A red letter on a black background

Description automatically generated



**PATEL BROTHERS BI TRANSFORMATION**

***Grocery to Insights***

**Team 10**

* **Jeswanth Adari**
* **Hitesh Mansukhlal Sancheti**
* **Noopura Vaidya**
* **Ujwal Vijay Kumar**

**Table of Contents**

[1 EXECUTIVE SUMMARY 5](#_Toc153226047)

[2 Introduction 7](#_Toc153226048)

[2.1 Geographic Market Area 7](#_Toc153226049)

[2.2 Size in Terms of Number of Employees 7](#_Toc153226050)

[2.3 Patel Brother’s Current Status on the Following Aspects: 7](#_Toc153226051)

[2.3.1 SWOT Analysis 9](#_Toc153226052)

[2.4 Case Study Summary 9](#_Toc153226053)

[2.4.1 Kroger 9](#_Toc153226054)

[2.4.2 Trader Joe’s 10](#_Toc153226055)

[3 The proposed BI solution 12](#_Toc153226056)

[3.1 Components of BI Solution 12](#_Toc153226057)

[3.2 Roles and Values of Each BI Component 13](#_Toc153226058)

[3.3 Proposed Dashboards 15](#_Toc153226059)

[3.3.1 Sales Overview Dashboard 15](#_Toc153226060)

[3.3.2 Customer Demographics Dashboard 15](#_Toc153226061)

[3.3.3 Sales Forecast Dashboard 15](#_Toc153226062)

[3.3.4 Branch Analytics Dashboard 15](#_Toc153226063)

[4 USE CASES/Prototype's 17](#_Toc153226064)

[4.1 Sales Overview 17](#_Toc153226065)

[4.2 Customer Demographics Dashboards 18](#_Toc153226066)

[4.3 Forecasting Dashboards 20](#_Toc153226067)

[4.3.1 Benefits 21](#_Toc153226068)

[4.3.2 Pitching the BI Solution 21](#_Toc153226069)

[4.4 Branch Analytics: 21](#_Toc153226070)

[4.4.1 Benefits 26](#_Toc153226071)

[4.4.2 Pitching the BI solution 26](#_Toc153226072)

[5 Implementation 27](#_Toc153226073)

[5.1 Potential Challenges to Implementation 27](#_Toc153226074)

[5.2 Ethical Aspects of BI Implementation 28](#_Toc153226075)

[6 Summary and conclusion 32](#_Toc153226076)

[6.1 Recommendations for Patel Brothers 32](#_Toc153226077)

[6.2 Future Scope 33](#_Toc153226078)

[7 References 34](#_Toc153226079)

[8 Appendix 35](#_Toc153226084)

**List of Tables**

[Table 1: KPI Matrix 22](#_Toc153225457)

[Table 2: KPI Matrix 23](#_Toc153225458)

[Table 3: Kotter's 8 Steps to Implementation 28](#_Toc153225459)

## List of Figures

[Figure 1: BI Maturity Model Adoption Curve 6](#_Toc153224307)

[Figure 2: SWOT Analysis of Patel Brothers 7](#_Toc153224308)

[Figure 3: BI Framework for Patel Brothers 11](https://wpi0.sharepoint.com/sites/gr-BITeam10/Shared%20Documents/General/Project%20Report/Team10_PateL_Brothers_Project_Report_revised.docx#_Toc153224309)

[Figure 4: Data Warehouse Structure for Patel Brothers 12](https://wpi0.sharepoint.com/sites/gr-BITeam10/Shared%20Documents/General/Project%20Report/Team10_PateL_Brothers_Project_Report_revised.docx#_Toc153224310)

[Figure 5: Operational Dashboard - Visualization of key sales metrics 16](https://wpi0.sharepoint.com/sites/gr-BITeam10/Shared%20Documents/General/Project%20Report/Team10_PateL_Brothers_Project_Report_revised.docx#_Toc153224311)

[Figure 6: Tactical Dashboard - Know Your Customer 17](https://wpi0.sharepoint.com/sites/gr-BITeam10/Shared%20Documents/General/Project%20Report/Team10_PateL_Brothers_Project_Report_revised.docx#_Toc153224312)

[Figure 7: Sales Forecasting 18](https://wpi0.sharepoint.com/sites/gr-BITeam10/Shared%20Documents/General/Project%20Report/Team10_PateL_Brothers_Project_Report_revised.docx#_Toc153224313)

[Figure 8: Visualizing Relationship between Branch Attributes 20](#_Toc153224314)

[Figure 9: Feature Analysis 22](https://wpi0.sharepoint.com/sites/gr-BITeam10/Shared%20Documents/General/Project%20Report/Team10_PateL_Brothers_Project_Report_revised.docx#_Toc153224315)

[Figure 10: BI Transformation of Patel Brothers 30](#_Toc153224316)

# EXECUTIVE SUMMARY

As an emerging grocery chain in the growth phase, leveraging data and analytics is key to gaining competitive advantage. Currently grappling with decentralized operations and a heavy reliance on Microsoft Excel for analysis, the company faces challenges in obtaining a cohesive, enterprise-wide perspective, hindering optimal decision-making capabilities. The existing setup also lacks the capacity for advanced analytics, forecasting, and predictive modeling, necessitating a crucial upgrade for scalability. Upgrading to a robust BI solution is critical for supporting advanced analytics, forecasting, and predictive modeling as Patel Brothers scales. This team developed an integrated business intelligence solution comprising sales overview, customer demographics, sales forecasting, and branch analysis dashboards to provide data-driven insights.

The proposed BI solution comprises of central data warehouse converging transactional, customer, operational and financial data, business analytics platform enabling data visualization, predictive modeling, and interactive dashboards and four dashboard prototypes for sales analytics, customer intelligence, demand forecasts, and cross-branch comparisons. The solution will drive operational efficiencies, enhanced customer experience, increased revenues and profitability, and future scalability. The overarching goal is to empower data-driven actions across various functions, from enhancing inventory optimization and targeted marketing campaigns to identifying promising new branch locations and benchmarking branches for best practices.

The report unfolds in multiple phases, beginning with a meticulous assessment of Patel Brothers' current BI landscape and limitations. The blueprint for the BI solution outlines core components, designs high-level data architecture, and develops prototype dashboards aligned with business requirements. The implementation roadmap proposes a phased approach, emphasizing sales data consolidation and essential KPI monitoring in the initial stages, followed by the integration of other datasets, deployment of advanced analytics models, and the expansion of BI access.

The solution equips managers to base decisions on trends instead of gut feeling, driving growth through targeted assortments, customized customer engagement and anticipatory production. Implementation following Kotter's model, the roll-out focuses on quick wins first while building for cultural shift.

As a trusted analytics advisor, this team is committed to evolving the BI solution as the organic grocery chain's needs develop from single to multi-site. Additional functionality around pricing, distribution, and hiring can rapidly boost operations as part of the expansion roadmap while keeping the customer at the core.

The business case is quantified, highlighting expected benefits in terms of increased revenues, cost reduction, and improved customer experience. The expected benefits include operational efficiencies through optimized inventory management, lower supply chain costs, and increased productivity. Enhanced customer experiences will result from personalized marketing, improved satisfaction, and faster response times. Increased revenues and profitability are anticipated through higher sales conversion, new revenue streams, expanded market reach, and improved margins.

The future scalability aspects of the proposed BI solution are designed to ensure adaptability and growth through agile support for mergers and acquisitions, flexible architecture ingesting new data, and easy expansion into advanced analytics like machine learning.

In summary, implementing the proposed BI solution will significantly improve the Patel Brothers’ ability to analyze data, make informed decisions, and optimize operations that will lay solid foundation for BI transformation of Patel Brothers: From Grocery to Insights. This will ultimately lead to increased profitability, customer satisfaction, and competitive advantage.

