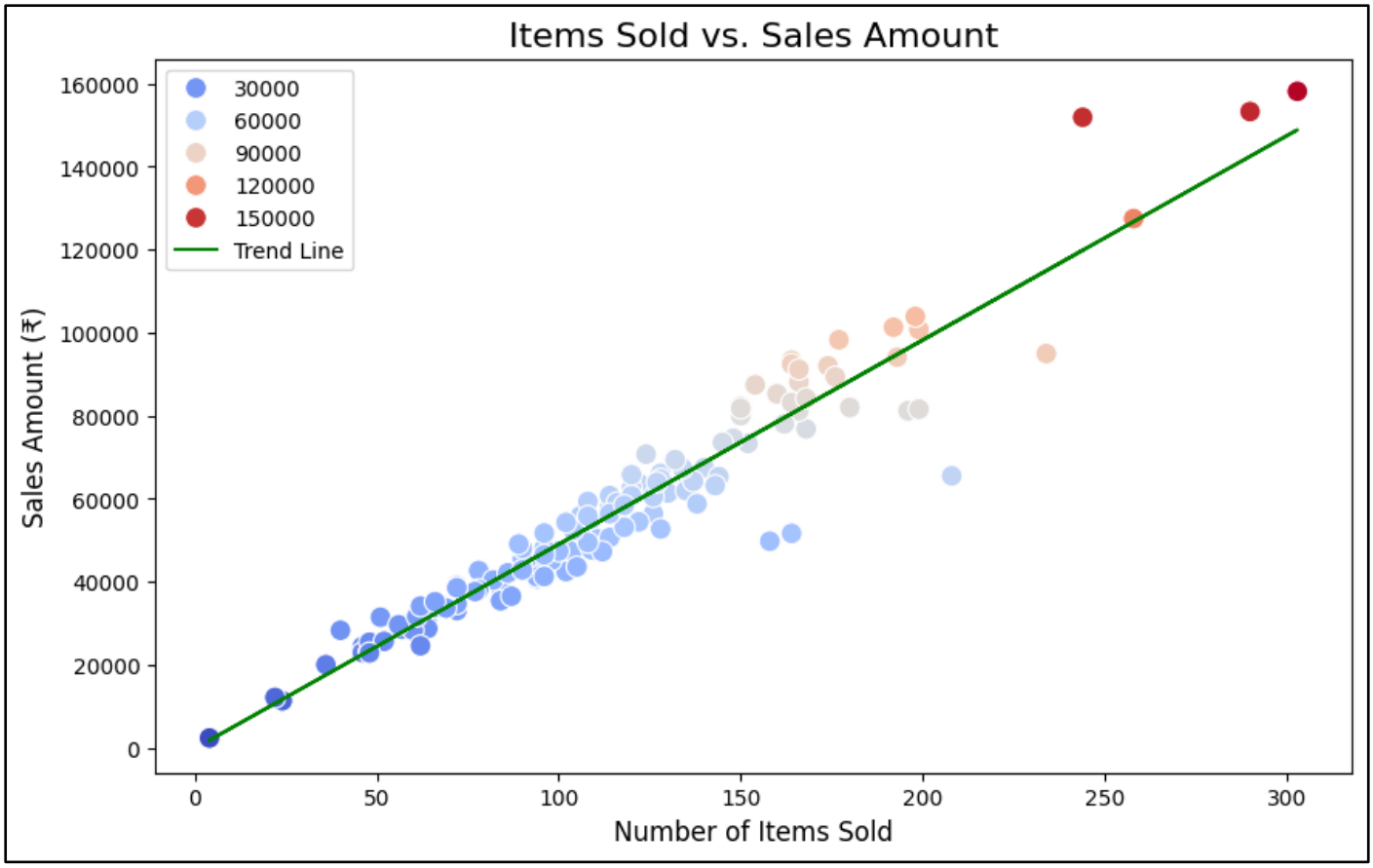
1. The line plot for the sales over time is comparatively shown in graph 5.1. The sales of the company from the period of August 2023 – December 2023 and April 2024 – August 2024 are compared. There are visible spikes, primarily in September 2023 and July 2024, and overall better sales during the festive season in India. The larger spike in mid-July 2024 (₹4,33,591) compared to mid-September 2023 (₹3,09,656) suggests sales have improved year-over-year
2. Graph 5.2 is a tree map that compares the best revenue-generating customer companies of Nalanda Dresses. Pulimoottil is the highest-value company for Nalanda having multiple warehouses in Kottayam, Thrissur, Tiruvallur, Kollam, etc. All the companies shown in the treemap are from various regions of southern parts of India. This shows that the company has minimal to no local clientele.
3. The scatter plot of Items sold vs Sales Amount is shown in graph 5.3. The graph shows a positive linear relationship between the items sold and the sales amount, indicating that the 'Items Sold' is a high-value feature for sales prediction. The plot also indicates the bulk sales contribution to the revenue.
4. The baseline Linear Regression model had an MSE of 0.20135755299324748. When using another ML model of Gradient Boosting, the MSE was significantly reduced to: 0.12506297480092418. This suggests that other tree models might give a better score and similarly higher accuracy in predicting the sales figures.



Graph 5.1: Line plot for sales over time



Graph 5.2: Treemap of the top 5 high-income companies.



Graph 5.3: Scatter plot for Items sold vs Sales Amount.