# Introduction

Patel Brothers, founded in 1974, is a grocery retail chain specializing in South Asian and international foods which is currently worth $ 140 million. With numerous locations across the United States, including states like Arizona, California, Connecticut, Florida, Georgia, Illinois, Indiana, Kentucky, Maryland, Massachusetts, Michigan, New Hampshire, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Tennessee, and Texas. It offers a wide range of products including pantry staples, produce, bakery items, and religious and health products.

## Geographic Market Area

Patel Brothers was founded in Chicago and is headquartered in New York; Patel Brothers has a strong presence in the United States. They have numerous stores located in different states across the USA, but they are majorly centered across the East coast and particularly the northeast. Patel Brothers currently has Over 57 store locations in the United States.

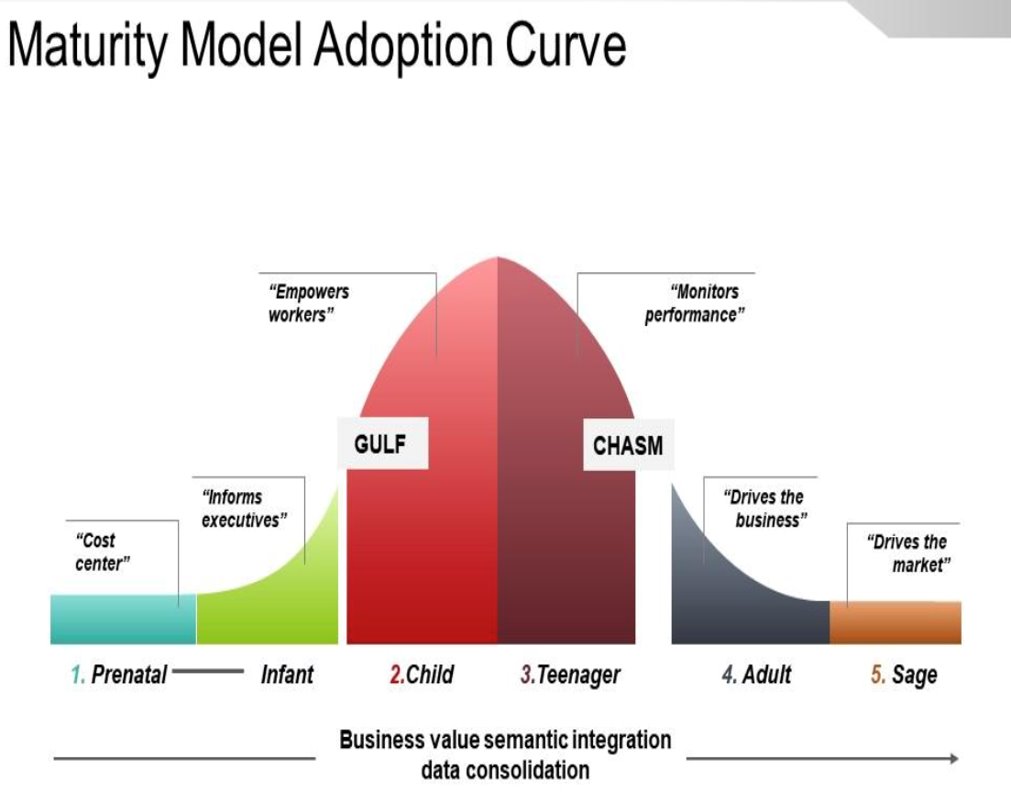
## Size in Terms of Number of Employees

Patel Brothers operates with a considerable workforce managing its extensive retail operations. and it may vary across various locations and seasons, but the estimated count is to be around 3000-5000 employees including cashiers, delivery partners, food workers, managers, etc. and They have various kinds of departments like store management, Customer Service, Inventory and Stock Management, Marketing and Promotions with 57 locations in 20 U.S. states. Typical store sizes range from 5,000 sq. ft to 30,000 sq. ft. and the biggest is in East Windsor, New Jersey.

## Patel Brother’s Current Status on the Following Aspects:

1. Maturity Model Adoption:

* Patel Brothers is currently positioned in the child phase of the BI maturity model adoption curve, indicating an early stage of BI implementation. This is as depicted in Figure 1.



Patel Brothers

Figure 1: BI Maturity Model Adoption Curve

2. Data Storage Structure:

* Separate data storage is implemented for each of the 57 stores.
* This decentralized structure facilitates localized data management.
* Microsoft Excel is identified as the primary tool for data analysis across various departments at Patel Brothers.

3. Primary BI Tool:

* Microsoft Excel is the predominant tool for data analysis across various departments.
* Excel is versatile, offering flexibility for data reporting and visualization.

4. Data Warehousing:

* The presence or absence of a data warehouse is not conclusively determined from the information available.

5. Software Integration:

Various software applications may be utilized for distinct operational aspects, including:

* Inventory management
* Customer Relationship Management (CRM)
* Supply chain management
* Security and surveillance

6. Business Analytics:

* Patel Brothers provides third-party services for business analytics.
* This approach suggests a collaborative strategy to enhance analytical capabilities, potentially supplementing in-house BI efforts.

7. Opportunities for Improvement:

* The current BI landscape presents opportunities for improvement and further integration.
* Exploring advanced BI solutions tailored to specific operational needs could enhance overall efficiency.

### SWOT Analysis

To better understand the current state of Patel Brothers, a SWOT analysis was conducted to understand the internal and external factors that are key to driving its success. This is shown in Figure 2.

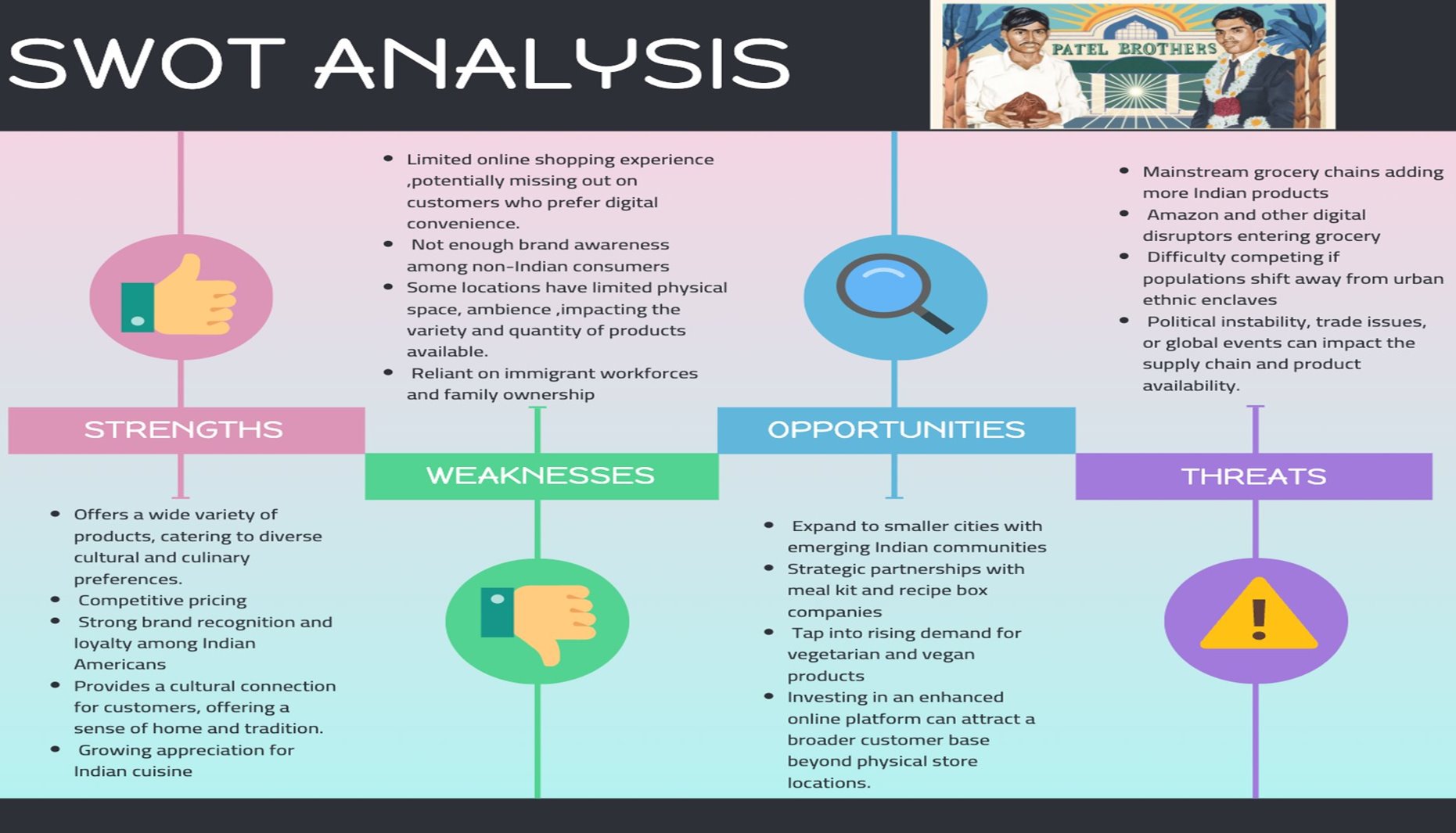


Figure 2: SWOT Analysis of Patel Brothers

## Case Study Summary

As part of this project, the team decided to refer to the BI implementation in two major grocery stores in the USA – Kroger and Trader Joe’s. A brief elaboration of their BI system is delineated in Sections 1.4.1 and 1.4.2 respectively.

### Kroger

Kroger is a grocery retail company that started in 1883 in Cincinnati, Ohio when Barney Kroger invested his life savings of 372$ to open a single grocery store. To adopt BI, they used self-checkout machines to reduce waiting time for the customers. Kroger approached IBM to find insights into implementing BI to increase their revenue. The IBM tool depicted a clear view of all the stores nationwide, insights by geography and even by store. Kroger uses BI to visualize and monitor and optimize its supply chain by analyzing data related to inventory levels, transportation which helps reduce costs, improve efficiency, and ensure products are available. It was evident from Kroger’s model that BI adoption helped them become the largest supermarket operator by revenue and it also one of the largest private employers in the US. Lastly, Kroger’s loyalty program has been an industry leader for more than a decade and highlights the largest source of highly advanced analytical talent.

### Trader Joe’s

Trader Joe’s is a national chain of neighborhood grocery stores. It was started in 1967 by Joe Columbe. Trader Joe does not stock a lot of well-known brands but, rather, its stores are stocked with unique and interesting products along with everyday groceries. This company prides itself on the quality of its private label products, which account for 70 percent of the product offerings.   
The main idea behind Trader Joe’s success is to keep the stores as minimal in size as possible as this makes it easy for people to find products. Trader Joe’s is not even at a 1/10 of the size of industry giants such as Walmart and Costco and yet, had a revenue of $13.2 Billion in 2021. This suggests that an often-overlooked KPI for retail sales is the Sales Per Square Foot (SPSF) as this is linked to its efficiency and inventory turnover. In an industry that stocks around 30,000 SKUs on average, Trader Joe’s bucks the trend by stocking less than a quarter of that at around 4000. The smaller SKU allocation also ensures that only the fastest moving products are stocked in its stores. This also helped them to streamline the supply chain, hence providing benefits even before the products reach the stores.

The case studies reveal that grocery leaders like Trader Joe's and Kroger owe much of their success to consistent innovation embracing new technologies while optimizing retail fundamentals. Patel Brothers can emulate certain proven approaches within the store network to drive growth. For one, adapting compact store sizes of around 5,000 - 10,000 sq ft like Trader Joe's formats focused sharply on fastest moving products will boost sales per square feet metrics through space productivity and inventory turns. Additionally, rolling out self-checkout kiosks leveraging scan-&-go options can significantly cut down waiting times during peak hours – a tactic Kroger leveraged to elevate customer experience. Lastly, increasing private label presence from 20% to 50-60% by heavily marketing the existing "SWAD" brand of spices and sauces can unlock higher profit margins following the Trader Joe's playbook. Thus, by combining store space optimization, checkout automation and private brand build-up, Patel Brothers can pave the way for surpassing $250M in annual revenues in the next 4-5 years. The case studies validate that blend of retail fundamentals and technology disruption as a winning formula.

# The proposed bi solution

To begin with, BI (Business Intelligence) refers to a collection of several different technologies, processes and practices that can be used together or individually to gather, analyze, and visualize data to gain valuable insights to make informed business decisions. BI helps an organization transform raw data into actionable intelligence and thus allows the company to identify trends, predict future outcomes, and make data-driven decisions across the various departments and functions of the organization.

## Components of BI Solution

Some of the main components of a BI solution would typically be:

* **Data Sources:** These could be either internal or external data. Internal data often refers to transactional data such as sales records or customer information, operational data such as inventory levels, and financial data. On the other hand, external data could be market research reports, competitor analysis or economic indicators
* **Data Warehouse and Data Lake:** A data warehouse would typically store historical data in a structures format suitable for analysis and reporting while a data lake would store all types of data both structured and unstructured that could be used for future exploration and analysis
* **Data Integration and Management:** Define processes for ETL (Extracting, Transforming, and Loading) data from several diverse sources into either the data warehouse or data lake. Also, data governance policies should be delineated as it ensures data consistency, accuracy, and security
* **Business Analytics Tools:** These could be either Data analytics tools, Reporting tools or Predictive analytics tools. The uses of each one of them is elaborated below:
  + Data Analysis Tools: Used to explore, analyze, and visualize data for identifying trends, and patterns
  + Reporting Tools: Used to create reports and dashboards that give clear and crisp insights concisely for different audiences (Top management/ Technical management/ Associates/ Client, etc.)
  + Predictive Analysis Tools: Uses advanced techniques such as AI/ML or statistical methods to forecast future outcomes and trends
* **User Interface and Dashboard**: Interactive dashboards and reports that allow the user to access and analyze data, drill-down into finer details that allows the user to share valuable insights with stakeholders which assists in real-time decision making

## Roles and Values of Each BI Component

This section describes the significance of each BI component more elaborately. Firstly, the data sources provide raw material for analysis and to generate insights. Next, Data Warehouse/Lake paves the way for efficient data storage, and retrieval for analysis of either historical or current data. Data Integration and Management is essential to ensure data accuracy, consistency, and security for reliable decision making. Business Analytics Tools is what allows the stakeholders to derive meaningful insights from data that allows them to make informed decisions and take strategic actions for the benefit of their organization. Lastly, User Interface and Dashboards allow the user to communicate the derived insights effectively and empower the stakeholders to act based on data-driven information.

**A diagram of data processing

Description automatically generated**Succinctly, BI empowers organizations to make data-driven decisions that reduces risks, increases efficiency that in-turn paves the way for organizations to identify growth opportunities that helps them increase profitability and stay ahead of competition by adapting to market changes. Therefore, companies can leverage this technology to achieve their business goals.

Figure 3: BI Framework for Patel Brothers

Figure 3 shows the BI Framework for Patel Brothers that highlights the several types of data that they could use along with the type of analytics or monitoring which would help them scale their business to greater heights. To begin with, they could use Data Warehouses and Data Lakes to store historical data (facilitates complex analysis) and all types of structured or unstructured data for future analysis. Once they are comfortable using the BI solution, they could explore the options for cloud storage as this would provide them scalability, cost-effectiveness, and accessibility for geographically dispersed stores. To make the ETL process efficient, the top management of Patel Brothers shall define policies and procedures to extract, transform and load data from various sources into a single repository as this will not only ensure data accuracy and completeness but also provide measures to implement data security and privacy policies. The users of this BI solution are:

* Senior Management: They can gain insights into which attributes drive success and identify areas for improvement
* Operations Team: Optimize resource allocation based on the performance potential (peak hours, best time of the month, popular branches, etc.)
* Marketing Team: Tailor marketing campaigns and promotions to specific branch segments based on the customer demographics
* Branch Managers: Benchmark their branch performance against similar branches
* Supply Chain Team: Plan for inventory management by identifying top selling product, least selling product, peak sales time, best time of the month, etc.

A computer screen shot of a computer

Description automatically generated

Figure 4: Data Warehouse Structure for Patel Brothers

Figure 4 shows a multi-dimensional data model for Patel Brothers with sales as the Fact table and dimensions have been restricted to Location, Time, Product and Customer for simplicity. However, additional dimensions such as Promotions, Employee or Inventory could be included to get a better visualization as the business needs evolve and additional data sources become available. With the simplified data model as shown, the BI user can analyze several aspects of Patel Brothers’ business, such as sales performance, customer behavior, branch performance, product performance, etc.

## Proposed Dashboards

The team decided on developing 4 dashboards for visualization and analytics. This section elaborates on each of the four dashboards.

### Sales Overview Dashboard

This is an operational dashboard that provides real-time insights into the overall sales performance highlighting the sales across the branches in Massachusetts, USA, identifying the top-selling product, least popular product, overall customer satisfactory rating, and allows for proactive adjustments to inventory and promotions. A more detailed description of the usage and benefits of this dashboard is illustrated in Section 2.1.

### Customer Demographics Dashboard

This is a tactical dashboard that displays valuable information about customer segments, customer satisfactory ratings, sales distribution, and hourly sales for the three branches in Massachusetts, USA enabling targeted marketing campaigns and product development strategies. A more detailed description of the usage and benefits of this dashboard is illustrated in Section 2.2

### Sales Forecast Dashboard

This is a predictive dashboard that assists in predicting future sales with granular detail that allows for informed planning of inventory, resource allocation, promotion, etc. A more detailed description of the usage and benefits of this dashboard is illustrated in Section 2.3.

### Branch Analytics Dashboard

This is a strategic dashboard that visualizes the geographic spread of branches, identifying potential expansion areas. With the Branch Attribute Correlation, the BI user can identify relationships between branch attributes like size, product mix, location, and performance metrics that help analyze regional performance variations and identify potential expansion areas. A more detailed description of the usage and benefits of this dashboard is illustrated in Section 2.4.

By leveraging the BI solution effectively, Patel Brothers can unlock valuable data insights, improve operational efficiency, enhance customer satisfaction, and achieve sustained success in the competitive US grocery market.

# USE CASES/Prototype's

As mentioned in Section 2.3, 4 prototypes were developed to help the Senior Management at Patel Brothers understand the significance and relevance of using BI or BA to improve their sales and inventory that will in turn pave the way for increased Return on Investment, Reduced losses due to overstocked inventory and heighted customer satisfaction. This section delineates in detail about each of the dashboards, their use case, benefits, KPIs, Managerial Implications and target audience. To develop the prototypes, Sections 3.1 and 3.2 use a dataset containing a historical record of sales data for the three stores in Massachusetts, USA: North Attleboro, Shrewsbury, and Waltham. This dataset gives information about the customer demographics, product lines, sales, timestamp, and customer satisfaction rating. Sections 3.3 and 3.4 use a branch wise dataset that contains information regarding Store ID, Store Area, Store Sales and Products Stocked.

## Sales Overview

An operational value with a focus on understanding the sales metrics like:

* Total sales
* Average Customer Satisfactory Rating
* Sales By Product Lines
* Sales By Branch
* Top Selling Product
* Least Selling Product
* Distribution of Gender across the three branches
* Distribution of Customer Type across the three branches

Figure 5 shows the operational dashboard created for Patel Brothers. On top of the dashboard are 5 summary tiles highlighting the Top Selling Product, Least Selling Product, Product Lines, Customer Satisfactory Rating and Sales. Below are two bar graphs that show customer distribution by Gender and Customer Type. The last row shows a map where the three branches are located along with the average rating of each of those branches. The branches for which this analysis is conducted are in the state of Massachusetts in Waltham, Shrewsbury, and North Attleboro. The Sales Overview is depicted using a sunburst chart with the inner layer depicting the sale distribution by branches and the outer layer shows the sale distribution by product line. Sales Overview chart is provided on the right. Interactive filters are provided for the map as well as the sunburst chart for drill-down to specific branch or product line or time-period for further analysis.

Figure 5: Operational Dashboard - Visualization of key sales metrics

A close-up of a chart

Description automatically generatedWith this dashboard, we can answer the following managerial questions such as:

* What is the comprehensive sales performance?
* Which product is driving the highest revenue?
* Are there any seasonal trends or patterns in sales?
* How does the performance of branches compare to each other?
* What is the level of customer satisfaction?

The target audience would be Store Managers, Regional Managers, or the Executive Management Team as this helps them gain real-time insights into sales performance and identify areas for improvement by making data-driven decisions.

## Customer Demographics Dashboards

A Tactical dashboard with a focus on understanding customer segments, their buying behavior, and preferences through features such as:

* Demographics breakdown (Gender, Customer Type, Payment Type)
* Product preferences
* Customer satisfactory rating for each branch and their respective product line
* Daily Sales Distribution
* Hourly Sales Distribution

**A screenshot of a computer

Description automatically generated**Figure 6 shows the tactical dashboard created for Patel Brothers. The first row gives the sales distribution by Gender and the Customer Satisfactory Rating (CSR) for the three branches in Massachusetts. The middle charts show the diversification of customers by Customer Type and Gender followed by their preferred mode of payment. Lastly, the chart depicts the hourly sales and daily sales distribution. A filter for Month is included to understand the distribution of sales for a specific month and the legend for the Product Line, City, Customer Type and Gender is depicted on the right. Interactive filters are provided for Customer Type Diversification and Gender Diversification for drill down to specific customer segments based on demographics and loyalty program status. This helps us understand their purchasing habits by product line and their preferred payment method along with comparing customer satisfaction scores across different branches. This also helps us understand the busiest time of the day and busiest time of the month for sales. With this dashboard, we can answer the following managerial questions:

Figure 6: Tactical Dashboard - Know Your Customer

Figure 6:

* Who are the core customer segments?
* What are the customers purchasing the most?
* What is the busiest time of the day?
* How can customer loyalty be improved?
* What are the customer satisfaction levels?
* What time of the day requires staff assistance most?
* What time of the month requires most staff assistance?
* What time of the month requires maximum stocked inventory?
* What time is best to run promotions to increase sales?
* How can the payment method be incentivized to offer benefits to the customer?
  + For example, if a store sees a greater number of customers paying by cash, should they be given a discount or included in a reward system to ensure more business?

The target audience would be the Marketing Department, Customer Service Team, Executive Management Team, Inventory Team, and Supply Chain Management Team as this helps them develop marketing campaigns that resonate with specific customer segments, improve customer service based on identified trends and feedback, and increase customer loyalty through personalized offers and rewards.

## Forecasting Dashboards

A predictive dashboard that assists in forecasting future sales based on historical data.

Figure 7: Sales Forecasting

Figure 7 shows the Forecasting dashboard created for Patel Brothers.A screenshot of a computer

Description automatically generated The first row shows the sales by customer type which predicts sales by both members and non-members that also focuses on predicting future sales based on historical data. And moving to the second row, it details us the revenue generated for each product for each month and leveraging this information as a reference, it is going to predict the upcoming months revenue for the respective product aiding in the estimation of stock replenishment. And coming to the third row that gives us details about the sales by branch this analysis will aid in discerning whether they need any changes to be made within the staff or the inventory where it clearly shows which branch is falling behind and require additional attention to enhance sales performance. With this dashboard, we can answer the following managerial questions:

* What are our projected sales for the upcoming period?
* How much inventory do we need to order to meet demand?
* How many staff members do we need to schedule to meet customer needs?
* What is the expected impact of our planned promotions?
* Are there any potential risks of stock shortages?

The target audience would be inventory management team, Operations department, Marketing department, Executive management department as it helps them with the ability to maximize the return on investment for marketing campaigns through targeted promotions and increased efficiency and customer satisfaction derived from proactive decision making. Lastly, anticipating the market trends and adapting strategies accordingly.

### Benefits

With this analysis, Patel Brothers can precisely forecast future sales for individual products and branches, preventing stock shortages and overstocking through optimized inventory levels that help reduce their costs. Also, they can anticipate future customer traffic and optimize staffing schedules, ensuring efficient service that reduces labor costs and improves customer satisfaction.

### Pitching the BI Solution

This dashboard underscored the potential for cost savings through optimized inventory levels and staffing through accurate forecasting. It also establishes the ability to maximize the return on investment for marketing campaigns through targeted promotions. Furthermore, it highlights the competitive advantage gained by anticipating market trends and adapting strategies accordingly. Lastly, it demonstrates the increased efficiency and customer satisfaction derived from proactive decision-making.

## Branch Analytics:

A strategic dashboard that focuses on the geographic spread of branches and their expansions.

A screenshot of a computer screen

Description automatically generated

Figure 8: Visualizing Relationship between Branch Attributes

Figure 8 shows the Branch Analytics dashboard created for Patel Brothers. The key features include interactive map visualization of branches, market sales analysis, also under and over performing locations etc. It focuses on the hidden patterns and relation between branch attributes i.e., (size, location, demographics, product mix) and both financial and operational performances. Coming to the key features, it has advanced analytics and data visualizations to utilize statistical tests to identify relations between attributes and performances, and to visually represent complex relationships and identify outliners. we also used spatial analysis to identify connections between local demographics and branch performance and lastly the interactive dashboards which allows user to experience to explore and filter data by specific attribute for their deep understanding of individual branch segments. With this dashboard, we can answer the following managerial questions:

* Which branch attributes are most positively correlated with high sales growth?
* What combination of attributes contributes to optimal profitability for different branch segments?
* How do specific branch attributes relate to customer satisfaction levels?
* Are there branches with low inventory turnover that require adjustments in product mix or ordering strategies?
* Which branch segments have the highest staff productivity, and what are their key characteristics?
* How can we tailor marketing campaigns to specific branch segments based on their unique attributes and customer demographics?
* Can we identify any underperforming branches based on their attribute profile and performance metrics?
* What specific interventions can be implemented for underperforming branches based on their identified weaknesses?
* How can we leverage attribute insights to optimize resource allocation (staffing, inventory) across the branch network?
* What are the ideal attributes for new branch locations based on our understanding of successful branch profiles?

The target audience would be senior management, operations team, marketing team, real estate team, and branch managers for a comprehensive understanding and cohesive decision-making process. through identifying areas for improvement across entire branches. optimizing resource allocation based on branch performance and designing optimal locations for new branches based on data-driven insights. Lastly, benchmarking their branch performances against similar branches within their segments.

To demonstrate the effectiveness of this dashboard, a machine learning model is built that can estimate the given task by learning the correlations between the input features and estimate the target for a new data point. For this scenario, the key features are:

* Store\_Area
* Items\_Available
* Daily\_Customer\_Count
* Store\_Sales

To provide some instances to illustrate, having known the Store\_Area, Items\_Available, Daily\_Customer\_Count, the model can predict the Store\_Sales or, if one knows the Items\_Available, Daily\_Customer\_Count, Store\_Sales, the model could be used to predict the Store\_Area. Figure 9 shows the feature analysis for the aforementioned key features.

A screenshot of a computer screen

Description automatically generated

Figure 9: Feature Analysis

This work implements regression analysis for Store\_Area and Items\_Available as the variables are linearly correlated as depicted in Figure 8. This information can be verified using Table 1. Below is more information on the regression:

1. KPIs Used:
   1. Mean Absolute Error (MAE)
   2. Mean Squared Error (MSE)
   3. R-squared (R2)
2. Algorithm:
   1. Decision Tree Regressor Model (Max. depth of 5)
   2. Train, test split: 90% and 10%
   3. Forward feature selection
3. Results:

Table 1: KPI Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sl.No.** | **Target Variable** | **Features** | **MAE** | **MSE** | **R2** |
| 1 | Store\_Area | Items\_Available | 14.33 | 445.67 | 0.99 |
| 2 | Items\_Available | Store\_Area, Store\_Sales | 20.17 | 592.5 | 0.99 |
| 3 | Daily\_Customer\_Count | Items\_Available, Store\_Sales | 526 | 301700 | 0.9 |
| 4 | Store\_Sales | Store\_Area. Daily\_Customer\_Count | 13151 | 205271416 | 0.7 |

1. Inference:
   1. Due to high variance of the variable, it is extremely challenging to predict the value precisely
   2. Even with a given prediction, the value is bounded by upper and lower bounds of variations
   3. This kind of approach leads to classification task - Where a class is category with start and end bins (Similar to box plots with variance)

For Daily\_Customer\_Count and Store\_Sales, a classification approach is used as it has good non-linear characteristics as shown in Figure 8. This information can be verified using. Below is more information on classification:

1. KPIs Used:
   1. Overall Accuracy (OA)
   2. Overall Precision (OP)
   3. Overall Recall (OR)
   4. Overall F1-Score (O-F1)
2. Algorithm:
   1. Binning Strategy is based on interquartile range of the data
   2. K-Neighbor Classifier (Neighbors = 6)
   3. Train, test split: 90% and 10%
   4. Forward feature selection

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sl.No.** | **Target Variable** | **Features** | **OA** | **OP** | **OR** | **O-F1** |
| 1 | Daily\_Customer\_Count | Store\_Area, Store\_Sales | 66.6% | 0.40 | 0.37 | 0.35 |
| 2 | Store\_Sales | Store\_Area, Items\_Available, Daily\_Customer\_Count | 63.3% | 0.54 | 0.55 | 0.52 |

1. Results:

Table 2: KPI Matrix

1. Inference:
2. The results provide a promising direction for modelling non-linear characteristics of Daily\_Customer\_Count and Store\_Sales
3. The use of automated binning strategy using interquartile ranges of the data helps in converting a numerical column to a categorical column. Thereby, translating the prediction task from regression to classification
4. Each predicted class can be translated to a numeric value by making use of a box plot that describes mean value, quartiles, and outliers

### Benefits

This analytics suggests means for strategic expansion planning by identifying optimal store area for maximum sales. It allows for comparing sales with branch attributes such as State, City, Store Area revealing areas for improvement and best practices for replication. Lastly, it paves the way for making informed decisions about branch allocation of resources, marketing efforts, and product offerings based on comprehensive data analysis.

### Pitching the BI solution

Through this analysis, the team emphasizes the potential for increased revenue and market share through targeted expansion into promising locations while underscoring the strategic advantage gained by making data-driven decisions about branch management and marketing.

By implementing a comprehensive BI solution with these two dashboards, the Indian grocery store can gain valuable insights into their sales trends, customer demographics, and branch performance. This will enable them to make data-driven decisions that optimize their operations, increase efficiency, improve customer satisfaction, and achieve their business goals.

# implementation

Given that Patel Brothers is in the Child stage of the BI Maturity model, the implementation of BI Analytics through dashboards will require a phased approach which is detailed below:

**Phase 1: Data Gathering & Cleaning**

The focus should be on readily available data such as sales figures, customer demographics and branch specific attributes (store area, no. of products in their inventory, no. of customers visiting the stores, no. of sales, etc.). To do so, the team could use the spreadsheet to gather data following which, the data must be pruned for inconsistencies such as missing data, format inconsistencies, etc. to ensure data integrity.

**Phase 2: Simple Analysis & Visualization**

Begin by focusing on easily calculated matrices or KPIs such as sales per branch, customer spending and customer satisfaction scores. Create simple visualizations to highlight these KPIs to understand the relationship between branch specific attributes and their performance. This will help identify obvious patterns and trends that help us understand the branches that are better performing and specific attributes.

**Phase 3: Advanced Analytics using BI Tools**

Use tools such as Tableau or Power BI or Google Data Studio to conduct advanced analytics and visualization. The cleaned data can be used with the selected tool for visualization and analytics. Next, simple dashboards such as sales overview, customer demographics distribution could be developed that will help them understand their current state better. This will help them make better decisions regarding their inventory, sale prices, promotions, loyalty programs, etc. Once the top management is convinced about the advantages of analytics, the senior management invest further in more advanced technology such as data modeling, AI/ML algorithms for deeper analysis. Forecasting techniques could be used to predict future sales based on historical data.

## Potential Challenges to Implementation

Some of the challenges in implementing BI for Patel Brothers would be the availability of quality data as they are in the child stage of the BI maturity model. In the child stage, comprehensive data across branches and relevant attributes might not be readily available. Since data is collected manually, it can be prone to errors and inconsistencies, hindering accurate analysis. Furthermore, data from various sources might be siloed and this requires additional manual effort to combine and analyze.

One other challenge could be the lack of expertise amongst their current staff to utilize BI tools effectively and conduct advanced analysis. Also, Managing and maintaining data quality can be difficult without proper data governance processes. Now, Patel Brothers does not have the necessary hardware and software infrastructure might not be available to support BI initiatives.

Also, there could be organizational challenges like resistance to change from the current staff. The staff might be very accustomed to traditional business practices and be reluctant to embrace data-driven decision making, which can create challenges. Communication is key to the success of implementing change in an organization and hence, it is essential that the management clearly communicates the project's purpose and benefits to stakeholders. Investment in BI tools and consulting services might be constrained due to budget limitations that could make the top management reluctant to invest in one such technology.

However, these challenges could be overcome by starting small and focusing on high-value KPIs and utilizing free or inexpensive BI tools to understand its importance in increasing the company’s Return on Investment (ROI) or reducing their losses. The team at Patel Brothers could first partner with external consultants who have the relevant expertise to guide Patel Brothers on the usage of such tools. To bring the top management on-board, the advantages of using BI tools should be documented and communicated clearly and effectively to create awareness and glean support from stakeholders and leadership. A data governance framework must be developed to ensure data quality and security. Lastly, the management must invest in training and development of staff to enhance BI skills.

## Ethical Aspects of BI Implementation

The implementation of BI solutions does come with ethical considerations that require careful attention. Some of the key aspects for Patel Brothers to consider are:

* **Data Privacy:** Transparency must be ensured in data collection practices and obtain informed consent from customers and employees. In addition, robust security measures should be implemented to protect sensitive data from unauthorized access, use, or disclosure. Further to that, clear guidelines for data retention should be established and secure disposal methods for outdated or unnecessary data must be implemented
* **Algorithmic Bias:** The team must be aware of potential biases inherent in the data used to train algorithms and take steps to mitigate them. When designing algorithms, it must be ensured that they are designed to be fair and unbiased in their decision-making processes. Additionally, explainable AI techniques could be employed to understand how algorithms reach decisions and identify potential biases.
* **Impact on Employees and Customers:** It is essential to be mindful of the potential for automation and job displacement caused by BI/BA tools and implement retraining programs if necessary. The top management must clearly Communicate how BI/BA tools are used in decision-making processes and be open to feedback from employees and customers. Also, take measures to ensure algorithms are not used in discriminatory ways that disadvantage specific groups of individuals. Prioritizing informed consent and providing easy opt-in/opt-out options would empower customer control. Following a minimalist approach by gleaning information or customer data essential to grocery sale will help gather customer support and confidence.
* **Accountability and Governance:** Establish a clear data governance framework with defined roles, responsibilities, and procedures. Also, regularly audit data collection, usage, and security practices to ensure compliance with ethical principles. The management should consider establishing an independent oversight body to monitor ethical use of BI/BA tools and address potential concerns as they will not be prone to bias

By acknowledging these ethical concerns and implementing appropriate safeguards, Indian grocery stores can leverage BI/BA tools responsibly and ethically, generating valuable insights while safeguarding data privacy, mitigating bias, and ensuring fairness and transparency in their operations.

In addition to the above, there could be some unintended consequences of using BI for Patel Brothers such as, ignoring human judgment and intuition in favor of solely data-driven decisions and spending more resources on technology than necessary for the store's current needs and capabilities. This could be overcome by utilizing data as a guide but maintain human oversight and critical thinking for decision-making and prioritize high-value dashboards and focus on the most impactful data insights starting with free or low-cost BI tools and gradually upgrade as needed and resources allow.

To better get the support of the senior management, we can use Kotter’s 8 step implementation plan which is described in Table 1.

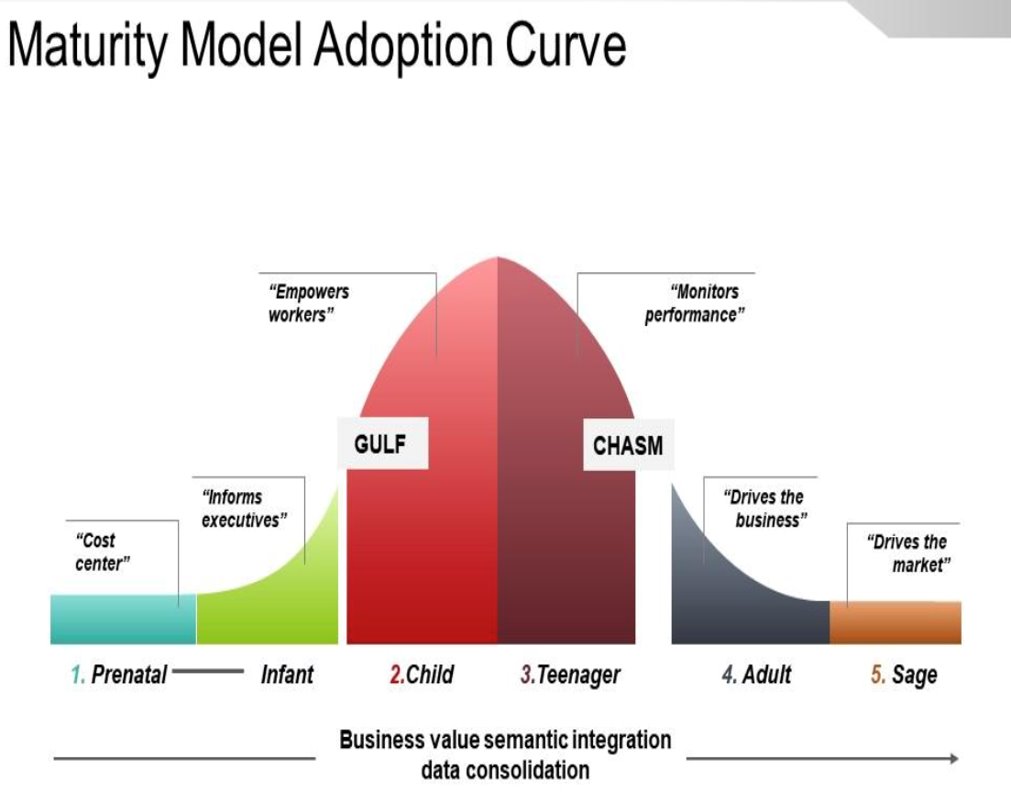
Table 3: Kotter's 8 Steps to Implementation

|  |  |
| --- | --- |
| **Kotter's Steps** | **Transforming Patel Brother’s** |
| 1. Establishing a Sense of Urgency | Need to stay ahead in the South Asian and international foods market in the US Increasing competition from ethnic supermarkets and internet retailers that provide tailored advice and sales. |
| 2. Form a Powerful Guiding Coalition | Appoint a Chief Data Officer or equivalent role and create a cross-functional team with members from the marketing, IT, customer service, and retail management divisions. |
| 3. Create a Vision | a) Expand stores and utilize data to enhance customer service, tailor marketing campaigns, and optimize inventories.  b) Become the top destination for South Asian groceries in the US with an omni-channel approach. |
| 4. Communicate the Vision | Present key metrics and trends visually, enabling data-driven decision-making using interactive dashboards at all organizational levels. |
| 5. Empower Others to Act on the Vision | Equip store managers with real-time inventory data, train cashiers to analyze sales trends, and enable marketing teams to leverage customer behavior insights for promotions. |
| 6. Plan for and Create Short-Term Wins | Focus on successful new store openings, reducing excess inventory by 15% via improved forecasting, and aiming for a reduction in stock-outs. |
| 7. Build on the Change | Regularly gather employee suggestions for improving customer service and work toward enhancing the average customer rating. |
| 8. Institutionalize New Approaches | Foster a culture valuing data-driven decision-making by recognizing and incentivizing staff who actively engage with and apply BI-driven insights. |

# Summary and conclusion

This report has explored the potential of BI for Patel Brothers, which is in the child stage of the BI maturity model. This work presents 4 dashboards as prototypes to the top management of Patel Brothers that provide valuable insights into sales, customer demographics, sales forecasting, and branch attributes. These dashboards will empower Patel Brothers to make data-driven decisions to improve their operational efficiency and achieve business goals. Through this BI transformation, Patel Brothers can traverse from ‘Grocery to Insights’ that will help them replace ‘gut feeling’ with data-driven insights for better strategic and operational decisions. satisfaction, help them optimize their resource allocation, improve their customer satisfaction which in turn will help them increase their sales and profitability.

Embracing BI strategies outlined in our plan, including optimized store sizes, self-checkout kiosks, and a focus on private label branding, can catalyze data-driven decision-making, operational efficiency, and customer satisfaction. This transition to the teenage stage signifies a maturation in BI capabilities, setting the stage for Patel Brothers to navigate the complexities of the business landscape with heightened agility and strategic insight. As we envision this transformation, we anticipate that Patel Brothers will not only enhance its competitive edge but also lay a solid foundation for continued growth and success in the dynamic market landscape. Figure 10 shows the transition of Patel Brothers to the next stage of the BI maturity model.



Patel Brothers

Figure 10: BI Transformation of Patel Brothers

## Recommendations for Patel Brothers

Through this work, the team recommends that Patel Brothers adopt BI into their businesses as it can be transformative, offering numerous benefits in various areas of its operation. Patel Brothers should start small and scale gradually as they expand their BI capabilities with growth in resources. To begin with, they should evaluate their data sources to identify all relevant data sources and assess their available resources prior to investing heavily in BI tools and solutions. Later, Patel Brothers could build effective essential BI dashboards and market the effectiveness of dashboards to every associate of Patel Brothers. Furthermore, they should adapt dashboards to reflect changing business needs, incorporate new data sources, and explore advanced analytics techniques. The team at Patel Brothers should leverage data insights to make informed decisions and improve customer experience. Through this, the store can ensure its long-term success and profitability.

## Future Scope

As Patel Brothers progress on their BI journey, they can enhance their analysis by incorporating data from loyalty programs, social media and external market sources and employ more sophisticated technology for forecasting, customer segmentation, and anomaly detection. Lastly, they can analyze customer reviews and social media interactions to gain deeper insights into customer preferences and satisfaction. By continuously expanding their BI capabilities and embracing recent technologies, Patel Brothers can unlock even greater value from data, solidify their competitive edge, and achieve long-term success.

# References

# *Indian marketplace for spices,* *groceries and puja items*. Patel Brothers. (2023, November 30). <https://www.patelbros.com/>

# *Kroger leverages AI, Data Analytics*. Specialty Food Association. (n.d.). [https://www.specialtyfood.com/news/article/kroger-leverages-ai-and-data-analytics/#:~:text=Spanner%3A%20With%20Google%20Cloud%20Spanner%27s,direct%20associates%20to%20be%20more](https://www.specialtyfood.com/news/article/kroger-leverages-ai-and-data-analytics/)

# Bean, J. (2023, June 29). *America’s Beloved Grocery Brand: The analytics behind the unexpected success of Trader Joe’s*. Medium. <https://jordanbean.medium.com/the-analytics-behind-the-unexpected-success-of-trader-joes-12318e715132>

* PRAKASH, S. (2022A, NOVEMBER 25). *MAFAT PATEL: BUILDING THE PATEL BROTHERS EMPIRE*. THE GLOBAL INDIAN. [HTTPS://WWW.GLOBALINDIAN.COM/STORY/GLOBAL-INDIAN-EXCLUSIVE/MAFAT-PATEL-BUILDING-THE-PATEL-BROTHERS-EMPIRE/](https://www.globalindian.com/story/global-indian-exclusive/mafat-patel-building-the-patel-brothers-empire/)

# *The patel brothers: An Indian-american success story*. Juno. (n.d.). <https://juno.finance/blog/the-patel-brothers-indian-american-success-story>

* *LAWTON, G. (2021). RETRIEVED FROM* [*HTTPS://WWW.TECHTARGET.COM/SEARCHBUSINESSANALYTICS/TIP/TOP-11-BUSINESS-INTELLIGENCE-CHALLENGES-AND-HOW-TO-OVERCOME-THEM*](https://www.techtarget.com/searchbusinessanalytics/tip/Top-11-business-intelligence-challenges-and-how-to-overcome-them)

# APPENDIX

The Python code for regression and classification described in Section 3.4.

*"""  
Regression: Store\_Area  
"""*# Import necessary libraries  
from sklearn.model\_selection import train\_test\_split  
from sklearn.tree import DecisionTreeRegressor  
import pandas as pd  
from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score  
  
# Load the data from the CSV file  
df = pd.read\_csv('PBStore\_branch\_analysis\_d\_FINAL.csv')  
  
X = df[[  
 "Store\_Area",  
 # "Daily\_Customer\_Count",  
 # "Store\_Sales"  
 ]]  
  
y = df['Items\_Available']  
  
# Split the data into training and testing sets  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1)  
  
## Decision Tree  
regressor = DecisionTreeRegressor(max\_depth=5)  
regressor.fit(X\_train, y\_train)  
  
# # # Make predictions on the test set  
y\_pred = regressor.predict(X\_test)  
  
# Calculate metrics  
mae = mean\_absolute\_error(y\_test, y\_pred)  
mse = mean\_squared\_error(y\_test, y\_pred)  
r2 = r2\_score(y\_test, y\_pred)  
  
# Print the metrics  
print(f'Mean Absolute Error (MAE): {mae:.2f}')  
print(f'Mean Squared Error (MSE): {mse:.2f}')  
print(f'R-squared (R2): {r2:.2f}')  
  
"""  
Regression: Items\_Available  
"""  
  
# Import necessary libraries  
from sklearn.model\_selection import train\_test\_split  
from sklearn.tree import DecisionTreeRegressor  
import pandas as pd  
from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score  
  
# Load the data from the CSV file  
df = pd.read\_csv('PBStore\_branch\_analysis\_d\_FINAL.csv')  
  
X = df[[  
 "Items\_Available",  
 "Daily\_Customer\_Count",  
 # "Store\_Sales"  
 ]]  
  
y = df["Store\_Area"]  
  
# Split the data into training and testing sets  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1)  
  
  
## Decision Tree  
regressor = DecisionTreeRegressor(max\_depth=5)  
regressor.fit(X\_train, y\_train)  
  
# # # Make predictions on the test set  
y\_pred = regressor.predict(X\_test)  
  
# Calculate metrics  
mae = mean\_absolute\_error(y\_test, y\_pred)  
mse = mean\_squared\_error(y\_test, y\_pred)  
r2 = r2\_score(y\_test, y\_pred)  
  
# Print the metrics  
print(f'Mean Absolute Error (MAE): {mae:.2f}')  
print(f'Mean Squared Error (MSE): {mse:.2f}')  
print(f'R-squared (R2): {r2:.2f}')  
  
"""  
Regression: Daily\_Customer\_Count  
"""  
  
# Import necessary libraries  
from sklearn.model\_selection import train\_test\_split  
from sklearn.tree import DecisionTreeRegressor  
import pandas as pd  
from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score  
  
# Load the data from the CSV file  
df = pd.read\_csv('PBStore\_branch\_analysis\_d\_FINAL.csv')  
  
X = df[[  
 "Items\_Available",  
 "Store\_Area",  
 "Store\_Sales"  
 ]]  
  
y = df["Daily\_Customer\_Count"]  
  
# Split the data into training and testing sets  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1)  
  
  
## Decision Tree  
regressor = DecisionTreeRegressor(max\_depth=5)  
regressor.fit(X\_train, y\_train)  
  
# # # Make predictions on the test set  
y\_pred = regressor.predict(X\_test)  
  
# Calculate metrics  
mae = mean\_absolute\_error(y\_test, y\_pred)  
mse = mean\_squared\_error(y\_test, y\_pred)  
r2 = r2\_score(y\_test, y\_pred)  
  
# Print the metrics  
print(f'Mean Absolute Error (MAE): {mae:.2f}')  
print(f'Mean Squared Error (MSE): {mse:.2f}')  
print(f'R-squared (R2): {r2:.2f}')  
  
  
"""  
Regression: Store\_Sales  
"""  
# Import necessary libraries  
from sklearn.model\_selection import train\_test\_split  
from sklearn.tree import DecisionTreeRegressor  
import pandas as pd  
  
# Load the data from the CSV file  
df = pd.read\_csv('PBStore\_branch\_analysis\_d\_FINAL.csv')  
  
X = df[[  
 # "Items\_Available",  
 "Store\_Area",  
 "Daily\_Customer\_Count"  
 ]]  
  
y = df["Store\_Sales"]  
  
# Split the data into training and testing sets  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1)  
  
  
## Decision Tree  
regressor = DecisionTreeRegressor(max\_depth=5)  
regressor.fit(X\_train, y\_train)  
  
# # # Make predictions on the test set  
y\_pred = regressor.predict(X\_test)  
from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score  
  
# Calculate metrics  
mae = mean\_absolute\_error(y\_test, y\_pred)  
mse = mean\_squared\_error(y\_test, y\_pred)  
r2 = r2\_score(y\_test, y\_pred)  
  
# Print the metrics  
print(f'Mean Absolute Error (MAE): {mae:.2f}')  
print(f'Mean Squared Error (MSE): {mse:.2f}')  
print(f'R-squared (R2): {r2:.2f}')  
  
  
"""  
Classification: Daily\_Customer\_Count  
"""  
from sklearn.neighbors import KNeighborsClassifier  
import pandas as pd  
import numpy as np  
from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix  
from sklearn.model\_selection import train\_test\_split  
from sklearn.preprocessing import StandardScaler  
  
def convert\_num\_to\_cat(data, bin\_override=None):  
 iqr = np.percentile(data, 75) - np.percentile(data, 25)  
 bin\_width = 2 \* iqr / np.power(len(data), 1/3)  
  
 # Calculate the number of bins  
 num\_bins = int(np.ceil((max(data) - min(data)) / bin\_width))  
 print(num\_bins)  
  
 hist, bin\_edges = np.histogram(data, bins=bin\_override or num\_bins)  
 bin\_edges  
  
 y\_cat = []  
 for i in data:  
 for e,\_ in enumerate(bin\_edges[:-1]):  
 if bin\_edges[e]<=i <= bin\_edges[e+1]:  
 y\_cat.append(e)  
 break  
 return pd.Series(y\_cat)  
  
df = pd.read\_csv('PBStore\_branch\_analysis\_d.csv')  
X = df[[  
 "Store\_Area",  
 # "Items\_Available",  
 "Store\_Sales"  
 ]]  
y = convert\_num\_to\_cat(df["Daily\_Customer\_Count"], 3)  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1)  
  
scaler = StandardScaler()  
X\_train = scaler.fit\_transform(X\_train)  
X\_test = scaler.transform(X\_test)  
  
classifier = KNeighborsClassifier(n\_neighbors=6)  
classifier.fit(X\_train, y\_train)  
  
# Make predictions on the test set  
y\_pred = classifier.predict(X\_test)  
  
# Evaluate accuracy  
accuracy = accuracy\_score(y\_test, y\_pred)  
print(f"Accuracy: {accuracy}")  
print(classification\_report(y\_test, y\_pred))  
print(confusion\_matrix(y\_test, y\_pred))  
  
  
"""  
Classification: Store\_Sales  
"""  
from sklearn.neighbors import KNeighborsClassifier  
import pandas as pd  
import numpy as np  
from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix  
from sklearn.model\_selection import train\_test\_split  
from sklearn.preprocessing import StandardScaler  
  
def convert\_num\_to\_cat(data, bin\_override=None):  
 iqr = np.percentile(data, 75) - np.percentile(data, 25)  
 bin\_width = 2 \* iqr / np.power(len(data), 1/3)  
  
 # Calculate the number of bins  
 num\_bins = int(np.ceil((max(data) - min(data)) / bin\_width))  
 print(num\_bins)  
  
 hist, bin\_edges = np.histogram(data, bins=bin\_override or num\_bins)  
 bin\_edges  
  
 y\_cat = []  
 for i in data:  
 for e,\_ in enumerate(bin\_edges[:-1]):  
 if bin\_edges[e]<=i <= bin\_edges[e+1]:  
 y\_cat.append(e)  
 break  
 return pd.Series(y\_cat)  
  
df = pd.read\_csv('PBStore\_branch\_analysis\_d.csv')  
X = df[[  
 "Store\_Area",  
 "Items\_Available",  
 "Daily\_Customer\_Count"  
 ]]  
y = convert\_num\_to\_cat(df["Store\_Sales"], 3)  
  
X["Items\_Available"] = convert\_num\_to\_cat(X["Items\_Available"], 10)  
X["Daily\_Customer\_Count"] = convert\_num\_to\_cat(X["Daily\_Customer\_Count"], 7)  
  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1)  
  
scaler = StandardScaler()  
X\_train = scaler.fit\_transform(X\_train)  
X\_test = scaler.transform(X\_test)  
  
classifier = KNeighborsClassifier(n\_neighbors=10)  
classifier.fit(X\_train, y\_train)  
  
# Make predictions on the test set  
y\_pred = classifier.predict(X\_test)  
  
# Evaluate accuracy  
accuracy = accuracy\_score(y\_test, y\_pred)  
print(f"Accuracy: {accuracy}")  
print(classification\_report(y\_test, y\_pred))  
print(confusion\_matrix(y\_test, y\_pred))